FORECASTING AGRICULTURAL COMMODITY PRICES WITH LIMITED DATA SETS

By

Andrew Eliel Jacque

A THESIS (Plan B paper)

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

MASTER OF SCIENCE

Department of Agricultural Economics

1991

DEDICATION

To Almarie and Alana for their immense patience.

ACKNOWLEDGEMENTS

I wish to gratefully acknowledge the invaluable assistance received from the members of my dissertation research committee. Thanks to Professors James Hilker and John Staatz of the Department of Agricultural Economics, Professor Roweena Pecchenino of the Department of Economics and Professor Paul Rubin of the School of Management, for assuring that quality of research was always of primary importance. Special thanks to Professor Robert Myers (formerly of the Department of Agricultural Economics, Michigan State University) for guidance and suggestions in the early stages of the thesis proposal development.

ABSTRACT

FORECASTING AGRICULTURAL COMMODITY PRICES WITH LIMITED DATA SETS

By

Andrew Eliel Jacque

Many developing countries wish to, but do not publish forecasts of commodity prices. One difficulty often cited is the length (shortness) of the data series. This study tried to determine whether this was a valid reason.

Price data for two commodities in a specific country, where the problem is acknowledged were utilized in a case study approach. Existing models were evaluated for applicability to the data series. The identified models were estimated with estimation data and then while keeping the model parameters constant, prices were forecasted in the out-of-sample period. The error associated with the forecast of each model was calculated and the relationship between quantity of data and model accuracy evaluated.

The error values of all the models were high. The results of the analyses were inconclusive in terms of the study objectives.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION AND PROBLEM DEFINITION

1.1 PROBLEM STATEMENT

The five-year National Agricultural Development Plan (NADP) of Trinidad and Tobago (accepted in 1990) provided a rationale and guidelines for the reorganization and refocussing of the Ministry of Agriculture and several major agricultural institutions in the country. One such institution, the Central Marketing Agency (CMA), had until then functioned as a marketing board for several agricultural commodities. The NADP mandated a radical transformation of the CMA, with its new role emphasizing responsibility for the provision of market information to market participants.

In order to meet the new expectations the CMA has reorganized and established a specialized Market Information Unit (MIU). The MIU has been in operation for the past three years, providing mainly historical market analysis information. It is yet to produce published forecasts. One difficulty often quoted is the relatively short data series on prices. Six years of monthly prices are available for the agricultural commodities of interest. This is perceived as insufficient to allow the generation of accurate forecasts.

This predicament is not unique to Trinidad and Tobago. Many developing countries, including those of Eastern Europe, face the same challenge. Specifically, the desire to forecast prices where the length of the data series is perceived as a major limiting factor.

The existence within a country of very short data-series can be attributed to either nonexistent/poor data collection efforts (which is the case in many developing countries) or changes in the price determination process. In the latter situation, two general cases can be distinguished. The change may affect all commodities within the country. This is the situation in many countries of Eastern Europe, where movement from command to market economy has caused tremendous upheavals in the prices of all commodities. Alternatively, the change may affect only one or a few commodities. This can occur with the removal of price controls on a specific commodity in an otherwise essentially market-oriented economy. This is the situation in many developing countries.

1.2 ANALYSIS OF THE PROBLEM

The problem addressed in this research is the inability of countries to publish to market participants, the forecasted prices of agricultural commodities even though there is a desire so to do. The four factors that can cause this problem are highlighted and then individually discussed. They are:

- inadequate data, or conversely inadequate models;
- inadequately trained professional staff;
- expectations of forecast accuracy;
- limitations associated with the use of the existing models.

1.2.1 Inadequate data/models

Forecasting models are classified as either qualitative or quantitative. A wide range of both types of models exist for use in forecasting.

Qualitative models can operate with little or no recorded data, since "experts" subjectively generate forecasts. They may thus prove invaluable in situations where very little or no data exists, or extensive changes are occurring in the price determination process. Quantitative models use historical quantitative data to generate forecasts. Generally, they operate under the assumption that the future will be like the past. Quantitative models are classed as either explanatory or extrapolative depending on the method used in analysis. Explanatory models include the single and multiple equation regression models. Three types of extrapolative models can easily be identified, based on the technique used. These are decomposition models, models using the smoothing technique and those using the Box-Jenkins methodology. More complete descriptions and working details of these models are provided in chapters 2 and 4.

Quantitative models provide better forecasting accuracy, given the availability of adequate data. Each has particular data needs. In general, extrapolative models require less data than explanatory models. The lower limit of data required by any model, in order to obtain accurate forecasts, would specify the minimum data requirement of that model.

This is a critical factor in model selection and applicability. Thus, a short data-series restricts model choice. If the data-series is very short, it may restrict the set of applicable models to include only qualitative models. But as the data-set expands extrapolative and explanatory models become applicable.

1.2.2 Training of the professional staff

The ability of staff to generate useful forecasts is closely related to training received and the ease with which the various models can be understood. The training of most professionals provides extensive practice in the concepts and practice of regression analysis, but little if any, in time-series methods. This situation may reflect perceptions about the relative applicability of time-series methods in data analysis. Additionally, time-series methods, particularly Box-Jenkins methodology, are more difficult to understand.

1.2.3 Expectations of forecast accuracy

All forecasts despite the quantity of data used as input, carry with them the possibility of error. The issues thus are:

- how accurate is the forecast, in terms of its variability from the true data points?
- whether the amount of associated error is acceptable to the users and forecasters; and

- whether the amount of error deemed acceptable is reasonable.

These issues highlight the subjectivity associated with acceptance of the point forecast. If unacceptable, the forecast can be modified or its publication withheld. Modification of the model-generated forecast is widely viewed as the art (as against science) of forecasting. Suppression of the forecast can be done by the forecaster or at any point along the decision chain. Thus it is possible that forecasting institutions in developing countries (e.g. the CMA) produce but do not publish forecast information, because the associated error may be unacceptable to the forecaster, politician or at some other point in the decision chain.

1.2.4 Model-associated Inadequacies

Limitations associated with use of the existing models include:

- the ease with which the models can be understood;
- the appropriateness of assumptions of normality, stationarity etc.
- accessibility to user-friendly software packages, such as menu driven as against command driven packages or packages that run the more complex methods, (e.g. Box-Jenkins) in an automatic mode; and

- accessibility to comprehensive forecasting software packages that allows the beginning forecaster to evaluate the output of a range of applicable models.

The last two points highlight the importance of the software selection process to the successful forecasting enterprise. An extensive selection of software packages is available. To select the one that best matches the situation and characteristics of the forecaster requires a process that aims to evaluate rather than arbitrarily select. The International Journal of Forecasting and books such as Compton and Compton (1990) can be of great assistance in this regard.

Inadequacies due to staff and model-associated factors reduce the analytical capacity within the country. Thus in any given situation the range of models considered and applied is reduced. Indeed it ties model selection and choice to the knowledge/skill level of the forecaster and to the models on the available software. This situation will be further worsened if the forecaster is biased toward or against particular models. As a consequence, the minimum data requirements of a wide range of models do not assume their critical role as a factor in model selection.

Thus short data-series, staff and model-associated deficiencies and accuracy expectations can affect a country's ability to forecast prices. The relative importance of these four factors within the problem set, will affect and define the solution strategy for each situation.

1.3 PURPOSE OF THE STUDY

This research is an exercise in applied short-term price forecasting. Its purpose is to determine whether accurate short term price forecasts can be obtained under the limitation of a short data-series. A case study approach is utilized. Data for two commodities from a specific country, Trinidad and Tobago, are used in the analyses. Trinidad and Tobago is a market-oriented economy. The study is exploratory in nature, in part because of the lack of research specific to this general problem.

Both qualitative and quantitative models can be applicable with short data-series. Indeed, qualitative models may hold an advantage if the data-set is very small. However qualitative models cannot be included in the study because of the constraint of time. Their inclusion would require at least a one-year period of research to ensure adequate field testing of the model and its output. Thus only quantitative models are applied in this study.

The unavailability of data on relevant exogenous variables excludes explanatory models from the study. Thus the class of extrapolative quantitative models are the focus of investigation.

The effect of the data limitation constraint on forecast accuracy can be evaluated by comparing, under varying quantities of data, the forecasting accuracy of applicable existing models. Use of a model selection method that allows evaluation of a wide range of existing models is emphasised. Hence a decision framework for model selection is presented and used in the analysis.

In this research effort applicable models are those that both satisfy an objective model-selection framework and are available in the MicroTSP and SPSS software packages. A review of relevant literature (e.g. Compton and Compton, 1990 and Rycroft, 1989) indicates the existence of a large selection of forecasting software. However few of these are comprehensive with respect to the forecasting models included. More importantly only a few of these are available on the Michigan State University (MSU) campus and in the Lansing area. The combination of MicroTSP and SPSS gives good coverage of the major forecasting models, and both are available in the Department of Agricultural Economics' computer room. These two software packages are not overly restrictive on the model selection process.

1.4 STUDY DEFINITIONS

Accuracy - model and forecast accuracy are both judged in a relative rather than absolute sense. The literature does not indicate the existence of an absolute measure or absolute values of the existing accuracy measures. In this study relative accuracy specifies accuracy relative to other models or forecasts in terms of minimizing forecasting errors, in the out-of-sample period. The magnitude of the forecasting error is measured by accuracy measures such as the Mean Absolute Percent Error (MAPE).

Applicable models - forecasting models which satisfy both an objective selection framework and are available in the MicroTSP and SPSS software packages.

Short-term forecast - given the use of monthly price-data in the analysis, this refers to forecasts extending up to three months ahead.

Short data-series - for this study a short data-series is a time-series with monthly observations and of length shorter than five years.

1.5 RESEARCH OBJECTIVES

The major objective of the study is to determine whether any applicable existing models can provide accurate forecasts given a short data-series. A secondary objective is to determine whether ARIMA models provide more accurate forecasts than other extrapolative models, given a short data-series.

1.6 SIGNIFICANCE OF THE STUDY

Many countries view the provision of forecast information as an important strategy for improving market transparency and stimulating improved market performance. But forecasting has much wider applicability. "Predictions provided by the various forecasting methods are used as inputs for all types of planning, strategy formulation, policy making, scheduling, purchasing, inventory control and a great majority of decision making activity. There is no question that the role of forecasting is becoming central and its necessity, indisputable" (Makridakis, 1984). Hopefully this study will aid in the achievement of forecasting objectives in countries desirous of so doing.

1.7 ASSUMPTIONS AND LIMITATIONS

Accuracy measures are used to identify the best applicable forecasting models based on the accuracy of forecasts in the out-of-sample period. Accuracy measures, however do not carry any information on the economic value/cost associated with using the best model as against one that is inferior in terms of accuracy. Thus use of the accuracy measures to select or rank models, in effect implies an assumption that economic value/cost does not constitute an important factor in model selection.

The accuracy definition, in effect, equates an accurate forecast and model, to the forecasts or applicable model with the lowest value of the accuracy criterion. As a consequence a major limitation of the study is the inability to determine whether the error associated with the forecast of the most accurate model is acceptable to the users and forecasters.

Focussing the analyses on the data limitation constraint, implies assumptions that:

- staff can be trained to interface effectively with whichever method is selected; and
- forecaster and user accuracy expectations are in line with the measures used to evaluate forecast accuracy in this study.

1.8 ORGANIZATION OF THE STUDY

Some important issues and concepts were raised in this chapter. Many of these are expanded and further developed in the literature review conducted in chapter 2. The methodology that guides the analyses and assures achievement of the study objectives is detailed in the third chapter. Chapter 4 contains preliminary analyses aimed at identifying the set of forecasting models appropriate for the case being studied. This chapter also contains equations for the identified models.

The results of the study are presented in chapter 5. The final chapter is dedicated to concluding, examining the implications and making recommendations.

CHAPTER 2

LITERATURE REVIEW

2.1 TAXONOMY OF FORECASTING MODELS

Three major forecasting approaches can be identified. These are qualitative, quantitative, and technological (Makridakis and Wheelwright, 1989, pp 13).

Technological methods "address long-term issues of a technological, societal, political or economic nature" (Wheelwright and Makridakis, 1985). There are four subcategories of technological methods; extrapolative, analogy-based, expert-based, and normative. Extrapolative technological methods "use historical patterns and relationships as a basis for forecasts." Analogy-based methods use historical and other analogies to make forecasts, while expert-based methods use the knowledge of experts. Normative-based methods use objectives, goals and desired outcomes as the basis for forecasting (Makridakis and Wheelwright 1989, pp 13).

Qualitative methods "generally use the opinion of experts to subjectively predict future events" (O'Donovan, 1983). These models can provide forecasts in situations where data are extremely limited or for the very long-run horizon where the existing data pattern is of limited applicability. Delphi is one well-known qualitative model. It is an iterative process wherein "experts" provide forecasts and receive feedback on its performance before the next forecast is made. More detailed discussions of this method can be found in Wright and Ayton (1987) and Stewart (1987).

Quantitative methods "involve the analysis of historical quantitative data in an attempt to predict future values of a variable of interest" (O'Donovan, 1983). These can be categorized into extrapolative, explanatory and monitoring methods.

Extrapolative methods analyze historical patterns and then forecast using a time-based extrapolation of these patterns. Based on the techniques used, four types of extrapolation models can be identified. These are naive, smoothing, decomposition and Box-Jenkins type.

Explanatory methods forecast, based on past relationships identified between the endogenous and exogenous variables. In forecasting it is assumed that these relationships will continue to hold. Explanatory models can be classified as single or multiple equation regression or vector autoregressive models (Makridakis and Wheelwright, 1989).

Extrapolation and (to a lesser extent) explanatory methods operate with the assumption that existing data patterns will continue into the future. This assumption is clearly more realistic in the short and medium term than in the very long term.

Monitoring methods "seek to identify changes in patterns and relationships to indicate when extrapolation of the past patterns or relationships is not appropriate" (Wheelwright and Makridakis, 1985).

"Each major approach includes several types of methods, many individual techniques and a number of variations of each technique," (Wheelwright and Makridakis, 1989, p 13). In addition, some forecasters emphasize techniques which combine the output from one or more models (composite forecasts).

This study focuses on the class of extrapolative quantitative models. The range of models within this class is highlighted in the next section.

2.2 EXTRAPOLATIVE QUANTITATIVE MODELS

2.2.1 Naive models

These models utilize simple rules, eg., forecast equals most recent actual value (random walk model). Some data patterns are best described by the naive model.

2.2.2 Decomposition models

In this approach a time series is thought of as having four components; trend, cyclical, seasonal and a residual random component. Usually a multiplicative relationship is assumed between these four components. Once the systematic components are identified and separated, they can be reintegrated to generate forecasts (Makridakis and Wheelwright, 1982). The major models in this group are the Classical and Census II decomposition models. Both are similar in principle. One important use is in estimating seasonals (Wheelwright and Makridakis, 1985). The entire historical data series is used each time computations are performed (Chambers et al, 1974). Decomposition models are good for intermediate range forecasting (three months to one year ahead) (Wheelwright and Makridakis, 1985).

2.2.3 Smoothing models

The method of smoothing is based on taking either averages or combinations of the past values or past errors of a time series. A formula is used to specify the weight of each data point within the smoothing segment. New data values are added into the computation as they become available and the oldest value simultaneously dropped. Thus the smoothing segment moves as new data is added.

Smoothing models can be further classified into Moving Average or Exponential Smoothing. The smoothing segment in the moving average models have a fixed number of equally weighted data points within them. The data points within the smoothing segment of the exponential smoothing models have exponentially decreasing weights assigned to them. The most recent data point is assigned the highest weight. The number of points within the smoothing segment of the exponential smoothing model is related to the weight assigned to the most recent data point. The higher the weight assigned to the most recent data point the shorter the smoothing segment of the exponential smoothing model.

2.2.3.1 Moving average models

The major use of moving averages is to smooth the data thus reducing fluctuations (random or systematic) in the data. This ability is used in time series analysis to eliminate trend and seasonality from the data. Seasonality is removed with a moving average of length equal to the length of the seasonal cycle.

Moving average models are poor at identifying turning points or forecasting seasonal data, but they are extremely low cost to operate (Makridakis and Wheelwright, 1985). However with seasonal data better results are obtained if the data used as input is seasonally adjusted. The single moving average models can be applied to data which contains no significant trend. Data exhibiting a significant linear trend are best analysed with a double moving average. The double moving average is in effect the single moving average twice applied. While only a few bits of data are required to forecast using these models, a larger data set is required in model-fitting to determine the optimal length of the moving average.

2.2.3.2 Exponential smoothing models

These models provide forecasts which are based on a weighted sum of the past observations and/or errors. The weights sum to one and their distribution depends on the value placed on the smoothing constant, alpha (Compton and Compton, 1990).

There are many variations of exponential smoothing. These differ on the method of weighting the data points and the mathematical form used to fit the trend, if it exists. These include single, double, and triple models. In addition there are several variations within each type.

The single exponential smoothing model can provide good results when applied to data not containing any significant trend (Chambers, Mullick and Smith, 1974). Double and triple exponential smoothing models can be applied to data exhibiting linear and quadratic trend respectively. Single, double or triple exponential smoothing models give poor forecasts when directly applied to seasonal data. However use of seasonally adjusted data and its accompanying seasonal index, allows fairly accurate forecasts.

Winter's exponential smoothing model can handle seasonal data. It uses three parameters, one each for trend (Gamma), seasonality (Beta) and the smoothing constant (alpha).

Exponential smoothing models use only a few data points to forecast. However a much larger data set is required in model-fitting. They are good for short term forecasting.

2.2.3.3 Adaptive response rate exponential smoothing

The adaptive models automatically specifies the value of the smoothing constant alpha, as the data changes. "When the forecast error increases, (i.e. forecasts are consistently under or over the actual values) a smaller or larger value of alpha respectively is specified. In this manner the forecasting method is truly responsive to changes in the pattern of the data and completely automatic, requiring no input from the user. The disadvantage is that they might over-react to changes in the data in such a way that random fluctuations are mistakenly identified as changes in the pattern of the data giving poor performance" (Makridakis and Wheelwright, 1978). These models are good for short-term forecasting.

2.2.4 Adaptive filtering models

These are an extension of exponential smoothing. Forecasts are expressed explicitly as functions of past actual values and/or errors. However no fixed weighting scheme is assumed. Three variations of adaptive filtering are Adaptive Estimation Procedure (AEP), Generalized Adaptive Filtering (GAF) and Kalman filters. In all these techniques, the weights are not arbitrarily chosen as in exponential smoothing, but instead are determined by some iterative procedure that tries to optimize them (Makridakis Wheelwright and McGee, 1983)

2.2.5 ARMA models

This is the most comprehensive time series extrapolation technique. "The method is popular because it provides a wide choice of forecasting models that can theoretically fit any type of data" (Makridakis and Wheelwright, 1989). Basically the forecaster fits a time series with a mathematical model.

Essentially three different categories of time-series regression models are assumed to exist. Autoregressive (AR) models are essentially regression models which explore the correlation relationships between successive values of the data. Moving Average (MA) models explore correlation relationships between successive values of the error term. Combination AR and MA (ARMA) models are used when the data is best fit by a model that specifies a relationship between past successive values and past successive error terms of the series. The parameter p is used to note the order of the AR model. That is the number of past terms of the series included in the equation. The parameter q denotes the order of the moving average model.

ARMA models assume that the data are generated by a stationary process. Stationarity in the data is obtained when all means, variances and covariances are time-invariant. If a data-series is non-stationary, stationarity can sometimes be achieved by differencing the data one or more times. The differenced series can then be used in analysis. The order of differencing required is noted by the parameter d. Models using either stationary or differenced data are referred to as integrated(I). Integrated versions of the AR, MA and ARMA models are termed ARI, IMA and ARIMA models. The ARMA models allow use of seasonal data through seasonal differencing of the data. That is, the data are differenced at a length equal to the length of the seasonal cycle.

Given all this flexibility, basically the task of the forecaster is to fit the appropriate mathematical model to the time series. This model will be optimal, in the sense that it will have smaller errors or variability than any other model fitted to the data.

In addition to the univariate ARMA type model described above, there are also multi-variate models. Multi-variate ARMA models incorporate historical data on variables other than the variable being forecasted, thus providing some explanatory power. Like regression models, they are single output multiple input, in which the added variables are highly correlated with the variable being forecasted. Multivariate ARMA models are referred to as transfer functions.

ARMA models are computationally complex, and relatively difficult to relate to. They are good for short-term and medium-term forecasting.

2.3 DECISION FRAMEWORK FOR MODEL SELECTION.

"The person desiring to make a forecast can choose from a large number of forecasting models, but not all of them are equally effective for a given situation. Thus to maximise the success of a forecast it is usual that the forecaster select the most appropriate method" (Kress, 1985).

Cleary and Levenbach (1981), suggests use of a two-step process in model selection. First is a model identification phase aimed at pinpointing the set of appropriate models that can be applied to the problem. Then, the models so identified are subject to an evaluation phase, which aims to select the most appropriate model from this set.

2.3.1 Model identification phase

The following six decision criteria are highlighted by Cleary and Levenbach (1981):

- **Data Characteristics:** identify any data patterns, such as trends, cycles, seasonality and randomness.
- **Data Requirements:** minimum needs of applicable models.
- **Forecast horizon:** immediate, short, medium or long term.
- **Accuracy:** of each model given the time horizon of interest and the data characteristics.
- **Applicability:** to the needs of users.
- **Cost:** primarily of data collection and analysis.

To these six we can add:

- **Availability of data:** particularly on exogenous variables.
- **Capabilities of professional staff**
- **Availability of computer software**

These nine criteria establish a framework for the a-priori evaluation of the many models that exist. Cleary and Levenbach describe their first two criteria as relating to the "characteristics of the data," while the latter four "relate to the inherent characteristics of the various techniques and models." In their opinion the two groups can be distinguished because the latter four are "influenced by the requirements, resources and objectives of the project" while the former two "are influenced by the nature of the data." The three additional criteria can be included in the latter group.

Levenbach and Cleary highlight that in the process of model identification, little control can be exercised over the first group of criteria. Thus the "characteristics of the data" in effect prescribe use of specific approaches, models and techniques. The second group of criteria can then be applied to this specified group to identify further the set of appropriate models.

2.3.2 Model evaluation phase

"The ultimate test of any forecast is whether or not it is capable of predicting future events accurately" (Makridakis and Hibon, 1979). Thus accuracy can be used as the sole performance criterion of the identified models.

The above suggests need for an out-of-sample rather than in-sample evaluation process. Makridakis (1984) supports this: "Ample evidence exist to suggest that a model which fits historical data does not necessarily forecast well." To implement this, the available data set must be split into estimation and prediction data (Steece, 1982). The estimation data is used to estimate the model coefficients. Forecasting accuracy is evaluated using a second set of data not used in model estimation. Thus the prediction data is used for out-of-sample forecasting.

"The most common accuracy measures are the Mean Squared Error (MSE), the Mean Absolute Percent Error (MAPE) and Theil's inequality coefficient (Makridakis and Hibon, 1979). With the MSE and Theil accuracy measures, the effect of an error is proportional to the square of the error. This is referred to as a quadratic error loss function (Makridakis and Hibon, 1979). The MAPE assumes a linear error loss function. Thus measures built on the quadratic error loss function penalize techniques that make sizeable forecasting errors. "Consequently, they should be used when the forecaster wants a model that provides reasonable accuracy for each period", (Kress, 1985; p 32).

2.4 ACCURACY OF FORECASTING MODELS

There are conflicting results about the relative accuracies of the various models. Most studies indicate general agreement that "quantitative models are more accurate than qualitative in short-term forecasting". (Makridakis, 1984; p 4).

The issue of "the best quantitative model" does not bring such agreement (Makridakis and Hibon, 1979; p 37). In fact this issue seems to be clouded in a lot of erroneous perceptions and beliefs. Armstrong (1978), after a study of econometricians and economic methods concluded that the belief of "more complexity, greater the accuracy", was without empirical support.

The erroneous belief that "more complex is more accurate" is prevalent in the selection of extrapolative models. Wheelwright and Makridakis (1985; p 371) contend "empirical evidence does not support the assumption that sophisticated methods outperform simple ones".

Makridakis and Hibon (1979) used 111 time series and 22 methods to study the accuracy of various forecasting methods. In general, they found that the simpler methods performed as well as the more sophisticated methods. Specifically, deseasonalised exponential smoothing was the "best method overall", and "exponential smoothing methods are at least as good as the Box-Jenkins methodology" (p 41).

Makridakis et al (1982), report the results of a "competition" that involved the evaluation of 1001 time series and 24 methods. This study was a follow-up on the study by Makridakis and Hibon. Specifically the study showed that "deseasonalised single exponential smoothing when used on monthly data," does relatively better than Holt-Winter's, Automatic adaptive exponential smoothing, Bayesian, Box-Jenkins and Landowski" (p 127).

Both studies tried to determine the factors affecting forecasting accuracy of the models evaluated. Identifiable factors were the length of the forecasting horizon, the accuracy measure applied and the type of series. The type of series relates to the extent of trend, seasonality and randomness within the data. In particular "the greater the randomness of the data the less important is the use of statistically sophisticated methods" (Wheelwright and Makridakis, 1985).

2.5 MINIMUM DATA REQUIREMENTS

Quantitative models operate with the assumption that the future is not too unlike the past. Thus the minimum quantity of data used must capture this representative pattern. This quantity will be greater when seasonal data is used.

The nature of the model, whether mathematical or statistical, greatly affects its minimum data needs. The moving average and exponential models are mathematical in nature. These require sufficient data to estimate the model coefficients. Often much less data is required to generate a specific forecast.

Statistical models have additional requirements associated with the assumptions on which they are built. These assumptions for example require adequate degrees of freedom and the absence of small-sample bias. The explanatory methods and the Box-Jenkins models are statistical in nature.

Table 2.1 provides information on the minimum data requirements of the various forecasting methods. Column 7 presents the information provided in Makridakis and Wheelwright (1978, pp 295). Column 8 contains data gleaned from a number of sources.

2.6 HUMAN AND MODEL DEFICIENCY

Table 2.1 shows that the simple extrapolative methods can operate with much less than one year's data. However these models (eg. moving average or single exponential smoothing) do not forecast well with seasonal data. Deseasonalizing the data removes this difficulty. However, to provide one season of deseasonalized data requires at least two seasons of data. This with the least demanding technique of centered moving average.

Models that can handle seasonal data patterns generally require more than two years of data. The explanatory and more complex extrapolative methods, however, require much more data.

Literature reviewed in the previous section suggests that the simple models are at least as accurate as the more complex, when used in short-term forecasting.

Table 2.1 Characteristics of alternative forecasting models

Method	Type (Pattern) of Data			Horizon		Min. Data	
	No Trend	Trend	Seasonal	Short	Medium	I	II
Seasonal Naive	X			X		L	
Single Moving Average	X			X		5p	
Double Moving Average		X		X		10p	
Single Exponential Smoothing	X			X		2p	
Linear Exponential Smoothing		X		X		3p	
Winters' Exponential Smoothing	X	X	X	X		2L	
Adaptive Exponential smoothing	X			x		3p	
Classical Decomposition	X	X	X	X		5L	
Harrison's Harmonic Smoothing	X	X	X	X		3L	
Generalised Adaptive Filtering	X	X	X	X		5L	
Univariate Box-Jenkins	X	X	X	X		7L	3 yrs
Multivariate Box-Jenkins	X	X	X	X		9L	
Simple Regression	X	X	X	X	X	20p	4 yrs
Econometric	X	X	X		X	6L	4 yrs

Columns 1-7; Source - Makridakis and Wheelwright (1978; pp 295)
Column 8; Source - Cleary and Levenbach (1981)
where:
L - length of seasonality
p - data points
short-term - 1-3 months ahead
medium-term - 4-12 months ahead

This study does not investigate reasons for the inability to forecast if it can be shown that the available data is adequate. However potentially, staff deficiencies can be a key factor.

These deficiencies can be caused by:

- Erroneous Perceptions and Beliefs

The belief that more complex means more accurate adds bias to the model identification process, which may result in the simple models being overlooked. In this situation the forecaster operates under the belief that the minimum data requirements of the more complex models, is the quantity of data required for accurate forecasts to be obtained.

-Unrealistic Expectations about Accuracy.

Makridakis (1984; pp 2 & 3) makes the point well. "Forecasting is not a substitute for prophesy. Forecasters unfortunately do not have crystal balls allowing them to look into the future. **Forecasting errors are inevitable."** While the user of forecast information was the subject of these comments, they are equally valid when applied to the beginning forecaster. If a forecaster expects to be precise with his forecasts, none of the available models will be applicable.

