TIME SERIES MODELLING AND FORECASTING OF THE YIELD OF CASHEW (*Anacardium occidentale* L.) IN KERALA

BY

MINI K. G.

THESIS

submitted in partial fulfilment of the

requirement for the degree of

Master of Science (Agricultural Statistics)

Faculty of Agriculture

Kerala Agricultural University

DEPARTMENT OF AGRICULTURAL STATISTICS

COLLEGE OF HORTICULTURE

VELLANIKARA, THRISSUR

1996

DECLARATION

I here by declare that this thesis entitled " Time Series Modelling and Forecasting of the Yield of cashew *(Anacardium occidentale L.)*in Kerala" is a bonafide record of research work done by me during the course of research and that the thesis has not previously formed the basis for the award of any degree, diploma, associateship, fellowship or any other similar title, of any other University or Society.

Vellanikkara
23.8.1996

MINI, K. G

CERTIFICATE

We, the under signed members of the Advisory committee of **Miss. Mini, K. G**, a candidate for the degree of **Master of Science in Agricultural Statistics**, agree that the thesis entitled " Time Series Modelling and Forecasting of the Yield of Cashew (*Anacardium occidentale* L.) in Kerala" may be submitted by **Miss. Mini, K. G.** in partial fulfillment of the requirement for the degree.

Smt. Graceamma Kurien
Assistant Professor (Agricultural statistics)
College of Horticulture
Vellanikkara
(Chairperson)

Prof. P. V. Prabhakaran
Professor & Head (Agricultural Statistics)
College of Horticulture
Vellanikkara
(Member)

Sri. S. Krishnan
Assistant Professor(Agricultural Statistics)
College of Horticulture
Vellanikkara
(Member)

Dr. K. P. Mani
Assistant Professor
College of Co-operation & Banking
Mannuthy
(Member)

External Examiner

ACKNOWLEDGEMENT

*I express my deep sense of gratitude and indebtedness to **Smt. Graceamma Kurien**, Assistant Professor, Department of Agricultural Statistics, Kerala Agricultural University, Vellanikkara and Chairperson of the Advisory committee for suggesting the problem, excellent guidence, and overwhelming inspiration throughtout the course of the investigation and preparation of the thesis.*

*My sincere thanks to **Prof. P. V. Prabhakaran**, Professor & Head, Department of Agricultural Statistics, College of Horticulture, Vellanikkara, **Sri. S. Krishnan**, Assistant Professor, College of Horticulture, Vellanikkara, **Dr. K. P. Mani**, Assistant Professor, College of Co-operation & Banking, Mannuthy, Members of the Advisory Committee for the pains taking involvement throught out the course of the programme.*

I would like to express my sincere thanks to the Associate Dean, College of Horticulture, Kerala Agricultural University, Vellanikkara for providing facilities for the study. I will be failing my duty if I do not express my extreme gratitude to the staff members of the Department of Agricultural Statistics and my fellow students for their co-operation and sincere interest in my work.

*I thankfully acknowledge the assistance rendered for statisical analysis by **Prof. Nandan Bose** and **Miss. Geetha**, Department of Future Studies, Kerala University,* *Thiruvananthapuram and **Dr. Balakrishna,** Lecturer, Department of statistics, Cochin University of science of technology, Kalamessery.*

I am highly obliged to all my friends and well-wishers for their assistance and kind co-operation.

I am greatful to the Kerala Agricultural University for the award of Junior fellowship during the course of study.

*I express my heartfelt gratitude to my cousin **Smt. Sujatha, K. S.** , Assistant Professor, College of Co-operation & Banking, Mannuthy and her family members for the constant encouragement and affection rendered by them at all stages of study.*

I am forever beholden to my parents and brothers who have been the sourse of inspiration for me at all critical junctures. I acknowledge with a full heart, their bundless affection, incessent encouragement and stead fast confidence in me, which I hope to have justified in some measure by this work.

*Above all I bow my head before the **Almighty God** who blessed me with health and confidence to undertake the work successfully.*

MINI, K. G.

CONTENTS

LIST OF TABLES

LIST OF FIGURES

Introduction

1. INTRODUCTION

Time series is one of the most widely pursued areas of statistics. A time series is a set of statistical observations arranged in chronological order. The essential fact which distinguish time series data from other statistical data is the specific order in which observations are taken. While observations from areas other than time series are statistically independent, the successive observations from a time series may be dependent, the dependence based on the position of the observations in the series.

A general statistical model of the time series $Y(t)$, $t \in T$ is $Y(t)=f(t)+\epsilon(t)$ where $f(t)$ represents the systematic part and $\epsilon(t)$ represents the random part. These two components are also known as `signal'and `noise' respectively. The model is just theoretical and hence $f(t)$ and $\epsilon(t)$ are not seperately observable. While the model for $Y(t)$ gives the structure of the generating process, a series of observations (or time series data) is a realisation or a sample function of the process. The effect of time may be in both the systematic and random parts. A general model in which the effect of time is represented in the random part involves stochastic process. The systematic part is represented by a non-random or deterministic function of time. The forces at work affecting the systematic part may be broadly classified as (1) Trend or long term movement and (2) Periodic changes which include both seasonal and cyclical variations.

Analysis of time series is of great significance as it helps in understanding the past behaviour of economic and social phenomena. Also it helps in evaluating the current performance in comparison with the past behaviour. Perhaps, the most important advantage of time series analysis is that it enables one to predict effectively the future behaviour of the series.

Time series can be used as an effective tool of forecasting only if sufficiently large chronological series is available for analysis. A great deal of data in agriculture and allied fields are in the form of *Time Series* where observations are dependent and where nature of this dependence is of interest in itself. Hence time series technique can be used to forecast various agricultural phenomena such as crop output, crop acreage, agricultural wages, agricultural prices etc. Such forecasts are crucial for planning and policy making.

The present study is concerned with the building up of emperical stochastic models for discrete time series and the use of these models in forecasting of the production of cashew, an most important cash crop of India. Cashew has emerged as one of the most important dollar earning crops of the country.

Stochastic models tried in this study include Box-Jenkins model, Log-normal diffusion model, Distributed lag model and Markov chain model. The ARIMA model introduced by Box and Jenkins (1970) has been proved to be very general and effective in handling complex time series. Though it has been very popular, its application in the field of agriculture is still rare. The log-normal diffusion model assumes that the logarithm of production is normally distributed and that the value is a function of time and area under the crop. Price responsiveness of farmers, measured by the functional relationships between planned output and expected prices can be worked out using Distributed lag model. A Markov chain model incorporating the property of dependence in time series can be very useful for long term forecasting for a specific population. Obviously these models are different in structure and application. Hence a direct comparison is impossible. However, it is worthwhile to examine the applicability of these techniques for forecasting the production of agricultural crops.

Review of Literature

2. REVIEW OF LITERATURE

The literature on various aspects pertaining to the study were reviewed under four sections viz. Box- Jenkins model, Distributed lag model, Log normal diffusion model and Markov chain model.

2. 1 Box-Jenkins model

Box and Pierce (1970) studied the relationship between various pairs of economic series. They showed that after taking into account autocorrelation within each series very little correlation was left between the residual series.

Chatfield and Prothero (1973) worked out a step by step account of Box- Jenkins analysis of sales data showing high multiplicative seasonal variation. It was shown that Box-Jenkins procedure offered a good technique for short term sales forecasting.

Granger and Morris (1976) reported that the mixed Autoregressive Moving Average (ARMA) model was the one, most likely to occur when the model was generated by the sum of two or more series. As most of the economic series were aggregates of several components measured with error it followed that such mixed models would often be found in practice. In such models, the possibility of resolving the series into simple components was considered both theoretically and for simulated data.

David and Kenneth (1976) had done a comparitive study of two traditionally separate approaches to time series modelling namely Box-Jenkins methods and multivariate methods of econometrics. The Box-Jenkins model was compared with as existing econometric model of aggregate demand in U. K. In terms of residual variance, the econometric model provided a better fit.

Anderson (1977) made an effort to emphasize the need for the Box-Jenkins iterative cycle of identification, estimation and verification in time series modelling . The author made an attempt to reconcile the Box-Jenkins models with the classical decomposition of series into trend, seasonal and irregular components. Box- Jenkins approach generally produced forecasts of superior quality to those from other extrapolatory procedures.

Anderson (1977) discussed some do's and don'ts in univariate Box -Jenkins analysis. A balanced account of the Box-Jenkins methodology including an assessment of its value and a discussion of its development were given. The author insisted that any modelling inadequacy should lead to modification.

An approach to the modelling and analysis of multiple time series was proposed by Tiao and Box (1981). Properties of a class of vector autoregressive moving average models were discussed. Modelling procedures were outlined and illustrated by examples.

Ogallo (1986) fitted Autoregressive Integrated Moving Average (ARIMA) (p, d, q) model to annual rainfall data of two homogeneous regions in East Africa with rainfall records extending between the period 1922-1980. It was observed that ARIMA (3, 0, 1) model best fitted the annual rainfall data. However, it was noted that fitted models had low forecasting skill. In all cases the models accounted for less than 50 per cent of the total variation.

Abraham and Ledolter (1986) pointed out the importance of ARIMA models to determine the form of the corresponding forecast functions. In this paper, the authors described how the various methods update the coefficients in the forecast functions and discussed their similarities and differences. In addition, they compared the forecasts from seasonal ARIMA models against the forecast from Winter's additive and multiplicative smoothing methods. It was found that seasonal ARIMA models were less suitable compared to Winter's additive method when there existed a high multiplicative seasonal component.

Pino *et al.* (1987) showed that the linear combinations of univariate time series that follow ARIMA model had the same internal structure as those of the original series in both seasonal and nonseasonal cases. They compared two approaches. Apprach I - First form the linear combination, then model and forecast. Approach II - First model and forecast, then form linear combination. They suggested that in terms of mean squared error of forecasting, approach II should be preferred to approach I.

Chatfield (1988) reviewed several forecasting methods. It was recognised that there were many different types of forecasting problems, requiring different treatments. They may be classified into univariate, multivariate, judgemental methods and also by whether an automatic or nonautomatic approach was adopted. Chatfield compared the strength and weaknesses of different univariate methods both in automatic and nonautomatic models.

Anderson (1989) discussed the modifications and extensions to the well established Box-Jenkins methodology for analysing and forecasting time series domain. The author concentrated on improving interpretation of the serial correlation structure for the purpose of enhancing model identification.

Haines *et al.* (1989) presented the fitting of ARIMA models to time series data relating to births at Edenhale hospital in Natal, South Africa, over a sixteen year period. The model $(011 \times 011)_{12}$ provided an excellant fit of the monthly total of mothers delivered which included the series of monthly totals of caesarean sections performed.

Box-Jenkins ARIMA models was used by Mary (1991) to predict the demand and production of natural and synthetic rubber in Kerala. The results revealed that the method offered a good technique for predicting the rubber production.

Ray (1991) examined the supply and demand linkages between agriculture and industry and also availability of power as determinantes of the rate of growth of manufactured goods production in India during 1951-1984. Box-Jenkins methodology was used to examine the nature of the dynamic relationship between manufacturing growth rate, the internal terms of trade, availability of food grains, agricultural income and availability of power. The results revealed that the increase in food availability stimulates manufacturing growth with a two period lag by relaxing the wage goods constraint. The incresase in agricultural income had an immediate impact on manufacturing growth through demand stimulation.

Chan (1992) studied the usefulness of sample autocorrelation and partial autocorrelation as specification tools when the observed time series was contaminted by an outlier. The results indicated that the specification power of these statistics could be significantly jeoparadized by an additive outlier, on the other hand an innovational outlier seemed to cause no harm to them.

Gupta (1993) suggested that ARIMA model offered a good technique for predicting the magnitude of tea production using monthly data for January 1979 through July 1991 in India. The author stated that the method was suitable for any time series with any pattern of change and it did not require the forecaster to chose a priory value of any parameter.

Sarkar and Kartikeyan (1993) made an attempt to forecast monthly arrivals of a particular cotton variety in Raichur district of a major cotton producer, Karnataka in India. A particular AR model had been considered along with seasonal multiplicative model for forecasting a particular seasonal series. The forecasting performance of both models were compared. Results revealed that the subset AR model gave better forecasts since the identification procedure of subset AR model was simple, the authors suggested the use of such models when there existed complicasy in the identification of multiplicative ARIMA models.

Borah and Bora (1995) fitted a seasonal ARIMA model to the monthly rainfall data of Guwahati for the period 1956-1965. The model parameters were estimated using Marquardt's algorithm for nonlinear optimization. The model was used to predict the rainfall for the month ahead and monthwise rainfall for the year ahead. The forecast made by the model were compared with the observed values and were found to be reasonably good.

2. 2 Distributed lag model

Marc Nerlove (1956) developed a statistical model to estimate the elasticity of supply for corn, cotton and wheat in the United states over the period 1909-1932. The study aimed at evaluating the role of the farmers expectations of future prices, which played a role in shaping their decisions as to how much acres they should devote to each crop. The estimation equation included lagged prices and lagged area. The results revealed that price elasticities were positive and significant.

Rajkrishna (1963) estimated the price response of major crops in the pre-partition Punjab for the period 1914-1945. In addition to the relative price, the author used three shifter variables- relative yield, irrigation and rainfall. The elasticities for cotton and maize were positive. All crops except jower showed positive and significant response.

Devi and Rajagopalan (1965) examined the influence of relative prices of groundnut and its competing crops for the period 1934-1962 in North Arcot. A simple linear regression model was fitted to the data and concluded that an increase in relative price influence the acreage under groundnut in the following year, while its influence on productivity was not at all significant. The results also showed that increase or decrease of acreage was inversely associated with acreage under competing crops.

Madhavan (1972) studied the supply response of food crops such as rice, ragi and sorghum and cashcrops such as cotton, grondnut, sesamum and sugarcane for the period 1947-1965, using the Nerlovian adjustment model. Lagged relative price, lagged yield and acreage of the crops and its competitors and a rainfall index were the determinants in the function. The price coefficients were statistically significant in the supply of all crops except rice. The results revealed that price elasticities were high when both depending and competing crops, came from the cereal crop group.

Reddy (1977) studied the supply response of groundnut farmers of Kurnool in Andhrapradesh over the period 1931-'43 using a Nerlovian adjustment model. The determinants were relative price, yield lagged by one year, rainfall and trend. The results showed that the coefficients relating price and yield had the right sign and were statistically significant. The author concluded that the farmers in the area understudy were responsive to relative price changes and relative yield changes.