- Insufficient Training and Knowledge

Lack of awareness of the full range and capabilities of forecasting methods and models increases bias in the model selection process.

- Inaccessibility to Comprehensive Forecasting Software

The availability of a comprehensive forecasting software package would allow the forecaster to evaluate all models identified as appropriate. This would reduce bias in the model selection process.

CHAPTER 3

METHODOLOGY

3.1 INTRODUCTION

The methodology must be guided by the objectives of the study. In effect we need to evaluate the effectiveness of the models to forecast with limited data and of specific models to do so more effectively than others.

A case study approach is used. The unavailability of data on relevant exogenous variables excludes explanatory and multivariate autoregressive techniques. Thus, the set of quantitative extrapolative models are investigated for applicability to two specified data-series. Accuracy of the out-of-sample forecast of each model is measured.

The two data-series are the price series for two vegetable commodities in Trinidad and Tobago. Six years of monthly, current and unadjusted price data (quoted in Trinidad and Tobago dollars (TT $)) are available for cabbage and tomatoes. The country has been unable to provide published forecasts. One reason advanced is the limitation imposed by the length of the data series. Thus, use of these data sets provide an opportunity to investigate length of the data series as a valid explanatory factor in the country's inability to forecast prices.

3.2 ESSENTIALS OF THE METHODOLOGY

In brief, the methodology identifies the set of extrapolative quantitative models applicable to each of the two data sets, then evaluates the performance of these methods given the study objectives. The "decision framework for model selection", adapted from Cleary and Levenbach, which directs use of model identification and evaluation phases, is used in this process.

3.2.1 Model Identification Phase

Some of the nine decision variables in this phase will be explicitly taken account of, while the effect of others will be assumed away. To some degree the assumptions reflect the extent to which we are removed from the role of actual price forecaster with its very specific constraints.

Six variables are explicitly taken into account in the analyses. These variables and the specific aspects considered, are:

- **characteristics of the data:** price-data are available at monthly intervals and are so forecasted. The data are evaluated to determine the significance of seasonality, trend and stationarity;
- **minimum data requirements:** five years of estimation data are available for each commodity;
- **forecast horizon:** short-term (one, two and three steps (months) ahead) forecasts are required;

- **accuracy:** this variable is explicitly considered to the extent that it qualifies a method as being applicable to a specific horizon;
- **exogenous data:** none was available, consequently the explanatory models are not included in the analyses;
- **computer software availability:** the statistical software packages SPSS and MicroTSP and the spreadsheet Quattro Pro are used to analyse the data.

For this study the other three variables are assumed to have neutral effects in the process of model choice. Specifically we assume that:

- all models equally fulfill the needs of users;
- cost associated with each model is equally burdensome or insignificant;
- staff capability is equal for each model.

The analyses necessary for model identification are conducted in the next chapter of this report.

3.2.2 Model Evaluation

The evaluation decision criterion is, accuracy of the out-of-sample forecast. Three measures are used to evaluate accuracy. These are the MAPE, MSE and percentage turning point errors (%TPE).

Use of the out-of-sample forecasting procedure requires separation of each data-series into estimation and prediction data. The study objectives are evaluated by evaluating for each identified model, changes in the accuracy measures as the data limitation condition is progressively relaxed. Three, four and five years of estimation data are used when evaluating the models. In all cases one years prediction data is utilized. Consequently three data-sets were created from each of the data-series. Each data-set so developed consists of estimation data plus one year of prediction data. The characteristics of these data-sets are detailed in Table 3.1 below.

Thus the accuracy of each identified model will be evaluated:

- over varying forecasting horizons (1-3 steps ahead);
- with varying quantities of estimation data (3-5 years) and
- using one year of prediction data.

Table 3.1 **<u>The six primary data sets</u>**

NAME	CHARACTERISTICS	
	Estimation period	Prediction period
Cabb3	1985.01 - 1987.12	1988.01 - 1988.12
Cabb4	1985.01 - 1988.12	1989.01 - 1989.12
Cabb5	1985.01 - 1989.12	1990.01 - 1990.12
To3	1985.01 - 1987.12	1988.01 - 1988.12
To4	1985.01 - 1988.12	1989.01 - 1989.12
To5	1985.01 - 1989.12	1990.01 - 1990.12

The hypotheses will be tested by comparing for each forecasting horizon and across (the average of) forecasting horizons the MAPEs, MSEs and %TPEs of all models as the data limitation condition is progressively relaxed.

The above outlines the primary analysis conducted in this study. Possible weaknesses of this analysis include:

- use of a varying (1988, 1989 and 1990 are used) as opposed to a fixed prediction period. This may affect the results of the evaluation phase.
- use of forecasting periods containing varying numbers of months. As is shown later, the one, two and three steps ahead forecasts contain twelve, eleven and ten forecasted values respectively. This may affect the ability to compare forecasts for different steps ahead.
- use of additive as opposed to multiplicative seasonals in the model estimation and forecasting phases. This may decrease the accuracy of forecasts for the Winter's, Single Exponential Smoothing (SES), Naive and Moving Average models, if the process generating the data is better modelled with the use of multiplicative seasonals.

Thus a secondary analysis was conducted using a fixed prediction period (1990), each with twelve forecasted values, and analyzing with multiplicative seasonals where applicable. The secondary analysis is intended simply as a check on the primary analysis. A subset of the applicable models are used in this secondary analysis. These are the Winter's, SES and Naive models. The characteristics of these data-sets are detailed in Table 3.2.

Table 3.2 The six secondary data sets

NAME	CHARACTERISTICS	
	Estimation period	Prediction period
M-Ca3	1987.01 - 1989.12	1990.01 - 1990.12
M-Ca4	1986.01 - 1989.12	1990.01 - 1990.12
M-Ca5	1985.01 - 1989.12	1990.01 - 1990.12
M-To3	1987.01 - 1989.12	1990.01 - 1990.12
M-To4	1986.01 - 1989.12	1990.01 - 1990.12
M-To5	1985.01 - 1989.12	1990.01 - 1990.12

Analyses were conducted using the statistical packages MicroTSp and SPSS and the spreadsheet Quattro Pro. Quattro Pro was used to calculate, estimate and forecast the moving average models, and to calculate the value of the accuracy measures for each model. Graphs were produced using either the MicroTSP or the Quattro Pro software and then exported to the word processing software package, Word Perfect 5.1.

3.3 ANALYTICAL PROCEDURE

3.3.1 Model-fitting Phase

The estimation data of each created data set were used as input in the model fitting phase. The statistical packages were used to:

1) seasonally adjust the data and provide the associated seasonal factors;
2) test for stationarity of the data;
3) estimate the parameters of the Winters, SES and ARIMA models. A nonlinear estimation procedure aimed at minimizing either the MSE or the sum of squared errors (SSE) is used in the selection of parameter values.

The spreadsheet package was used to calculate the optimal length of the moving average for each of the data sets. The criterion used to determine optimal, was the minimum MSE of forecasts for the estimation data;

3.3.2 Forecasting Phase

Once the models were specified the parameters were held constant while forecasting in the out-of-sample period. For one-step ahead forecasts the last three points in the data set used to produce the forecast would consist of actual data. For the two-step ahead forecasts the last three points in the data set would consist of actual data, actual data and the last available one-step ahead forecasted value respectively, for the specific months. In the three-step ahead forecast it would be actual data and the last available one-step ahead and two-steps ahead forecasted values respectively for the specific months. The forecasted values are thus used as substitutes for actual data values when two and three-steps ahead forecasting.

3.3.3 Evaluation Phase

Once forecasts were obtained graphs of the actual versus forecasted values were generated . The accuracy measures were also computed and these results analyzed.

3.4 REPORTING THE RESULTS

The results are presented in the form of tables and graphs. Tables, located in the results chapter, present the MAPEs, MSEs and %TPE of each model as the forecasting horizon and quantity of estimation data is varied. This allows evaluation of the best model for one, two and three steps ahead forecast. Graphs are located in appendices A through D. These show the actual versus the forecasted output of each model.

The next two chapters contain the results of the analyses. Chapter 4 contains the analyses associated with the model identification process. It also contains model equations and working details for all identified models. Chapter 5 contains the results of the evaluation phase.

CHAPTER 4

FORECASTING MODELS

4.1 THE COMMODITIES AND DATA

Both cabbage and tomatoes are produced by a large number of small farmers, and sold competitively, in an environment relatively free of government intervention. A ban on imports and virtually no export or processing use effectively channel production onto the fresh market. Each crop requires approximately three months from time of transplanting to harvesting. Typically harvesting occurs over a period of several (2-3) weeks.

Examination of the price data for both commodities shows strong seasonal effects. Seasonality is almost exclusively due to variations in supply, driven by the occurrence of the wet and dry seasons. The dry season spans the period December to May, with the remainder of the year being wet. The petit-careme, a 2-3 week dry spell within the wet-season, occurs sometime within the period September to October. Both commodities give higher yields in the dry season. However this effect is much greater in tomato. Prices are thus higher in the wet as against the dry season.

4.2 ANALYSIS OF THE DATA

Model identification and selection is heavily influenced by patterns present in the data. Also importantly the statistical models are based on the assumption of stationarity of the process generating the data. Thus the unadjusted data were analyzed to determine the significance of seasonal effects, trend and stationarity. The raw data for cabbage and tomatoes are presented in tabular form (Tables 4.1 and 4.2) and graphical form (Figures 4.1 and 4.2).

Table 4.1 TOMATO - Monthly Average Wholesale Prices ($/kg)

	Jan.	Feb.	Mar.	Apr.	May	Jun.	Jul.	Aug.	Sep.	Oct.	Nov.	Dec.
1985	5.68	3:32	1.43	0.98	3.52	5.13	2.34	5.65	5.95	8.57	8.93	8.37
1986	5.59	2.88	2.03	1.44	2.96	3.24	4.1	2.76	1.93	5.02	9.52	9.17
1987	5.37	3.09	2.79	2.71	2.85	4.26	5.73	5.64	4.28	5.25	6.16	7.33
1988	5.04	2.75	2.45	2.31	1.99	3.88	5.11	3.78	3.59	6.78	7.82	6.03
1989	6.03	3.4	2.55	1.62	1.69	1.93	3.91	6.79	5.16	7.02	7.52	7.34
1990	3.84	3.02	2.9	2.9	4.31	8.11	7.04	2.65	5.03	8.27	7.45	6.63
av. 85-89	5.54	3.09	2.25	1.81	2.6	3.69	4.24	4.92	4.18	6.53	7.99	7.65

Source: Central Marketing Agency - Market Information Unit

Table 4.2 **CABBAGE -Monthly Average Wholesale Prices(TT$/kg)**

	Jan	Feb	Mar.	Apr.	May.	Jun.	Jul.	Aug.	Sep.	Oct.	Nov.	Dec.
1985	5.7	3.9	1.61	1.33	3.18	4.72	4.48	2.3	1.24	2.97	5.51	7.99
1986	6.12	5.39	2.2	1.65	3.61	3.26	2.3	1.59	2.32	3.78	7.11	5.85
1987	4.29	4.66	4.06	4.82	4.11	6.86	7.67	7.36	2.59	2.89	2.76	2.49
1988	3.84	3.41	2.27	1.95	1.48	1.58	3.35	3.99	4.31	4.95	5.99	3.26
1989	1.52	1.28	1.19	0.95	2.12	1.84	2.83	2.84	1.73	1.97	1.67	3.95
1990	2.74	3.57	2.04	1.04	1.3	2.72	2.39	2.04	2.9	5.53	6.37	4.56
av. 85-89	4.29	3.73	2.27	2.14	2.9	3.65	4.12	3.62	2.44	3.31	4.61	4.71

Source: Central Marketing Agency - Market Information Unit

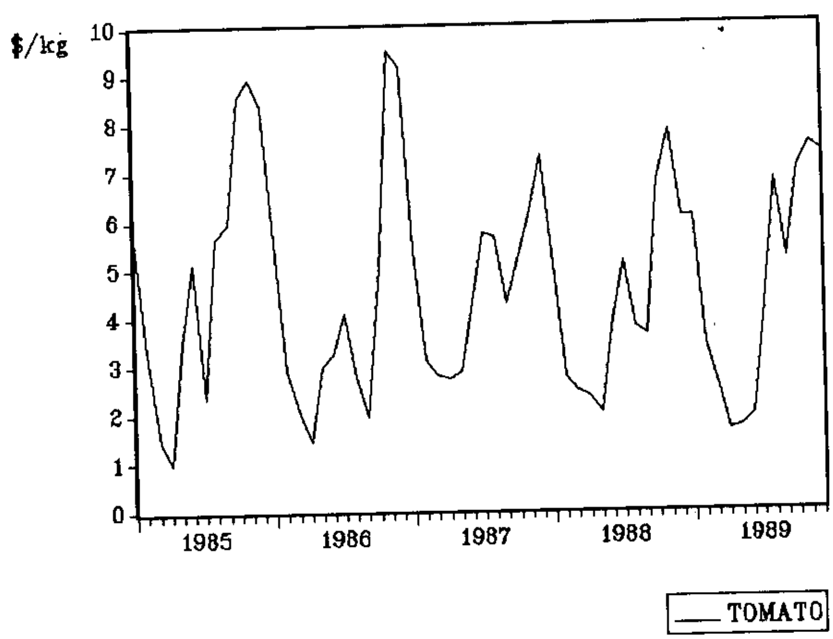

Figure 4.1 **Tomato - monthly average wholesale prices (TT $/kg)**

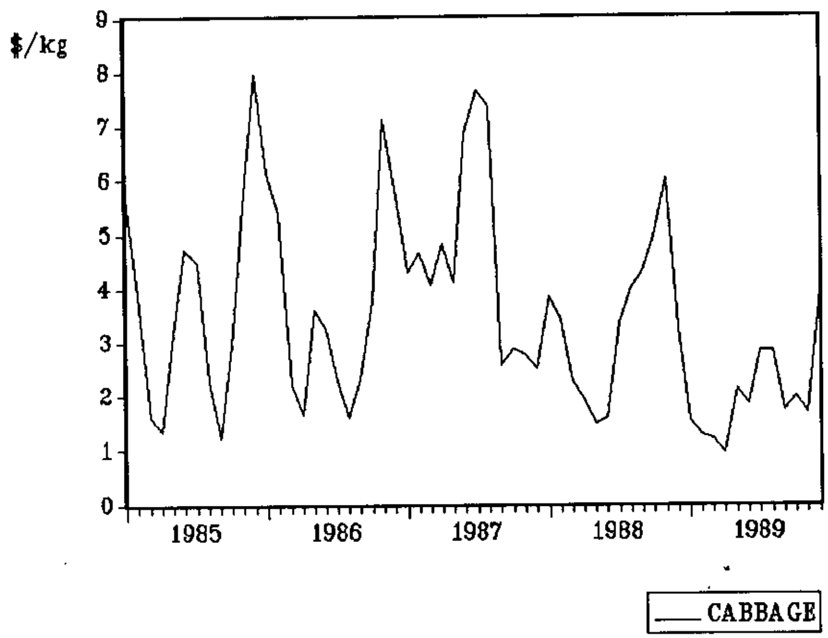

Figure 4.2 Cabbage - monthly average wholesale prices (TT $/kg)

4.2.1 Seasonality

The importance of seasonal effects was determined by examination of the raw data and the seasonal factors of the seasonally adjusted data. The seasonal factors for each data set are presented in Table 4.3. The seasonal factors confirm the presence of seasonality in both data-series.

The effects of seasonality can be either additive or multiplicative. An additive model is appropriate when the seasonal effects within the data series remains fairly constant over time (see Lillien (1990) and SPSS Inc (1990).

Table 4.3 Additive seasonal adjustment factors

	Cabbage			Tomato		
	Quantity of estimation data (years)					
	3	4	5	3	4	5
Jan	0.87	0.87	0.81	1.06	1.06	0.81
Feb	-1.69	-1.69	0.54	0.70	0.70	-1.63
Mar	-2.23	-2.23	-1.17	-1.32	-1.32	-2.05
Apr	-2.47	-2.47	-1.28	-1.25	-1.25	-2.26
May	-1.51	-1.51	-1.05	-0.56	-0.56	-1.78
Jun	-0.58	-0.58	-0.16	0.81	0.81	-0.53
Jul	0.61	0.61	0.51	0.90	0.90	0.68
Aug	-0.09	-0.09	0.50	0.48	0.48	-0.25
Sep	-1.19	-1.19	-0.67	-1.50	-1.50	-1.06
Oct	0.82	0.82	0.16	-0.63	-0.63	1.34
Nov	3.52	3.52	1.60	1.01	1.01	3.51
Dec	3.94	3.94	0.22	0.32	0.32	3.22

The seasonal factors resulting from analysis of the additive model assume positive and negative values. They represent the difference between the raw data values and the seasonally adjusted data values. A multiplicative model is appropriate when the seasonal effects are increasing or decreasing with time. The multiplicative model produces a seasonal index whose values are centered around a value of one. In this case the raw data series is equal to the seasonally adjusted data times the seasonality index. Section 4.4 outlines calculation of the seasonally adjusted data and the seasonal factors.

Visual inspection of the data plots indicate that an additive model is appropriate for both commodities. Thus the

Winters exponential smoothing model and seasonal adjustments conducted in this study are specified as additive. Note however that the secondary analysis was conducted using multiplicative seasonals.

4.2.3 Stationarity

The Adjusted Dickey-Fuller (ADF) test (David Lillien, 1990) was used to determine stationarity of the time series. This test also allows determination of the degree of differencing required to achieve stationarity, and the significance of the trend variable in the process generating the data. An absolute value of the t-statistic that is greater than the MacKinnon critical value implies stationarity of the time series. The reverse implies non-stationarity. Evaluation of the t-statistic of the trend variable included in the ADF regression allows us to determine whether trend is significant.

The testing procedure involves evaluation of the time series for evidence of non-stationarity. The tests are conducted at varying lag-lengths within the data set. If unable to reject the null hypothesis of non-stationarity, we conclude that the time series is non-stationary. We then test the first differences using the same hypothesis. Stationarity of the first differences, where the time series is non-stationary allows us to conclude that the data is integrated of order one, I(1).

ADF tests were conducted on the six study data sets. The hypothesis of non-stationarity could not be rejected at the 5% confidence level, for any of the data sets. Test t-statistics smaller than the critical MacKinnon t-statistic values (i.e. evidence of non-stationarity) often occurred at lag two or three. The first difference of each series, however, proved stationary. Thus we conclude that each of the data sets are integrated of order one, I(1). This knowledge is important in ARMA model identification.

Trend was not found to be a significant variable in any of the data sets. This of course has implications for the choice of smoothing model.

Thus, each of the six data sets are non-stationary (specifically, they are integrated of order one), does not contain any significant trend and exhibits seasonality.

4.3 MODEL IDENTIFICATION PROCESS

Knowledge of the data patterns, the quantity of data available and the forecasting horizon desired was then applied to the apriori framework for model selection. Specifically, this information was applied to Table 2.1 in order to identify models applicable to the situation. The Winter's exponential smoothing, adaptive filtering, univariate ARMA and Harrison's harmonic smoothing models were identified. A second set of models is identified when seasonally adjusted data are used as input. These are the single moving average, naive, single exponential smoothing and adaptive response rate exponential smoothing models.

However, use of the MicroTSP and SPSS software further constrains this set. Thus for this study the applicable models are the univariate ARMA, Winter's exponential smoothing, naive, single moving average and single exponential smoothing models. The last three models use seasonally adjusted data as input. The mathematical form of these models is explored in the next two sections.

4.4 FORECASTING WITH SEASONALLY ADJUSTED DATA

Models which utilize seasonally adjusted data as input forecast the seasonally adjusted series. Thus, the seasonal factors must be added to the forecast in order to obtain forecasts of the original data series.

Seasonally adjusting the raw data can be viewed as a two-step process, which involves first calculating the seasonal factors and then removing their effect from the time series. The procedure outlined below is in line with that used by the MicroTSP software.

An additive relationship between the components of the time series can be represented as follows:

$$X_t = S_t + T_t + C_t + R_t$$

Where S, T, C and R are the seasonal, trend, cyclical and random components respectively of the time series. The subscript t, specifies the current month or a specific month under review. Thus X_t refers to the price of the commodity in the current month.

First a centered moving average of the series is calculated. The moving average must be equal to the length of seasonality. This is 12 months in the cabbage and tomato data sets. This smoothing operation eliminates randomness and seasonality from the smoothed series (M_t). Thus

$$X_t - S_t - R_t = M_t = T_t + C_t.$$

Calculating the difference of the moving average (M_t) from the time series (X_t) isolates the component containing the seasonal effects and randomness. That is

$$X_t - M_t = S_t + R_t.$$

Randomness is eliminated by averaging separately the differences obtained for each month over all the years in the sample. These averages are the seasonal factors (S_t).

The seasonally adjusted series (D_t) is calculated by subtracting the seasonal factors from the time series. Thus

$$X_t - I_{t-L} = D_t.$$

The subscript t-L specifies that the seasonal factor calculated for the same month one cycle ago must be used in this process.

Thus in models using D_t as input, the forecasted output is D_{t+1}. To obtain a forecast of the time series (raw data) it is necessary to add the seasonal factor from the last cycle to this value. That is $X_{t+1} = D_{t+1} + I_{t-(L-1)}$.

4.5 WORKING DETAILS OF THE IDENTIFIED MODELS

The information contained in this section is to a large extent a summary and reproduction of material contained in Makridakis, Wheelwright and McGee (1983).

4.5.1 Notation

The following notation is used in the equations.

a - constant term

X - actual value of the variable of interest.

F - forecasted value for the variable of interest.

α - smoothing constant for randomness in the data

β - smoothing constant for seasonality

γ (gamma) - smoothing constant for trend

I - seasonal factors or seasonality index

L - length of seasonality

n - number of periods (months)

t - time period

e - error

4.5.2 Accuracy measures.

The applicable models are evaluated using three accuracy measures. The MAPE and MSE differ in the weight attached to the forecast error. In addition both these measures are used in the report evaluating the Makridakis data-set, allowing some comparisons to be made. Percentage Turning Point

Error(%TPE) evaluates on a different though important criterion.

The MAPE and MSE are calculated with the aid of the Quattro Pro spreadsheet. %TPE are computed from plots of the forecasted against the actual series, in the out-of-sample period. A mechanical computational procedure is used to decide whether the turning point is in error. The graphs are found in Appendices A and B.

$$\text{mape} = \frac{1}{n} \times \sum_{t=1}^{n} \left[\frac{|X_t - F_t|}{X_t} \right] \times 100 \qquad \textbf{(1)}$$

$$\text{MSE} = \frac{1}{n} \times \sum_{i=1}^{n} (X_t - F_t)^2 \qquad \textbf{(2)}$$

$$\%\text{TPE} = \left(\frac{\text{no. of incorrect turning points}}{\text{no. of out-of-sample data points}} \right) \times 100 \qquad \textbf{(3)}$$

4.5.3 Single moving average

This simple averaging process can only produce good results if the process underlying the observed values of X has no noticeable trend or seasonality.

The greater the number of terms (n) in the moving average, the greater the smoothing effect. When n=1, the model is equivalent to a random walk model. The use of a small value for n will thus allow the moving average to follow the pattern of the data, though it will always be behind it by one or more periods. The optimal length of the moving average was determined by minimizing the MSE of the estimation data. This was accomplished in the model fitting phase.

The forecasting equation for the single moving average can be written algebraically as:

$$F_{t+1} = \frac{X_t + X_{t-1} + X_{t-2} + \cdots + X_{t-(n-1)}}{n} = \frac{1}{n} \sum_{i=t-(n-1)}^{t} X_i \qquad (4)$$

This equation indicates that the forecast is an average of the present and past actual values contained in the moving average. Thus, once the model is estimated, only n data points are required to produce a forecast.

4.5.4 Single exponential smoothing

$$F_{t+1} = \alpha X_t + (1 - \alpha) F_t \qquad (5)$$

This equation is the general form used in computing a forecast with the method of exponential smoothing. The operational concept inherent in exponential smoothing can be easily demonstrated if equation 5 is expanded by replacing F_t with its components:

$$F_t = \alpha X_{t-1} + (1 - \alpha) F_{t-1} \tag{6}$$

Thus the first substitution yields:

$$\begin{aligned} F_{t+1} &= \alpha X_t + (1 - \alpha)(\alpha X_{t-1} + (1 - \alpha) F_{t-1}) \\ &= \alpha X_t + \alpha(1 - \alpha) X_{t-1} + (1 - \alpha)^2 F_{t-1} \end{aligned} \tag{7}$$

If this substitution process is continued we obtain the relationships:

$$\begin{aligned} F_{t+1} = {} & \alpha X_t + \alpha(1 - \alpha) X_{t-1} + \alpha(1 - \alpha)^2 X_{t-2} + \\ & (1 - \alpha)^3 X_{t-2} + \cdots + (1 - \alpha)^n F_{t-(n-1)} + \cdots \end{aligned} \tag{8}$$

Since alpha has a value between zero and one, the older observed values have exponentially decreasing values. Hence, the name exponential smoothing.

Equation 9 is an alternate form of equation 5.

$$\begin{aligned} F_{t+1} &= F_t + \alpha(X_t - F_t) \\ &= F_t + \alpha e_t \end{aligned} \tag{9}$$

Thus the new forecast is simply the old forecast plus alpha times the error in the old forecast. This implies that when alpha has a value close to one the new forecast consists of the previous forecast plus a large portion of the error associated with that old forecast. In this situation the new forecast taks a value close to the last observed value of X.

Conversely when alpha is close to zero, the new forecast will include very little adjustment. Thus it will have a value close to that of the previous forecast. A large value of alpha gives very little smoothing in the forecast, while a very small value gives considerable smoothing. An alpha value of 1 gives the random walk model.

The value of alpha is determined by a non-linear estimation procedure aimed at minimizing the MSE in the model estimation phase.

Equation 8 provides an alternative way of looking at the mechanics of the model. Here an alpha value of one in effect means that all the weight is placed on the last observed value of X. Thus the forecast for the next period is equal to the present value of X, i.e. the random walk model. An alpha value close to zero means that a large number of data points will be included in the smoothing segment to calculate the forecast. In this situation the forecast takes on the characteristics of an average of the past data. Thus the forecast in each succeeding period tends to be identical or very close to value forecasted for the previous period.

4.5.4 Seasonal exponential smoothing - Winters model

The single and double exponential smoothing methods may do a poor job of forecasting when seasonality is present in the data set. Winter's model explicitly takes account of seasonality and trend through inclusion of additional equations and parameters.

Winter's model uses four equations and three parameters. Three of the equations, and their associated parameters, smooth for randomness, trend and seasonality (equation 10, 11, and 12 respectively) in the data. The fourth equation (equation 13) is used to forecast.

$$S_t = \alpha (X_t - I_{t-L}) + (1 - \alpha)(S_{t-1} + b_{t-1}) \quad (10)$$

$$b_t = \gamma (S_t - S_{t-1}) + (1 - \gamma) b_{t-1} \quad (11)$$

$$I_t = \beta (X_t - S_t) + (1 - \beta) I_{t-L} \quad (12)$$

$$F_{t+m} = (S_t + b_t m) + I_{t-L+m} \quad (13)$$

In the above equations b,I and S estimate the trend, seasonal factors and present level of the deseasonalized data. Thus in operation this model removes seasonality from the data, using the seasonal factor calculated for the same month in the preceding seasonal cycle. In addition randomness is removed from the seasonally adjusted data. These two operations are accomplished in equation 10. The series that results, S_t, is used to calculate the seasonal factors and the trend in the data. In equation 12 since X_t contains seasonality and S_t does not, the difference of these two provides the seasonal adjustment factors. S_t is also used to calculate the trend (equation 11). This is calculated as the difference between two values of S_t. In equation 13, the trend and seasonal factor is added to the present level of the deseasonalized data to obtain the forecast.

4.5.5 Auto-regressive moving average (ARMA) models

Forecasts are expressed as functions of past values, but unlike exponential smoothing, no fixed weighting scheme is assumed.

As noted in Chapter 2 p, d and q denote the order of the AR model, the degree of differencing required for stationarity and the order of the MA model respectively. In addition, when seasonality is present in the data, seasonal differencing is required. The mathematical models are outlined below.

4.5.5.1 The Autoregressive model

$$X_t = a + \phi_1 X_{t-1} + \phi_2 X_{t-2} + \phi_3 X_{t-3} + \cdots + \phi_p X_{t-p} + e_t \qquad \textbf{(14)}$$

Equation 14 is an autoregressive(AR) scheme. It shows that future values are linear combinations of past values of the variable. This equation can be thought of as similar to the general equation in the Single Exponential Smoothing model (equation 5), but with the following substitutions:

$$\phi_1 = \alpha \ , \quad \phi_2 = \alpha(1-\alpha) \ , \quad \phi_p = \alpha(1-\alpha)^{p-1} \qquad \textbf{(15)}$$

Thus p can take different values. The equations below show the mathematical form of the AR model when p=1 and p=2 (equations 16 and 17 respectively). These are auto-regressive processes of order one AR(1) and order two AR(2).

$$X_t = a + \phi_1 X_{t-1} + e_t \qquad \textbf{(16)}$$

$$X_t = a + \phi_1 X_{t-1} + \phi_2 X_{t-2} + e_t \qquad \textbf{(17)}$$

We can see that the AR(1) process has the previous value of X included, while the AR(2) model has the two previous terms. Thus the order of the AR process determines the number of past terms included. Thus model specification is flexible allowing the most appropriate combination to be selected.

4.5.4.2 The Moving Average(MA) model

$$X_t = a + e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2} - \theta_3 e_{t-3} - \cdots - \theta_q e_{t-q} \qquad \textbf{(18)}$$

Moving Average models differ from autoregressive ones, as they assume that future values of X_t are linear combinations of past values of the errors or noise(e_t). Equation 18 is an MA model of degree q. Like the parameter p in the AR model, the parameter q can take on various values. Equations 19 and 20 show the mathematical form of the Moving Average model when q=1 MA(1) and q=2 MA(2) respectively.

$$X_t = a + e_t - \theta_1 e_{t-1} \qquad \textbf{(19)}$$

$$X_t = a + e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2} \qquad \textbf{(20)}$$

4.5.4.3 Autoregressive Moving Average Models

$$X_t = a + \phi_1 X_{t-1} + \phi_2 X_{t-2} + \cdots + \phi_p X_{t-p} + e_t - \theta_1 e_{t-1} - \cdots - \theta_q e_{t-q} \quad \textbf{(21)}$$

Equation 21 is a combination of an AR(p) model and an MA(q) model. An ARMA model can thus be specified as being of order p and q. Thus equations 22 and 23 are ARMA(1,1) and ARMA(2,2) models respectively.

$$X_t = a + \phi_1 X_{t-1} + e_t - \theta_1 e_{t-1} \quad \textbf{(22)}$$

$$X_t = a + \phi_1 X_{t-1} + \phi_2 X_{t-2} + e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2} \quad \textbf{(23)}$$

The advantage of an ARMA scheme is that it includes different AR models and uses whatever error remains in an MA equation in attempting to further improve forecasting. This can be done until the errors have been reduced to randomness (white noise), at which point no further improvements in the model fitted are possible.

The ARMA model operates on the assumption of stationarity in the process generating the data. When this assumption is violated, differencing the data and working with the differenced series can provide stationarity. The Augmented Dickey-Fuller test was used to determine stationarity of the data and the degree of differencing required to achieve stationarity. Differenced models are referred to as integrated (I). When seasonal differencing is also necessary then the models are known as seasonal ARIMA models.

The ARMA approach is powerful because in addition to p, q, and d, the coefficients ϕ_i and θ_i of an integrated ARMA model can vary depending upon the data. These are determined in such a way that the sum of the squared errors are minimized.

When the data are seasonal, there is added complexity involved in achieving stationarity. We must achieve stationarity among the same month of different years, or specifically among months which are L periods apart. The degree of differencing required for seasonal stationarity can be denoted by the parameter s. The seasonal component, depending on the process generating the data, is placed in either the MA or AR part of the model. The seasonal component (in parenthesis) is in the AR part of the ARMA model of equation 24 and the MA part in equation 25.

$$\hat{X}_t = a + \phi_1 X_{t-1} + (\phi_s X_{t-L} + \phi_1 \phi_s X_{t-L-1}) + e_t - \theta_1 e_{t-1} \qquad \textbf{(24)}$$

$$\hat{X}_t = a + \phi_1 X_{t-1} + e_t - \theta_1 e_{t-1} - (\theta_s e_{t-L} - \theta_1 \theta_s e_{t-L-1}) \qquad \textbf{(25)}$$

In the equations above the seasonal coefficient is ϕ_s. $\phi_1\phi_s$ exists because the model is multiplicative and therefore shows the combined effects of the AR (or MA in equation 25) non-seasonal and seasonal parameters.

Given the wide range of models, the major difficulty is in model selection. The Box-Jenkins methodology however provides guidelines for this choice.

4.5.5.4 Box-Jenkins methodology

The Box-Jenkins methodology is an efficient and practical procedure for identifying and testing the adequacy of an appropriate ARMA model for use in forecasting.

The autocorrelation coefficients and the partial auto correlation coefficients are two important statistics in the Box-Jenkins methodology. The autocorrelation function (ACF) describes the association among values of the same variable but at different time periods. The partial autocorrelation function (PACF) measures the degree of association between X_t and X_{t-j} when the effects of other time lags (1,2,3,...k-1) are partialled out.

The Box-Jenkins procedure consists of three stages. These are model identification, parameter estimation and diagnostic checking.

Model identification consists of first evaluating the data for stationarity, and then if needed, transforming the data to obtain it. The ADF test is used to test for stationarity. Differencing the data is one method of achieving stationarity.

Once stationarity is achieved the tentative model is identified. This consists of determining the order of the parameters p and q. The parameter d denotes the degree of differencing required to achieve stationarity. The order of the parameters p and q can be determined through examination of the PACF and ACF.

As a general rule, when the autocorrelations drop off exponentially to zero, the model is AR, and its order is determined by the number of partial autocorrelations that are significantly different from zero. If the partial autocorrelations drop off exponentially to zero, the model is MA and its order is determined by the number of statistically significant autocorrelations. When both auto correlations and partial autocorrelations drop off exponentially to zero, the model is ARMA (Makridakis and Wheelwright 1989, p 138). (See texts such as Makridakis and Wheelwright 1989, for specific guidelines).

Parameter estimation consists of determining the values of the ϕ and θ variables included in the model.

Once the model is specified it must be checked to determine its adequacy. This is accomplished through examination of the ACF and PACF of the residuals. If the model is adequate, then the AC and PAC coefficients of the residuals will not be significantly different from zero, nor will it have any pattern. If adequate the model can be used in forecasting. If not then the procedure must be repeated.

CHAPTER 5

RESULTS

5.1 FORECASTING MODELS

The following models were specified as a consequence of the model-fitting operations. The Naive model is not listed because its one step ahead forecast is always equal to the present-month actual value.

PRIMARY ANALYSIS

CABBAGE

Single Moving Average	3 yrs	n=2		
	4 yrs	n=2		
	5 yrs	n=2		
Single Exp. Smoothing	3 yrs	α=.956		
	4 yrs	α=.999		
	5 yrs	α=.993		
Winter's	3 yrs	α=1	ß=0	y=0
	4 yrs	α=1	ß=0	y=0
	5 yrs	α=1	ß=0	y=0
ARIMA	3 yrs	(0; 1; [2,3]);	sma 12.	
	4 yrs	(0; 1; [1,2])		
	5 yrs	(0; 1; [1,2])		

TOMATOES

Single Moving Average	3 yrs	n=2		
	4 yrs	n=2		
	5 yrs	n=3		
Single Exp. Smoothing	3 yrs	α=0.001		
	4 yrs	α=0.001		
	5 yrs	α=0.001		
Winter's	3 yrs	α=0.510	ß=0	y=0
	4 yrs	α=0	ß=0	y=0
	5 yrs	α=0.20	ß=.010	y=0
ARIMA	3 yrs	([1,2]; 1; 0);	sma 12	
	4 yrs	([1,2]; 1; 1);	sma 12	
	5 yrs	([2,3];1;[7,8]);	sar 12	

SECONDARY ANALYSIS

CABBAGE

Single Exp. Smoothing	3 yrs	α=.999		
	4 yrs	α=.930		
	5 yrs	α=.983		
Winter's	3 yrs	α=.780	β=0	y=0
	4 yrs	α=.920	β=0	y=0
	5 yrs	α=1.0	β=0	y=0
ARIMA	3 yrs	(2; 1; 2)		
	4 yrs	(0; 1; [10, 11])		
	5 yrs	(0; 1; [1,2])		

TOMATOES

Single Exp. Smoothing	3 yrs	α=.972		
	4 yrs	α=.001		
	5 yrs	α=.001		
Winter's	3 yrs	α=0	β=.180	y=0
	4 yrs	α=1.0	β=0	y=0
	5 yrs	α=.010	β=.110	y=0

In the notation for the ARIMA model above, the first term denotes the order of the AR model. The second term specifies the degree of differencing required for achieving stationarity, while the third term specifies the order of the MA term. The last part of the ARIMA model notation indicates the presence of seasonality and whether it is best included in the AR or MA scheme. Each term in the ARIMA model notation is separated by a semi-colon.

5.2 INTERPRETING THE RESULTS

In general we expect forecast accuracy to successively increase as the quantity of data used in model estimation is increased. This effect should be readily apparent if data is a limiting factor in achieving forecast accuracy. In a like manner we expect forecast accuracy to decrease as the number of steps-ahead forecast is increased. Whether these patterns in fact occur can be determined through examination of the tables displaying the results. The results of the primary analysis are presented in tables 5.1 - 5.5 and 5.9 - 5.13 for cabbage and tomatoes respectively. Tables 5.6 - 5.8 and 5.14 - 5.16 display the results of the secondary analysis for cabbage and tomatoes respectively.

5.3 RESULTS - CABBAGE

The values of the accuracy measures obtained, are high. For example the lowest MAPE value in the primary analysis is 32.5%. In particular the values obtained for the ARIMA model are in the extreme.

5.3.1 Most accurate model

Tables 5.1- 5.3 display the values of the accuracy measures (MSE, MAPE and %TPE respectively) resulting from the primary analysis. The asterisks in each table identify the most accurate model for the specific number of steps-ahead and a given quantity of estimation data. Tables 5.4 and 5.5 identify the most accurate model in providing forecasts for the one through three steps-ahead period, evaluating with the MSE and MAPE measures respectively. These last measures are in effect mean MSE and mean MAPE respectively.

The results indicate that selection of the "most accurate model" is influenced by the accuracy measure used. A comparison of the results obtained using the MSE, MAPE and %TPE measures (tables 5.1, 5.2 and 5.3) readily shows this. The variations are a result of the differing evaluating criteria used by each measure. For example the MAPE is a linear as opposed to the "quadratic" MSE measure. The %TPE on the other hand evaluates differences in changes of direction.

Table 5.1 reports the result of evaluating with the MSE measure. It indicates that the Naive model was most accurate at one step ahead forecasting when using three and five years estimation data, and at two steps ahead forecasting using three years data. The moving average model was best at two steps ahead forecasts using five years data and three steps ahead using three years data. The Winter's model was the most accurate model when forecasting with four years data, irrespective of the number of steps ahead. The Winter's model was also best at three steps ahead forecasting using five years estimation data.

In essence therefore the models which used deseasonalised data as input (the Naive, SES and Moving average models) all performed relatively well. So too did the Winter's model. The Arima model however was never able to achieve the distinction of being most accurate. Somewhat similar results occur when evaluating with the MAPE, mean MSE and mean MAPE. The situation is a little different when evaluating with the

%TPE measure. The ARIMA model is most accurate on two occasions. These are when forecasting two steps ahead with five years data and when forecasting three steps ahead with three years data. Even so the basic distribution of the "most accurate model" results is little affected.

The results of the secondary analysis does not indicate any great variations with the primary analysis in terms of either the magnitude of the values of the accuracy measures or the basic distribution of the "most accurate model" results.

5.3.2 Quantity of data and model accuracy

The relationship between quantity of data and model accuracy can be evaluated from the results posted in tables 5.1 - 5.3. Given the study objectives, four types of relationships can be identified when we evaluate changes in the values of the accuracy measure, as the quantity of estimation data is increased. These relationships can be described as:

- **Increasing** : describes successive decreases in the value of the accuracy measure (thus increases in accuracy).
- **Decreasing** : describes successive increases in the value of the accuracy measure (decrease in accuracy) as the quantity of estimation data is increased.
- **Inconsistent** : That is it is inconsistent in terms of being difficult to draw any meaningful conclusions given the objectives of the study. This category includes five

types of changes in the **value of the accuracy measure** in response to successive increases in the quantity of estimation data. The last two are classed as inconsistencies rather than as decreasing or increasing relationships respectively because of the inability to determine whether sustained change is occurring. These five are:

increase and then decrease

decrease and then increase

increase and then stabilize

stabilize and then increase

stabilize and then decrease

- **stabilized** : where the value of the accuracy measure either remains unchanged or at first decreases but then stabilizes.

These movements are denoted by insertion of the letters I, D, C or S respectively in the last column of the appropriate tables.

Most of the data-quantity/model-accuracy relationships reveal inconsistent movements. Few show sustained increases in accuracy as the quantity of data is increased. The instances where increasing relationships occur are greatest in the MAPE as opposed to the MSE and %TPE measures. In the MAPE results the ARIMA model posted increasing relationships at each step ahead. It accounted for more than half the number of increases in this table. However we must also note that in each case the initial MAPE value for the ARIMA model was extremely high.

The secondary analysis does not reveal any positive change in the relationships. In fact an even greater number of models posted inconsistent relationships under all three evaluation measures.

5.3.3 Accuracy of the ARIMA model

The ARIMA models specified had the highest MSE values (hence lowest accuracy) of all the models evaluated (Table 5.1). This was true at all levels of data use and steps ahead forecasted, except when four years of data were used for the one and three steps ahead forecast. There was not a single instance in the primary analysis where the ARIMA model was the most accurate model under the MSE evaluation measure.

The ARIMA models specified also had the highest MAPE values (Table 5.2) at all levels of data use and steps ahead forecasted. Exceptions were for the two and three steps ahead forecast using five years data. A somewhat similar story holds with respect to the %TPE.

Table 5.1 Primary Analysis: Cabbage MSE values

	Model	Estimation data (yrs)			
		3	4	5	
One-Step					
	ARIMA	6.34	1.94	3.83	C
	Winter's	2.46	*1.74	1.28	I
	SES	1.54	2.31	*1.21	C
	Naive	*1.51	2.31	*1.21	C
	Moving Average	2.19	2.59	1.24	C
Two-Step					
	ARIMA	10.65	2.65	5.24	C
	Winter's	5.55	*1.29	2.97	C
	SES	3.69	2.32	2.33	S
	Naive	*3.67	2.32	2.34	S
	Moving Average	3.88	2.27	*2.07	I
Three-Step					
	ARIMA	10.34	1.58	4.40	C
	Winter's	8.40	*1.13	*1.20	C
	SES	5.81	1.83	2.54	C
	Naive	5.89	1.83	2.55	C
	Moving Average	*5.39	2.19	2.46	C

* Lowest value for the number of steps ahead and quantity of data

Table 5.2 Primary Analysis: Cabbage MAPE values

	Model	Estimation data (yrs)			
		3	4	5	
One-Step					
	ARIMA	91.9	84.6	43.2	I
	Winter's	44.4	*47.4	32.9	C
	SES	37.1	66.7	*32.7	C
	Naive	36.9	66.7	32.8	C
	Moving Average	*32.5	59.4	43.0	C
Two-Step					
	ARIMA	117.8	98.0	46.8	I
	Winter's	71.4	*51.7	50.1	I
	SES	57.2	85.4	*44.5	C
	Naive	57.3	85.4	44.7	C
	Moving Average	*35.8	64.4	48.3	C
Three-Step					
	ARIMA	120.4	70.7	45.2	I
	Winter's	82.1	*39.7	37.9	I
	SES	65.7	69.3	*33.7	C
	Naive	66.0	69.3	34.0	C
	Moving Average	*41.8	60.4	54.4	C

* Lowest value for the number of steps ahead and quantity of data

Table 5.3 Primary Analysis: Cabbage %TPE values

	Model	Estimation data (yrs) 3	4	5	
One step					
	ARIMA	72.7	54.5	54.5	S
	Winter's	*45.5	*45.5	*36.4	I
	SES	*45.5	54.5	45.5	C
	Naive	*45.5	*45.5	45.5	C
	Moving average	54.5	54.5	*36.4	I
Two step					
	ARIMA	60.0	80.0	*50.0	C
	Winter's	*40.0	50.0	*50.0	C
	SES	*40.0	50.0	60.0	D
	Naive	50.0	50.0	*50.0	C
	Moving average	50.0	*40.0	*50.0	C
Three step					
	ARIMA	*44.4	55.5	55.5	D
	Winter's	66.7	55.5	*11.1	I
	SES	66.7	*33.3	44.4	C
	Naive	66.7	*33.3	44.4	C
	Moving average	66.7	44.4	33.3	I

* Lowest value for the number of steps ahead and quantity of data

Table 5.4 Primary Analysis: Cabbage mean MSE values

Data (yrs)	Models ARIMA	Winter's	SES	Naive	M Avg.
3	9.11	5.47	*3.68	3.69	3.82
4	2.06	*1.39	2.15	2.15	2.35
5	4.49	*1.82	2.03	2.03	1.92

* Lowest value for the specified quantity of data.

Table 5.5 Primary Analysis: Cabbage mean MAPE values

Data (yrs)	Models				
	ARIMA	Winter's	SES	Naive	M avg.
3	110.00	65.97	53.33	53.40	*36.70
4	84.43	*46.27	73.80	73.80	61.40
5	45.07	40.30	*36.94	37.17	48.57

* Lowest value for the specified quantity of data

Table 5.6 Secondary Analysis: Cabbage MSE values

	Model	Estimation data (yrs)			
		3	4	5	
One-Step					
	ARIMA	2.86	4.75	3.83	C
	Winter's	2.02	*1.21	1.38	C
	SES	3.07	2.72	*4.94	C
	Naive	3.07	3.07	5.07	C
Two-Step					
	ARIMA	3.25	3.21	4.81	C
	Winter's	3.85	2.86	3.25	C
	SES	*2.22	*2.05	*2.63	C
	Naive	*2.22	2.22	2.67	C
Three-Step					
	ARIMA	2.58	2.89	3.74	D
	Winter's	4.52	3.36	3.37	S
	SES	*2.32	*2.25	*2.86	C
	Naive	*2.32	2.32	2.88	C

* Lowest value for the number of steps ahead and quantity of data

Table 5.7 Secondary Analysis: Cabbage MAPE values

	Model	Estimation data (yrs) 3	4	5	
One-Step					
	ARIMA	67.0	71.3	43.2	C
	Winter's	*42.4	*31.3	*31.2	S
	SES	48.7	46.7	47.4	C
	Naive	48.7	48.7	48.3	C
Two-Step					
	ARIMA	68.4	*42.4	*43.1	C
	Winter's	58.0	51.9	52.6	C
	SES	*50.5	47.8	50.0	C
	Naive	*50.5	50.5	50.8	C
Three-Step					
	ARIMA	59.3	*39.2	*39.9	S
	Winter's	*49.6	40.5	42.5	C
	SES	52.3	51.6	57.5	C
	Naive	52.3	52.3	57.9	C

* Lowest value for the number of steps ahead and quantity of data

Table 5.8 Secondary Analysis: Cabbage %TPE values

	Model	Estimation data (yrs) 3	4	5	
One step					
	ARIMA	54.6	*36.4	54.5	C
	Winter's	*36.4	*36.4	*36.4	S
	SES	45.5	45.5	54.5	C
	Naive	45.5	45.5	54.5	C
Two step					
	ARIMA	*36.4	*27.3	*54.5	C
	Winter's	81.8	54.5	*54.5	S
	SES	63.6	63.6	63.6	S
	Naive	63.6	63.6	63.6	S
Three step					
	ARIMA	*36.4	63.6	54.5	C
	Winter's	63.6	45.5	54.5	C
	SES	*36.4	*36.4	*36.4	S
	Naive	*36.4	*36.4	*36.4	S

* Lowest value for the number of steps ahead and quantity of data

5.4 <u>RESULTS - TOMATO</u>

5.4.1 Most accurate model

The asterisk in each table identifies the most accurate model when forecasting for a specified number of steps ahead and a given quantity of data. Tables 5.9 - 5.13 and 5.14 - 5.16 report the results of the primary and secondary analyses respectively. Tables 5.12 and 5.13 in effect report mean MSE and MAPE values.

Examination of tables 5.9 and 5.10 reveal that in almost all cases of the primary analyses, the most accurate model was the Winter's or the SES model. The one exception was the result for three steps ahead forecasts using three years of data where the ARIMA model was best under both accuracy measures. Further, models found to be most accurate with one measure tended to hold this position when evaluated with the other measure. Evaluating with the %TPE measure results in a much more dispersed distribution of the "most accurate model" results. The Winter's and SES models were most accurate when evaluating with the mean MSE and MAPE measures.

The secondary analysis conducted using the Winter's, SES and Naive models little affected the above reported results.

5.4.2 Quantity of data and model accuracy

The last column in each table classifies the relationship that exists between model accuracy and quantity of estimation data. The notation used and the relevant definitions are set out in section 5.3.2.

The primary results indicate that most of the relationships can be described as either decreasing or inconsistent. Instances of increasing relationships are provided by the ARIMA model in its one step ahead forecasts evaluated with the MSE measure and its one and two steps ahead forecasts evaluated with the MAPE measure. The ARIMA model in these cases however had extremely high initial (i.e. with three years data) values. When evaluated with the %TPE measure all the relationships were classified as either decreasing or inconsistent.

The secondary analysis showed one model with a relationship classified as increasing. This was the Winter's model when three steps ahead forecasting using the MAPE measure. All other models had decreasing, inconsistent or stabilized relationships.

5.4.3 Accuracy of the ARIMA model

The ARIMA models specified had the largest MAPE values for the one step ahead and two steps ahead forecasts (Table 5.10). However it had the smallest MAPE value for the three step ahead forecast using three year of data.

Essentially the same results are obtained when the evaluation is based on the MSE measure (Table 5.9). The ARIMA model was not included in the secondary analysis.

Table 5.9 Primary Analysis: Tomato MSE values

	Model	Estimation data (yrs)			
		3	4	5	
One-Step					
	ARIMA	18.83	7.99	7.57	I
	Winter's	*1.10	*1.50	4.08	D
	SES	1.38	1.71	*2.95	D
	Naive	1.66	2.41	4.05	D
	Moving Average	1.47	2.96	4.56	D
Two-Step					
	ARIMA	9.26	10.17	6.40	C
	Winter's	*1.30	*1.53	4.50	D
	SES	1.47	1.84	*2.99	D
	Naive	2.73	5.08	7.63	D
	Moving Average	1.89	5.04	6.15	D
Three-Step					
	ARIMA	*0.62	6.75	4.15	C
	Winter's	0.97	*1.61	4.95	D
	SES	1.61	1.91	*3.29	D
	Naive	1.79	7.01	4.76	C
	Moving Average	1.23	6.04	3.70	C

* Lowest value for the number of steps ahead and quantity of data

Table 5.10 Primary Analysis: Tomato MAPE values

	Model	Estimation data (yrs)			
		3	4	5	
One-Step					
	ARIMA	128.2	63.2	61.7	I
	Winter's	24.6	*26.1	34.5	D
	SES	*22.6	33.2	*27.0	C
	Naive	27.3	26.4	36.4	C
	Moving Average	22.7	45.7	40.5	C
Two-Step					
	ARIMA	88.5	69.6	52.7	I
	Winter's	28.8	*26.8	34.2	C
	SES	*22.8	35.4	*25.7	C
	Naive	33.3	49.3	47.9	C
	Moving Average	25.2	49.2	45.6	C
Three-Step					
	ARIMA	*14.7	41.8	37.4	C
	Winter's	23.4	*27.0	36.7	D
	SES	24.0	37.5	*28.2	C
	Naive	29.9	66.7	43.0	C
	Moving Average	23.9	56.9	37.0	C

* Lowest value for the number of steps ahead and quantity of data

Table 5.11 Primary Analysis: Tomato %TPE values

	Model	Estimation data (yrs)			
One-step		3	4	5	
	ARIMA	45.5	36.4	45.5	C
	Winter's	45.5	*00.0	*36.4	C
	SES	27.3	18.2	*36.4	C
	Naive	*18.2	18.2	45.5	D
	Moving average	*18.2	27.3	*36.4	D
Two step					
	ARIMA	*20.0	40.0	*40.0	C
	Winter's	50.0	*00.0	*40.0	C
	SES	30.0	20.0	*40.0	C
	Naive	70.0	30.0	70.0	C
	Moving average	40.0	40.0	70.0	D
three step					
	ARIMA	*22.2	77.8	55.6	C
	Winter's	44.4	*00.0	55.6	C
	SES	33.3	22.2	44.4	C
	Naive	44.4	22.2	44.4	C
	Moving average	33.3	33.3	*33.3	C

* Lowest value for the number of steps ahead and quantity of data

Table 5.12 Primary Analysis: Tomato mean MSE values

Data (yrs)	Models				
	ARIMA	Winter's	SES	Naive	M avg.
3	9.57	*1.12	6.05	2.06	1.53
4	8.30	*1.55	1.85	4.83	4.68
5	6.04	4.51	*3.08	5.48	4.80

* Lowest value for the specified quantity of data

Table 5.13 Primary Analysis: Tomato mean MAPE values

Data (yrs)	Models				
	ARIMA	Winter's	SES	Naive	M avg.
3	77.13	25.60	*23.13	30.17	23.93
4	58.20	*26.63	35.37	47.47	50.60
5	50.60	35.13	*26.97	42.43	41.03

* lowest value for the specified quantity of data

Table 5.14 Secondary Analysis: Tomato MSE values

	Model	Estimation data (yrs)			
		3	4	5	
One-Step					
	Winter's	*3.82	5.59	4.06	C
	SES	4.17	*3.13	*3.18	C
	Naive	6.92	5.97	6.92	C
Two-Step					
	Winter's	*3.77	6.96	4.18	C
	SES	5.77	*3.13	*3.18	C
	Naive	7.85	7.00	7.85	C
Three-Step					
	Winter's	*3.73	8.63	4.12	C
	SES	9.55	*3.13	*3.17	C
	Naive	9.22	9.81	9.22	C

* Lowest value for the number of steps ahead and quantity of data

Table 5.15 Secondary Analysis: Tomato MAPE values

	Model	Estimation data (yrs)			
		3	4	5	
One-Step					
	Winter's	*30.3	41.0	32.8	C
	SES	92.0	*27.2	*28.9	C
	Naive	93.2	93.2	100.0	C
Two-Step					
	Winter's	*29.9	58.8	32.7	C
	SES	72.1	27.2	*28.9	C
	Naive	73.6	73.6	71.0	C
Three-Step					
	Winter's	99.3	53.6	32.5	I
	SES	36.6	27.2	28.9	C
	Naive	37.4	37.5	38.9	D

* Lowest value for the number of steps ahead and quantity of data

Table 5.16 Secondary Analysis: Tomato %TPE values

	Model	Estimation data (yrs)			
		3	4	5	
One step					
	Winter's	*54.6	*45.5	*45.5	S
	SES	63.6	*45.5	54.5	C
	Naive	63.6	63.6	*45.5	C
Two step					
	Winter's	*54.5	54.5	45.5	C
	SES	72.7	*45.5	*27.3	I
	Naive	72.7	72.7	81.8	C
Three step					
	Winter's	54.5	*45.5	*36.4	I
	SES	*45.5	*45.5	*36.4	C
	Naive	*45.5	54.5	45.5	C

* Lowest value for the number of steps ahead and quantity of data

CHAPTER 6

CONCLUSION

6.1 DISCUSSION

All models have relatively high error values under all three accuracy measures used. In particular the error values for the ARIMA model were in the extreme, sometimes above 100%.

Several of the models specified (both primary and secondary) are close to the Naive model. This may indicate that both data sets are adequately represented by a model of this type. The Moving average model where n=1 and the SES model with an alpha value of one ($\alpha = 1$) will both produce results identical to the Naive model when the same data are used as input. In this study the Moving average models have values of n=2 or n=3, but then specification of these as n=1 models was not attempted. Many of the SES models specified have high alpha values (above .9), consequently forecasts and error values for these SES and Naive models tend to be close. In this regard the SES and Naive models for the cabbage data-sets provide good examples (see Tables C1 - C3 and Tables 5.1 - 5.8).

In the great majority of cases the simpler models had the distinction of being declared most accurate for the given quantity of data and steps-ahead. The results show that the ARIMA model was hardly ever most accurate.

The relationship between the value of the accuracy measure and the quantity of estimation data as it is successively increased, has implications for evaluating the objectives of the study. The main objective was to determine if "existing models/techniques can provide accurate forecasts given a short data-set". To test this the relationships between forecast accuracy and quantity of data were evaluated for each identified model.