Kumar and Srivastava (1982) examined the short run and long run elasticities in hectareage allocation under a crop for major staple food (wheat and rice) in Allahabad district during the period 1961-'62 to 1977-'78. The model included current planted area under the crop as the dependent variable and one year lagged hectareage, price, yield, presowing / sowing period rainfall, competing crop's price and price variability as the independent variables. The variables which affected significantly the supply variations were presowing rainfall and lagged per hectare yield in case of wheat and sowing period rainfall in case of rice.

Prabhakaran (1987) studied the impact of price variability in acreage allocation of five important crops of Kerala viz. rice, tapioca, coconut, pepper and cashewnut using Nerlovian model over a period of 30 years starting from the year 1952- '53. Lagged area and farm price were taken as the independent variable. The values of the Nerlovian coefficient of adjustment for the five crops are comparitively low and nearer to zero indicating that in general Kerala farmers were less responsive to price fluctuations and were slow in adjusting their acreage according to expectations. The results also revealed that cashewnut growers were least sensitive to price movement.

Lakshmi and Pal (1988) carried out the decomposition analysis of aggregate crop output of Kerala into its component elements using a seven factor additive model for the 1952-'53 to 1984-'85 period. The study covered crops such as rice, cassava, pepper, arecanut, cashew, ginger, coconut, rubber, tea and coffee which together cover more than 80 per cent of the gross cropped area in Kerala. The results revealed that nearly 50 per cent of the change in crop output in Kerala is due to the change in the total area under the ten crops and 42 per cent through the change in the yield of the concerned crop. The changes in the crop pattern accounted for only 8.4 per cent which was much less than the contribution by the interaction of the changes in area and yield which explained 15.3 per cent.

Reddy (1989) analysed the supply response of paddy in Andhrapradesh over the period 1963-'64 to 1983-'84. Nerlovian partial adjustment model had been adopted to study the farmer's responsiveness. The short run and long run elasticities of price and non-price variables for three regions seperately and as a whole were estimated. The study revealed that in addition to price incentive, non-price incentives like provision of assured irrigation and HYV seeds are equally important, since they helped to increase the yield in achieving the targets of paddy output.

Indiradevi *et al.* (1990) analysed the growth supply response of banana in Kerala over the period 1970-' 71 to 1986- '87, both in terms of area and yield by using Nerlovian lagged adjustment models of linear and double log forms. The results obtained from linear model revealed that the regression coefficient as well as their level of significance were superior over the double log model. Neither lagged absolute price nor the rainfall during planting months was found to exercise any significance on acreage allocation decisions on this crop.

Thomas *et al.* (1990) estimated the acreage and yield response of ginger in Kerala over the period 1968-'69 to 1986-'87 by fitting response function of the Nerlovian type. The results revealed that the lagged price in ginger did not hold any significant inluence on the state and the farmers were found to be good risk bearers.

Chandrabhanu (1991) analysed the supply response on sesamum and groundnut both at the district and state levels of Kerala using time series data for the 1961-'62 to 1987-' 88 period. Supply responsiveness (both in terms of area and yield) was studied using response functions of Nerlovian lagged adjustment type fitted to the whole period. Response of aggregate sesamum area to price appeared positive though not significant. However non-price factors like pre-sowing rainfall and lagged yield seemed to exert much more stronger influence on aggregate acreage. Adjustment was slow indicating the existence of techno-institutional constraint. For groundnut there was no significant relationship between area and price movements. The Yield movements were found to be in a direction opposite to that of price. The results revealed that the farmers in general were found averse to bearing yield risk.

Janaiah *et al.* (1992) studied area response of cotton, sugarcane and tobacco using Nerlovian lagged adjustment model in double log form over the peiod 1956- '57 to 1985-'86. The independent variables were farm harvest price of concerned crop lagged by one year, price risk and total rainfall in the current year. Among the crops, cotton exhibited encouraging response followed by tobacco and sugarcane. The results revealed the price consciousness of farmers and strengthened the hypothesis that a remunerative price favoured more area allocation under the crop.

Tripathy and Gowda (1993) examined the area response of groundnut using Nerlovian lagged adjustment model in Orissa. The independent variables were the lagged area, lagged price, price risk, irrigation and rainfall in the planted period. The results revealed that the lagged area, price, price risk and irrigation had positive and significant impact on the area under the groundnut. The effect of rainfall turned out to be negative but statistically non significant, implying that excess rain fall during sowing time inhibited area expansion of groundnut.

2. 3 Log normal diffusion model

Tintner and Patel (1965) applied log normal diffusion model to the data on national income of India, using the government expenditure as the exogeneous variable. Tintner and patel (1969) also utilised the same model to explain trend in per hectare yield of crops viz. rice, wheat and sugarcane in India, taking the proportion of irrigated area under the crop as exogeneous variable. They also studied the trend in real agricultural production per capita in India, taking the real per capita government expenditure as the exogeneous variable.

Saraswathi and Thomas (1975) used log normal diffusion process for the explanation of trend in production of rice in Kerala for the period 1957-'58 to 1971-'72. Five different models were tried to explain the production of rice in terms of area and period. The models fitted the data satisfactorily.

Saraswathi and Thomas (1976) adopted the same method to explain trends in production of the rice, tapioca, coconut, arecanut, pepper, tea, coffee, rubber and cashewnut, taking area under the crop as the exogeneous variable. The base year was taken as 1952-'53. It was found that coefficient of determination reported by the authors were very high and forecast values were very satisfactory. They reported that the stochastic model used namely log normal diffusion model offered a reasonably close fit to the data and hence these models could forecast the preharvest production of crops.

2. 4 Markov chain model

Gabriel and Newmann (1962) fitted a two state Markov chain model to daily rainfall occurrence at Tel Aviv. The various properties of Markov chain model applicable to rainfall were also discussed.

Waghmare *et al.* (1977) fitted a four state Markov probability model to a time series. The data consisted of area under cotton crop for the years from 1925 to 1972. The four states of the Markov model were classified according to first, second and third quartile and a transition probability matrix P was obtained. Further equilibrium state was estimated by powering the matrix P. The results revealed that the equilibrium state of Marathwada region as a whole will be attained after eleven years with an estimated area of 5,45,324 hectares.

Asan *et al.* (1981) tried a three state Markov chain model to compare four dietary treatments. Three body weight groups were represented by the three states of the Markov chain model and transition probabilities were calculated. The expected first passage times were worked out.

Santhosh and Prabhakaran (1988) applied a two state Markov chain model of first order to daily rainfall data of five selected reporting stations of Northern Kerala. It was found that the model was adequate in representing the rainfall pattern in Northern Kerala.

Krishnan and Narayanikutty (1993) worked out a two state Markov chain model of first and second order for the daily rainfall data. The chi-square test for adequacy revealed that the first order was the best fit.

Materials and Methods

3. MATERIALS AND METHODS

In order to develop a time series model for the purpose of forecasting the yield of a crop it is inevitably needed to collect the data from the area under study. The present study is aimed at developing statistical forecasting models for cashew for the state of Kerala. The secondary data required for the present investigation were collected from the various publications of Directorate of Economics and Statistics, Thiruvananthapuram, Kerala state. Observations on acreage, production, productivity, annual rainfall and price of the raw cashew kernal were collected over the period 1956-1992. The following models were tried.

3.1 Box Jenkins model

Autoregressive Integrated Moving Average (ARIMA) models were developed by Geroge Box and Gwilym Jenkins (1970) and their names frequently have been used synonymously with general ARIMA process applied to time series analysis, forecasting and control. Box and Jenkins method of forecasting is one that is particularly well suited to handling complex time series and other forecasting situations in which a variety of patterns exist. The real power and attractiveness of this method is that it can handle complex patterns using a relatively well specified set of rules. Though it has become quite popular in west, its application to Indian data is still rare. This is basically because it is quite complicated and its appropriate use requires long time series data and the availability of requisite softwares.

The reasoning behind the development of Box-Jenkins approach is that the existing methods of forecasting always assume or are limited to specific kinds of pattern in the data.

In this method there is no need to assume initially a fixed pattern : rather, the approach begins by assuming a tentative pattern that is fitted to the data so that the error will be minimized. In general Autoregrssive Integrated Moving Average (ARIMA) model is expressed as follows.

ARIMA (p, d, q) (P, D, Q)^s

where p=order of non-seasonal autoregression (AR)

d = order of non-seasonal difference

q = order of non-seasonal moving average (MA)

P = order of seasonal autoregression (AR)

D = order of seasonal difference

Q = order of seasonal moving average (MA)

s = length of season (=4 in quarterly data, 12 in monthly data and so on)

If X denotes the variable, the model could be expressed in the form of an equation as below.

$$(1-\phi_1 B-\phi_2 B^2- \text{----------} -\phi_p B^p)(1-\phi_{1s} B^s-\phi_{2s} B^{2s}- \text{----}\phi_{ps} B^{ps})$$

$$(1-B)^d(1-B^s)^D X_t =(1-\theta_1 B-\theta_2 B^2- \text{-----}\theta_q B^q)(1-\theta_{1s} B^s-\theta_{2s} B^{2s}- \text{----}\theta_{Qs} B^{Qs})e_t \qquad \text{--------(1)}$$

which can be condenced as

$$\phi_p(B)\,\phi_{Ps}(B^s)(1-B)^d\,(1-B^s)^D\,X_t=\theta_q(B)\theta_{Qs}(B)^s\,e_t \qquad \text{---------(2)}$$

where X = variable under study

B = lag operator

e = error term

t = time subscript

$\phi_p(B)$ = non-seasonal AR

$\phi_{Ps}(B)$ = seasonal AR

$(1-B)^d$ = non-seasonal difference

$(1-B^s)^d$= seasonal difference

$\theta_q(B)$ = non-seasonal moving average

θ_{Qs} (B) = seasonal moving average

ϕ_p , ϕ_{Ps}, θ_q, θ_{Qs} are the parameters. The model as expressed in equation (1) contains p+q+P+Q parameters,which need to be estimated.

Since seasonal component is absent in the present study Autoregressive integrated moving average of the form ARIMA (p,d,q) is considered. The equation representing the model is

$$\phi_p(B)(1-B)^d X_t = \theta_q(B)e_t \qquad \text{-------(3)}$$

where

$\phi_p(B)$ = non-seasonal Autoregression

$\theta_q(B)$ = non-seasonal Moving average

$(1-B)^d$ = non-seasonal difference

X = variable under forecasting

B = lag operator

e_t = error term

t = time subscript

Box and Jenkins fit models of the form (3) to a given set of data by an iterative three step cycle of identification, estimation and diagnostic checking. At the first stage, a specific model that can be tentatively used as the forecasting method best suited to that situation is identified. Second stage consists of estimating the parameters using fully efficient statistical techniques. Finally diagnostic checking is made to determine whether or not the estimated model, describes the given time series. Any inadequacies discovered may suggest an alternative form of the equation and the whole iterative three step cycle is repeated until a satisfactory model is obtained. Each of these steps is now explained.

3.1.1 Model Identification

An ARIMA model is identified by the constants p, d and q. `d' is the degree of differencing required to produce stationarity, there by reducing the series to a mixed Autoregressive Moving Average (ARMA) process. The resulting ARMA process is identified by p and q, the orders of AR and MA operators.

The main tool used for identification purpose are the autocorrelation (ac) and the partial autocorrelation (pac) functions. To determine the value of d, the autocorrelation function of the original series as well as that of the differenced series are made use of. It is assumed that stationarity is reached as the autocorrelation function of the differenced series, $\Delta^d X_t$ dies out fairly quickly.

Having tentatively decided what `d' should be, the general appearence of the estimated autocorrelation function and partial auto correlation functions of the appropriately differenced series is studied to obtain clues about the choice of p and q. As a general rule, when the autocorrelations drop off exponentially to zero implies an autoregressive model whose order is determined by the number of partial autocorrelations which are significantly different from zero. If partial autocorrelations drop off exponentially to zero, the model is moving average and its order is determined by the number of statistically significant autocorrelations. When both autocorrelation and partial autocorrelation drop off exponentially to zero, the model is ARIMA.

3.1.2 Model Estimation

Once the decision to fit an ARIMA model has been made, the next task is to estimate the value of the parameters. That involves finding the values of $\phi_1, \phi_2, \ldots, \phi_p, \theta_1, \theta_2, \ldots, \theta_q$ in equation (3) that minimises mean square error (MSE)

Instead of employing a simple trial and error procedure to find the best ϕ's and θ's, it is usually more efficient to apply a method based on the Gauss - Newton constrained optimization approach (such as Marquardt algoritham) . A major difficulty in applying ARIMA models is that because of the iterative model and error procedure involved, considerable computations are required when the order of the model increases beyond one.

3.1.3 Diagnostic checking of the estimated model

Once the equation has been estimated, its appropriatness in describing the data is determined by examining the residuals. There are two possible findings (a) The errors are random, which means the fitted model has eliminated all patterns from the data, and that what remains are random errors (b) This is not the case and the tentatively identified model has not removed all patterns as indicated by the fact that the 'e_t' are not random. The AC's can tell us how successive values of the residuals relate to each other. If they are random then no autocorrelation should be significantly differnt from zero. Box-Pierce test is conducted to see if a number of autocorrelations together are significantly different from zero. The Q statistic is given by

$$Q = n \sum_{i=1}^{m} \gamma_k^2$$

where n = sample size

m = lenth of the lag considered

γ_k = autocorrelation coefficient of order k

The Q statistic has a chi-square distribution with `m' degrees of freedom. The calculated value of Q is compared with the tabulated value. A non-significant value of Q confirms that the identified model is quite appropriate.

3.1.4 Forecasting with ARIMA model

Once a model has been identified, the parameters are estimated and the residuals have been shown to be random, forecasting with that model is just a straight forword and mechanical matter. The computer programme, which is inevitably necessary to carry out the tedious calculations of identification and estimation can provide as many forecasts together with 95 percent or 99 percent confidence intervals for forecasts.

The computations for fitting the Box-Jenkins model is carried out using the RATS (Regressoion Analysis on Time Series) database. A sample programme given in appendix 1 illustreates the detailed output of Box-Jenkins methodology. Thirtyseven data points were used for the analysis. Univariate ARIMA models of production, area,productivity, annual rainfall and price of raw cashew kernal were considered separately.

3.1.5 Transfer function models

Multivariate models attempt to capture and measure the influence of external, independent factors on the dependent variable. The multivariate ARIMA models can be used to combine the concepts of multiple regression with those of univariate time series models. A general multivariate ARIMA model with the dependent variable `Y' and the independent variables say X, Z and W is given by the following equation.

$$Y_t = \delta_1 Y_{t-1} - \delta_2 Y_{t-2} - \cdots - \delta_r Y_{t-r} + \omega_0 X_{t-b} - \omega_1 X_{t-b-1} - \cdots - \omega_s X_{t-b-s} + \psi_0 Z_{t-c} - \psi_1 Z_{t-c-1} - \cdots - \psi_r Z_{t-c-r} + \xi_0 W_{t-d} - \xi_1 W_{t-d-1} - \cdots - \xi_u W_{t-d-u} + e_t \qquad (4)$$

To apply equation (4) several steps are required. First appropriate independent variables (such as X, Z, W) should be identified. These variables should influence the dependent variable Y in a way that is statistically significant. Second, the number of terms of Y, X, Z and W should be identified. This is determined by the values of b, c and d. These values indicate the number of periods before t that X, Z and W respectively lead the dependent variable Y. For instance, if b=2, c=1 and d=3, a change in X will influence Y two periods later, a change in Z will influence Y one period later, and a change in W will influence Y three periods later. The values of b ,c and d must be greater than zero in order to take real advantage of equation (4). Finally, the parameters δ , ω , ψ and ξ should be estimated.

There are several procedures for applying equation (4) to actual data. One among the best known approach is the method of fitting the transfer function model. Box and Jenkins (1976) popularized this approach which consists of three steps. First, the dependent variable and each of the independent variables are prewhitened. This is done by using univariate ARIMA model on each variable in such a way that the residuals of each of the variables to be included in the transfer function model are random. Second, a model relating the dependent variable and independent variables are identified using the residuals (prewhitened values) of each series. The principle for doing so is that the real relationship between indpendent variables and dependent variable can be found only after spurious correlations caused by trend, seasonality and other external patterns have been eliminated. Box and Jenkins provide guidelines for identifying appropriate models using the crosscorrelations or the impulse response function. In this study the cross correlations are used for identifying the appropriate model. The values of r,s,v and u as well as b,c and d are determined.

3. 2 Distributed lag model

The underlying aim of all supply response studies is to find out how farmers intend to react to movements in the price of the crop that he produces. There are serious difficulties in measuring the degree of response of producers to price changes. They arise mainly from the difficulties approximating theoretical formulations of functional relationship to observe real world situations. These difficulties are further compounded because of the time lag between changes in production capacity and changes in output. Problems in adequate representation of risk, producers expectations, changing technologies and government policies thus assume importance.

Due to time lag between changes in agricultural production capacity and changes in output, in any attempt to measure the price responsiveness, the functional relationship should ideally be worked out between planned output and expected prices. Producers base their decisions on prices of their own expectations. However it does not seem proper to assume that expectations of farmers in all areas are identical. Yield is prone to much more variation than area since yield can be influenced significantly not only by manmade factors but by natural factors as well.

In the past, researchers used two types of models in estimating supply response. The first is called as traditional price lag model, which assumes that farmers instantaneously and fully adjust their acreage allocations in response to changes in lagged prices. Thus second is called adjustment lag model often referred to Nerlovian adjustment model. The lagged adjustment model is said to present a more realistic picture by incorporating distributed lags and there by introducing a realistic assumption about the farmers adjustment behaviour. The advantage of this model compared to the traditional model is that it explains the data better by yielding coefficients more reasonable in sign and magnitude, there by producing better estimate of supply elasticities. Further it eliminates the incidence of serial correlation in the residuals (Nerlove, 1958)

The Nerlovian adjustment model, in its simplest form is based on the relation

$$A_t^* = A + b\,P_{t-1} + u_t \qquad \text{------------(5)}$$

$$A_t - A_{t-1} = B\,(A_t^* - A_{t-1})\ ;\ 0 < B < 1$$

where

A_t^* = Desired planted area in the year `t'

A_t = Actual planted area in the year `t'

A_{t-1} = Actual planted area under the crop in year `t-1'

P_{t-1} = Farm harvest price in period `t-1'

B = Nerlovian coefficient of adjustment and its value lies between 0 and 1

The reduced form becomes

$A_t = a_0 + b_0 P_{t-1} + c_0 A_{t-1} + v_t$ ---------(6)

where

A_0 = aB

b_0 = bB

c_0 = 1-B and

v_t = Bu_t

Additional variables can be easily incorporated into the structural equation. The output response function was measured in terms of area and yield, since the farmers respond to economic stimuli initially by altering the productivity through intensifying the cultivation practice and thereafter the area under the crop. The planned output is the product of intended cultivated area and planned yield. The elasticity of output can be easily be determined, once the area and yield response models were developed seperately on the basis of Nerlovian lagged adjustment model(Ramesh, 1988).

The chosen model should explain important expectations influencing the decision making process regarding the area allocation and adoption of yield increasing techniques. The variables chosen were lagged price, annual rainfall lagged by one year and risk factor. The price which farmers take into account for their decision making process is called the expected price. The price expectation implied in the Nerlovian adjustment lag model is previous years price. It must influence the farmers resource allocative decisions. Also one can expect the farmers to be responsive to the risk factor. In the lagged price models price risk in the period (t) was represented by the standard deviation of price in the past three years from the period `t'. Annual rainfall lagged by one year (w_{t-1}) was included in the area /yield response model as an independent variable.

Introducing all chosen variables into the Nerlovian lagged adjustment model, the final estimating equations are obtained as follows.

(i) Area response model

$$A_t^* = a_0 + a_1 P_{t-1} + a_2 R_t + a_3 W_{t-1} + U_t \quad \text{-------(7)}$$

$$A_t - A_{t-1} = B(A_t^* - A_{t-1}) \; ; \; 0<B<1$$

The reduced form becomes

$$A_t = b_0 + b_1 A_{t-1} + b_2 P_{t-1} + b_3 R_t + b_4 W_{t-1} + u_t \quad \text{-------(8)}$$

where A_t^* = Desired area in hectares under the crop in the year `t'

A_t = Actual area in hectares under the crop in year `t'

A_{t-1} = Actual area in hectares under the crop in year `t-1'

P_{t-1} = Farm harvested price in the year `t-1'

R_t = Price risk as measured by the standard deviation of price in the past three years from the year `t'

W_{t-1} = Annual rainfall in the year `t-1'

B = Nerlovian coefficient of adjustment

$b_0 = a_0 B$

$b_1 = 1\text{-}B$

$b_2 = a_1B$

$b_3 = a_2B$

$b_4 = a_3B$ and

$v_t = Bu_t$

(ii) Yield response model

$$Y_t^* = c_0 + c_1P_{t-1} + c_2R_t + c_3W_{t-1} + u_t \qquad \text{-------(9)}$$

$$Y_t - Y_{t-1} = \gamma\ (y_t^* - Y_{t-1})\ ;\ 0<\gamma<1$$

The final estimating equation was

$$Y_t = d_0 + d_1Y_{t-1} + d_2\ P_{t-1} + d_3R_t + d_4W_{t-1} + v_t \qquad \text{-------(10)}$$

where Y_t^* = Expected yield in year `t'

Y_t = Yield of the crop in year `t'

Y_{t-1} = Yield of the crop in year `t-1'

P_{t-1} = Farm harvest price in year `t-1'

R_t = Price risk

W_{t-1} = Annual rainfall in year `t-1'

γ = Coefficient of yield adjustment

$d_0 = c_0\gamma$

$d_1 = 1\text{-}\gamma$

$d_2 = c_1\gamma$

$d_3 = c_2\gamma$

$d_4 = c_3\gamma$ and

$v_t = \gamma u_t$

The functions were estimated in linear form by the method of ordinary least squares, for the whole period and two subperiods, viz. period 1 (from 1956-'57 to 1974-'75) and period 2 (from 1974-'75 to 1991-'92). The regression coefficients were tested for their significance using t - test. The short run and long run elasticities were computed as under

$$\text{short run elastcity} = \frac{\text{Regression coefficient of price x mean of price}}{\text{mean of area/yield}}$$

$$\text{long run elsticity} = \frac{\text{short run elasticity}}{\text{coefficient of area/yield adjustment (B)}}$$

Speed of adjustment was estimated using the relation

$(1 - B)^N = 0.05$

where B = coefficient of area/yield adjustment

N = number of years required to realise 95 percent of the price effect.

The major estimation problems arising out of the use of time series data are multicollinearity and autocorrelation. When the exogeneous variables in a relation are closely correlated it becomes difficult to isolate their separate influences to obtain reasonably precise estimates of their relative effects. Serial correlation of the random term `u' violates the assumption of the method of ordinary least squares. Though unbiased estimates of parameters can be obtained, their sampling variance will be unduely large and further, there will be serious under estimation of the variances(Johnston, 1972).

Durbin and Watson d-statistic is commonly employed for testing the incidence of autocorrelation

$$d = \frac{\sum_{t=2}^{n} (e_t - e_{t-1})^2}{\sum_{t=1}^{n} e_t^2}$$

where e_t and e_{t-1} are residual terms of current and lagged dependent variables respectively. The Durbin Watson statistic will lie in the 0 to 4 range with a value near 2 indicating no first order serial correlation (Durbin and Watson, 1951).

However, the `d' statistic is not an appropriate measure of autocorrelation if among the explanatory variables there are lagged values of the endogeneous variable. For such cases Durbin (1970) suggested `h' statistic

$h = (1-d/2)\sqrt{n} /(1-n(V(b_1))$

where $V(b_1)$= estimate of the sampling variance of b_1

n = sample size

d = computed Durbin - watson d statistic.

For large sample size, Durbin has shown that if $\rho = 0$ the `h' statistic follows the standardized normal distribution, that is the normal disttribution with zero mean and unit variance. Hence the statistical significance of an observed h can be determined from the standardized normal distribution table. The test involving `h' statistic breaks down when n $v(b_1) \geq 1$, for such cases the ` d' statistic was employed to check the incidence of serial correlation.

3. 3 Log-Normal Diffusion Model

The log-normal process (Tintner and Patel -1965) is used to represent the data. In the log-normal model, the logarithm of production at time 't' assumed to be normal. The influence of an exogeneous factor like area under the crop is considered by taking the parameters of the process as a function of such variable. The probability density function $P(Y_t)$ of the diffusion process is

$P(Y_t) = (1/Y_t\sqrt{2\pi\gamma t}) \exp(-1/2\gamma t)\{\log Y_t - \log Y_0 - \beta_0 t - \beta_1 \Sigma X_j\}^2$ where Y_0 is the known and fixed value of production at some previous period of time and time is measured as deviation from this base period , $\gamma = a_0$, $\beta_0 = b_0 - a_0/2$ and $\beta_1 = b_1$ are to be the estimated. The maximum likelihood estimate for γ, β_0 and β_1 are obtained as

$$\beta_0 = \Sigma \{\log Y_t - \log Y_{t-1} - \beta_1 \Sigma X_j\}$$
$$\beta_1 = \mathrm{cov}\{X_t, \log(Y_t/Y_{t-1})\}/V(X_t) \quad \text{and} \quad \gamma = 1/n\,\{\Sigma[\log(Y_t/Y_{t-1}) - \beta_0 - \beta_1 X_t]^2$$

The expected value of y_t is $Y_0 \exp[b_0 t + b_1 \Sigma X_j]$ and variance of Y_t is $Y_0^2 \exp 2[b_0^t + b_1 \Sigma X_j][\exp a_0 t - 1]$. The estimate of X_t based on the exponential model for the area have been used for predicting annual prodution of the crop.

3.4 Markov chain model

In most of the stochastic process, the probability that the system will be in a given state at a given time may be deduced from a knowledge of its state at an earlier time and does not depend on the history of the system before that time. The process satisfying this condition are called Markov process. A stochastic process is said to be a Markov proess if for any set of n time points t> t-1> t-2>-----> t-n+1 in the index set of the proess, the conditional distribution of X(t) for given values of X(t-1), X(t-2),----, X(t-n+1) depends only on X(t-1), the most recent known value. Thus the fundamental principle underlying Markov process is the independence of the future from the past if the present is known.

A class of Markov process in discrete time and whose state space is discrete is called a Markov chain. The Markov chain is described by its transition probability P_{ij} (n, n+1), the conditional probability that the system is in state `j' at time `n+1' given that it was in state `i' at time `n'. The transition probabilities are most conventiently handled in matrix form $P=(P_{ij})$ called transition probability matrix. The elements in the transition probability matrix will all be non-negative and the elements in each row sum to unity. The n^2 transition probability can be represented by the nxn matrix P given by

$$P = (P_{ij}) = \begin{bmatrix} p_{11} & p_{12} \text{------} & p_{1n} \\ p_{21} & p_{22} \text{------} & p_{2n} \\ \text{------} & \text{------} & \text{------} \\ p_{n1} & p_{n2} \text{------} & p_{nn} \end{bmatrix}$$

The notation $P_j^{(n)}$ denotes the probability that the chain is in state `j' at step n. Then it can be easily proved that

$$P_j^{(n)} = \sum_{j=1}^{m} P_{ij} \, p_j^{(n-1)}$$

which in matrix notation can be written as $P^{(n)} = P^{(0)} P^n$

where $P^{(0)}$ is the unconditional probability vector at time '0'. As the Markov chain advances in time, $P_j^{(n)}$ becomes less and less dependent as $P^{(0)}$. That is to say that the probability of being in state 'j' after a large number of steps becomes indepindent of the initial state of the chain. When this occurs, the chain is said to have reached a steady state.

The transition probability matrix P can be estimated from the observed data by tabulating the number of times the observed data went from state i to state j , that is n_{ij} . Thus an estimate of p_{ij} would be $p_{ij} = n_{ij} / \sum_j n_{ij}$

Stochastic models are also used for analysing data concerned with a flow of events in time. In the present study, a four state Markov chain model is used. The four states of the model is identified based on the quartiles of the time series distribution of yield. The four states of the model are given by $< Q_1$, Q_1-Q_2, Q_2-Q_3 and $>Q_3$, where Q_1, Q_2 and Q_3 are the quartiles. n_{ij} denote frequency by which the system moves from state i to state j, i=1, 2, 3, 4 ; j= 1, 2, 3, 4. The estimates of the transition probabilities are given by $p_{ij} = n_{ij}/n_{i.}$, where $n_{i.} = \Sigma\, n_{ij}$. To check the appropriatness of a Markov chain model, it is required to test the independence of the four states. This is done using χ^2 test statistic as follows.

$H_0 : P_{ij} = \Pi_j$, j= 1, 2, 3, 4 where Π_j is the unconditional probability of being in state j.