An increasing relationship (ie. forecast accuracy increases as the quantity of data is increased) implies that data is limiting. A stabilized relationship implies that data is not a limiting factor in forecast accuracy. This of course rests on the assumption that five years of data is sufficient for accurate forecasts. A decreasing relationship is difficult to interpret. In general it is felt that the greater the quantity of estimation data the more accurate the forecast obtained. However as was pointed out earlier, good fit in the model estimation phase does not guarantee accurate forecasts. Notwithstanding this we do not expect accuracy to decrease as the quantity of estimation data is increased. Inconclusive results can be explained for the simpler (non-statistical methods) if three years of data is more than adequate for their specification and operation. Then random movements in accuracy could be expected based on movements of the time series. The number of years over which the test were conducted are inadequate to test this hypothesis.

The majority of accuracy/quantity relationships were inconclusive, decreasing or stabilized. Very few relationships were classified as increasing. Thus the results of our analysis, in terms of determining the first objective, are inconclusive.

The second objective was to determine whether the ARIMA models would provide more accurate forecasts than the other applicable extrapolative models, given a short data series. The results indicate that this did not occur. In fact the simple methods of Moving average, Naive, SES and Winter's all performed in most cases much better than the ARIMA model. While there is a certain amount of specificity to the data in studies of this kind, the results are in accord with those obtained in the accuracy studies conducted by Makridakis. The studies by Makridakis, however did not take into account limitations on the length of the data series.

One weakness in the study is that the methodology made no attempt to investigate reasons for this poor performance of the ARIMA model. Poor performance could be due to inadequate model specification and/or the effects of the short data series. The ARIMA model had the highest number of increasing relationships (i.e the highest error values occurred when the least amount of data was used). We can conclude from this that length of the data series is a factor in the poor performance of the ARIMA model. The methodology used to specify the ARIMA model (see Appendix E) allowed acceptable random residuals to be obtained. However unlike the other models, the ARIMA model involved the greatest subjective input. Tests were not conducted to determine if model specification was a factor in the poor performance of the ARIMA model. The methodology does not allow us to conclusively make such a statement, thus we cannot conclude that the ARIMA model performs worse than the other models <u>when</u> the data series is short.

Given the main objective of this study the results are inconclusive. This result is in part due to weaknesses in the study methodology. These include:

- the inability to state an absolute value for each accuracy measure thus allowing us to judge whether model performance is acceptable. However no such absolute values exist, in part because an acceptable level of accuracy depends on the user, the use and the forecaster.
- inability to prove conclusively that the ARIMA models were well specified and thus argue that its poor performance is due solely to limitations imposed by the quantity of estimation data.

6.2 RECOMMENDATIONS

Further study is needed as this study was exploratory in nature. Further work must however take account of the weaknesses of this study. That may well require:

- definition of absolute values for the accuracy measures used, or the development of new measures;
- use of ARIMA models specified by experts in the field. This is the approach often adopted by Makridakis and others when conducting comparison studies of this kind;
- use of an approach (other than the case study) that allows statistical determination of the significance of the data/accuracy relationships.

There is a need for training of forecasters and beginning forecasters, with the emphasis on an evolutionary rather than arbitrary model selection process. This training must highlight the range of models available and appropriate user friendly software packages.

BIBLIOGRAPHY

Armstrong, Scott J., (1985) Long-Range Forecasting From Crystal Ball to Computer, John Wiley & Sons, New York, 2nd edn.

Armstrong, Scott J., (1978) "Forecasting with Econometric Methods: Folklore versus Fact", Journal of Business, vol. 51. (Reprinted in Makridakis et al. (1984) The Forecasting Accuracy of Major Time Series Methods, John Wiley & Sons, New York.)

Chambers, John C., Mullick, Satinder K., and Smith, Donald D., (1974) An Executives Guide To Forecasting, John Wiley & Sons, New York.

Cleary, James P., and Levenbach, Hans, (1982) The Professional Forecaster: The Forecasting Process Through Data Analysis, Lifetime Learning Publications, Belmont, California.

Compton, Joan C., and Compton, Stephen B., (1990) Successful Business Forecasting, Liberty Hall Press, Blue Ridge Summit, Pennsylvania.

Fildes, Robert, (1982) "Forecasting: The Issues", in Makridakis, Spyros, and Wheelwright, Steven, (eds), The Handbook of Forecasting: A Managers Guide, John Wiley & Sons, New York.

Kress, George, (1985) Practical Techniques of Business Forecasting, Quorum Books, Westport, Connecticut.

Lilien, David, Hall, Robert E., johnson, Jack, (1990) MicroTSP User's Manual version 7.0, Quantitative Micro Software, California.

Makridakis, Spyros, (1984) "Forecasting: State of the Art", in Makridakis et al., The Forecasting Accuracy of Major Time Series Methods, John Wiley & Sons, New York.

Makridakis, Spyros, and Hibon, Michele, (1979) "Accuracy of Forecasting: An Empirical Investigation (with Discussion)", Journal of the Royal Statistical Society, Series A, vol.142, Part 2. (Reprinted in Makridakis et al., (1984) The Forecasting Accuracy of Major Time Series Methods, John Wiley & Sons, New York.)

Makridakis et al., (1982) "The Accuracy of Extrapolation (Time Series) Methods: Results of a Forecasting Competition", Journal of Forecasting, vol. 1. (Reprinted in Makridakis et al., (1984) The Forecasting Accuracy of Major Time Series Methods, John Wiley & Sons, New York.)

Makridakis, Spyros, and Wheelwright, Steven C., (1978) Interactive Forecasting: Univariate and Multivariate Methods, Holden-Day, Inc., San Francisco, (2nd edn.)

O'Donovan, Thomas, (1983) Short Term Forecasting: An Introduction to the Box-Jenkins Approach, John Wiley & Sons, New York.

Rycroft, Robert S., (1989) Microcomputer Software in Comparative Review, International Journal of Forecasting 5 (pp 437 - 462), Elsevier Sciences Publications (North Holland).

SPSS Inc., (1990) SPSS/PC + Trends, SPSS Inc. USA.

Steece, Bert, (1982) "Evaluation of Forecasts", in Makridakis, S., and Wheelwright, Steven C. (eds), The Handbook of Forecasting: A Managers Guide, John Wiley & Sons, New York.

Stewart, Thomas, (1987) "The Delphi Technique and Judgemental Forecasting", in Forecasting in the Social and Natural Sciences pp 97 - 109, ed. Land, Kenneth C., and Schneider, Stephen, D. Reidel Publishing Company, Boston. (Reprinted from Climatic Change, Vol II, nos. 1-2 (1987).

Wheelwright, Steven, and Makridakis, Spyros, (1989) Forecasting Methods For Management, John Wiley & Sons, New York.

Wright, George, and Ayton, Peter, (1987) Judgemental Forecasting, John Wiley and Sons, New York.

APPENDIX A

PRIMARY ANALYSIS - TOMATO

Plots of actual and forecasted values

APPENDIX A

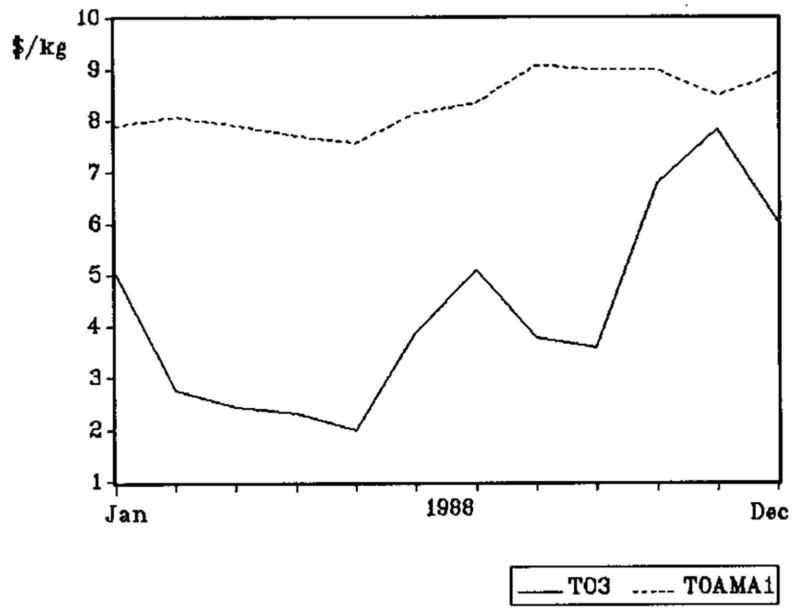

Figure A1 1-step ahead ARIMA forecasts (3 yrs estimation data)

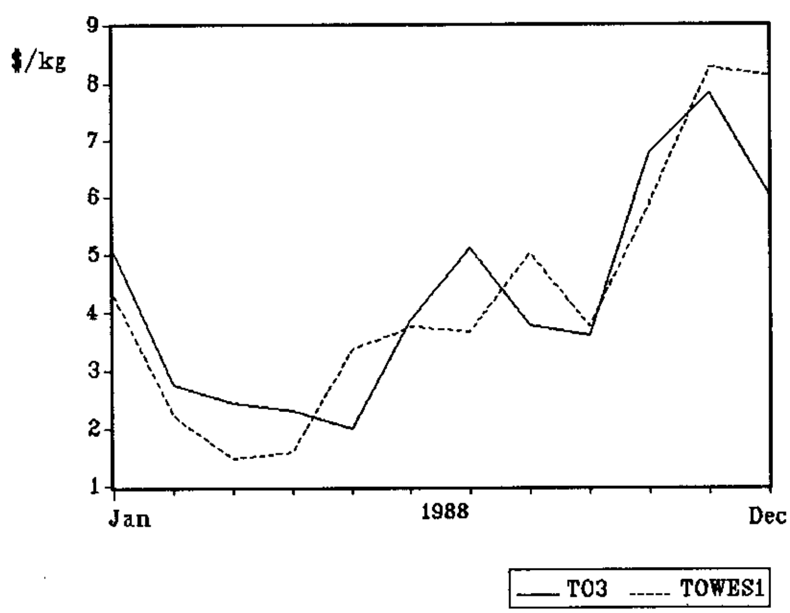

Figure A2 1-step ahead Winters forecasts (3 yrs estimation data)

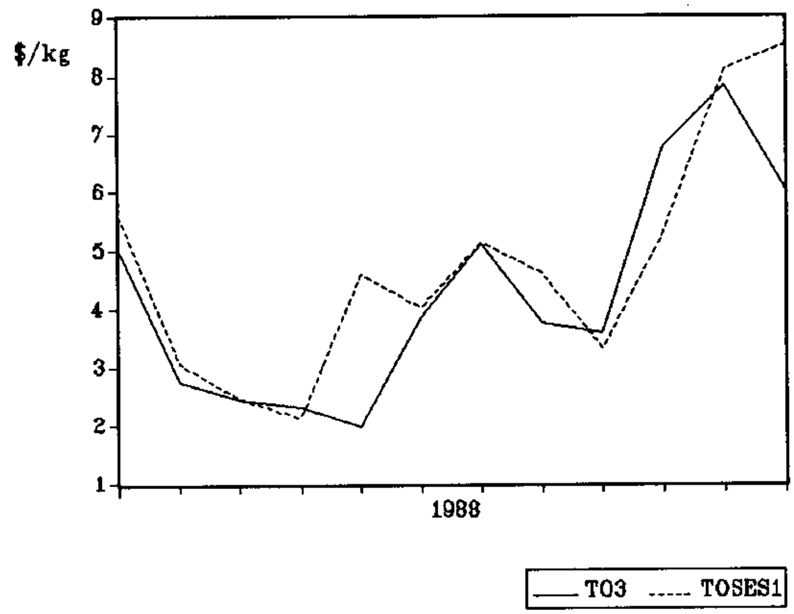

Figure A3 1-step ahead SES forecasts (3 yrs estimation data)

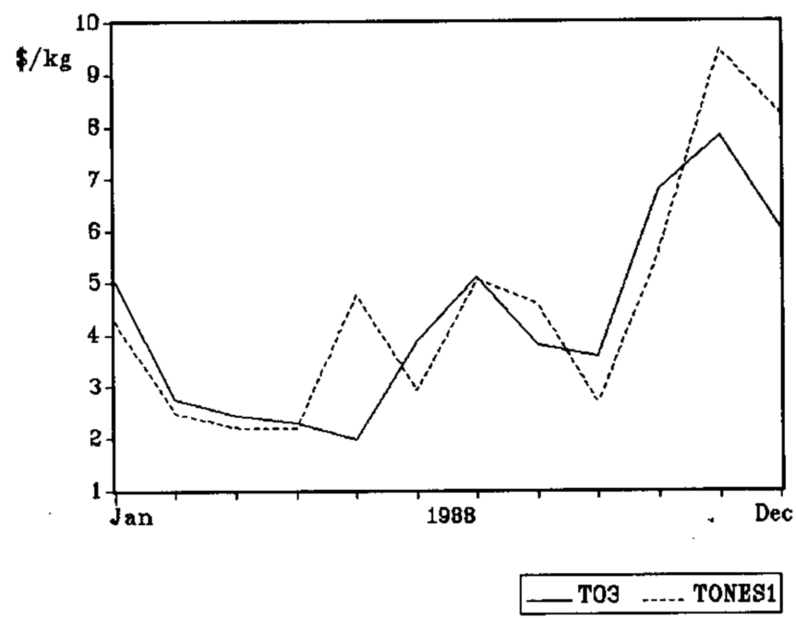

Figure A4 1-step ahead Naive forecasts (3 yrs estimation data)

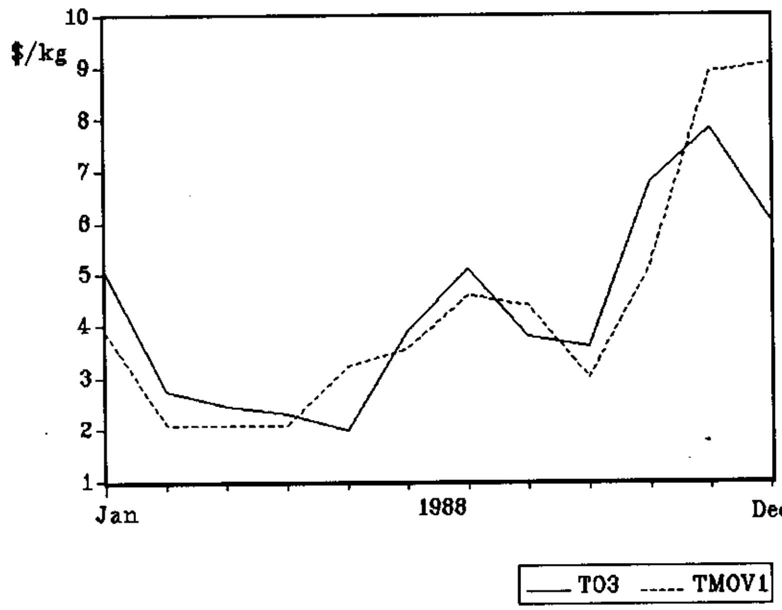

Figure A5 1-step ahead M avg. forecasts (3 yrs estimation data)

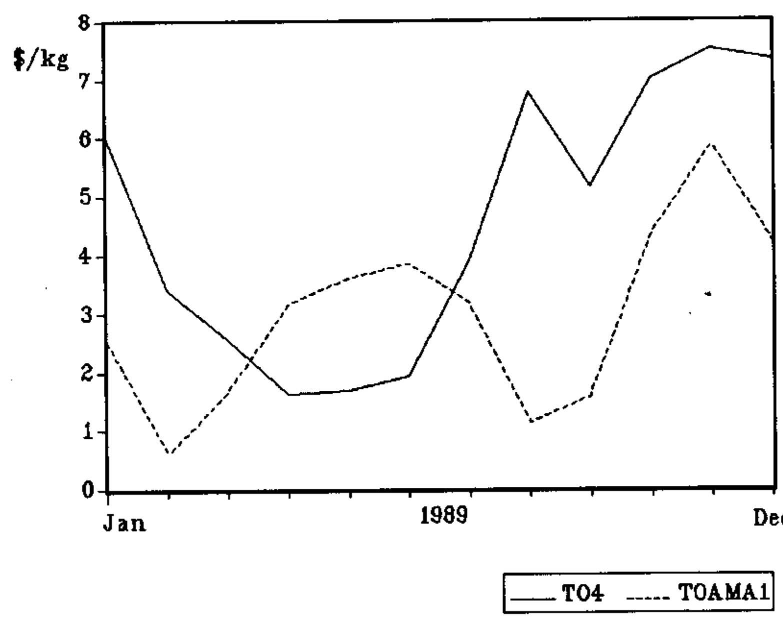

Figure A6 1-step ahead ARIMA forecasts (4 yrs estimation data)

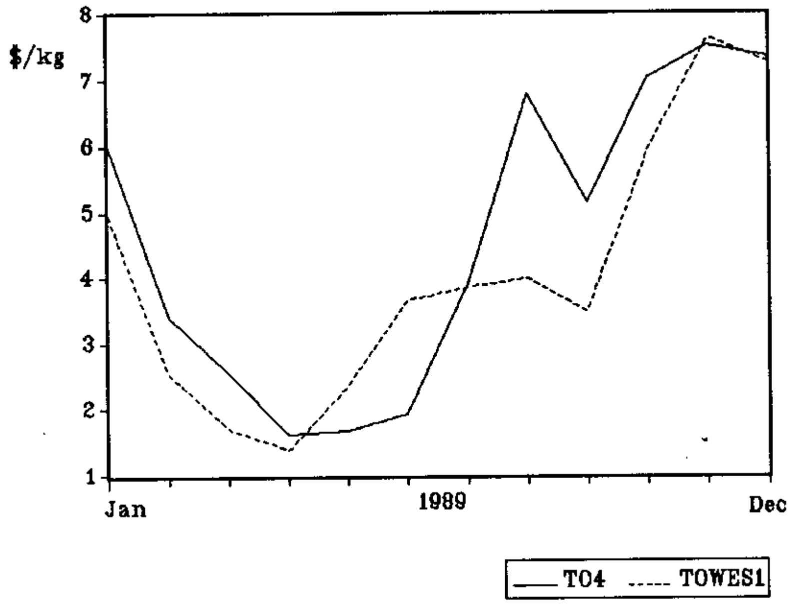

Figure A7 1-step ahead Winters forecasts (4 yrs estimation data)

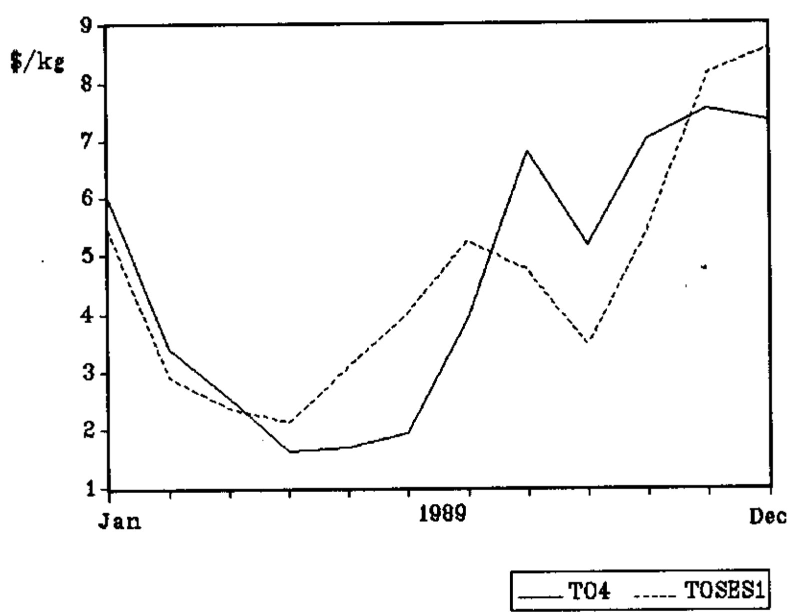

Figure A8 1-step ahead SES forecasts (4 yrs estimation data)

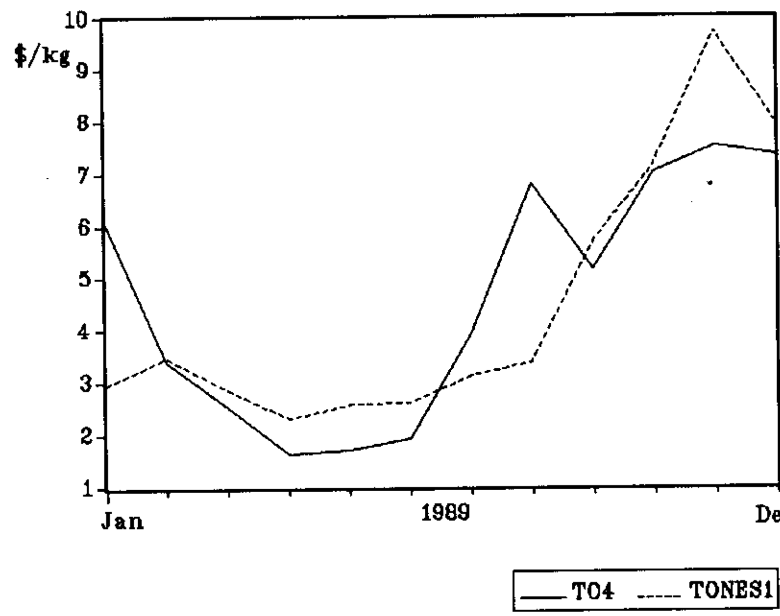

Figure A9 1-step ahead Naive forecasts (4 yrs estimation data)

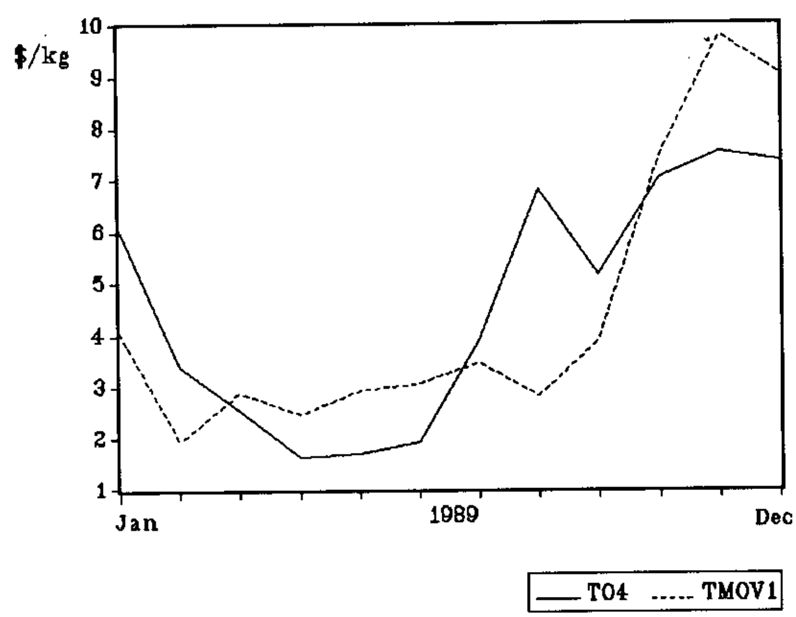

Figure A10 1-step ahead M avg. forecasts (4 yrs estimation data)

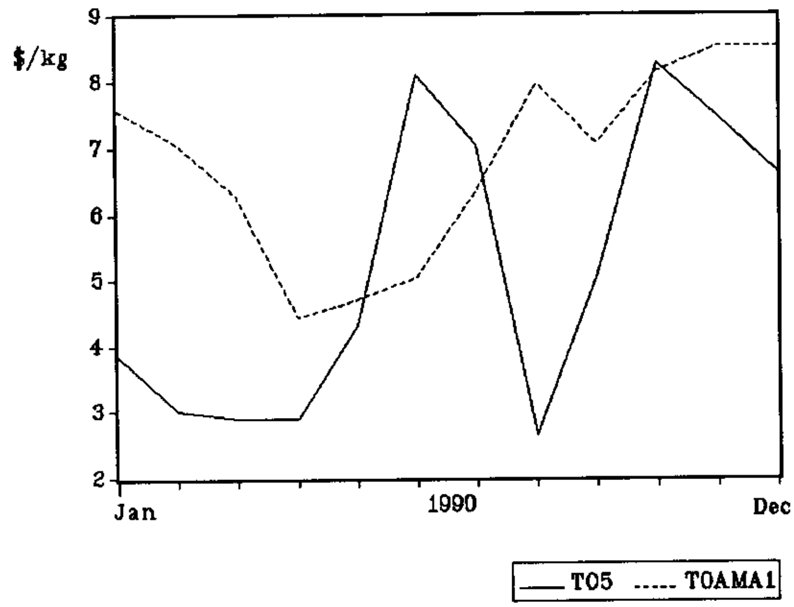

Figure A11 1-step ahead ARIMA forecasts (5 yrs estimation data)

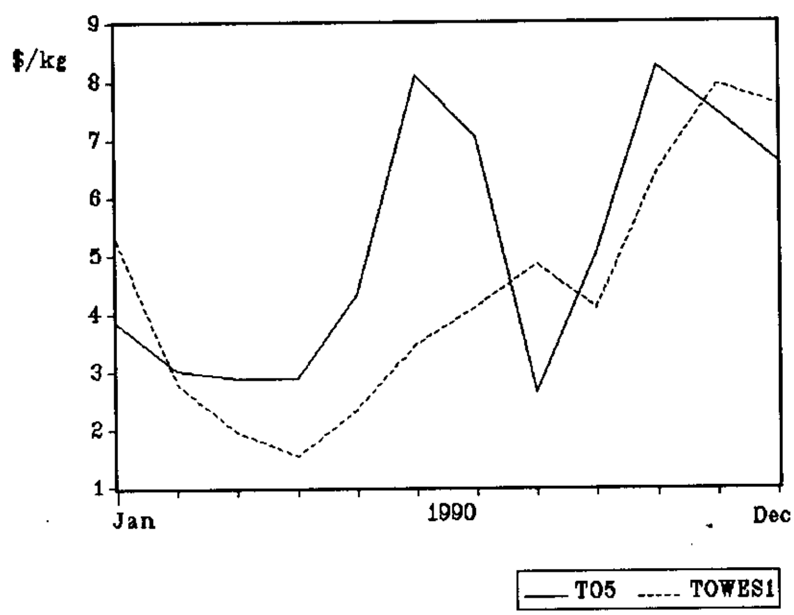

Figure A12 1-step ahead Winters forecasts (5 yrs estimation data)

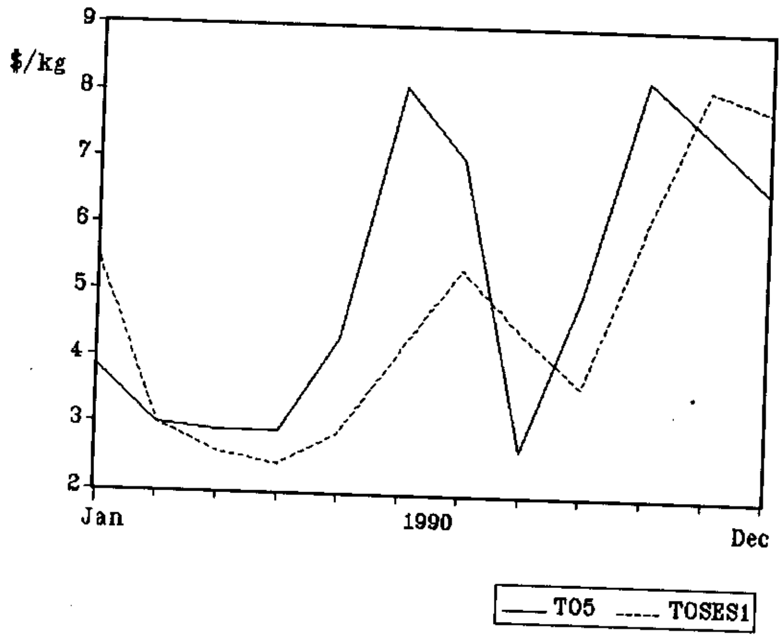

Figure A13 1-step ahead SES forecasts (5 yrs estimation data)

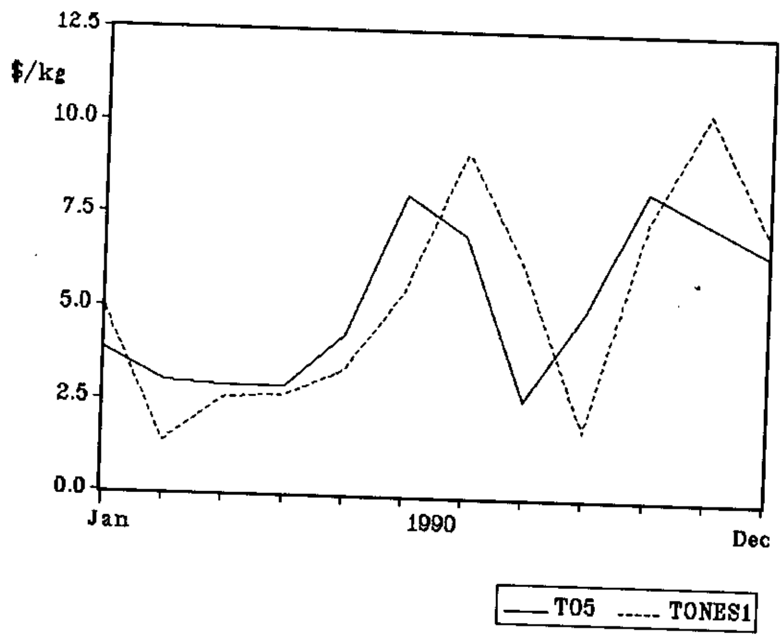

Figure A14 1-step ahead Naive forecasts (5 yrs estimation data)

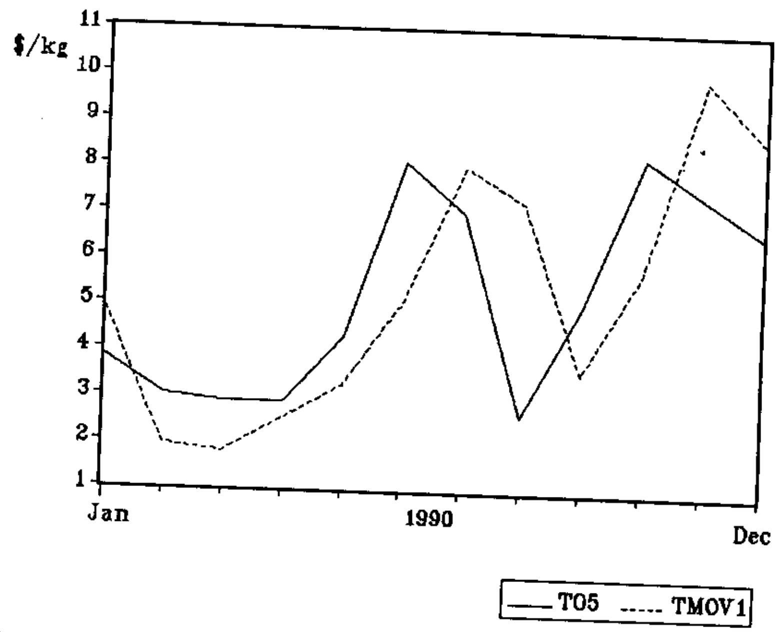

Figure A15 1-step ahead M avg. forecasts (5 yrs estimation data)

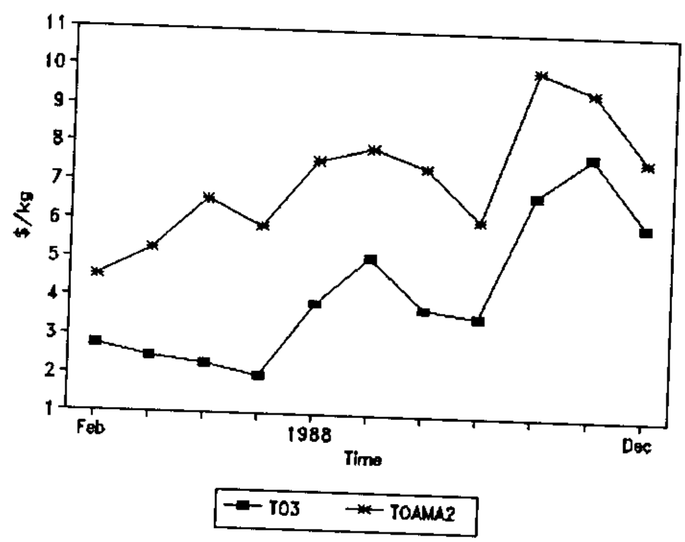

Figure A16 2-steps ahead ARIMA forecasts (3 yrs estimation data).

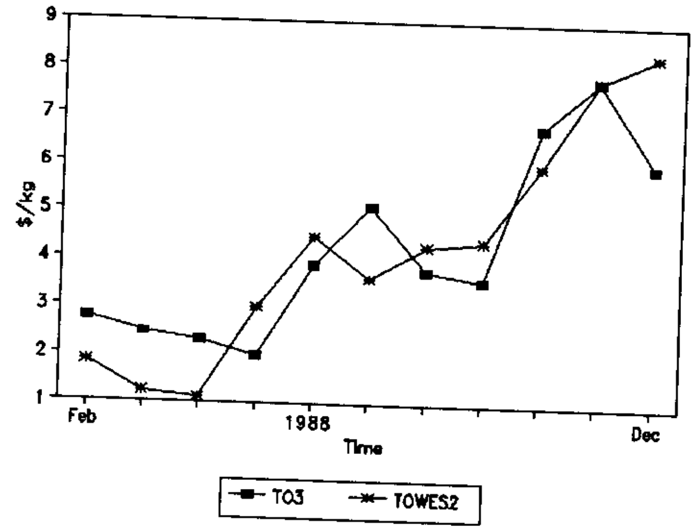

Figure A17 2-steps ahead Winters forecasts (3 yrs estimation data)

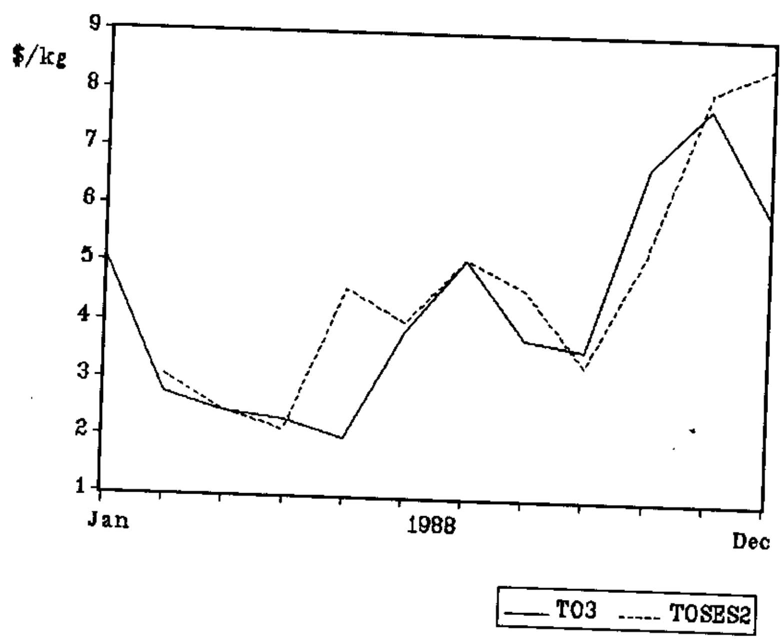

Figure A18 2-steps ahead SES forecasts (3yrs estimation data)

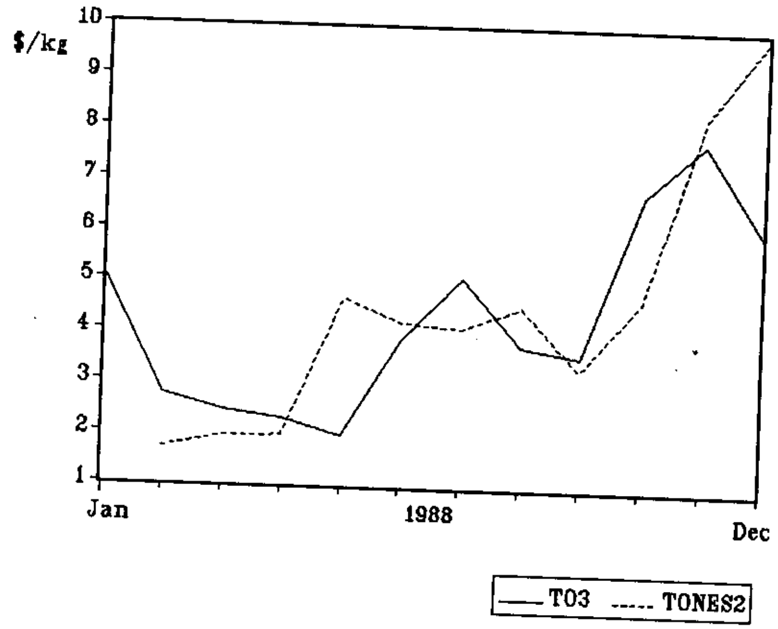

Figure A19 2-steps ahead Naive forecasts (3 yrs estmation data)

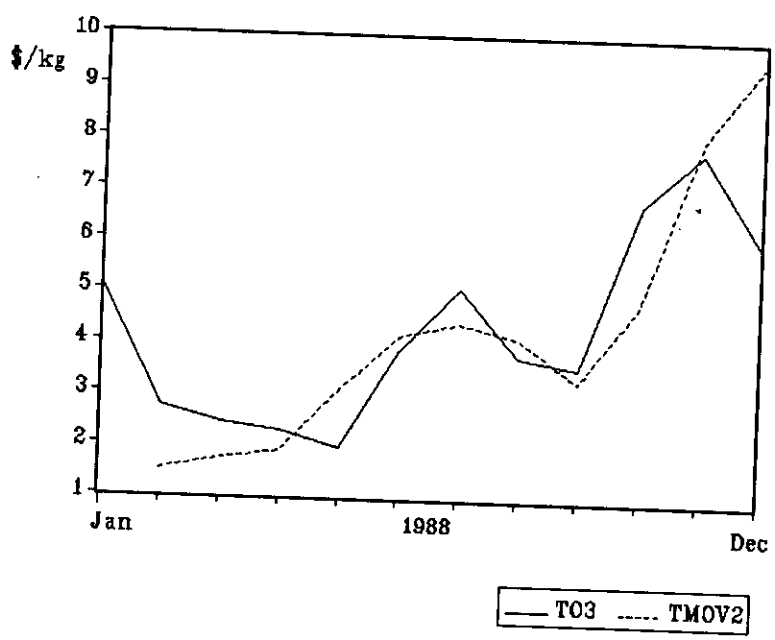

Figure A20 2-steps ahead M avg. forecasts (3 yrs estimation data)

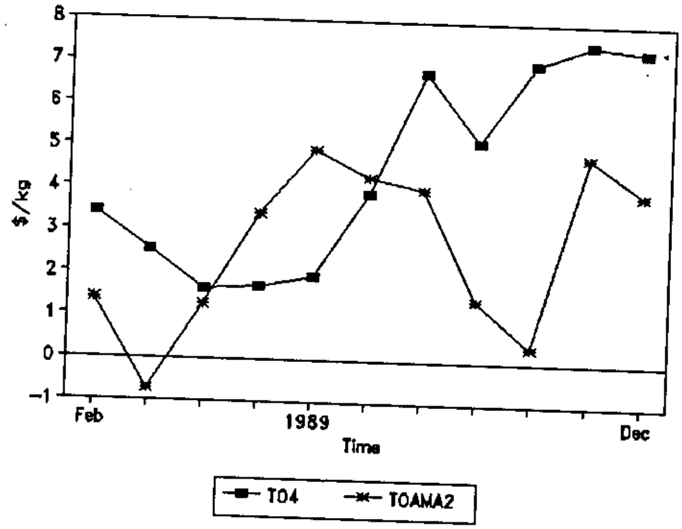

Figure A21 2-steps ahead ARIMA forecasts
(4 yrs estimation data)

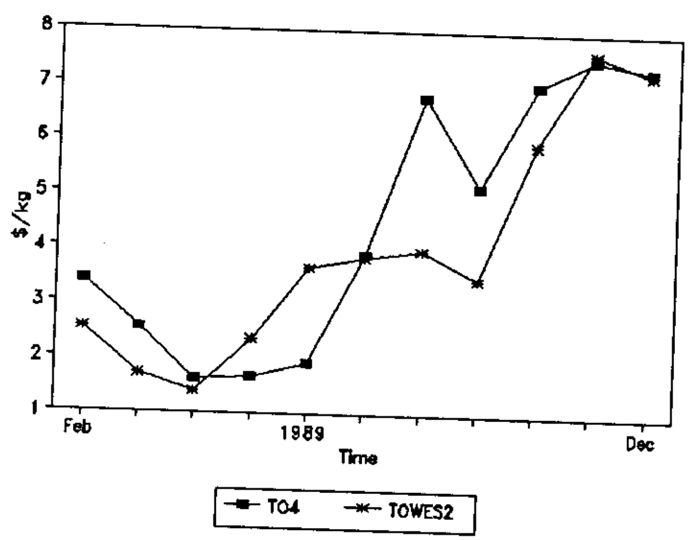

Figure A22 2-steps ahead Winters forecasts
(4 yrs estimation data)

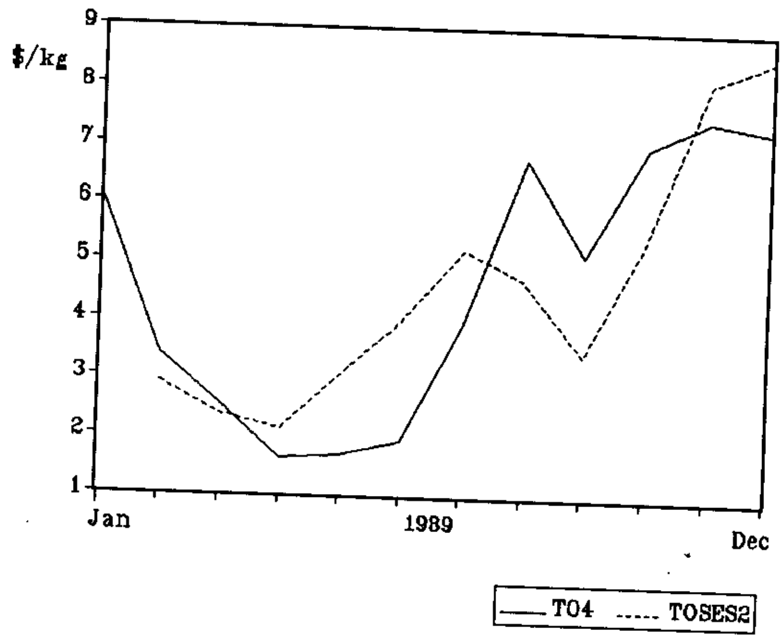

Figure A23 2-steps ahead SES forecasts (4 yrs estimation data)

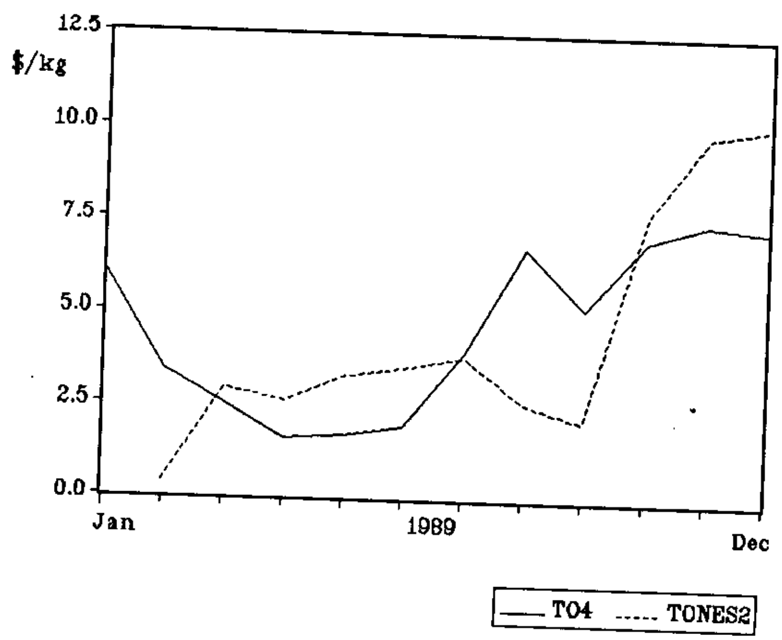

Figure A24 2-steps ahead Naive forecasts (4 yrs estimation data)

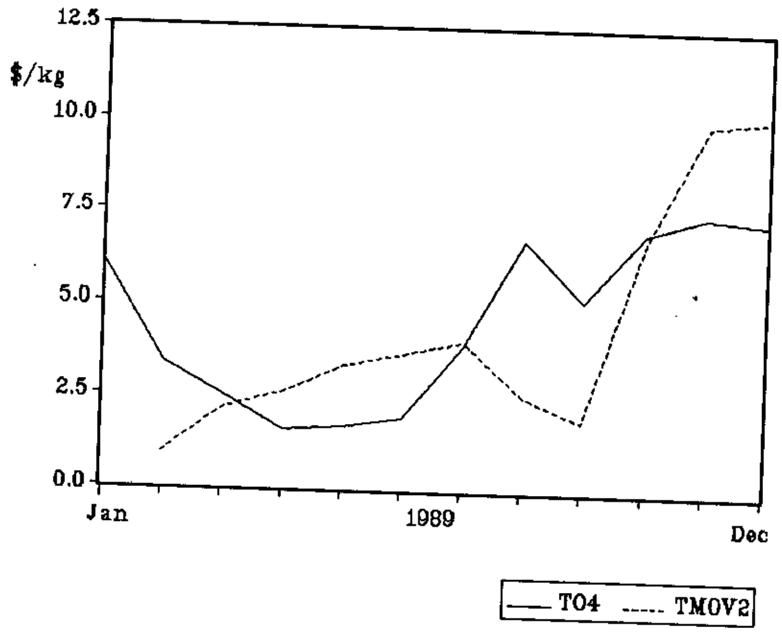

Figure A25 2-steps ahead M avg. forecasts (4 yrs estimation data)

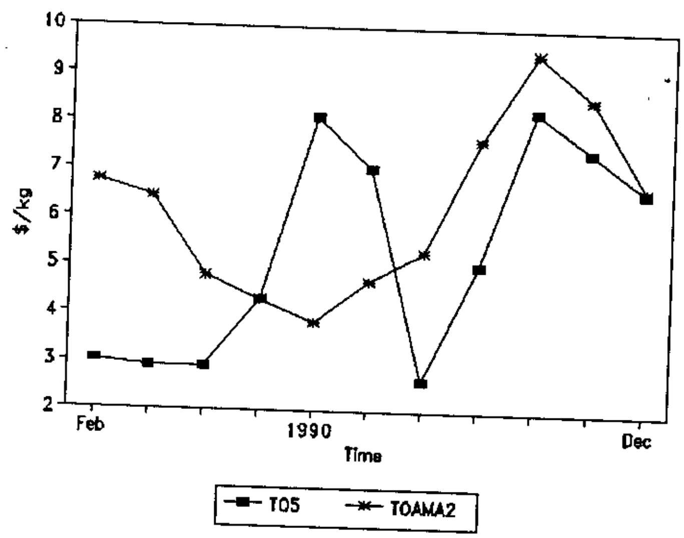

Figure A26 2-steps ahead ARIMA forecasts (5 yrs estimation data)

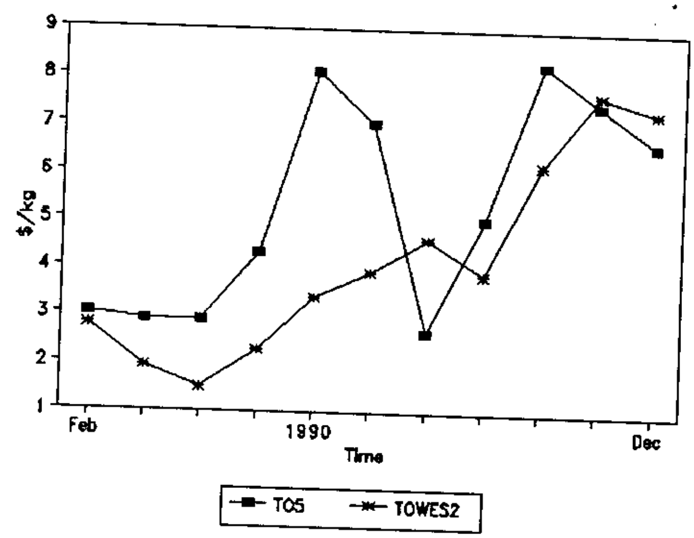

Figure A27 2-steps ahead Winters forecasts (5 yrs estimation data)

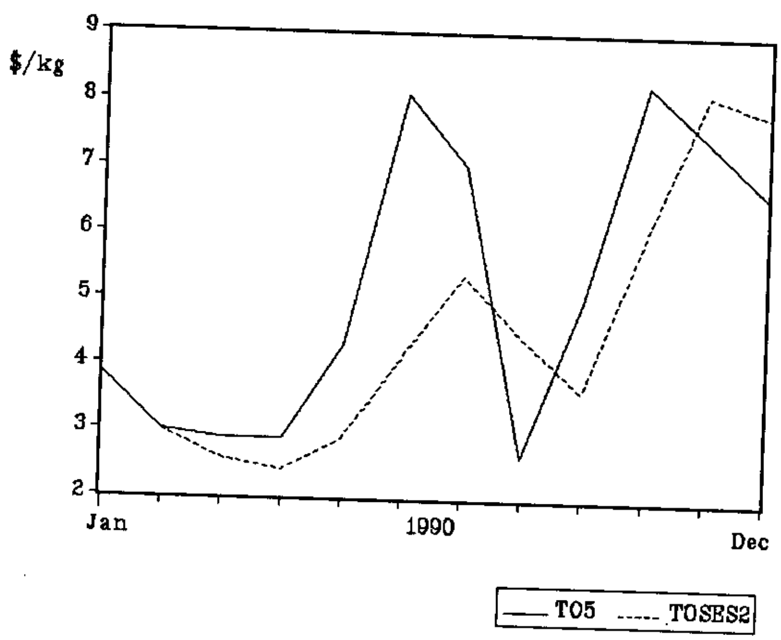

Figure A28 2-steps ahead SES forecasts (5 yrs estimation data)

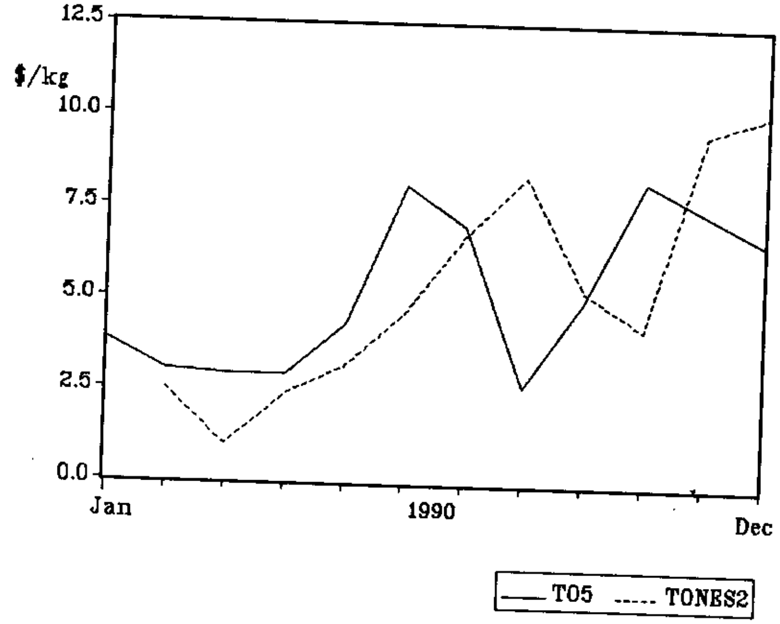

Figure A29 2-steps ahead Naive forecasts (5 yrs estimation data)

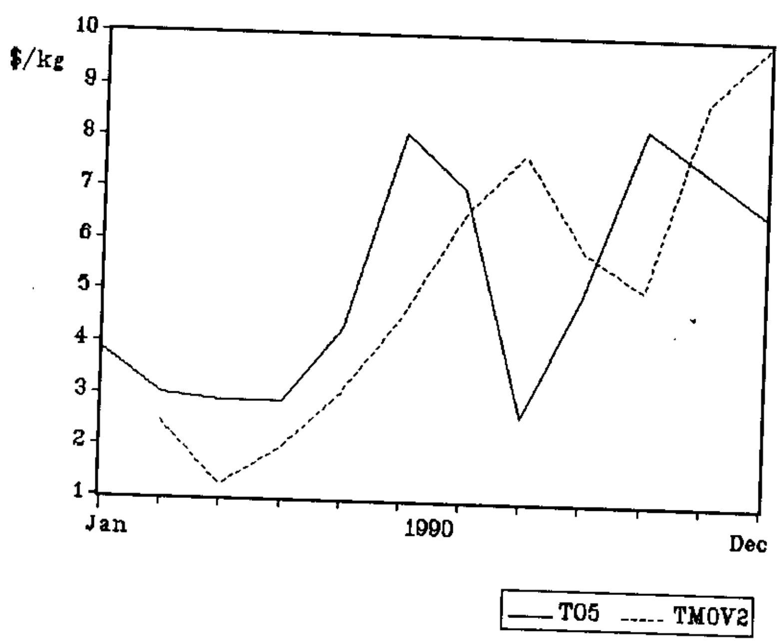

Figure A30 2-steps ahead M avg. forecasts (5 yrs estimation data)

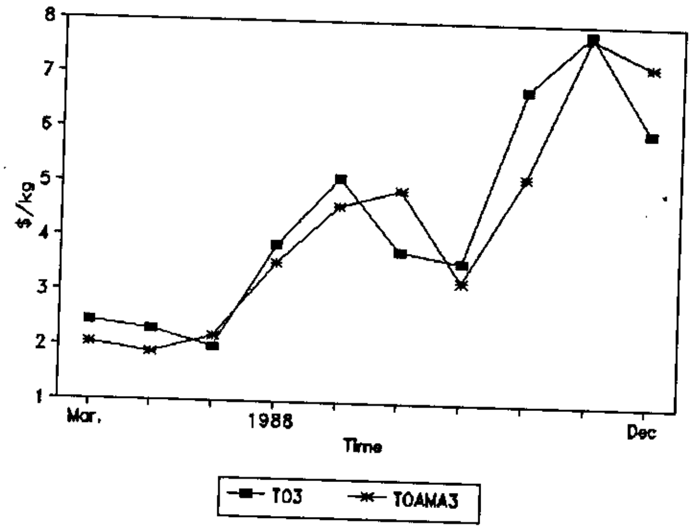

Figure A31 3-steps ahead ARIMA forecasts (3 yrs estimation data)

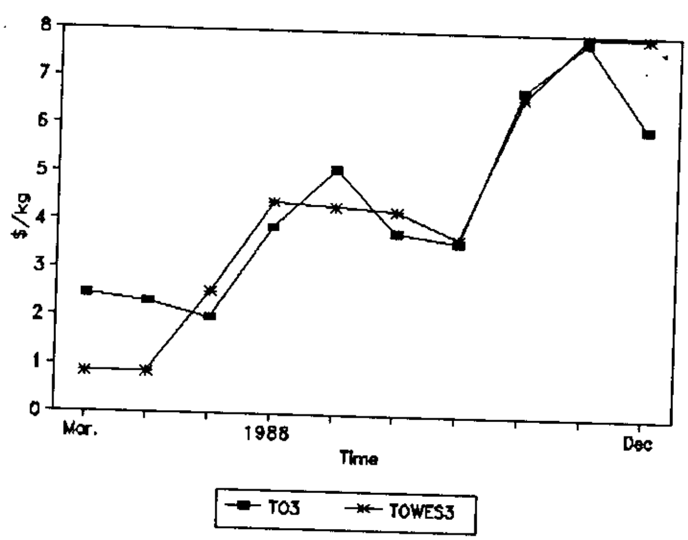

Figure A32 3-steps ahead Winters forecasts (3 yrs estimation data)