The test statistic $\chi^2 = \sum_i \sum_j \dfrac{\{n_{ij} - n_{i.}\, n_{.j}/n\}^2}{n_{i.}\, n_{.j}/n}$ follows chi square distribution with 9 degrees of freedom. To make a decision about H_0, calculated value of chi square is compared with the table value at the required level of significance. If the calculated value is greater than table value, H_0 is rejected, which means that the states are not independent. This implies that the data satisfies the basic criterian of a Markov chain model.

The transition probability matrix is given by $P = [p_{ij}]_{4x4}$

It is known that after a sufficiently long period of time, the system settles down to a condition of statistical equilibrium at which state occupation probabilities become independent of the initial condition. The steady state probabilities are obtained by powering the transition matrix P. The system was reached equilibrium after n steps if $p^n = p^{n+1}$ and the steady state probabilities are given by the elements of p^n. The steady state probabilities are used to forecast the yield. This is done as follows. Forecast = $\hat{\Pi}_1 Q_1^* + \hat{\Pi}_2 Q_2^* + \hat{\Pi}_3 Q_3^* + \hat{\Pi}_4 Q_4^*$ ----(11), where $\hat{\Pi}_j$ is the steady state probability of state j, j = 1, 2, 3, 4 and Q_1^*, Q_2^*, Q_3^* and Q_4^* are the midvalues of the range which are represented by the four states of the Markov chain model.

Results and Discussion

4. RESULTS AND DISCUSSION

The results obtained from the analysis of the time series data gathered in the study are presented in the appended tables and discussed under the following headings

1. Box-Jenkins model
2. Distributed lag model
3. Log normal diffusion model
4. Markov chain model

4. 1 Box-Jenkins model

4.1.1 Production

The data on production of cashew is given in appendix II. Over the years the series showed a trend pattern as illustrated in figure 1. The cashew production in the state showed a steady increase during the period from 1962 to 1975, afterwards showed declining trend. Box-Jenkins methodology of model identification, estimation and diagnostic checking was applied to the series of cashew production in Kerala.

Differencing is carried out thrice in order to attain stationarity. Hence the value of d is identified as three. In order to choose the appropriate values for the orders of AR and MA, the plot of the autocorrelations and partial auto correlations of the transformed series is used. Figure 4.1.1(a) and 4.1.1(b) respectively shows the plots of autocorrelation(AC) and partial autocorrelation(PAC) of the differenced series. The graphs revealed a mixed process containing a second order autoregressive component and a second order moving average component. Thus the ARIMA model (2, 3, 2) is identified for the production series.

Figure 1.. Production of cashew in Kerala

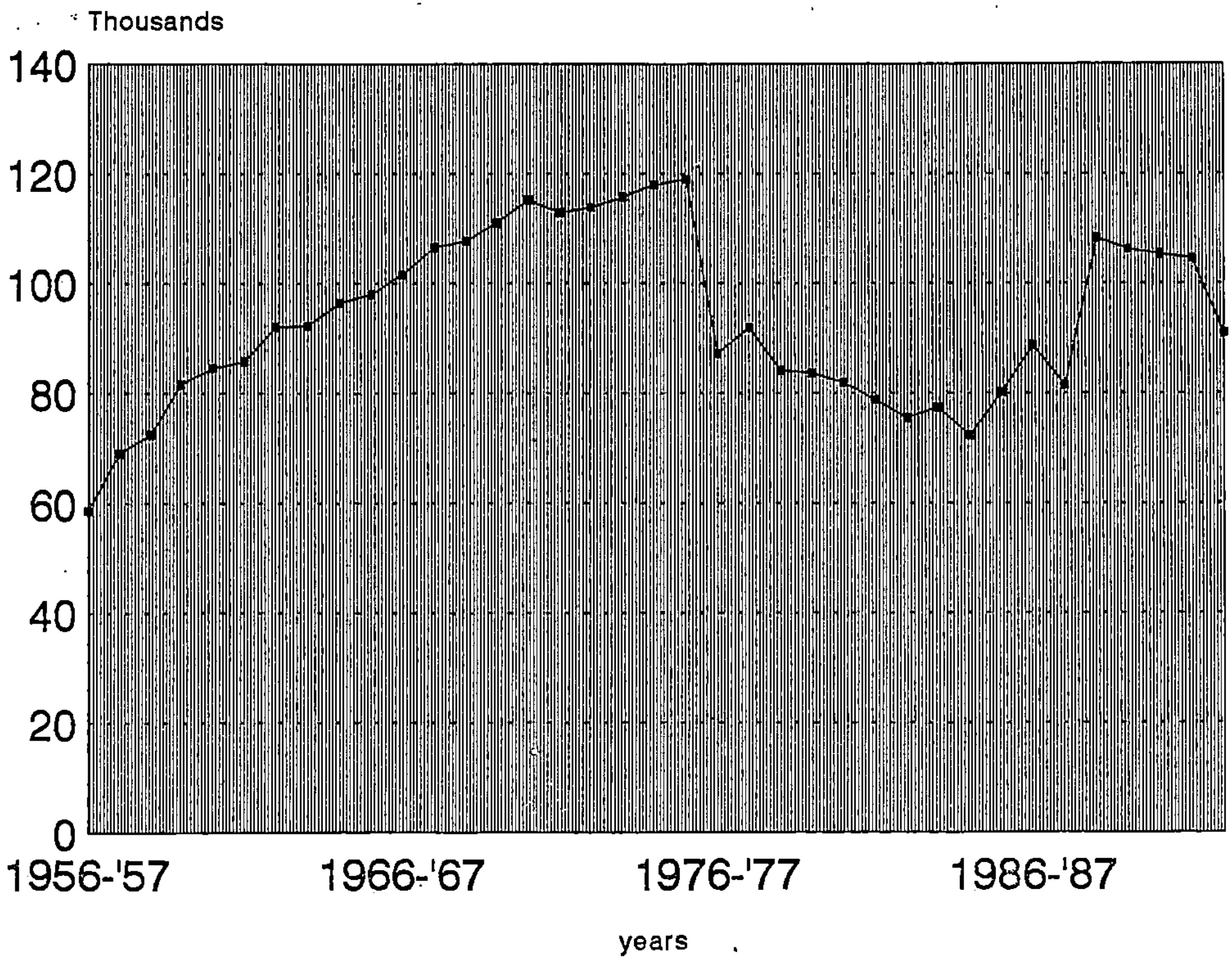

Figure 4.1.1(a). Auto correlation plots for the stationary series of Production.

```
  ++++++++++++++++++++++++++++++++++++++++++++++++++
 0+                C       +      B                A
 1+     A          C       +      B                +
 2+                C       +     AB                +
 3+                C       +A     B                +
 4+                C  A    +      B                +
 5+                C       +  A   B                +
 6+                C      A+      B                +
 7+                C    A  +      B                +
 8+                C       +  A   B                +
 9+                C      A+      B                +
10+                  A     +                       +
 ++++++++++++++++++++++++++++++++++++++++++++++++++
```

Figure 4.1.1(b). Partial auto correlation plots for the series of Production

```
 ++++++++++++++++++++++++++++++++++++++++++++++++++
 0+                 C       +      B               A
 1+     A           C       +      B               +
 2+                 C       +     AB               +
 3+                 C       +A     B               +
 4+                 C  A    +      B               +
 5+                 C       +   A  B               +
 6+                 C      A+      B               +
 7+                 C    A  +      B               +
 8+                 C       +   A  B               +
 9+                 C      A+      B               +
10+                   A     +                      +
 ++++++++++++++++++++++++++++++++++++++++++++++++++
```

```
OBSERVATIONS              35     DEGREES OF FREEDOM      30
R**2               .68204425     RBAR**2           .63965015
SSR           .22366610E+10     SEE            8634.5450
DURBIN-WATSON   1.92187427
Q(  15)=    13.9735        SIGNIFICANCE LEVEL  .527543
```

The ARIMA model (2, 3, 2) identified above have been estimated through the RATS package and the results are provided in table 4.1.1(a). The coefficient of determination of the model was 0.64, which was fairly good. The Durbin-Watson statistic d = 1.92 and this rules out the presence of serial correlation among the residuals. The estimated model is given by the equation $\hat{Y}_t = -6364387 + 1.449Y_{t-1} - 0.44Y_{t-2} + e_t - 0.6069e_{t-1} + 0.3226e_{t-2}$

----------(4.1)

where $\hat{Y}_t$ = production of cashew in year \`t'

Y_{t-1} = production of cashew in year \`t-1'

Y_{t-2} = production of cashew in year \`t-2'

e_t = error component in year \`t'

e_{t-1} = error component in year \`t-1'

e_{t-2} = error component in year \`t-2'

The model verifications is concerned with checking the residuals of the model to see if they contain any systematic pattern which can still be removed to improve the chosen ARIMA model. This is done through examinig the AC with different lags of the residuals. Box-Pierce Q test is used to see if a number of autocorrelations together are significantly diferent from zero. Q follows a chi-square distribution with fifteen degrees of freedom. The calculated value of Q is 13.97 for lags 15. This is compared with the theoretical value of chi-square for fifteen degrees of freedm and is found to be insignificant. This indicates that the group of autocorrelations is insignificant and thus confirms that the selected ARIMA model is an appropriate model.

Table 4.1.1(a). Estimates of the Coefficients of ARIMA model (2, 3, 2) of Production.

NO.	LABEL	VAR	LAG	COEFFICIENT	STAND. ERROR	T-STATISTIC
1	CONSTANT	1	0	-6364387.	.1341243E+08	-.4745143
2	AR	2	1	1.449847	.3888364	3.728681
3	AR	3	2	-.4498223	.3889367	-1.156544
4	MA	4	1	-.6069089	.3856856	-1.573585
5	MA	5	2	.3222556	.1860122	1.732443

Table 4.1.1(b). Actual and estimated values of cashew production

Year	Actual	Estimated
1976	87260	93661.0
1977	91930	93949.5
1978	84190	94238.1
1979	83700	94526.7
1980	81900	94815.3
1981	78898	95103.9
1982	75495	95392.5
1983	77375	95681.1
1984	72294	95969.8
1985	80203	96258.4
1986	88710	96547.1
1987	81481	96835.8
1988	108264	97124.5
1989	106258	97413.2
1990	105369	97701.9
1991	104601	97990.6
1992	90979	98279.4

Table 4.1.1(c). Post sample forecasts for cashew production

Year	Forecast (Y)
1997	99723.3
1998	100012.0
1999	100301.0
2000	100590.0
2001	100879.0
2002	101168.0
2003	101456.0
2004	101745.0
2005	102034.0

ARIMA models are developed basically to forecast the corresponding variable. There are two kinds of forecasts, sample period and post sample period forecasts. The former are used to develop confidence in the model and the latter to generate genuine forecasts for use in planning and other purposes. The ARIMA model can be used to yield both these kinds of forecasts. The sample period forecasts are obtained simply by substituting the actual values of the explanatory variables in the estimated equation(4.1) and are presented in the table 4.1.1(b). To judge the forecasting ability of the fitted ARIMA model, the Mean Absolute Percentage Error (MAPE) is calculated (Makridakis and Wheelwright, 1978). It turned out to be 13.16% indicating that the forecasting inaccuracy is low. Post sample forecasts for nine years from 1997 to 2005 are given in table 4.1.1(c).

4.1.2 Area

The time series data on area under cashew is given in appendix II. Trends in area under cashew is given in figure 2. It shows that the area under the crop increased rapidly from 1975-'76 tp 1983-'84 and declined thereafter.

In order to make it stationary the series is differenced once. Hence value for `d' is one. The AC and PACs of the differenced series are found out and are given in figure 4.1.2(a) and 4.1.2(b) respectively. Figure 4.1.2(a) reveals moving average component of order 2 and figure 4.1.2(b) reveals a autoregressive component of order 2. Thus the appropriate model was identified as ARIMA (2, 1, 2).

Figure 2. Area under Cashew

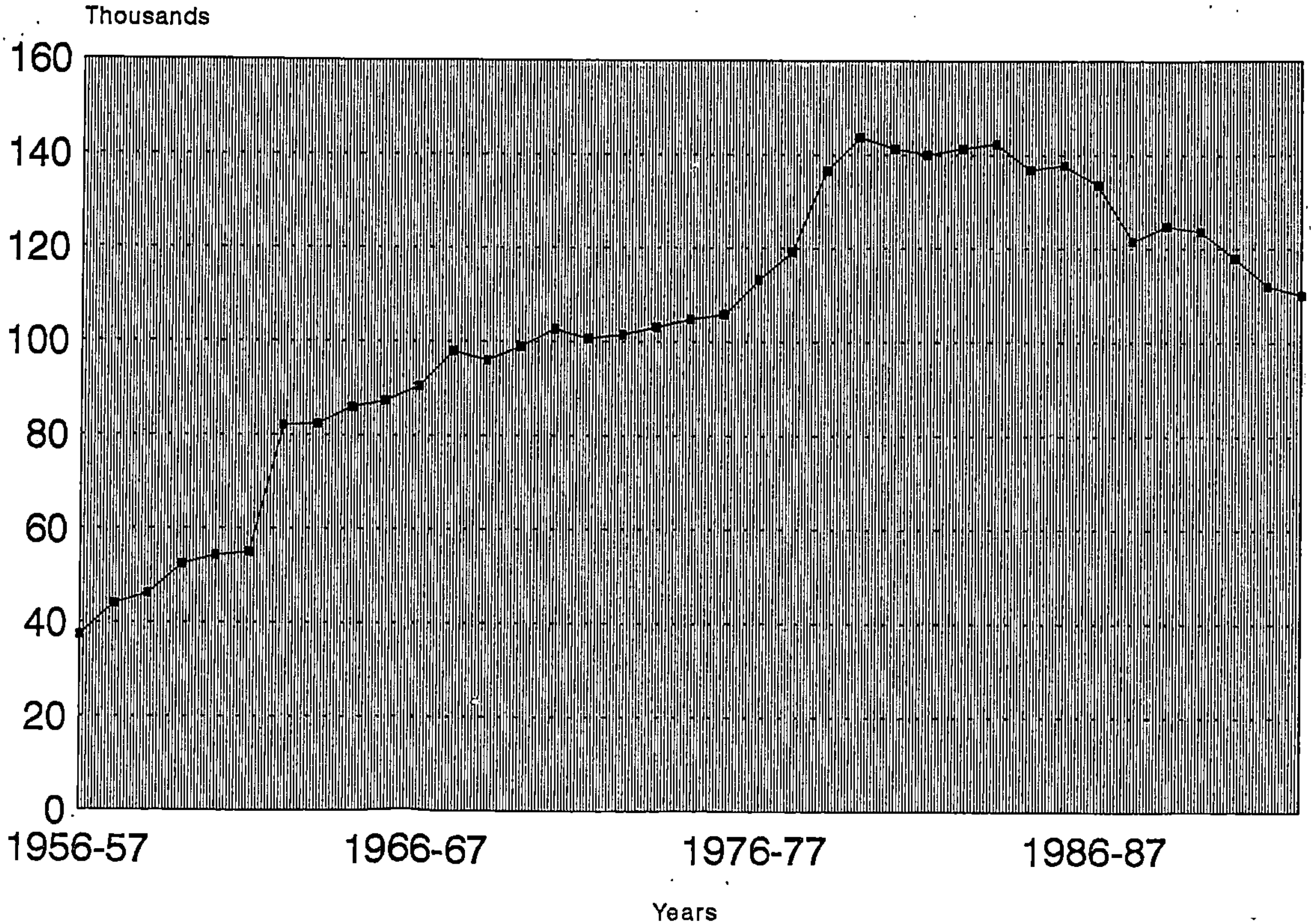

Figure 4.1.2(a). AC plots for the stationary series of area

```
   ++++++++++++++++++++++++++++++++++++++++++++++++++++
 0+                 C       +      B                 A
 1+                 C       A      B                 +
 2+                 C       +A     B                 +
 3+                 C       A      B                 +
 4+                 C       A      B                 +
 5+                 C       + A    B                 +
 6+                AC       +      B                 +
 7+                 C       A      B                 +
 8+                 C     A +      B                 +
 9+                 C     A +      B                 +
10+                        A+                        +
   ++++++++++++++++++++++++++++++++++++++++++++++++++++
```