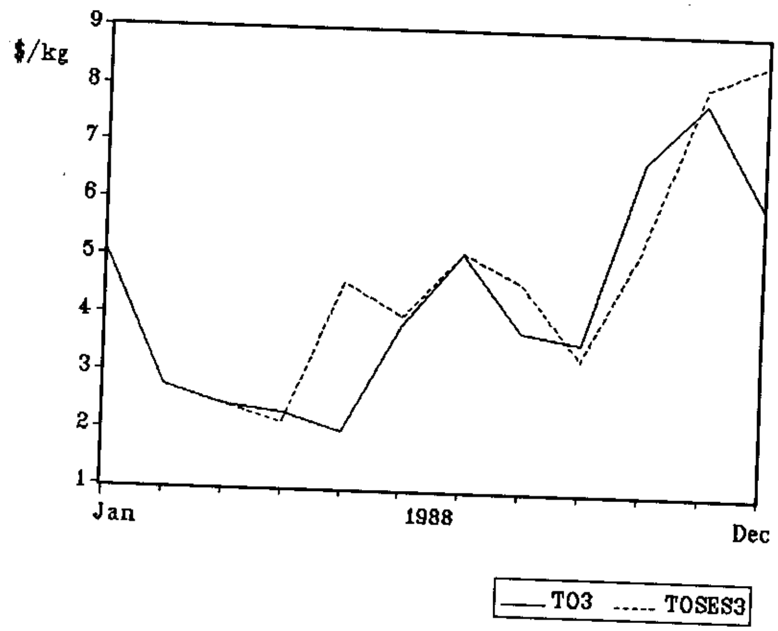

Figure A33 3-steps ahead SES forecasts (3 yrs estimation data)

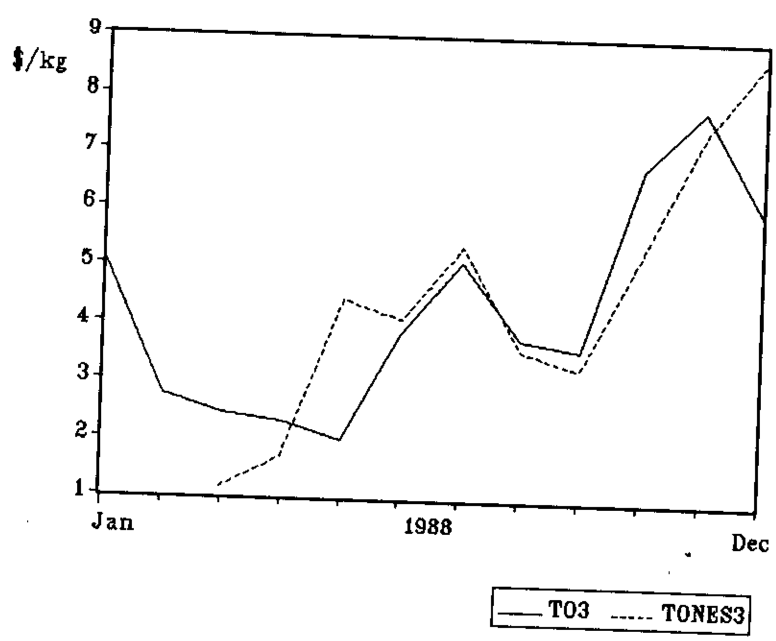

Figure A34 3-steps ahead Naive forecasts (3 yrs estimation data)

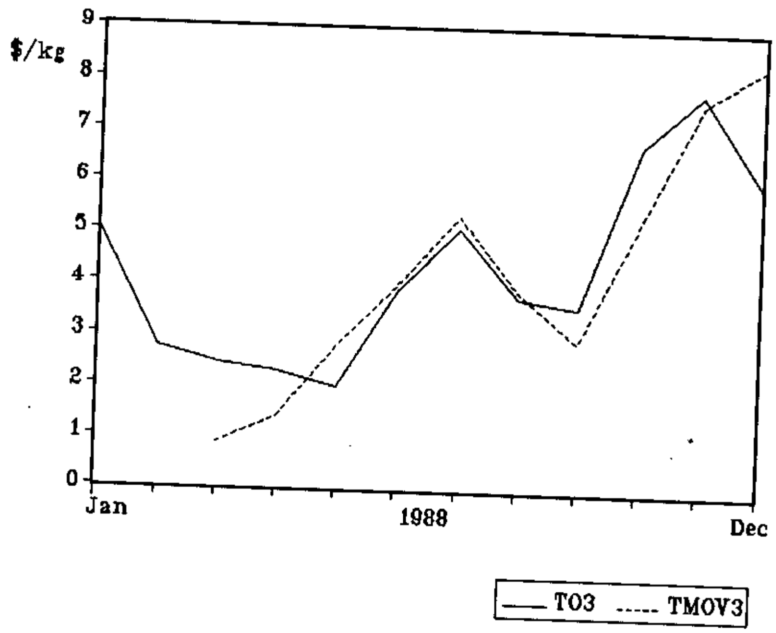

Figure A35 3-steps ahead M avg. forecasts (3 yrs estimation data)

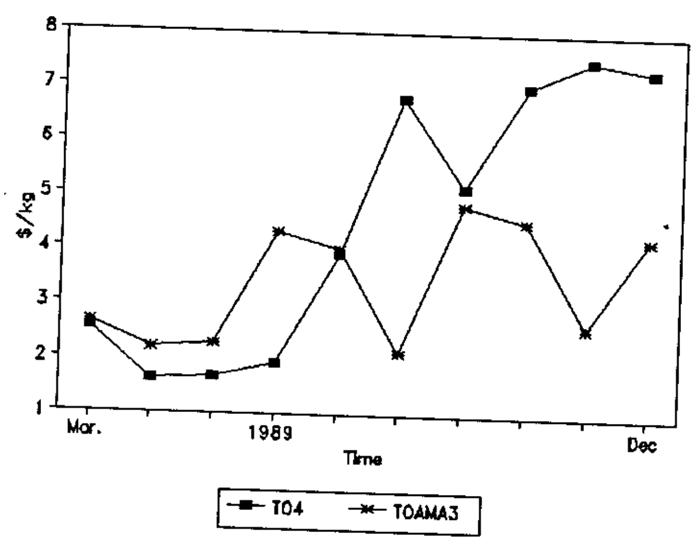

Figure A36 3-steps ahead ARIMA forecasts (4 yrs estimation data)

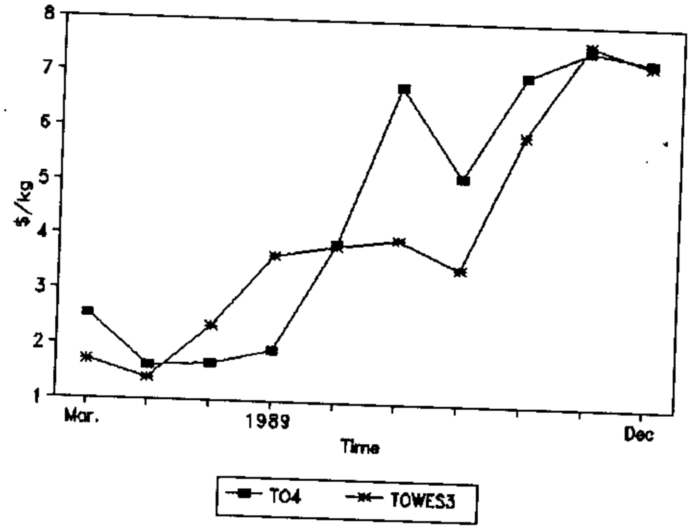

Figure A37 3-steps ahead Winters forecasts (4 yrs estimation data)

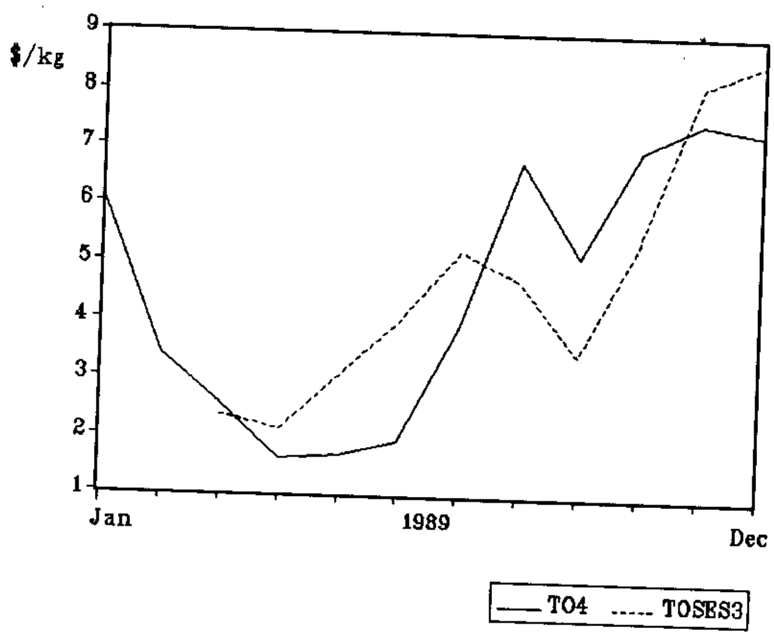

Figure A38 3-steps ahead SES forecasts (4 yrs estimation data)

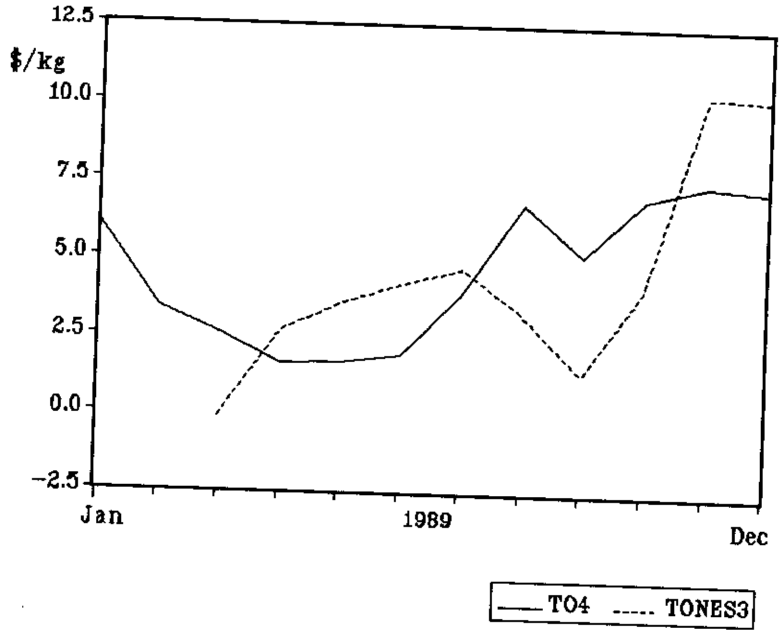

Figure A39 3-steps ahead Naive forecasts (4 yrs estimation data)

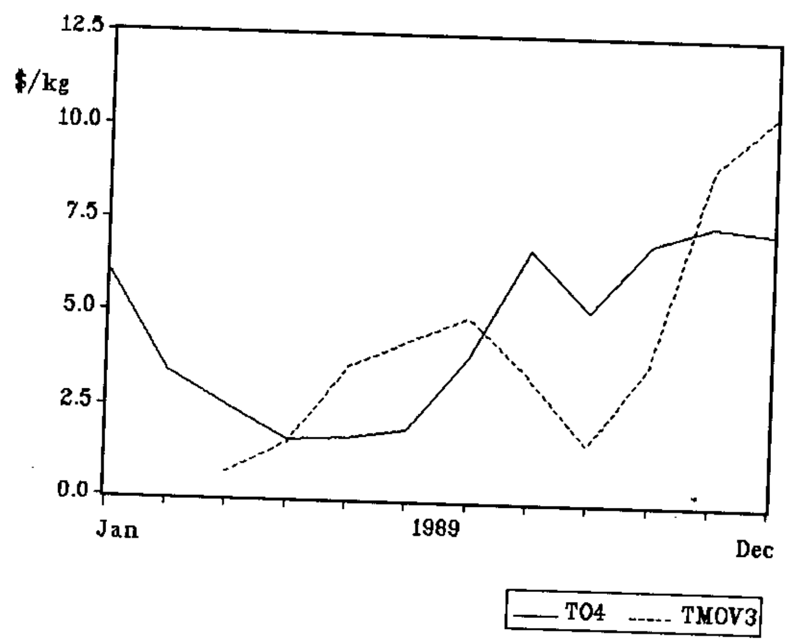

Figure A40 3-steps ahead M avg. forecasts (4 yrs estimation data)

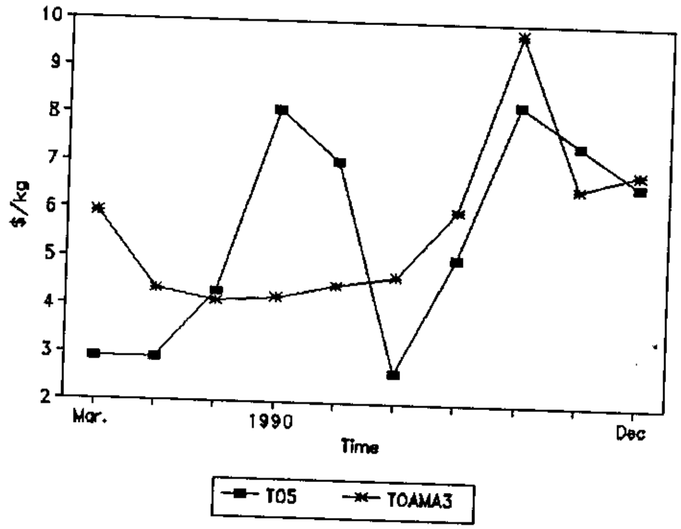

Figure A41 3-steps ahead ARMA forecasts (5 yrs estimation data)

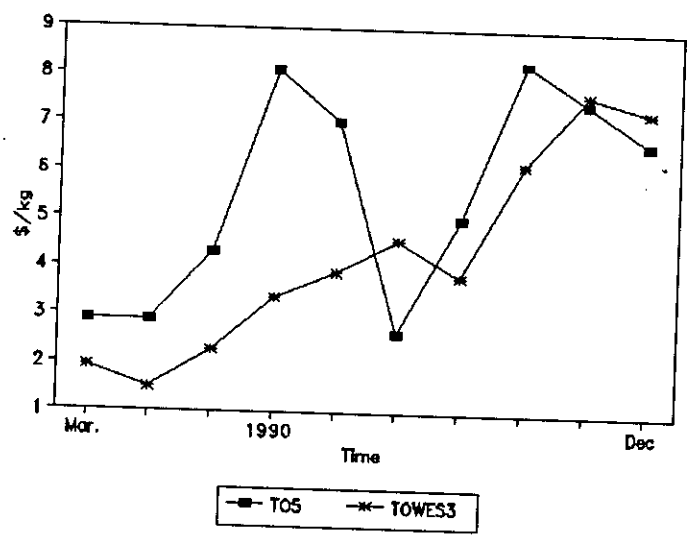

Figure A42 3steps ahead Winters forecasts (5 yrs estimation data)

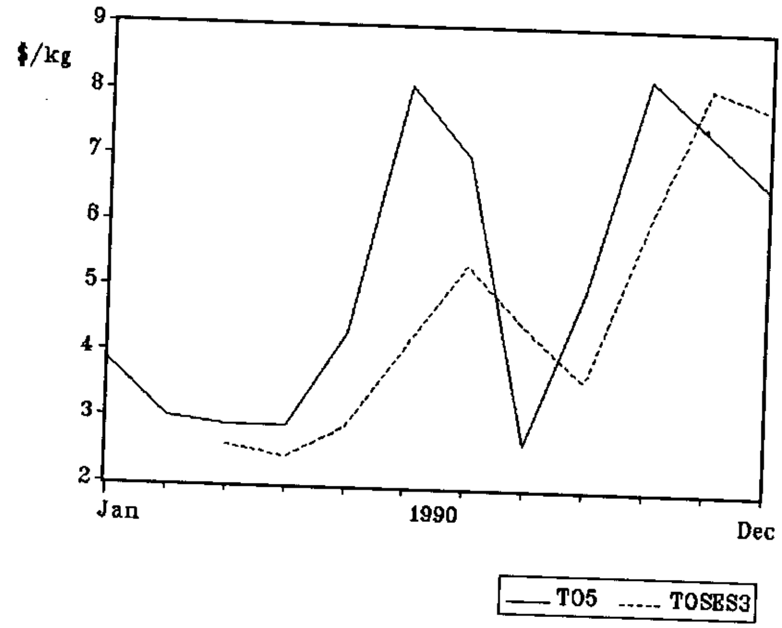

Figure A43 3-steps ahead SES forecasts (5yrs estimation data)

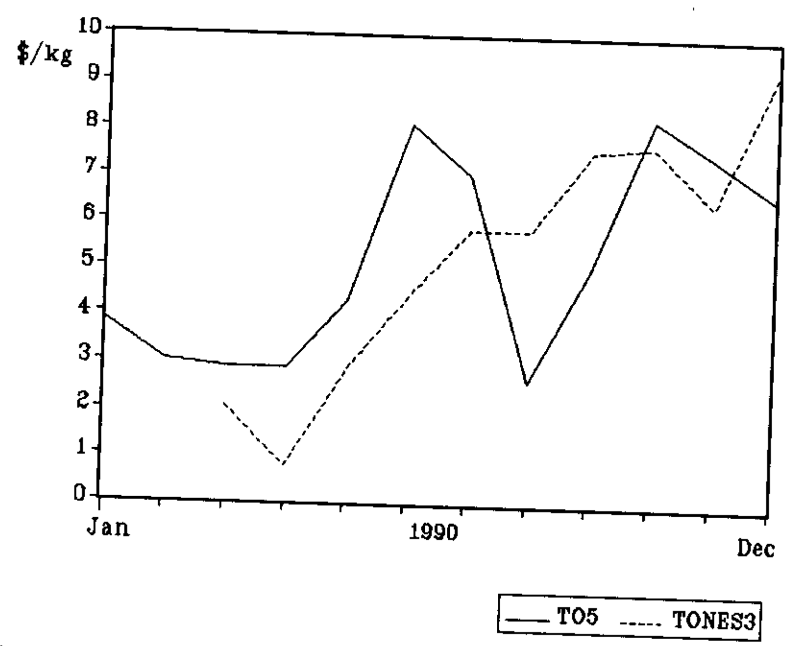

Figure A44 3-steps ahead Naive forecasts (5 yrs estimation data)

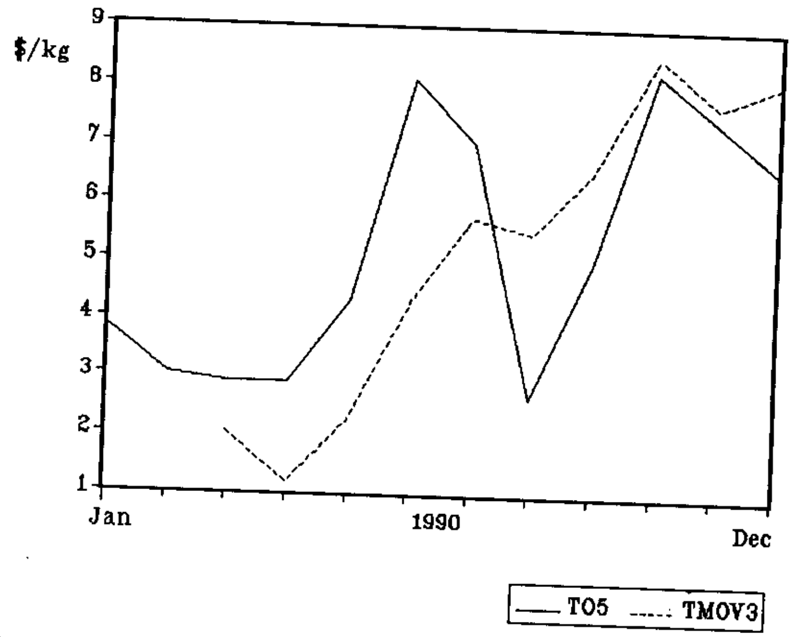

Figure A45 3-steps ahead M avg. forecasts (5 yrs estimation data)

APPENDIX B

PRIMARY ANALYSIS - CABBAGE

Plots of actual and forecasted values

APPENDIX B

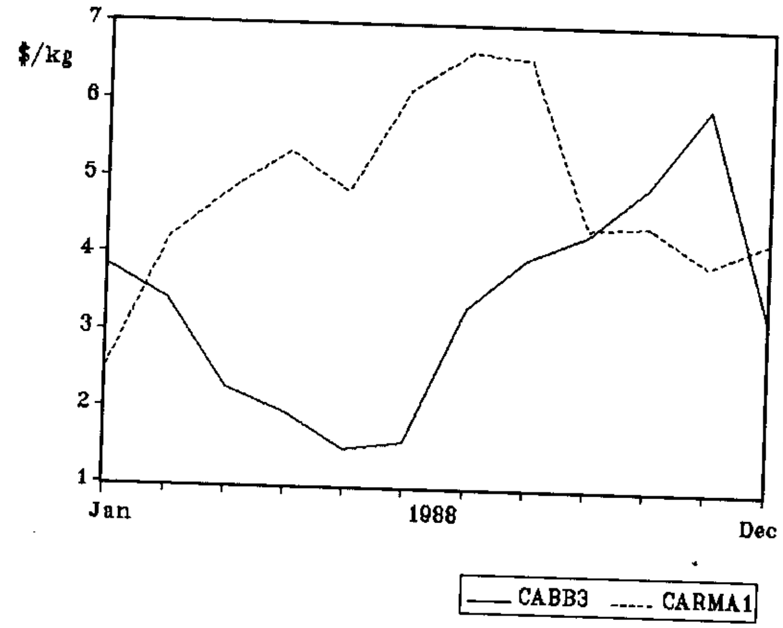

Figure B1 1-step ahead ARMA forecasts (3 yrs estimation data)

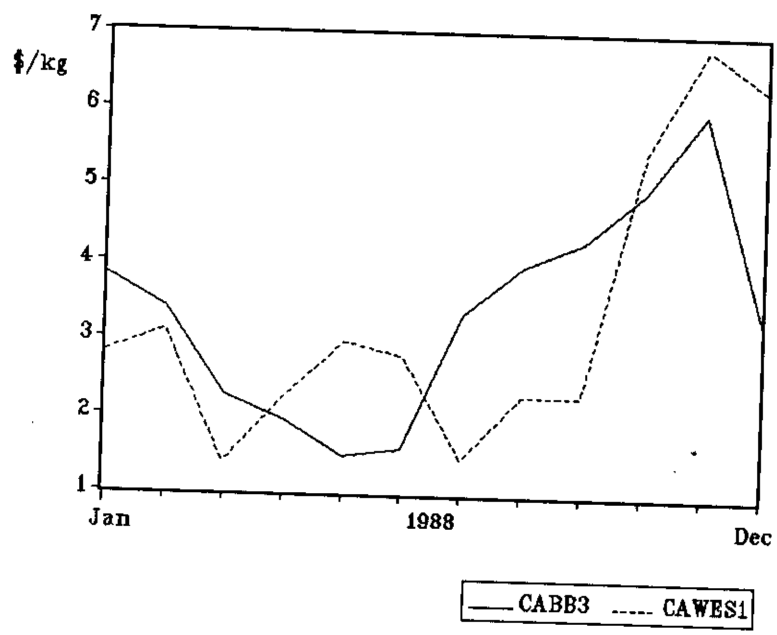

Figure B2 1-step ahead Winter's forecasts (3-yrs estimation data)

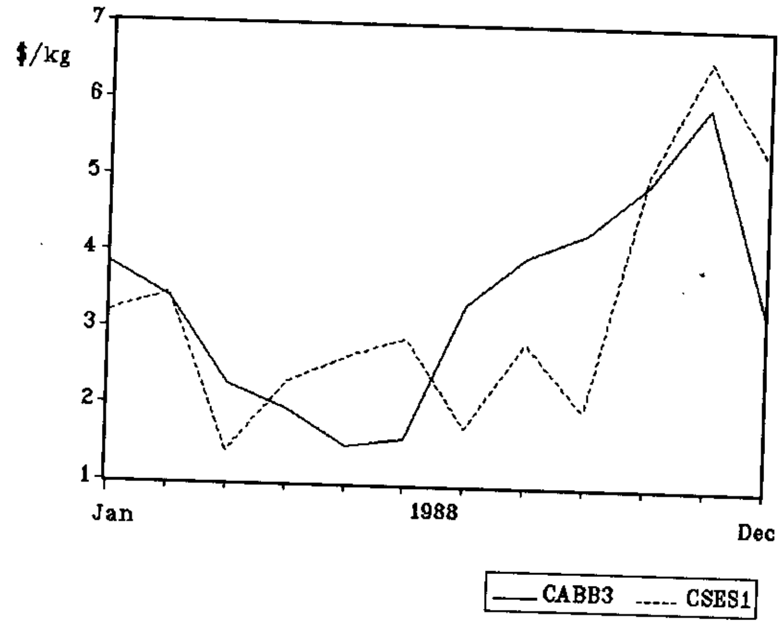

Figure B3 1-step ahead SES forecasts
(3 yrs estimation data)

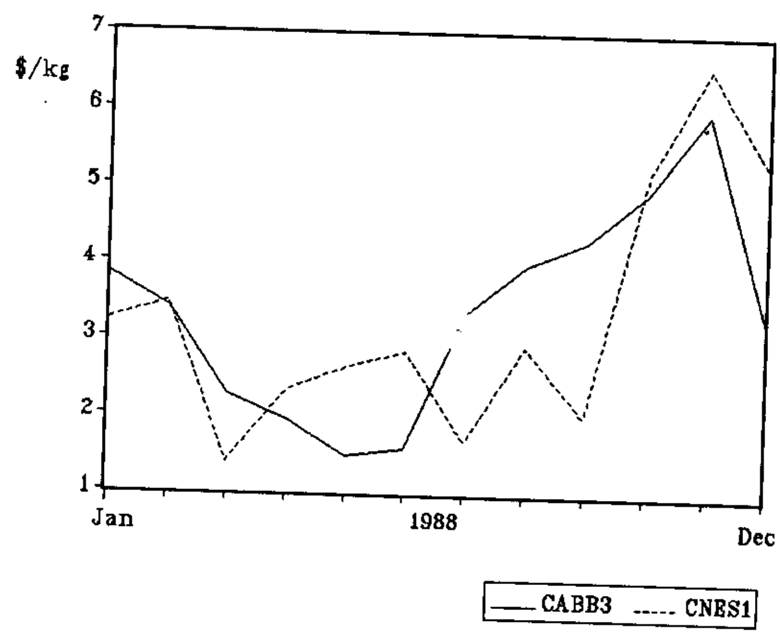

Figure B4 1-step ahead Naive forecasts
(3 yrs estimation data)

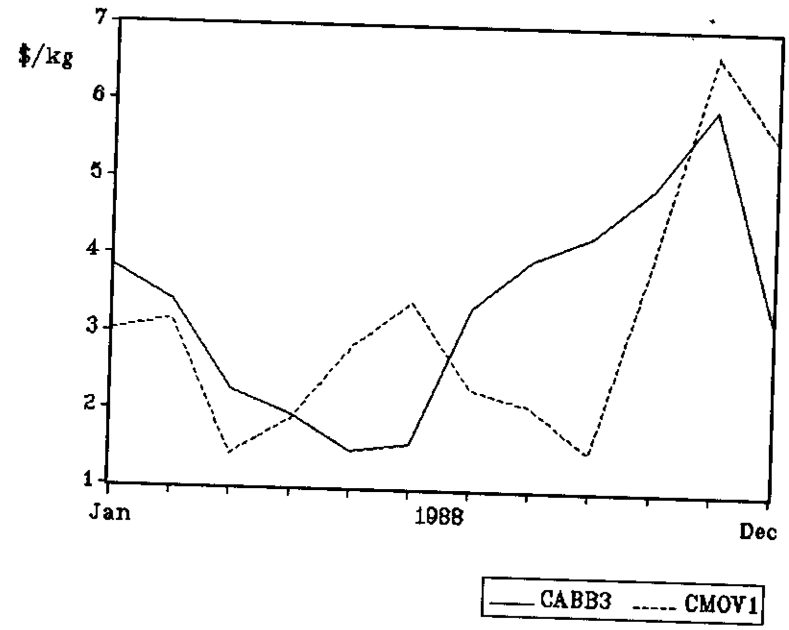

Figure B5 1-step ahead M avg. forecasts (3 yrs estimation data)

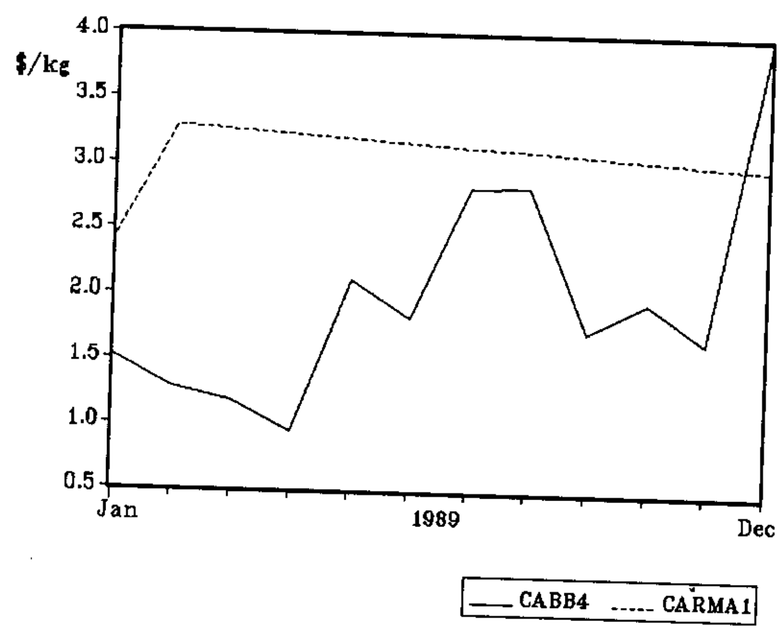

Figure B6 1-step ahead ARMA forecasts (4 yrs estimation data)

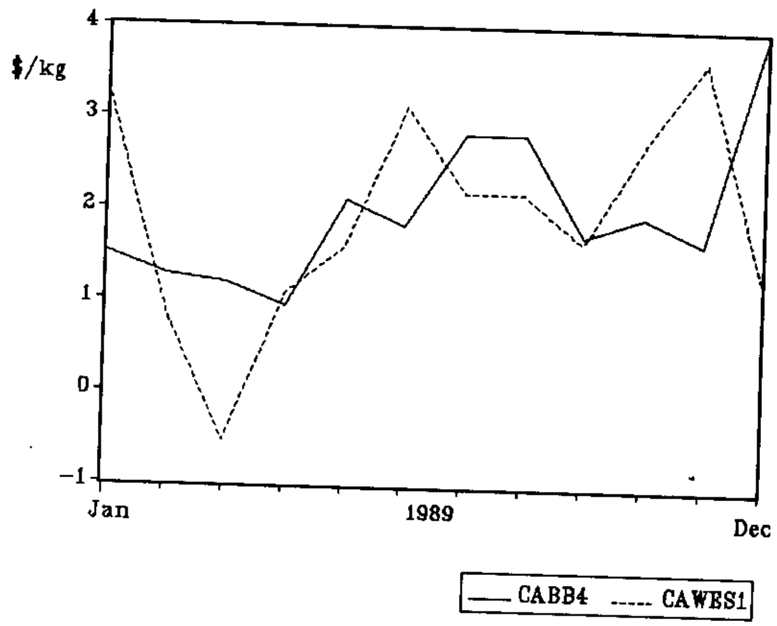

Figure B7 1-step ahead Winter's forecasts (4 yrs estimation data)

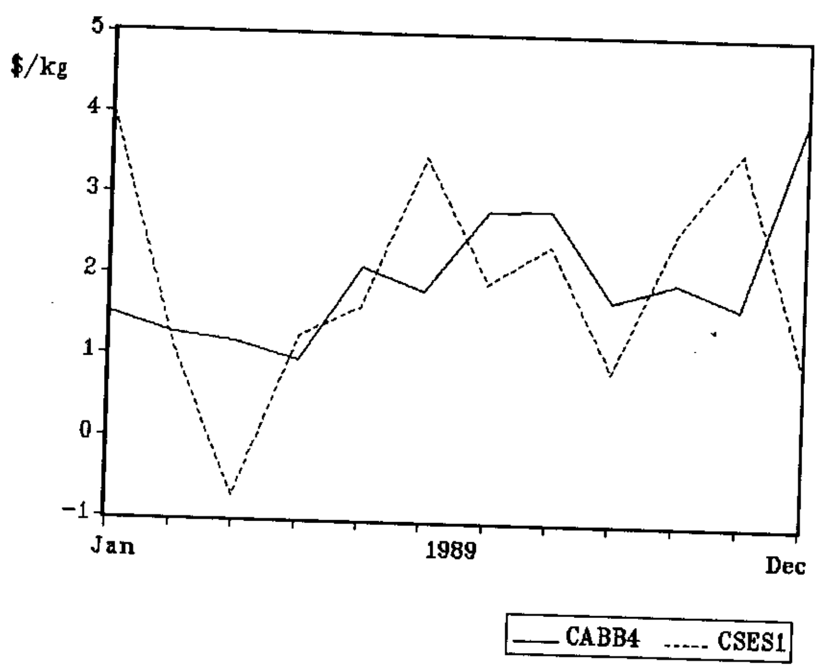

Figure B8 1-step ahead SES forecasts (4 yrs estimation data)

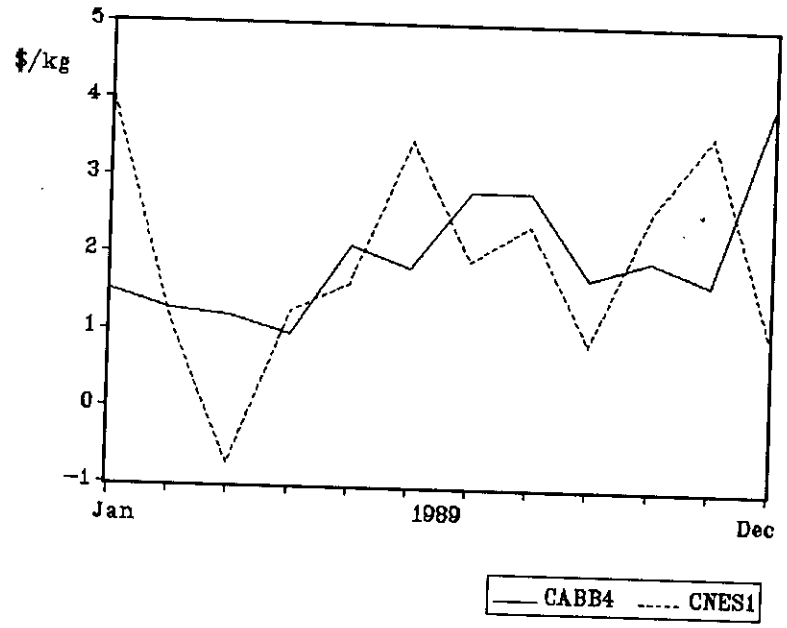

Figure B9 1-step ahead Naive forecasts (4 yrs estimation data)

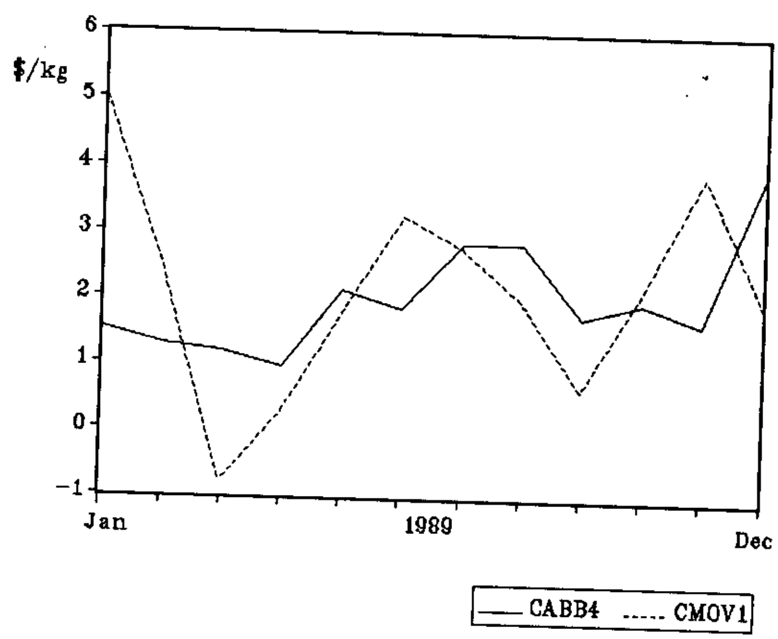

Figure B10 1-step ahead M avg. forecasts (4 yrs estimation data)

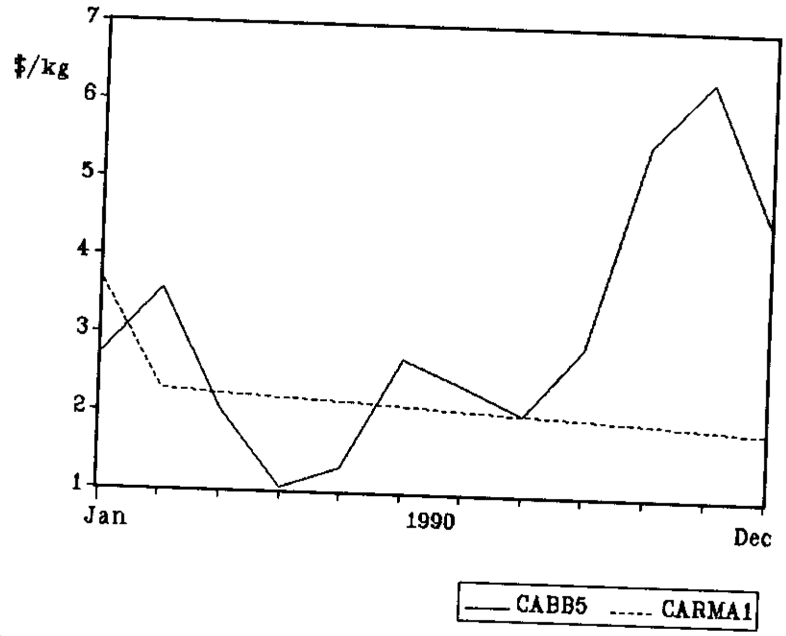

Figure B11 1-step ahead ARMA forecasts (5 yrs estimation data)

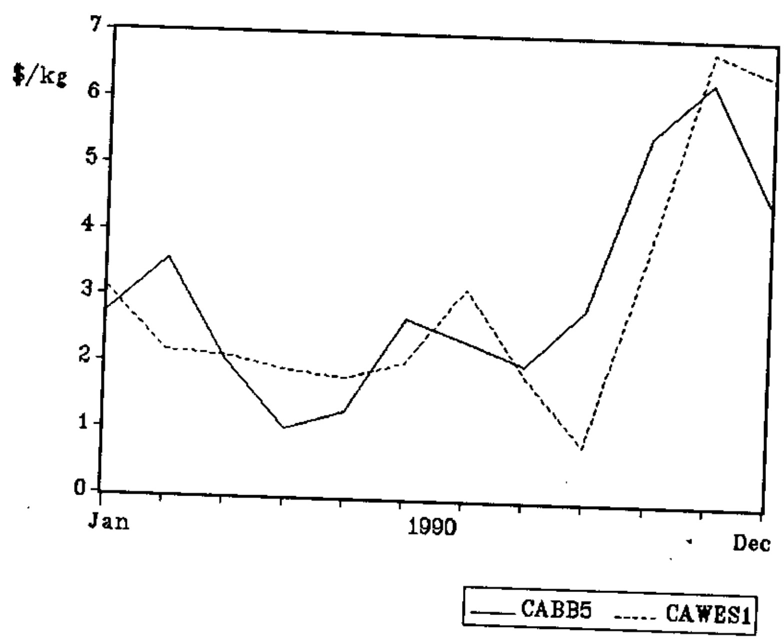

Figure B12 1-step ahead Winter's forecasts (5 yrs estimation data)

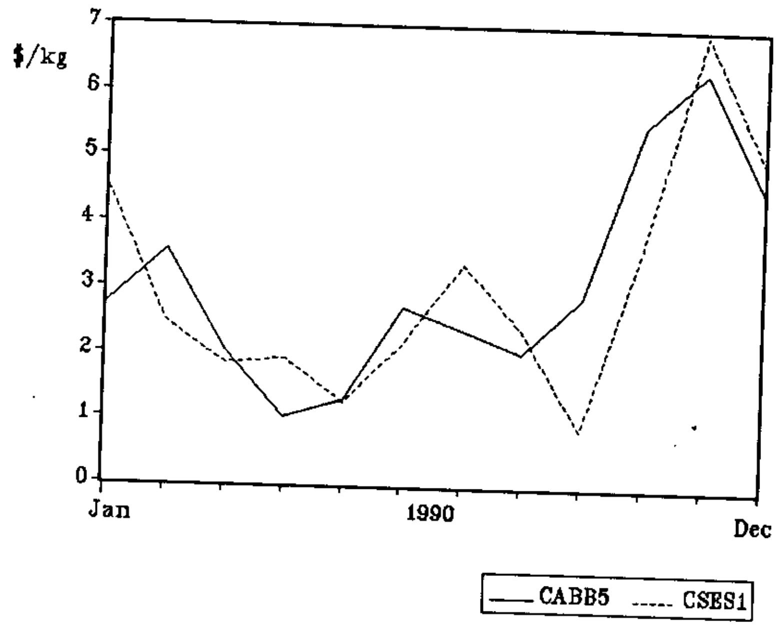

Figure B13 1-step ahead SES forecasts (5 yrs estimation data)

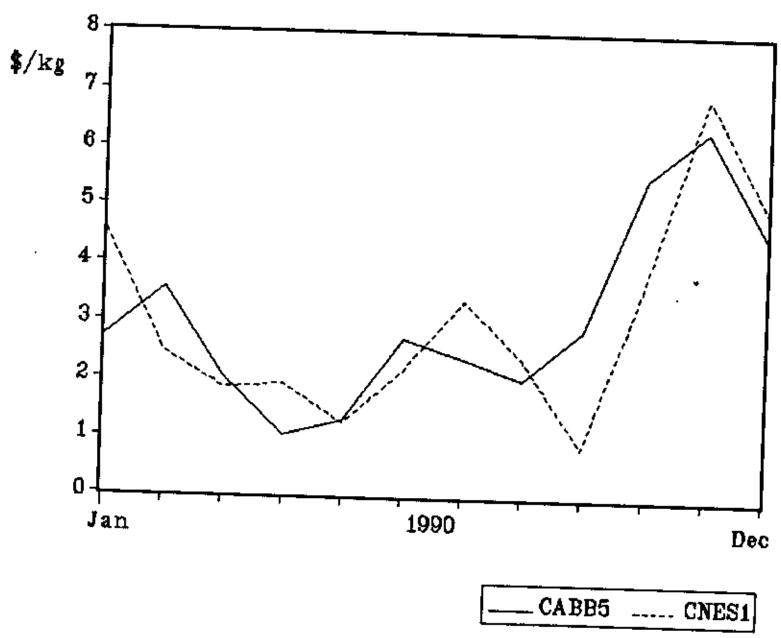

Figure 14 1-step ahead Naive forecasts (5 yrs estimation data)

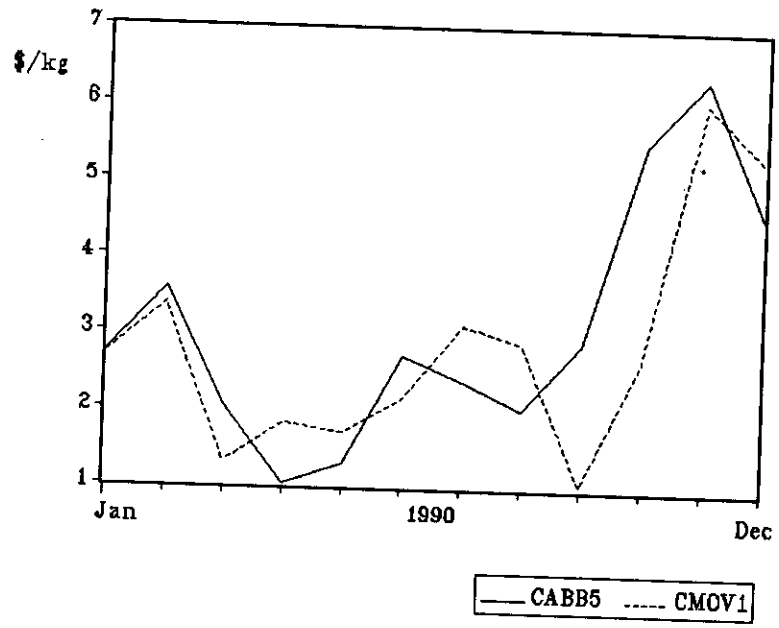

Figure 15 1-step ahead M avg. forecasts (5 yrs estimation data)

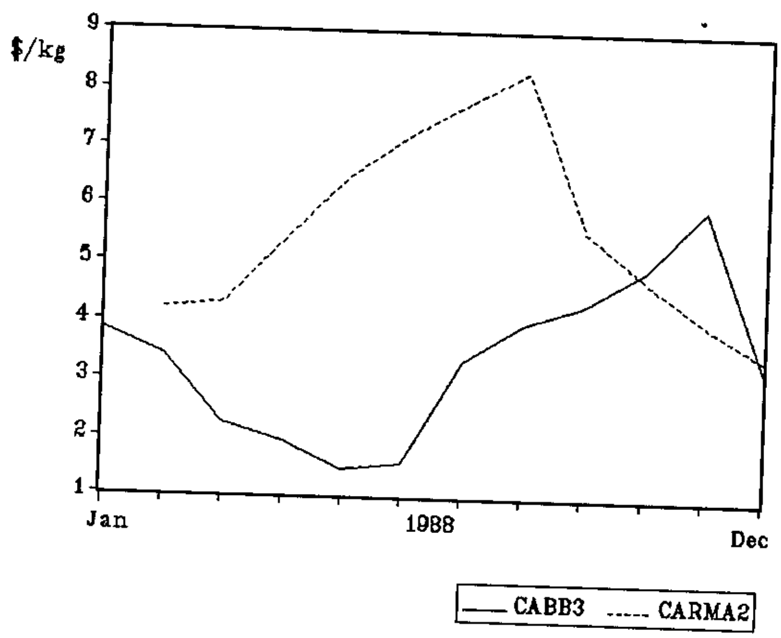

Figure B16 2-steps ahead ARMA forecasts (3 years estimation data)

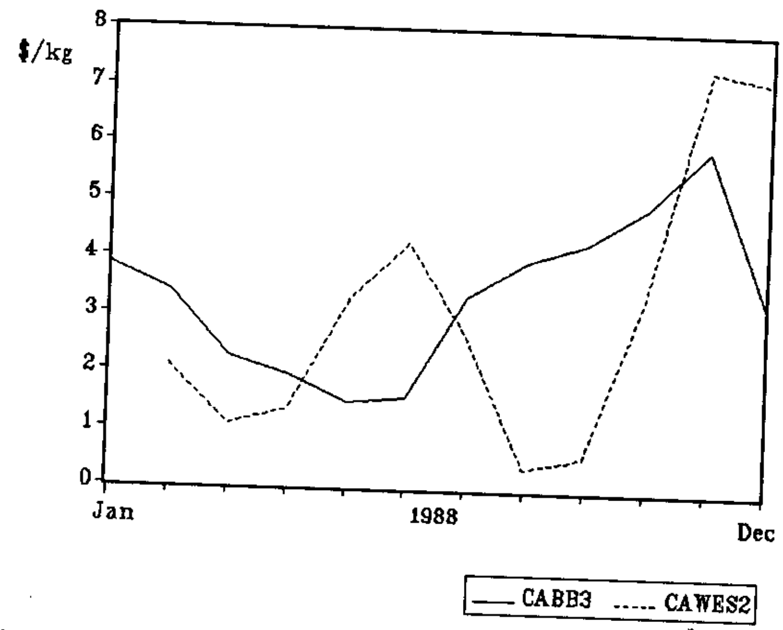

Figure B17 2-steps ahead Winters forecasts (3 yrs estimation data)

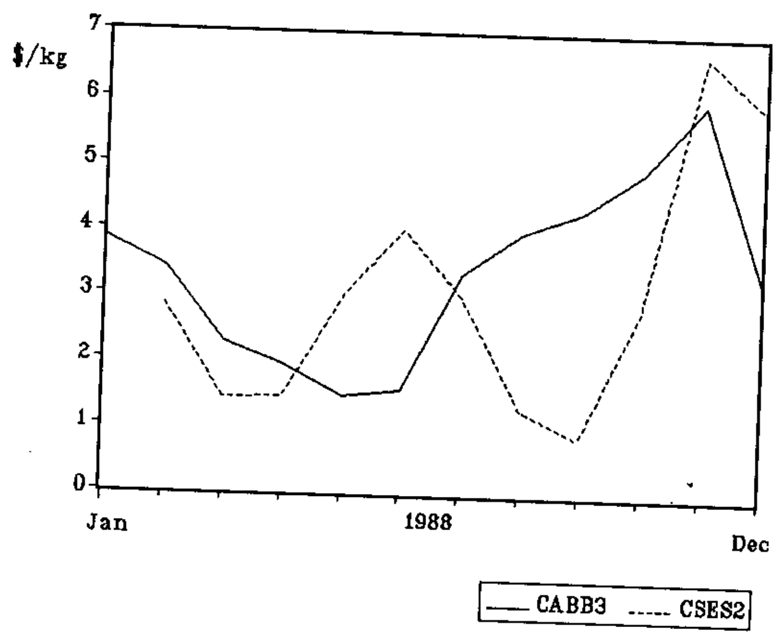

Figure B18 2-steps ahead SES forecasts (3 yrs estimation data)

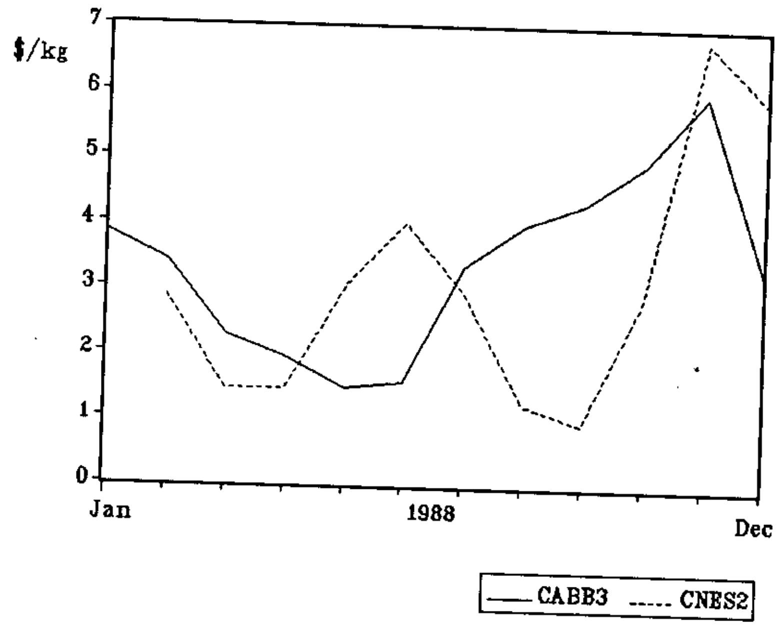

Figure B19 2-steps ahead Naive forecasts (3 yrs estimation data)

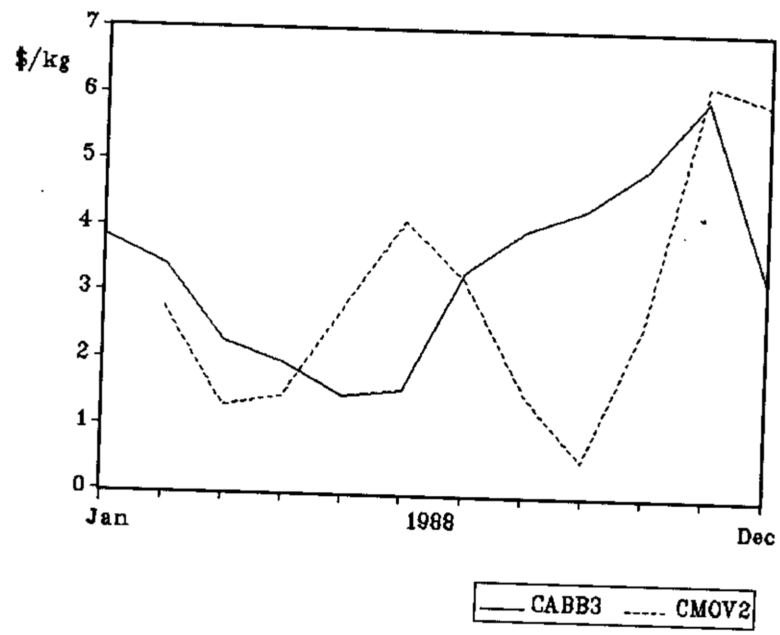

Figure B20 2-steps ahead M avg. forecasts (3 yrs estimation data)

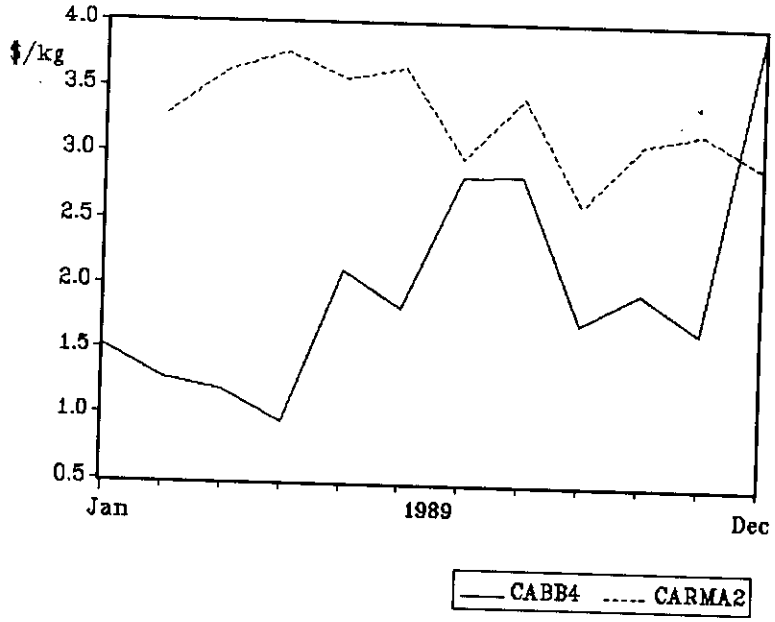

Figure B21 2-steps ahead ARMA forecasts (4 yrs estimation data)

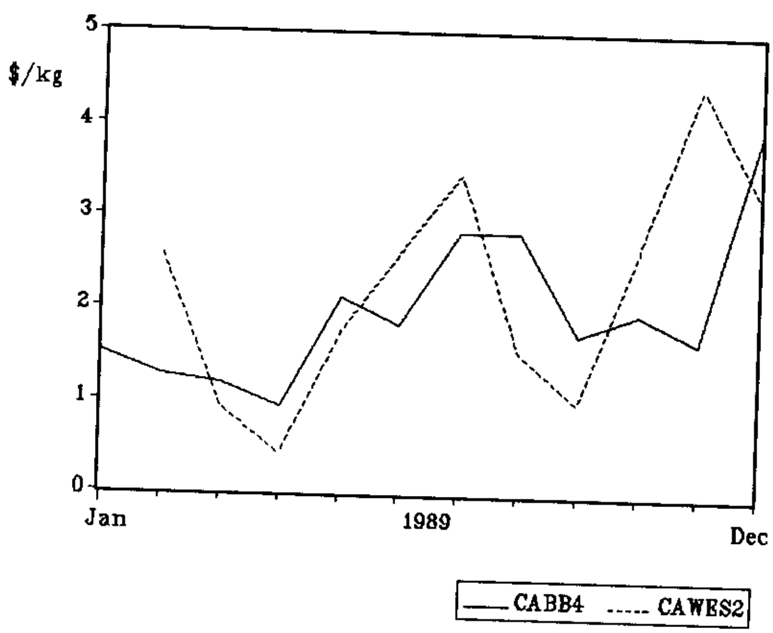

Figure B22 2-steps ahead Winters forecasts (4 yrs estimation data)

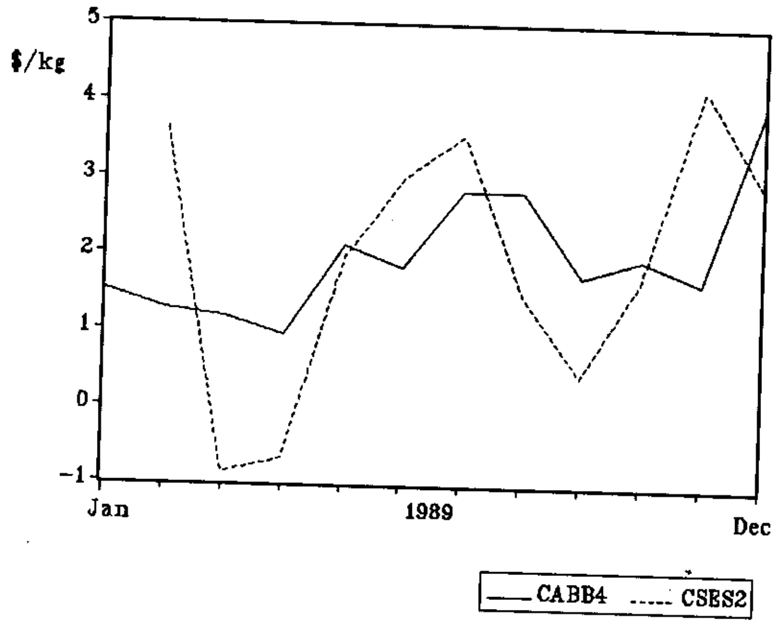

Figure B23 2-steps ahead SES forecasts (4 yrs estimation data)