Figure 4.1.2(b) PAC plots for the stationary series area

```
   +++++++++++++++++++++++++++++++++++++++++++++++++++
 0+                 C       +      B               A
 1+                 C       A      B               +
 2+                 C       +A     B               +
 3+                 C       A      B               +
 4+                 C       A      B               +
 5+                 C       +  A   B               +
 6+                AC       +      B               +
 7+                 C       A      B               +
 8+                 C     A +      B               +
 9+                 C     A +      B               +
10+                        A+                      +
   +++++++++++++++++++++++++++++++++++++++++++++++++++
```

```
OBSERVATIONS              35     DEGREES OF FREEDOM     30
R**2               .95953222     RBAR**2           .95413651
SSR           .10299931E+10     SEE           5859.4455
DURBIN-WATSON  1.73179602
Q(  15)=    6.62696         SIGNIFICANCE LEVEL  .967187
```

The ARIMA model (2, 1, 2) identified above have been estimated and the results are given in table 4.1.2(a). $R^2 = 0.95$ indicating a good fit. The Durbin-Watson statistic indicated absence of autocorrelation among the error terms. The coefficients obtained from the analysis are given in table 4.1.2(a). The estimated ARIMA model for area under cashew in Kerala is given by $\hat{A}_t = 120331.4 - 0.0764A_{t-1} + 0.888A_{t-2} + e_t + 1.22e_{t-1} + 0.196e_{t-2}$ ---------- (4.2)

where $\hat{A}_t$ = Area under the cashew at time 't'

A_{t-1} = Area under the cashew at time 't-1 '

A_{t-2} = Area under the cashew at time 't-2'

e_t = Random error component at time 't'

e_{t-1} = Random error component at time 't-1 '

e_{t-2} = Random error component at time 't-2'

To verify the appropriatness of the model AC and PAC coefficients of the residual terms for various lags were considered. Box-Pierce Q statistic is calculated upto fifteen lags. This value is found to be insignificant on comparison with the theoretical value of chi-square. Hence the test indicates that the autocorrelation among the residuals is absent. This proves that the selected ARIMA model is an appropriate one.

The estimated model is used for sample and post sample period forecasts which are given in table 4.1.2(b) and 4.1.2(c) respectively. The forecasting ability of the fitted model is judged by the Mean Absolute Percentage Error (MAPE) and it turns out to be 15.37% indicating that the forecasting inaccuracy is low.

Table 4.1.2(a). Estimates of the coefficients in ARIMA model (2, 1, 2) of area

NO.	LABEL	VAR	LAG	COEFFICIENT	STAND. ERROR	T-STATISTIC
1	CONSTANT	1	0	120331.4	13441.00	8.952565
2	AR	2	1	-.7639834E-01	.1433014	-.5331303
3	AR	3	2	.8878266	.1578044	5.626119
4	MA	4	1	1.224742	.2622093	4.670858
5	MA	5	2	.1959146E-01	.2492812	.7859180E-01

Table 4.1.2(b). Actual and estimated values of area under cashew (A

Year	Actual	Estimated
1976	113329.	108289.
1977	119310.	106022.
1978	136550.	110733.
1979	143700.	108361.
1980	141277.	112724.
1981	139960.	110285.
1982	141307.	114345.
1983	142339.	111869.
1984	136863.	115663.
1985	137747.	113175.
1986	133562.	116734.
1987	121550.	114253.
1988	124740.	117602.
1989	123661.	115143.
1990	118036.	118304.
1991	112059.	115880.
1992	110168.	118872.

Table 4.1.2(c). Post sample period forecsast for area under cashew

Year	Area Forecast
1997	117421
1998	119990
1999	117773
2000	120223
2001	118068
2002	120408
2003	118316
2004	120554
2005	118525

4. 1. 3 Productivity

The productivity showed a declining trend in the late seventies and early eighties (Figure. 3).

Box-Jenkins method of fitting an ARIMA model was applied to the series. An ARIMA model (2, 1, 2) was found to provide the most satisfactory fit to the data. The model is $\hat{P}_t = 873.881 + 1.60P_{t-1} - 0.708P_{t-2} + e_t - 0.8e_{t-1} + 0.149e_{t-2}$ -------(4.3)

where

$\hat{P}_t$ = Productivity of cashew at time `t'

P_{t-1}= Productivity of cashew at time `t-1'

P_{t-2}= Productivity of cashew at time `t-2'

e_t = Error component at time `t'

e_{t-1}= Error component at time `t-1'

e_{t-2}= Error component at time `t-2'

The relevant parameter estimates are given in table 4.1.3(a). A plot of AC's and PAC's are shown in figure 4.1.3(a) and 4.1.3(b) respectively. The coefficient of determination of the model is 0.89, indicating a good fit to the data. The Durbin- Watson statistic, d = 2. This rules out the presence of autocorrelation problem. The statistic Q is calculated to be 15.87. This is found to be insignificant on comparison with the theoretical value. Hence the model is acceptable.

The estimated model ARIMA (2, 1, 2) is used for sample and post sample period forecasts. They are given in table 4.1.3(b) and 4.1.3(c) respectively. For measuring the

Figure 3. Productivity of cashew

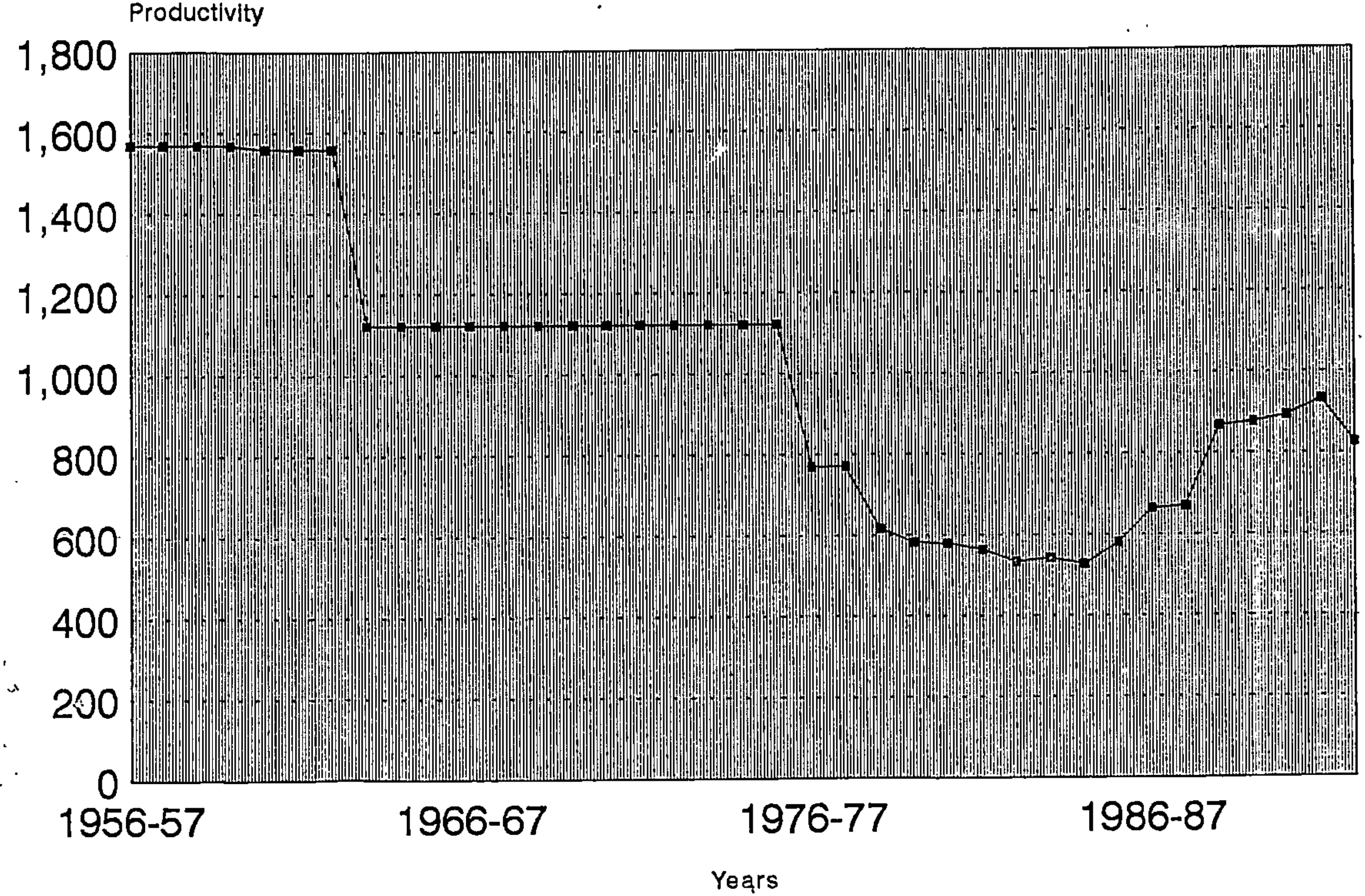

Figure 4.1.3(a). AC plots for the stationary series, productivity

```
   +++++++++++++++++++++++++++++++++++++++++++++++++++
  0+                C       +      B                 A
  1+                C       A      B                 +
  2+                C       +A     B                 +
  3+                C       A      B                 +
  4+                C       A      B                 +
  5+                C       + A    B                 +
  6+               AC       +      B                 +
  7+                C       A      B                 +
  8+                C     A +      B                 +
  9+                C     A +      B                 +
 10+                       A+                        +
   +++++++++++++++++++++++++++++++++++++++++++++++++++
```

Figure 4.1.3(b). PAC plots for the stationary series, productivity.

```
   ++++++++++++++++++++++++++++++++++++++++++++++++++
  0+                C       +      B                A
  1+                C       A      B                +
  2+                C       +A     B                +
  3+                C       A      B                +
  4+                C       A      B                +
  5+                C       + A    B                +
  6+               AC       +      B                +
  7+                C       A      B                +
  8+                C     A +      B                +
  9+                C     A +      B                +
 10+                       A+                       +
   ++++++++++++++++++++++++++++++++++++++++++++++++++
```

```
OBSERVATIONS              35     DEGREES OF FREEDOM      30
R**2               .90375222     RBAR**2          .89047647
SSR                319600.59     SEE          104.97957
DURBIN-WATSON     2.01160459
Q(  15)=  15.8789             SIGNIFICANCE LEVEL  0.390142
```

Table 4.1.3(a). Estimates of the coefficients in ARIMA model (2, 1, 2) of Productivity.

NO.	LABEL	VAR	LAG	COEFFICIENT	STAND. ERROR	T-STATISTIC
1	CONSTANT	1	0	873.8881	161.9271	5.3967
2	AR	2	1	1.660115	.359312	4.6204
3	AR	3	2	-.7088394	.322812	-2.1958
4	MA	4	1	-.8001709	.402645	-1.9872
5	MA	5	2	.1492575	0.19566	.762827

Table 4.1.3(b). Actual and estimated values of productivity

YEAR	ACTUAL	ESTIMATED
1976	770.000	844.024
1977	771.000	844.026
1978	617.000	845.482
1979	582.000	847.898
1980	580.000	850.877
1981	564.000	854.110
1982	534.000	857.365
1983	543.000	860.478
1984	528.000	863.337
1985	582.000	865.879
1986	664.000	868.070
1987	670.000	869.907
1988	868.000	871.403
1989	877.000	872.585
1990	893.000	873.486
1991	933.000	874.144

Table 4.1.3(C). Forecasts for productivity

YEAR	FORECAST
1997	875.001
1998	874.896
1999	874.772
2000	874.641
2001	874.511
2002	874.389
2003	874.278
2004	874.181
2005	874.097

forecasting ability of the fitted model, MAPE is used and it turns out to be 21.5 per cent indicating that the forecasting inaccuracy is low.

4. 1. 4 Price

During the sample period of the study, price in all periods witnessed an increasing trend. The variable increased almost monotonically over time (Appendix III). To make the series stationary, differencig is done three times. Therefore the value for d=3. Identification of the appropriate orders for AR and MA are attained through an examination of the autocorrelation and partial autocorrelation of the stationary series. Examination of figures 4.1.4(a) and 4.1.4(b) revealed a mixed autoregressive moving average process of order (4, 4). Hence the identified model for the series is ARIMA (4, 3, 4).

The parameter estimates of the model are given in table 4.1.4(a). The Durbin-Watson statistic indicated absence of serial correlation among the residual terms. The value of R^2 = 0.94, indicating a good fit to the data. The fitted model is given by the following equation.

$$\hat{F}_t = 4.87 + 0.498F_{t-1} - 0.477F_{t-2} + 0.4967F_{t-3} + 0.86F_{t-4} + e_t - 0.23e_{t-1} + 0.67e_{t-2} - 0.5e_{t-3} - 1.79e_{t-4} \qquad \text{-------(4.4)}$$

where $\hat{F}_t$ = Farm harvest price of cashew at time `t'

F_{t-1}= Farm harvest price of cashew at time `t-1'

F_{t-2}= Farm harvest price of cashew at time `t-2'

F_{t-3}= Farm harvest price of cashew at time `t-3'

F_{t-4}= Farm harvest price of cashew at time `t-4'

e_t = Error component at time `t'

e_{t-1}= Error component at time `t-1'

Figure 4.1.4(a). AC plots for the stationary series of price

```
  ++++++++++++++++++++++++++++++++++++++++++++++++++
 0+                C       +      B                A
 1+            A   C       +      B                +
 2+                C   A   +      B                +
 3+                C       +      B A              +
 4+              A C       +      B                +
 5+                C       +  A   B                +
 6+                C       A      B                +
 7+                C      A+      B                +
 8+                C       +A     B                +
 9+                C     A +      B                +
10+                        +A                      +
  ++++++++++++++++++++++++++++++++++++++++++++++++++
```

Figure 4.1.4(b). PAC plots for the stationary series price

```
   ++++++++++++++++++++++++++++++++++++++++++++++++++
  0+                C       +      B                A
  1+            A   C       +      B                +
  2+                C   A   +      B                +
  3+                C       +      B A              +
  4+              A C       +      B                +
  5+                C       +  A   B                +
  6+                C       A      B                +
  7+                C      A+      B                +
  8+                C       +A     B                +
  9+                C     A +      B                +
 10+                        +A                      +
   ++++++++++++++++++++++++++++++++++++++++++++++++++
```

```
R**2                .95468008      RBAR**2          .93957344
SSR             444781.68          SEE          136.13438
DURBIN-WATSON   1.76883633
Q(  15)=    12.3111        SIGNIFICANCE LEVEL   .655347
```

Table 4.1.4(a). Estimates of the coefficients in the ARIMA model (4, 3, 4) for Price

NO.	LABEL	VAR	LAG	COEFFICIENT	STAND. ERROR	T-STATISTI
1	CONSTANT	1	0	4.871749	33.56473	.1451449
2	AR	2	1	.4986933	.4650170	1.072419
3	AR	3	2	-.4770514	.8390071	-.5685904
4	AR	4	3	.4976776	.9105289	.5465808
5	AR	5	4	.8677333	.6427959	1.349936
6	MA	6	1	-.2334169	.4706674	-.4959274
7	MA	7	2	.6785413	.7247735	.9362115
8	MA	8	3	-.5062671	1.153847	-.4387646
9	MA	9	4	-1.795676	1.031453	-1.740919

e_{t-2}= Error component at time `t-2'

e_{t-3}= Error component at time `t-3'

e_{t-4}= Error component at time `t-4'

To verify the appropriateness of the model AC and PAC coefficients of the residual terms for various lags were examined. Box-Pierce Q test indicated that the group of auto correlations among the residuals is insignificant.

The forecasting ability of the model was tested by examining the accuracy of the within sample period forecasts. This was attempted on the basis of Mean Absolute Percentage Error(MAPE), the measure of the average absolute error in percentage. It is turned out to be 19.37 per cent and establishes the credibility of the model for forecasting the series. Sample and post sample period forecast for the variable are given in table 4.1.4(b) and 4.1.4(c) repectively.

4. 1. 5 Rainfall

Univariate ARIMA model is ftted to the annual rainfall data in Kerala. The data is given in appendix IV. To make the series stationary differencing is carried out twice. Therefore the value of d = 2. In order to choose appropriate values for the orders of AR and MA, the plot of AC and PAC's (Fig 4.1.5(a) & 4.1.5(b)) of the differenced series are found out. It is revealed that the differenced series is a mixed ARMA (5, 5) process. Thus the identified model of the rainfall series is ARIMA (5, 2, 5).

The ARIMA model(5, 2, 5) have been estimated and the results are provided in table 4.1.5(a). The coefficient of determination was found to be 0.38 only. The estimate of

Table 4.1.4(b). Actual and estimated values of price of cashew

Year	Actual	Estimated
1976	503.100	370.082
1977	535.300	389.586
1978	407.100	390.708
1979	582.700	454.949
1980	731.900	553.627
1981	655.530	589.674
1982	479.790	593.521
1983	869.890	683.096
1984	843.340	829.499
1985	1059.50	892.970
1986	1348.10	902.698
1987	1094.90	1027.86
1988	1154.70	1244.26
1989	1177.50	1352.39
1990	1379.90	1373.81
1991	2037.70	1549.21
1992	1961.50	1868.06

Table 4.1.4(c). Forecasts for price of cashew

Year	Forecast
1997	3100.17
1998	3183.59
1999	3532.44
2000	4219.25
2001	4691.59
2002	4845.50
2003	5341.44
2004	6346.38
2005	7097.42

Figure 4.1.5(a). AC plots for the stationary series rain

```
 +++++++++++++++++++++++++++++++++++++++++++++++++
 0+               C        +       B               A
 1+         A     C        +       B               +
 2+               C       A+       B               +
 3+               C        + A     B               +
 4+               C        +     A B               +
 5+             A C        +       B               +
 6+               C        +A      B               +
 7+               C        +       BA              +
 8+              AC        +       B               +
 9+               C        +A      B               +
10+                        +  A                    +
  ++++++++++++++++++++++++++++++++++++++++++++++++++
```

Figure 4.1.5(b). PAC plots for the stationary series of rain

```
  ++++++++++++++++++++++++++++++++++++++++++++++++++
 0+               C        +       B               A
 1+         A     C        +       B               +
 2+               C       A+       B               +
 3+               C        + A     B               +
 4+               C        +     A B               +
 5+             A C        +       B               +
 6+               C        +A      B               +
 7+               C        +       BA              +
 8+              AC        +       B               +
 9+               C        +A      B               +
10+                        +  A                    +
  ++++++++++++++++++++++++++++++++++++++++++++++++++
```

```
R**2               .58644562     RBAR**2          .37966844
SSR            2477790.4         SEE          351.97943
DURBIN-WATSON  1.96030399
Q(  15)=    9.11269        SIGNIFICANCE LEVEL  .871552
```

parameters are also found insignificant. The Durbin-Watson statistic indicated absence of autocorrelation among the residual terms. The estimated ARIMA model for rainfall in Kerala is $\hat{R}_t = 2703.627 + 0.409R_{t-1} + 0.109R_{t-2} + 0.575R_{t-3} + 0.126R_{t-4} - 0.535R_{t-5} + e_t + 0.386e_{t-1}$

$- 0.206e_{t-2} - 0.849e_{t-3} - 0.43e_{t-4} - 0.295e_{t-5}$ ________ (4.5)

where $\hat{R}_t$ = Estimated annual rainfall in Kerala at time 't'

R_{t-1} = Annual rainfall in Kerala at time 't-1'

R_{t-2} = Annual rainfall in Kerala at time 't-2'

R_{t-3} = Annual rainfall in Kerala at time 't-3'

R_{t-4} = Annual rainfall in Kerala at time 't-4'

R_{t-5} = Annual rainfall in Kerala at time 't-5'

e_t = Error component at time 't'

e_{t-1} = Error component at time 't-1'

e_{t-2} = Error component at time 't-2'

e_{t-3} = Error component at time 't-3'

e_{t-4} = Error component at time 't-4'

e_{t-5} = Error component at time 't-5'

Presence of systematic pattern among the residuals is checked by the Q statistic. The calculated value of Q is 9.11, which is insignificant on comparison with the theoretical value. Hence the model is acceptable.

The sample period and post sample period forecasts are calculated and are given in table 4.1.5(b) and 4.1.5(c) respectively. To judge the forecasting ability MAPE is calculated and the value of which is 13.16 per cent indicating that the forecasts are quite good.

Table 4.1.5(a). Estimates of the coefficients in the ARIMA model (5, 2, 5) for the rainfall

NO.	LABEL	VAR	LAG	COEFFICIENT	STAND. ERROR	T-STATISTIC
1	CONSTANT	1	0	2703.627	267.2377	10.11694
2	AR	2	1	.4090693	.4111331	.9949802
3	AR	3	2	.1090354	.3702182	.2945166
4	AR	4	3	.5757322	.2929397	1.965361
5	AR	5	4	.1260649	.4235386	.2976468
6	AR	6	5	-.5354515	.3655366	-1.464837
7	MA	7	1	-.3869156	.5969234	-.6481829
8	MA	8	2	-.2067308	.5404564	-.3825115
9	MA	9	3	-.8495492	.5278872	-1.609339
10	MA	10	4	-.4307208	.4961710	-.8680895
11	MA	11	5	-.2958896	.4290278	-.6896746

Table 4.1.5(b). Actual and estimated values of rainfall in Keral

Year	Actual	Estimated
1976	2050.30	2623.76
1977	2913.10	2639.68
1978	3209.60	2638.57
1979	3082.80	2687.13
1980	2668.50	2717.38
1981	3037.60	2704.70
1982	2171.60	2722.11
1983	2214.70	2751.98
1984	2349.50	2736.61
1985	2460.80	2725.80
1986	2091.40	2745.89
1987	2237.40	2738.52
1988	2653.00	2713.54
1989	2642.00	2720.95
1990	2780.00	2725.34
1991	3106.00	2701.88

Table 4.1.5(c). Forecasts for the rainfall in Kerala

YEAR	FORECAST
1997	2701.33
1998	2688.42
1999	2696.06
2000	2706.06
2001	2696.50
2002	2695.93
2003	2708.29
2004	2704.95
2005	2698.05

4. 1. 6 Transfer function model

An attempt is made to develop a multivariate ARIMA model. This is done by combinig the concepts of multiple regression with those of univariate ARIMA models. The independent variables identified for the purpose are area, price and annual rainfall.

ARIMA model was identified for each input variable and the adequacy of each was cofirmed. Next, the white noise residuals of the input series is cross correlated with the filtered residuals of the output series. The cross correlation analysis is presented in tables 4.1.6(a), 4.1.6(b) and in 4.1.6(c). The tables are not in support of any input-output relationship as the cross correlations with all the input variables were not found to be significant for lags less than nine. Hence it is concluded that transfer function modelling is not possible in this case.

4. 2 Distributed lag model

4. 2. 1 Area response analysis

With a view to estimating the response of producers in terms of area towards price and non-price factors, the actual area in the current year was expressed as a function of area under the crop lagged by one year (t-1), farm harvest price lagged by one year (t-1), price risk and weather represented by annal rainfall lagged by one year. Area response models were fitted for the whole period and two sub periods, viz. period I (from 1956-'57 to 1975-'76) and period II (from 1976-'77 to 1992-'93). The estimated area response functions for cashew at the state level are presented in table 4.2.1(a).

Table 4.1.6(a) Croos correlation analysis with production and area under cashew

SMPL 1956 - 1991
36 Observations

COR{PRODN, AREA (- i)}	COR{PRODN, AREA (- i)}	i	lag	lead
. |** .	. |** .	0	0.125	0.125
. | .	. |* .	1	-0.000	0.097
. *| .	. |* .	2	-0.074	0.095
. **| .	. |** .	3	-0.135	0.121
. **| .	. |** .	4	-0.169	0.151
. **| .	. |** .	5	-0.185	0.185
.***| .	. |** .	6	-0.218	0.191
.***| .	. |***.	7	-0.226	0.249
.***| .	. |****	8	-0.248	0.305
****| .	. |****	9	-0.272	0.341
****| .	. |*****	10	-0.285	0.365
****| .	. |*****	11	-0.313	0.377
****| .	. |*****	12	-0.327	0.370
*****| .	. |*****	13	-0.345	0.354
****| .	. |****	14	-0.349	0.333
	. |****	15	-0.273	0.383

S.E. of Correlations 0.166

Table 4.1.6 (b) : Cross correlation analysis with production and price of Cashew

SMPL 1956 - 1991
36 Observations

COR(PRODN,PRICE(-i))	COR(PRODN,PRICE(+i))	i	lag	lead
. | .	. | .	0	-0.001	-0.001
. *| .	. *| .	1	-0.030	-0.044
. *| .	. *| .	2	-0.089	-0.073
. **| .	. **| .	3	-0.118	-0.126
. **| .	.***| .	4	-0.148	-0.241
. **| .	. **| .	5	-0.145	-0.175
. **| .	. **| .	6	-0.181	-0.171
. **| .	. **| .	7	-0.188	-0.1[illegible]1
. **| .	. *| .	8	-0.175	-0.071
. **| .	. | .	9	-0.157	0.005
. **| .	. |* .	10	-0.146	0.102
. **| .	. |***.	11	-0.129	0.219
. *| .	. |****	12	-0.078	0.287
. *| .	. |*****	13	-0.004	0.355
. *| .	. |******	14	-0.056	0.427
. | .	. |******	15	-0.022	0.460

S.E. of Correlations 0.166

Table 4.1.6 (c) : Cross correlation analysis with production and annual rainfall series

BMPL 1956 - 1991
36 Observations

COR(PRODN,RAIN(-i))	COR(PRODN,RAIN(+i))	i	lag	lead
¦	¦	0	0.030	0.030
*¦	*¦	1	-0.070	-0.055
*¦	*¦	2	-0.043	-0.056
*¦	*¦	3	-0.096	-0.062
*¦	¦	4	-0.071	-0.001
*¦	¦	5	-0.081	-0.019
*¦	¦**	6	-0.045	0.120
¦*	¦**	7	0.082	0.121
¦*	¦	8	0.080	0.027
¦**	¦	9	0.168	0.000
¦***	**¦	10	0.243	-0.146
¦*****	***¦	11	0.356	-0.229
¦***	****¦	12	0.264	-0.287
¦*****	***¦	13	0.355	-0.225
¦****	***¦	14	0.276	-0.247
¦**	***¦	15	0.128	-0.253

S.E. of Correlations 0.166

In the model(Ia) (Area response model for the whole period) the independent variable could together explain 96.9 per cent of the variation in area. The Durbin `d' statistic indicated absence of serial correlation among the error terms. Among the parameter estimates lagged area, price and rainfall were significant. Price risk was found to be nonsignificant. The adjustment coefficient worked out to be 0.0288 indicating that only 2 per cent of the desired acreage changes could be affected in one year. Coefficient of lagged price was negative while that of annual rainfall lagged by one year and price risk were positive. The results indicate that weather conditions had a positive influence on acreage allocation. The estimated short run and long run elasticities (Table 4.2.1(b)) were -0.0438 and -1.521 repectively indicating very low negative response. Similar results were obtatined by Prabhakaran (1987) in his study on cashew.

In model(IIa) (Area response model for the sub-period I), the percent variation attributed to regression was significant at 5 per cent level indicating that the four independent variables included in the function could explain the variation in the dependent variable significantly to the extent of 93.9 per cent. The Durbin statistic indicated presence of serial correlation among the error terms. However, price coefficient continued to be negative and it became non significant. But the coefficient of annual rainfall continued to be positive and significant. Price risk was positively related to area, but it continued to be non significant. The estimated elasticities were -0.0023 and -0.0547 respectively for the short run and long run. As regards the overall explanatory power, this model was on par with the Ia model.

Table 4.2.1(a). Area response functions for cashew in Kerala

Models	Period	constant term	A_{t-1}	P_{t-1}	R_t	W_{t-1}	R^2
I a	1956-1992	-12482.9	0.9712**	-9.62*	39.00	5.933**	0.97
II a	1956-1975	-13709.5	0.9587**	-1.58	4.311	6.850*	0.94
III a	1976-1992	31085.98	0.7622**	-13.9**	33.48	3.066	0.87

** Significant at 1 per cent level

* Significant at 5 per cent level

Table 4.2.1(b). Price elasticities of the Area response and Yield response functions

Models	Price elasticity			
	S.R	L.R	d	h
I a	-0.044	-1.521	2.28	-0.875
II a	-0.002	-0.055	2.78	-1.930
III a	-0.096	-0.405	1.95	0.096
I b	0.021	0.1067	2.22	-0.7567
II b	0.033	0.2751	2.56	-0.3694
III b	0.185	0.2879	2.10	-0.3578

When the model(IIIa) (Area response model for the sub-period II) was used, it continued to exhibit the negative price acreage relationship. Here also the coefficient of lagged area was significant. The regression coefficient of rainfall was positive and significant for the whole period and subperiod I. On the contrary, during period II this variable was found to be non-significant. However, coefficient of price risk continued to be positive and non-significant for the whole period and two subperiods. The explanatory power of model III a was 0.868, which was significant at 5 per cent level. The Durbin statistic indicated the absence of serial correlation among the error terms. The price coefficient was negative and significant at 5 per cent level. The short run and long run elasticities of price were -0.0962 and -0.4048 respectively.

4.2.2 Yield response Analysis

The analysis was done by fitting a yield response equation with actual yield in the current year as dependent variable and lagged (t-1) yield, lagged (t-1) farm harvest price, price risk and weather represented by annual rainfall lagged by one year as independent variable, in order to examine the response of cultivators, towards price and non price factors, reflected in yield. The estimated equation is given in table 4.2.2(a).

The coefficient of determination R^2 was found to be 0.769. The significance of R^2 value was indicative of the fact that the variables included in the yield response model for the whole period (Model I b) were capable of capturing sizeable proportion of the variation in the dependent variable. The coefficient of lagged yield was found to be significant at 5 per cent level. The price factor represented as farm harvest price lagged by one year got

Table 4.2.2(a). Yield response functions for cashew in Kerala

Models	Period	constant term	Y_{t-1}	P_{t-1}	R_t	W_{t-1}	R^2
I b	1956-1992	35690.97	0.8002**	4.22	-60.7*	-5.17	0.77
II b	1956-1975	12155.12	0.8795**	2.774	-87.3	0.083	0.97
III b	1976-1992	61844.79	0.3550*	18.41*	-88.1	-4.19*	0.49

** Significant at 1 per cent level

* Significant at 5 per cent level

significant influence on current yield. But the price risk negatively influenced the yield significantly.

The coefficient of yield adjustment was 0.1998 indicating that only 2 per cent of the desired acreage changes could be affected in one year. It would take nearly 13 years for 95 per cent of the effect of price to be realized.

The price elasticites were 0.0213 and 0.1067 in short run and long run respectively. It means that with every one rupee increase in farm harvest price, there was an increase of yield by 0.0213 and 0.1067 kg in short run and long run respectively. Thus the response of yield to price was positive but low. That means there was no much change in yield in response to prices. This result was also confirmed by the coefficient of yield adjustment which indicated that there was not much scope to increase the yield.

In the model(IIb) (Yield response model for sub-period I) the independent variables could together explain 97.2 per cent of the variation in the yield. The Durbin statistic indicated the presence of serial correlation among the error terms. Among the parameter estimates all of them except lagged yield were non significant. The adjustment coefficient worked out to be 0.1205 indicating that only 12 per cent of the desired changes could be affected in one year. At such slow rate of adjustment it would take nearly 23 years for 95 per cent of the effect of price to be realized. Coefficient of price risk was negative while that of annual rainfall lagged by one year and lagged price was positive. The estimated short run and long run elasticities of price (Table 4.2.1(b)) were 0.0332 and 0.2756 respectively indicating very low positive response.

When model (IIIb) (Yield response model for sub-period II) was used, about 49 per cent of the yield variation was explained by the determinants. In the case of rate of adjustment it would take 3 years for 95 per cent of the effect of price to be realized. Durbin statistic indicated absence of serial correlation among the error terms. Coefficient of lagged yield was significant at 5 per cent level suggesting that past yield level significantly influenced present yield.

The yield adjustment was 0.6444 indicating that 64 per cent of the desired yield changes could achieved in one year. The estimated coefficient of price risk was negative and non-significant for the two sub-periods. But it is significantly negative during the whole period of study. The regression coefficient for the annual rainfall was non-significant for the whole period and period I. But significantly negative for period II. Coefficient of price lagged by one year turned out to be positively significant. The price elasticities were 0.1856 and 0.2879 in short run and long run respectively. It means that with every one rupee increase in farm harvest price there was an increase of yield by 0.1856 and 0.2879 kg in short run and long run respectively.

The result of the analysis clearly indicated that area was not responsive to prices. Cashew growers are least sensitive to price movements and they prefer to grow the crop in all types of soil due to its wide adaptability and ease of management. The coefficient of determination of all functions were relatively high, indicated that the proposed model was satisfactory in describing acreage fluctuatons. None of the partial regression coefficient of acreage on lagged price were significant. In general it can be concluded that deflated price of the producct has no significant effect on the acreage allocation of any of the crops.

The result is in confirmity with the findings of Prabhakaran (1987) who proposed, cultivators of horticultural crops are less responsive to price fluctations.

4.3 LOG NORMAL DIFFUSION MODEL

The log normal model is adopted to explain the trends in production of cashew. The base period is taken as 1956-'57 and data from 1956-'57 to 1992-'93 have been utilized for the crop.

Using the log normal model, production can be estimated by the expected value. Expected value of Y_t is

$$Y_t = 58680 \exp \{0.04673\ t - 3.875 \times 10^{-7} \sum_{1}^{t} X_j \} \text{ ------(4.6)}$$

with $R^2 = 0.924$

The estimate of production using the fitted model (4.7) is given in table (4.3(a)) along with actual data. The coefficient of determination computed for the model reveals a good fit to the data. The estimate of X_t based on the exponential model for area have been used for predicting annual production of the crop. The geometric growth rate of crop area is 2.83. In general it was found that the model gave a satisfactory fit to the data and hence result was in confirmity with the findings of Saraswathy and Thomas (1976), who proposed stochastic models namely log normal model to explain the trend in production of crops in Kerala.

Table 4.3(a). Year wise area and production of cashew in Kerala and the expected trend in production using log normal diffusion model

Year	t	Area(X_t) in Ha.	Production (y) in tonnes	Estimate of y_t in tonnes
1956	0	37390	58680	60445.45
1957	1	44040	69100	62211.62
1958	2	46210	72510	63875.31
1959	3	52420	81670	65535.16
1960	4	54320	84630	67219.62
1961	5	55030	85800	68226.47
1962	6	82120	92040	69241.68
1963	7	82370	92310	70173.64
1964	8	85980	96460	71079.80
1965	9	87370	98030	71908.60
1966	10	90560	101610	72908.50
1967	11	97990	106580	72537.65
1968	12	96020	107730	73228.16
1969	13	98960	111030	73840.96
1970	14	102710	115240	74350.63
1971	15	100660	112940	74923.38
1972	16	101419	113880	75476.24
1973	17	103160	115750	75983.92
1974	18	104900	118000	76443.40
1975	19	105906	119000	76875.80
1976	20	113300	87260	77088.12
1977	21	119310	91930	77121.88
1978	22	136550	84190	76641.27
1979	23	143700	83700	75952.65
1980	24	141277	81900	75341.61
1981	25	139960	78898	74772.50
1982	26	141307	75495	74169.50
1983	27	142339	77375	73541.91
1984	28	136863	72294	73074.73
1985	29	137747	80203	72585.62
1986	30	133562	88710	72216.95
1987	31	121550	81481	72185.80
1988	32	124740	108264	72065.41
1989	33	123661	106258	71975.35
1990	34	118036	105369	72042.47
1991	35	112059	104601	72277.07
1992	36	110168	90979	72565.65

4. 4 Markov chain model

A four state Markov chain model is used to represent the time series on yield. Quartiles of the yield series were located as $Q_1 = 81575.5$, $Q_2 = 91930$, $Q_3 = 107155$. Using these values of the quartiles, four states of the attempted model are defined as follows.

State I : Yield less than Q_1
State II : Yield between Q_1 and Q_2
State III : Yield between Q_2 and Q_3
State IV : Yield greater than Q_3

Values of n_{ij}, the frequency by which the system moves from state `i' to state `j' were found out and are given in the following table (4.4(a)).

Table 4.4(a). Table of frequencies classified according to four states.

i/j	1	2	3	4	n_i
1	6	2	0	1	9
2	2	4	2	0	8
3	0	2	7	1	10
4	0	1	1	7	9
n_j	8	9	10	9	36

The dependence of the four states is tested using a χ^2 statistic. The test statistic is

$$\chi^2 = \sum_i \sum_j \frac{\{n_{ij} - n_{i.}\, n_{.j}/n\}^2}{n_{i.}\, n_{.j}/n}$$

where $n_{i.} = \sum_j n_{ij}$ and $n_{.j} = \sum_i n_{ij}$. The calculated value of χ^2 is found to be 38.532. This is tested against the tabulated value for 9 degrees of freedom. The calculated value is significant at 0.001 level. Hence it is concluded that the identified states are not independent and thus the system satisfies the basic criterian of a Markov chain model.

The transition probabilities from state i to state j is estimated as $p_{ij} = n_{ij}/ n_{i.}$. The estimated transition probability matrix of first order is as follows.

$$p = \begin{bmatrix} 0.667 & 0.222 & 0 & 0.111 \\ 0.250 & 0.500 & 0.25 & 0 \\ 0 & 0.200 & 0.70 & 0.100 \\ 0 & 0.111 & 0.111 & 0.778 \end{bmatrix}$$

The second and higher order transition probabilities are estimated through applying the stated property of a Markov chain, namely $P^{(n)} = p^n$, n = 2, 3,.-------. It is seen that the transition probabilities reach an equilibrium state after 20 steps. The twentieth order transition probabilities are given by

$$p^{(20)} = \begin{bmatrix} 0.1968 & 0.2620 & 0.3054 & 0.2358 \\ 0.1968 & 0.2620 & 0.3054 & 0.2358 \\ 0.1968 & 0.2620 & 0.3054 & 0.2358 \\ 0.1968 & 0.2620 & 0.3054 & 0.2358 \end{bmatrix}$$

From this result it is concluded that the production series reaches steady state after 20 years. The steady state probabilities are Π_1 = 0.1968, Π_2 = 0.260, Π_3 = 0.305 and Π_4 = 0.236. Q_1^* = 70127.75, Q_2^* = 86752.75, Q_3^* = 99542.5 and Q_4^* = 113077.469. Applying equation (11) in chapter 3, the steady state value of production is forecasted as 93594.316 tonnes.

Summary

5. SUMMARY

The present study focusses on time series modelling and forecasting of the yield of cashew in Kerala. For this purpose, secondary data were collected from the various publications of Directorate of Economics and Statistics, Government of Kerala, Thiruvananthapuram, Kerala state. The data on acerage, production, productivity, annual rainfall and price of raw cashew kernal were collected over the period 1956-1992.

Four different types of stochastic models namely (1) Box-Jenkins model (2) Distributed lag model (3) Log-normal diffusion model and (4) Markov chain model were tried.

Univariate ARIMA models of all the variables viz. yield, area under cultivation, price and annual rainfall were considered separately. The modelling procedures consisting of identification, estimation and diagonastic checking were done as per the guidelines proposed by Box and Jenkins (1970)

The following models were developed

Sl. No.	Variable	Estimated model	Coefficient of determination (R^2)
1.	Production	ARIMA (2, 3, 2)	0.64
2.	Area	ARIMA (2, 1, 2)	0.95
3.	Productivity	ARIMA (2, 1, 2)	0.89
4.	Price	ARIMA (4, 3, 4)	0.94
5.	Rainfall	ARIMA (5, 2, 5)	0.38

To judge the forecasting ability of the model, the Mean Absolute Percentage Error (MAPE) was calculated in each case. In all the models MAPE values were small indicating that the forecasting inaccuracy is low. Thus the univariate ARIMA models offered a good technique for predicting the magnitude of all the variables.

With a view of fitting a multivariate ARIMA model, cross correlation analysis of the series was done with production as dependent variable and area, price and rainfall as independent variables. But the results were not in favour of trying a transfer function model.

Distributed lag models of varying types involving selected exogeneous variables were tried. To estimate the response of producers in terms of area towards price and non-price factors, the actual area in the current year was expressed as a function of area under the crop lagged by one year, farm harvest price lagged by one year, price risk and annual rainfall lagged by one year. Response to aggregate cashew area to price appeared to be negatively significant. The results indicate that rainfall had a positive significant effect on acreage allocation. Coefficient of price risk was found to be nonsignificant. The yield response analysis was done with actual yield in the current year as dependent variable and yield price and rainfall lagged by one year and price risk as independent variables. The significance of coefficient of detemination (0.77) was indicative of the fact that the variables included in the yield response model for the whole period was capable of capturing sizeable proportion of variation in the dependent variable. The price factor got significant influence on current yield. But the price risk negatively influenced the yield. The results of the analysis indicated that area was not responsive to price risk. Cashew growers were least sensitive to price movements and they prefer to grow the crop in all types of soil due to its wide adaptability and ease of management.

The log normal diffusion model was fitted to the data on production of cashew in Kerala. The coefficient of determination (0.924)computed for the model revealed a good fit to the data. The model was used for estimating the production of cashew as implied by time and the major exogeneous variable namely area under the cashew.

A four state Markov chain model was used to represent the time series on production. Four states of the model were identified based on the quartiles of the series. The transition probability matrix was estimated and the steady state probabilities were worked out. The analysis showed that production reached equilibrium after twenty years. The steady state probabilities were used to forecast the production.

References

REFERENCES

Abraham, B. and Ledolter, J. 1986. Forecast functions implied by autoregressive integrated moving average models and other related forecast procedures. *Int. statist.Rev.***54**:51-66

Anderson, O. D. 1977a. The interpretation of Box-Jenkins time series models. *The Statistician* **26**(2):127-144

Anderson, O. D. 1977b. Time series analysis and forecasting : Another look at the Box-Jenkins approach. *The Statistician* **26**(4):285-303

*Anderson, O. D. 1989. Analysing time series for forecasting. *Rairo Rech. Operat.***23**:113-150

Asan, R. B., Elizabeth, V. K. and Venugopalan, C. K. 1981. Markov chain model for analysing weekly body weight of broilers. *Kerala J. Vet. Sci.***12**(2):325-331

Borah, D. K. and Bora, P. K. 1995. Predicting the monthly rainfall around Guwahati using a seasonal ARIMA model. *J. Indian soc. Stat.***47**(3):278-287

Box, G. E. P. and Jenkins, G. M. 1970. *Time Series Analysis, Forecasting and Control.* Holden-Day, San Francisco.pp.285-377

Box, G. E. P. and Pierce, D. A. 1970. Distribution of residual autocorrelation in autoregressive integrated moving average time series models. *J. Am. statist. Ass.***65**:1509-1526

Chan, W. 1992. A note on time series model specification in the presence of outliers. *J. appl Statist.***19**(1):117-124

Chandhrabhanu, P. 1991. Supply behaviour of sesamum and groundnut in Kerala. Unpublished MSc.(Ag.) thesis. Kerala Agricultural University, Thrissur.

Chatfield, C. 1988a. What is the best method of forecasting. *J. appl. Statist.*15:19-38

Chatfield, C. and Prothero, D. L. 1973. Box-Jenkins seasonal forecasting:Problems in a case study. *Jl R. statist.Soc.***A136**:295-336

David, L. P. and Kenneth, F. W. 1976. Modelling macroeconomic time series. *Jl R. statist. Soc. A***139**(4):468-485

Devi, P. K. and Rajagopalan, R. 1965. Price and acreage response. *Indian J. agric. Econ.***20**(1):31-35

Durbin, J. and Watson, G. S. 1951. Testing for serial correlation in least squares regression. *Biometrika*.**38**:159-177

Durbin, J. 1970. Testing for serial correlation in least squares regression when some of the regressors are lagged dependent variables. *Econometrica*.38:410-421

Gabriel, K. R. and Newmann, J. 1962. A Markov chain model for daily rainfall occurrence at Tel Aviv. Q. *J. R. Met. Soc.***88**(375):90-95

Granger, C. W. J. and Morris, M. J. 1976. Time series modelling and interpretation. *Jl R. statist.Soc.A***132**(2):246-257

Gupta, G. S. 1993. ARIMA model for and forecast on tea production in India. *Indian econ. J.***41**(2):88-110

Haines, L. M., Munoz, W. P. and Gelderen, C. J. V.1989. ARIMA modelling of birth data. *J. appl. Statist.***16**(1):55-67

Indiradevi, P., Thomas, E. K. and Thomas, J. K. 1990. Growth and supply response of banana in Kerala. *Agric. situ. India.* 45(4):239-242

Janaiah, A., Raju, K. S. and Krishnajah, J. 1992. Area response of major commercial crops in Andhrapradesh. *Agric. Situ.India* **47**(4):257-260

Jonston, J. 1972. *Econometric method,* Mc GrawHill, New York, 2nd ed.p.121-166

Krishnan, S. and Narayanikutty, U. 1993. Estimation of distribution of cumulative rainfall for a specfic period within a Njattuvela. *J. trop. Agric.***31**:107-111

Kumar, A. A. and Srivastava, I. L. 1982. An econometric analysis in estimation of short run and long run elasticities for wheat and rice. *J. Indian Soc. agric. Statist.***34**(3):133

Lakshmi, K. R. and Pal, T. K. 1988. Growth of crop output in Kerala. *Agric. Stiu. India* **43**(9):767-771

Lim, D. 1975. *Supply Response of Primary Producers.*Penerbit University, Malaya, pp.10-39

Madhavan, C. 1972. Acreage response of Indian farmers: A case study of Tamil Nadu. *Indian J. agric. Econ.***27**(1):67-87

Makridakis, S. and Wheelwright, S. C. 1978. *Interactive Forecasting,* Holden-Day, San Francisco, pp.613

Mary, P. 1991. Forecsasting of the production of natural rubber in Kerala. Unpublished MPhil thesis, University of Kerala, Thiruvananthapuram.

Nerlove, M. 1956. Estimates of elasicities of supply of selected agricultural commodities. *J. Fm Econ.***38**(2):496-509

Nerlove, M. 1958. Distributed lags and estimation of long run supply and demand elasticities, theoretical consideration. *J.Fm Econ.***40**(1):301-311

Ogallo, A. J. L. 1986. Stochastic modelling of regional annual rainfall anomalies in East Africa. *J. appl. Statist.* **31**(1):49-56

Pino, F. A., Morettin, P. A. and Mentz, R. P. 1987. Modelling and forecasting linear combinations of time series. *Int. statist.Rev.* 55:295-313

Prabhakaran, P. V. 1987. Impact of price variability in acreage allocation of selected crops of Kerala state. *South Indian Hort.*35(4):318-319

Rajkrishna. 1963. Farm supply response in India-Pakistan: A case study of the Punjab region. *Econ. J.*73(291):441-487

Ramesh, Y. S., Ramanna, R. and Achoth, L. 1988. Supply response of rice in Karnataka. *Agric. Situ. India*43(1):9-13

Ray, C. S. 1991. Growth of manufacturing production in India: A time series analysis. *Journal of Quantitative Economics* 7(1):23-42

Reddy, A. N. 1989. Farm supply response of paddy : A case study of Andhrapradesh. *Indian J. agric. Econ.* 44(4):444-447

Reddy, J. M. 1977. Estimation of farmers supply response: A case study of groundnut. *Indian J. agric. Econ.*25(4):57-63

Santhosh, K. and Prabhakaran, P. V. 1988. Characterization of the pattern of rainfall in Northern Kerala. *Agric. Res. J. Kerala* 26(2):266-276

Saraswathy, P., Thomas, E. J. 1975. Stochastic models for theexplanation of trend in production of rice in Kerala for the period 1957-1958 to 1971-1972. *Agric. Res. J. Kerala* **13**(1):22-26

Saraswathy, P., Thomas, E. J. 1976. Trends in production of certain agricultural crops in Kerala. *Agric. Res. J. Kerala* **14**(2):149-152

Sarkar, A. and Karthikeyan, B. 1993. Extending AR models for seasonal series. *J. Indian Soc. Stat.* **45(1)**:144-162

Thomas, E. K., Thomas, J. K. and Indiradevi, P. 1990. Growth and output response of ginger of Kerala. *Spice India* 3(10):12-17

Tiao, G. C. and Box, G. E. P. 1981. Modelling multiple time series with applications. *J. Am. statist. Ass.***76**:802-816

Tintner, G. and Patel, M. 1969. A log normal diffusion process applied to the growth of yields of some agricultural crops in India. *Indian J. Dev. Stud.***6**:49-57

Tintner, G. and Patel, R. C. 1965. A log normal diffusion process applied to the economic development of India. *Indian econ.J.* **13**:465-474

Tripathy, B. and Gowda, M. V. 1993. An analysis of growth, instability and area response of groundnut in Orissa. *Indian J. agric. Econ.***48**(3):345-350

Waghmare, P. R., Kulkarni, B. S. and Despande, B. S. 1977. Markov process and shifts in cotton area. *J. Indian Soc. agric. Statist.***30**(2):144-145

*Originals not seen

Appendices

APPENDIX I

Programme for ARIMA model

```
bma global 1000
env noundefinederrors
cal 1956 1 1
all 0 2005:1
open data \lotus\prod.wk1
data(for=wks,org=obs) 1956:1 1992:1 prod
diff prod / 3 prod d
corre prod d 1956:1 1992:1 10 auto1
par prod d 1956:1 1992:1 10 part1
set upper 1 10 = 1.96/sqrt(37)
set lower 1 10 = -upper(t)
plot(num=0,max=1.0,min=-1.0) 3
# auto1
# upper
# lower
gra(num=0,max=1.0,min=-1.0) 3
# auto1
# upper
# lower
plot(num=0,max=1.0,min=-1.0) 3
# auto1
# upper
# lower
gra(num=0,max=1.0,min=-1.0) 3
# auto1
# upper
# lower
boxjenk(constant,ar=2,ma=2,def=eq1) prod / resids
smpl 1961:1 2005:1
for 1
# eq1 prodf
print 1976:1 1992:1 prod prodf
print 1992:1 2005:1 prodf
end
```

APPENDIX II

Trends in area, production and productivity of cashew in Kerala

Year	t	Area(X_t) in Ha.	Production (y) in tonnes	Productivity in Kg/Ha.
1956	0	37390	58680	1569.0
1957	1	44040	69100	1569.0
1958	2	46210	72510	1569.0
1959	3	52420	81670	1559.0
1960	4	54320	84630	1558.0
1961	5	55030	85800	1558.0
1962	6	82120	92040	1122.0
1963	7	82370	92310	1122.0
1964	8	85980	96460	1122.0
1965	9	87370	98030	1122.0
1966	10	90560	101610	1122.0
1967	11	97990	106580	1122.0
1968	12	96020	107730	1122.0
1969	13	98960	111030	1122.0
1970	14	102710	115240	1122.0
1971	15	100660	112940	1122.0
1972	16	101419	113880	1122.0
1973	17	103160	115750	1122.0
1974	18	104900	118000	1122.0
1975	19	105906	119000	1122.0
1976	20	113300	87260	770.0
1977	21	119310	91930	771.0
1978	22	136550	84190	617.0
1979	23	143700	83700	582.0
1980	24	141277	81900	580.0
1981	25	139960	78898	564.0
1982	26	141307	75495	534.0
1983	27	142339	77375	543.0
1984	28	136863	72294	528.0
1985	29	137747	80203	582.0
1986	30	133562	88710	664.0
1987	31	121550	81481	670.0
1988	32	124740	108264	868.0
1989	33	123661	106258	877.0
1990	34	118036	105369	893.0
1991	35	112059	104601	933.0
1992	36	110168	90979	826.0

APPENDIX III

Farm harvest price of cashew in Kerala

Year	Price of cashew (Rs/Quintal
1956	57.200
1957	48.300
1958	53.600
1959	62.400
1960	77.300
1961	63.700
1962	56.500
1963	74.700
1964	83.500
1965	92.100
1966	106.100
1967	116.700
1968	123.900
1969	146.400
1970	139.900
1971	158.200
1972	219.000
1973	328.600
1974	276.420
1975	244.000
1976	503.100
1977	535.300
1978	407.100
1979	582.700
1980	731.900
1981	655.530
1982	479.790
1983	869.890
1984	843.340
1985	1059.500
1986	1348.100
1987	1094.900
1988	1154.700
1989	1177.500
1990	1379.900
1991	2037.700
1992	1961.500

APPENDIX IV

Annual rainfall in Kerala

Year	Annual rainfall (in mm.)
1956	2637.400
1957	3079.500
1958	3095.400
1959	3404.600
1960	3371.000
1961	4007.000
1962	3347.000
1963	2588.400
1964	2822.200
1965	2364.600
1966	2565.900
1967	2636.400
1968	3438.000
1969	2595.600
1970	2626.900
1971	3097.100
1972	2688.900
1973	2388.900
1974	2720.900
1975	3409.900
1976	2050.300
1977	2913.100
1978	3209.600
1979	3082.800
1980	2668.500
1981	3037.600
1982	2171.600
1983	2214.700
1984	2460.800
1985	2091.400
1986	2237.800
1987	2653.400
1988	2411.400
1989	2642.000
1990	2780.000
1991	3106.000
1992	3014.300

TIME SERIES MODELLING AND FORECASTING OF THE YIELD OF CASHEW (*Anacardium occidentale* L.) IN KERALA

BY

MINI K. G.

ABSTRACT OF A THESIS

submitted in partial fulfilment of the

requirement for the degree of

Master of Science (Agricultural Statistics)

Faculty of Agriculture

Kerala Agricultural University

DEPARTMENT OF AGRICULTURAL STATISTICS

COLLEGE OF HORTICULTURE

VELLANIKARA, THRISSUR

1996

ABSTRACT

The present investigation, Time series modelling and forecasting of the yield of cashew in Kerala was undertaken with the following objectives.

1. To formulate a suitable model for the forecast of production of cashew crop in Kerala.
2. To work out the major determinants of yield variations.

For this purpose, secondary data were collected from the Directorate of Economics and statistics, Government of Kerala, Thiruvananthapuram, for a period of thirtseven years starting from the year 1956-'57. The data on average, production, productivity, price of raw cashew kernal and annual rainfall were collected. The stochastic models viz. Box-Jenkins model, Distributed lag model, Log-normal diffusion model and Markov chain model were tried on the time series.

Univariate ARIMA models of all the variables were considered seperately. Diagnostic checking was done to ascertain the adequacy of the model. Then the fitted models were used to obtain the sample period and post sample period forecasts. To judege the forecasting ability of the model the Mean Absolute Percentage Error (MAPE) was calculated. The results showed that the univariate ARIMA models offered a good technique for predicting the magnitude of all the variables. Cross correlation analysis of the series was done with yield as the dependent variable and area, price and rainfall as the independent variables. But the results were not in favour of trying a transfer function model.

Distributed lag models of varying types invoving selected exogeneous variables were developed. The area response models had lagged area, price risk, lagged price and lagged rainfall as the explanatory variables, while yield, price , rainfall lagged by one year and price risk served as the determinants of the yield response function. The result of the analysis

clearly indicated that area was not responsive to prices. Cashew growers are least sensitive to price movements and they prefer to grow the crop in all types of soil due to its wide adaptability and ease of management. The coefficient of determination of all functions were relatively high indicating that the proposed models were satisfactory in describing yield and acreage fluctations.

The log normal diffusion model was fitted to the data on production of cashew in Kerala. It was found that the model gave a satisfactory fit to the data. Yield forecasts for the period from 1997 to 1999 were obtained using this model.

A four state Markov chain model was used to represent the time series distribution of production. The four states of the model were identified based on the quartiles of the series and a transition probability matrix was calculated. Equilibrium probabilities were estimated. It was found that the yield reached equilibrium position after twenty years. The steady state probabilities were estimated and used to forecast the production.