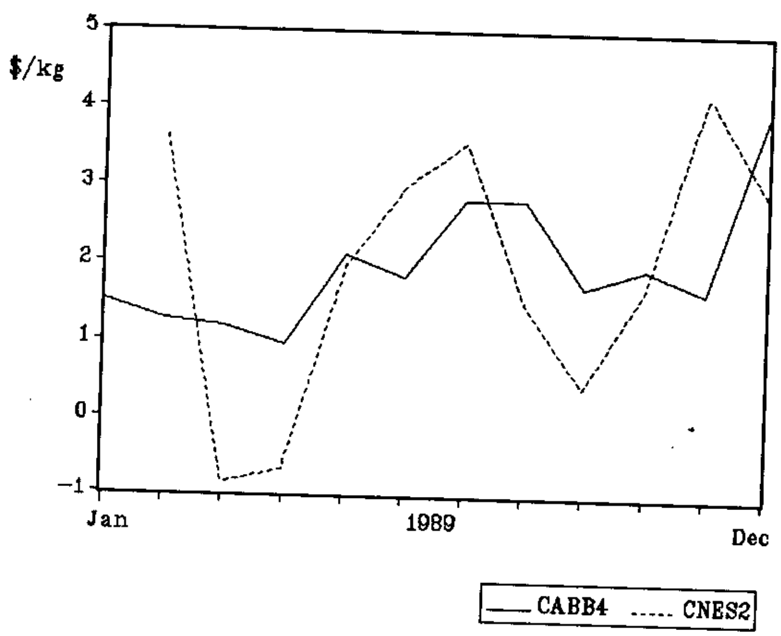

Figure B24 2-steps ahead Naive forecasts (4 yrs estimation data)

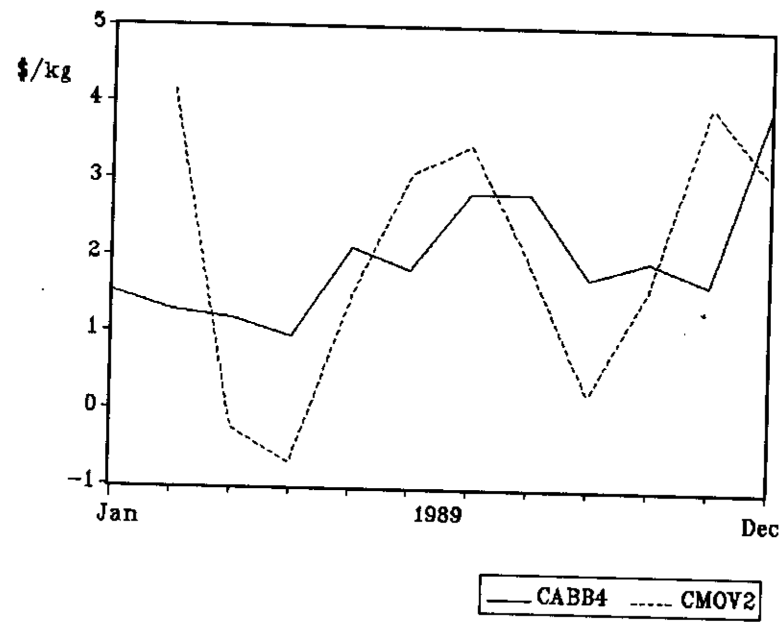

Figure B25 2-steps ahead M avg. forecasts (4 yrs estimation data)

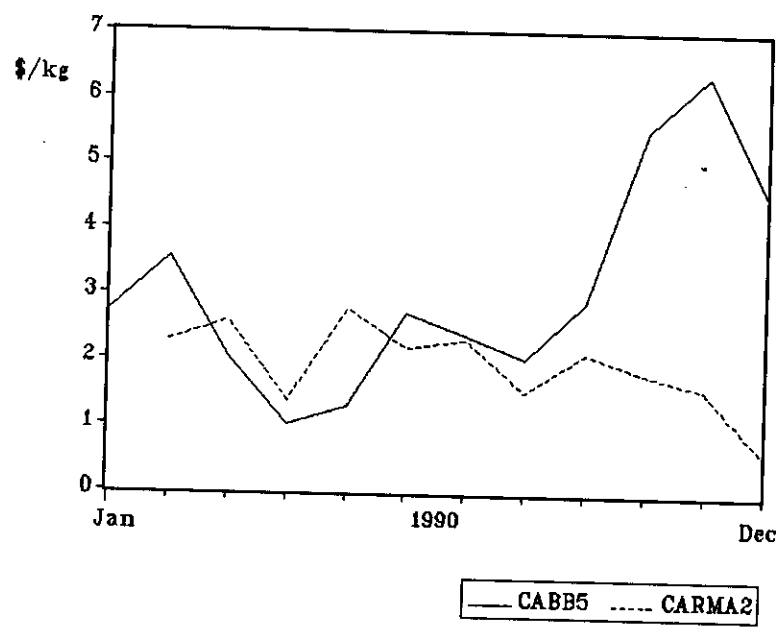

Figure B26 2-steps ahead ARMA forecasts (5 yrs estimation data)

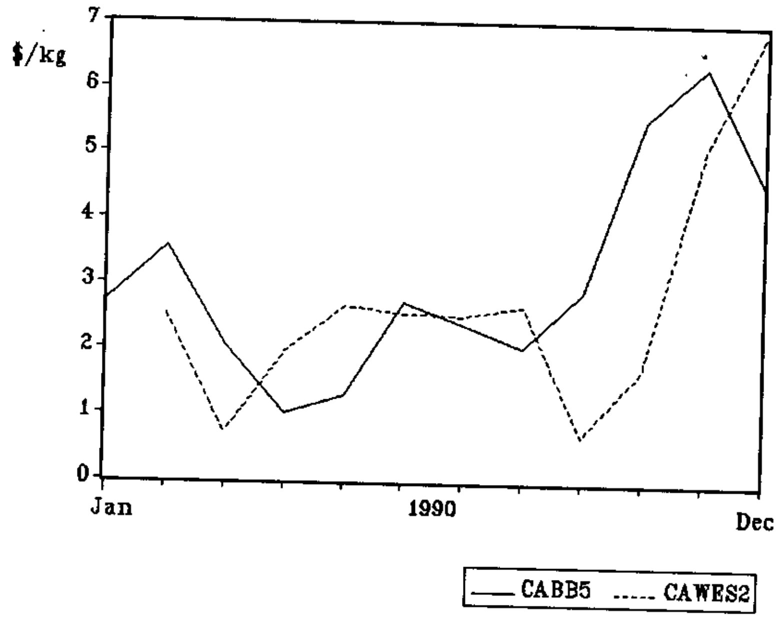

Figure B27 2-steps ahead Winters forecasts
(5 yrs estimation data)

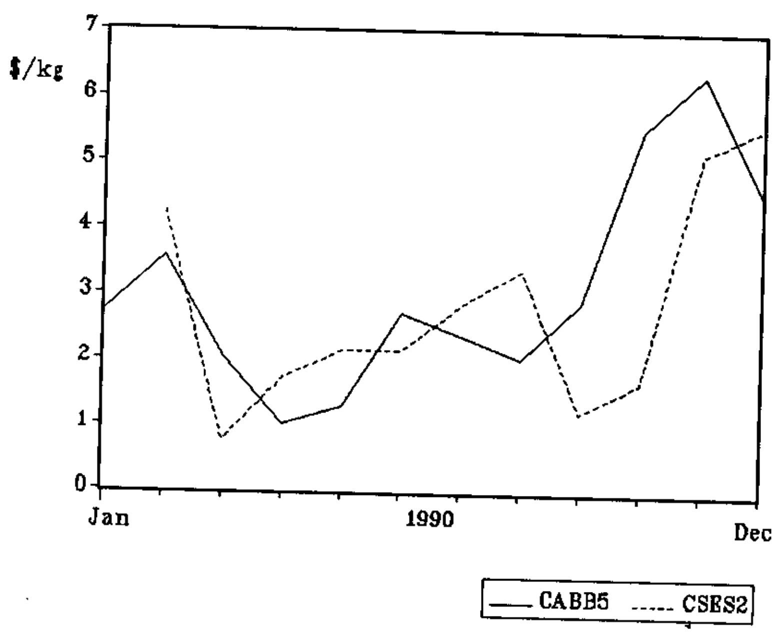

Figure B28 2-steps ahead SES forecasts
(5 yrs estimation data)

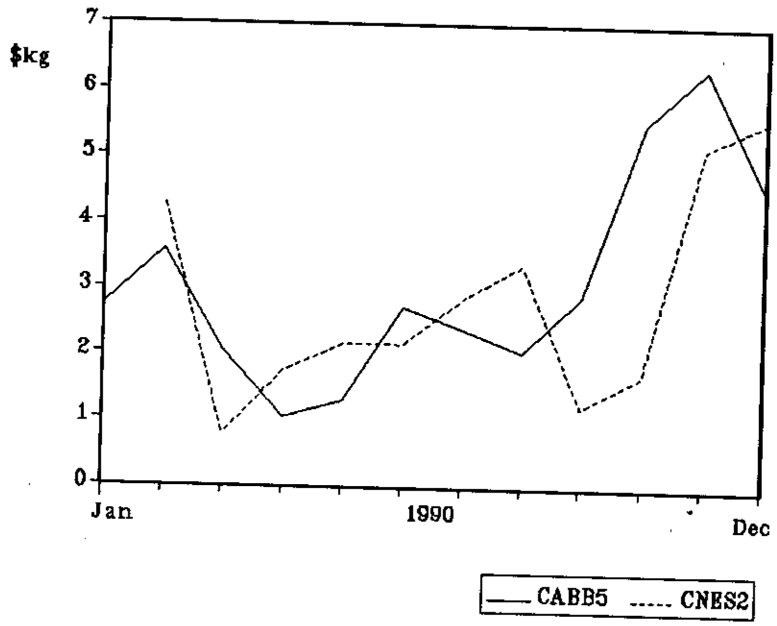

Figure B29 2-step ahead Naive forecasts (5 yrs estimation data)

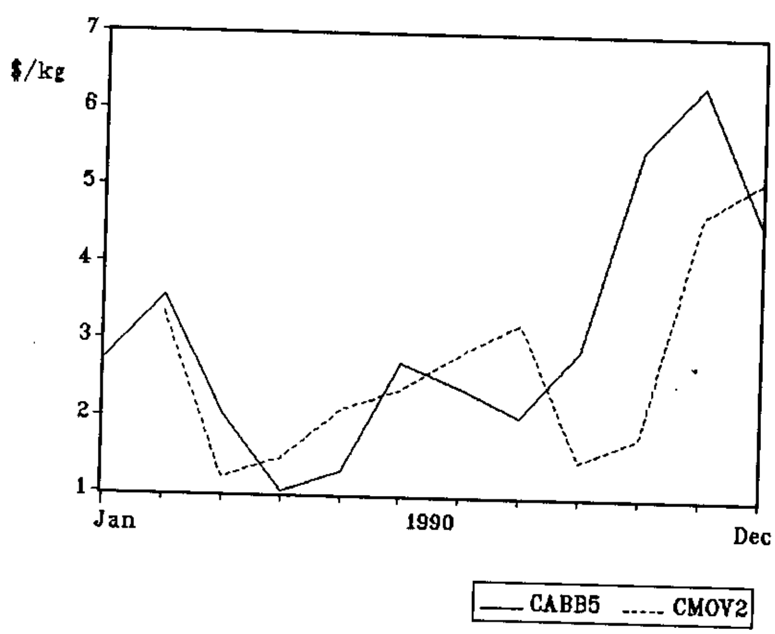

Figure 30 2-steps ahead M avg. forecasts (5 yrs estimation data)

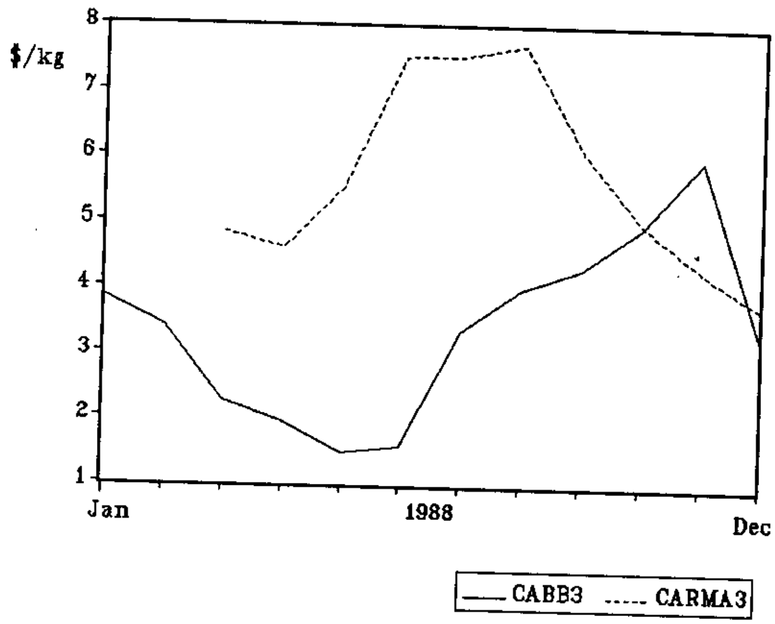

Figure B31 3-steps ahead ARMA forecasts (3 years estimation data)

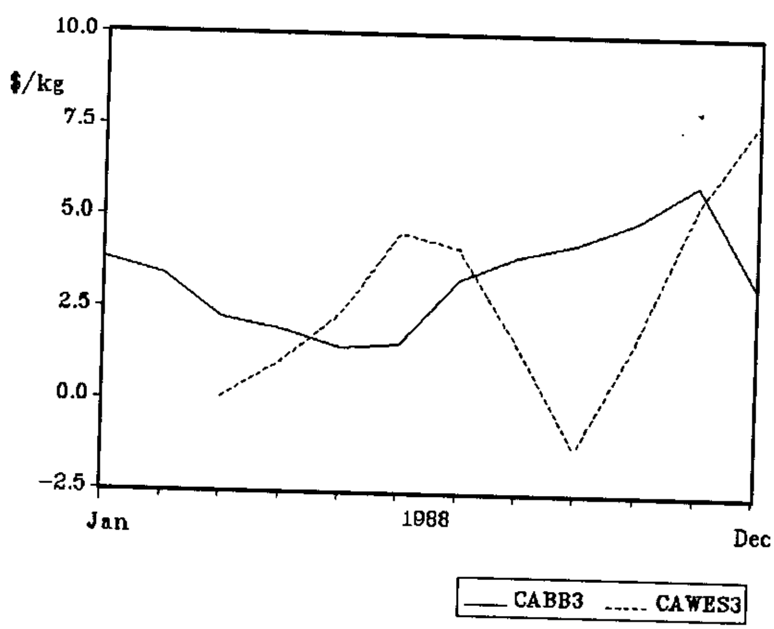

Figure B32 3-steps ahead Winters forecasts (3 yrs estimation data)

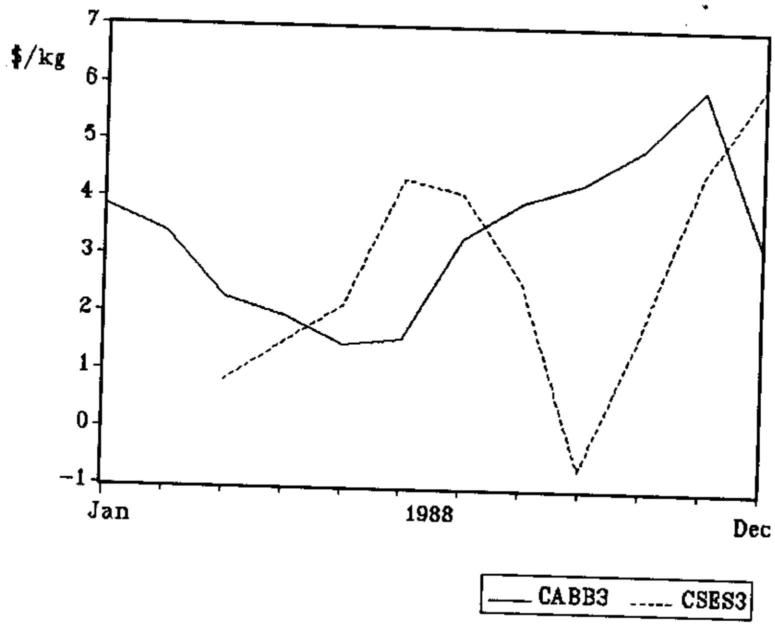

Figure B33 3-steps ahead SES forecasts (3 yrs estimation data)

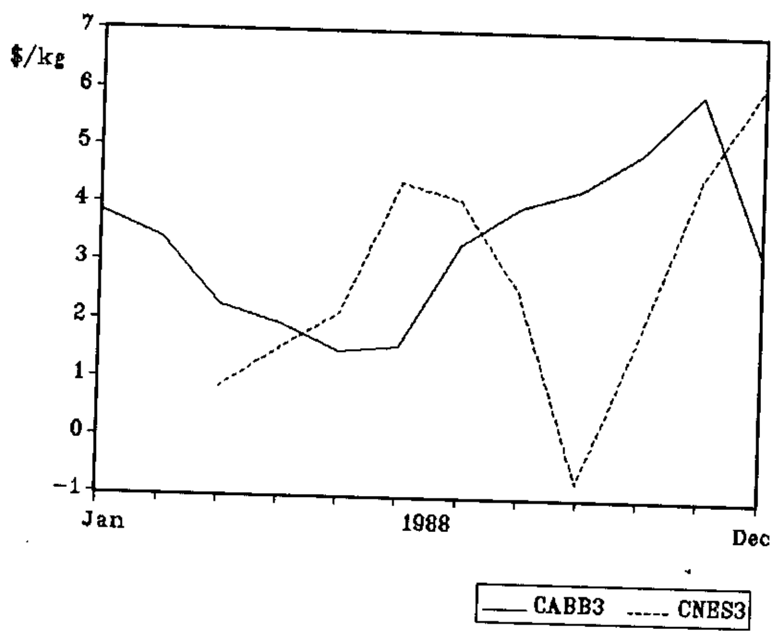

Figure B34 3-steps ahead Naive forecasts (3 yrs estimation data)

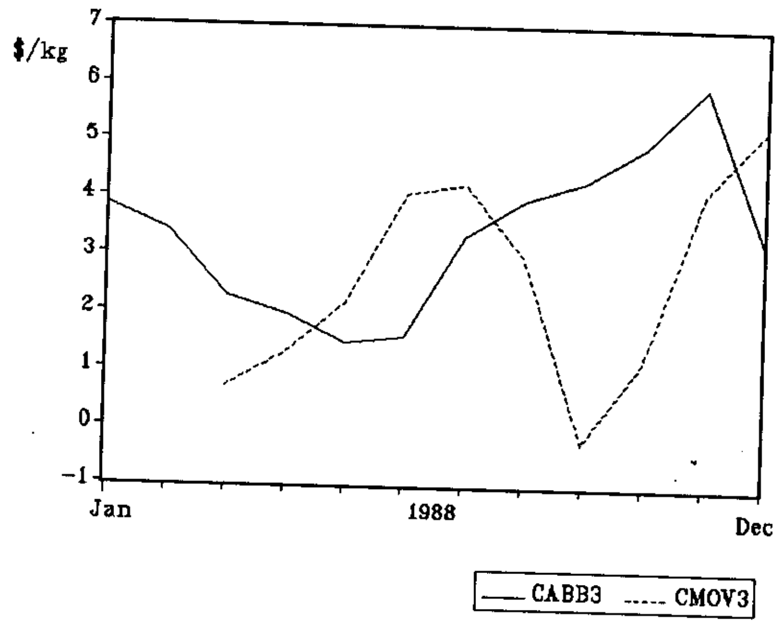

Figure B35 3-steps ahead M avg. forecasts (3 yrs estimation data)

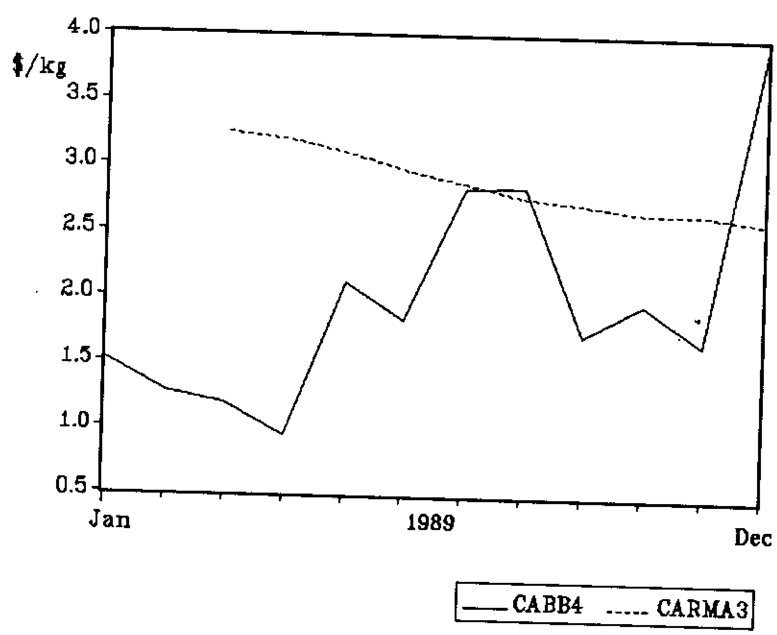

Figure B36 3-steps ahead ARMA forecasts (4 yrs estimation data)

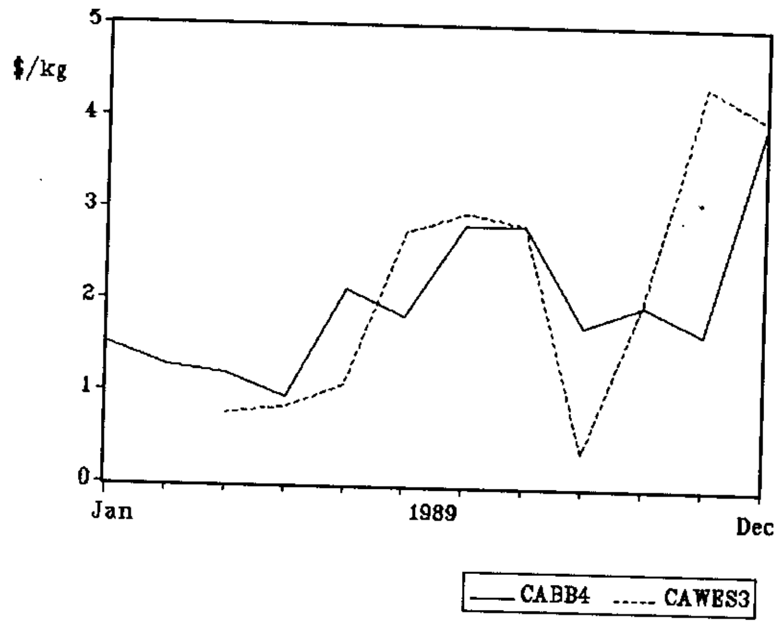

Figure B37 3-steps ahead Winters forecasts (4 yrs estimation data)

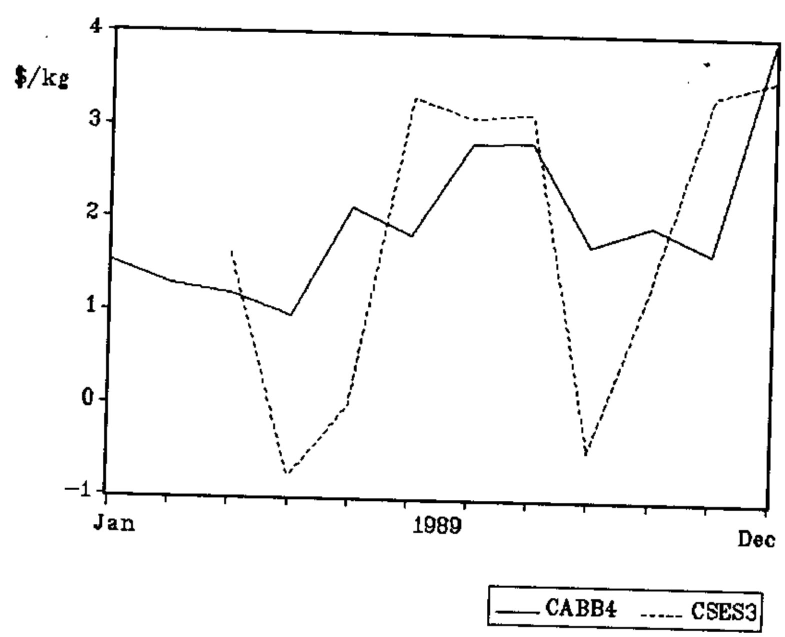

Figure B38 3-steps ahead SES forecasts (4 years estimation data)

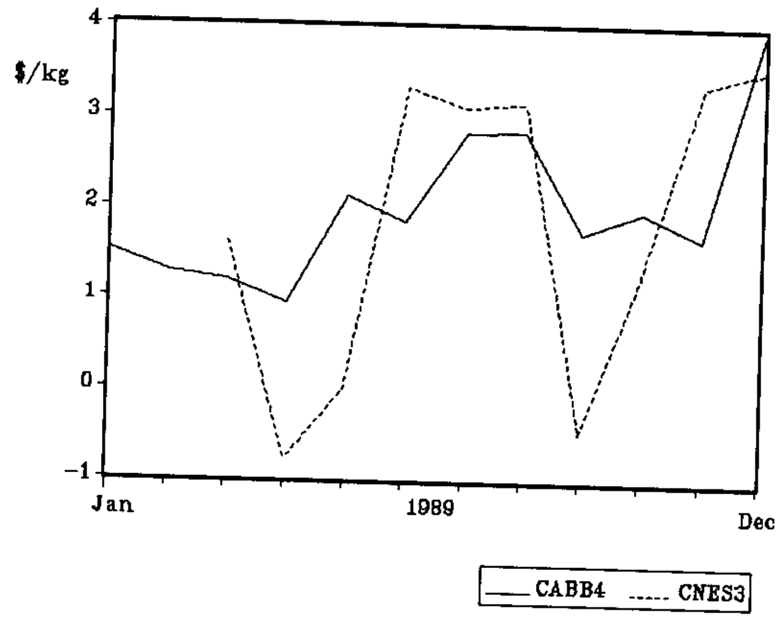

Figure B39 3-steps ahead Naive forecasts (4 yrs estimation data)

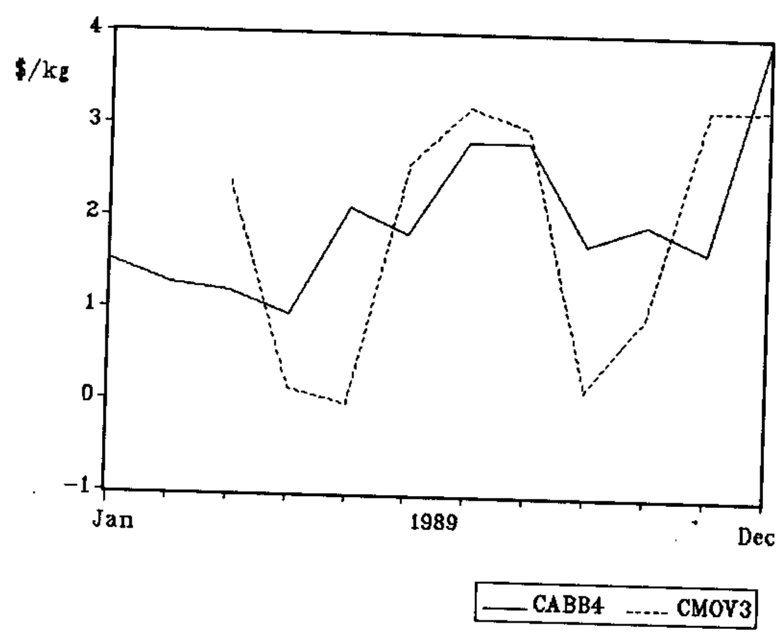

Figure B40 3-steps ahead M avg. forecasts (4 yrs estimation data)

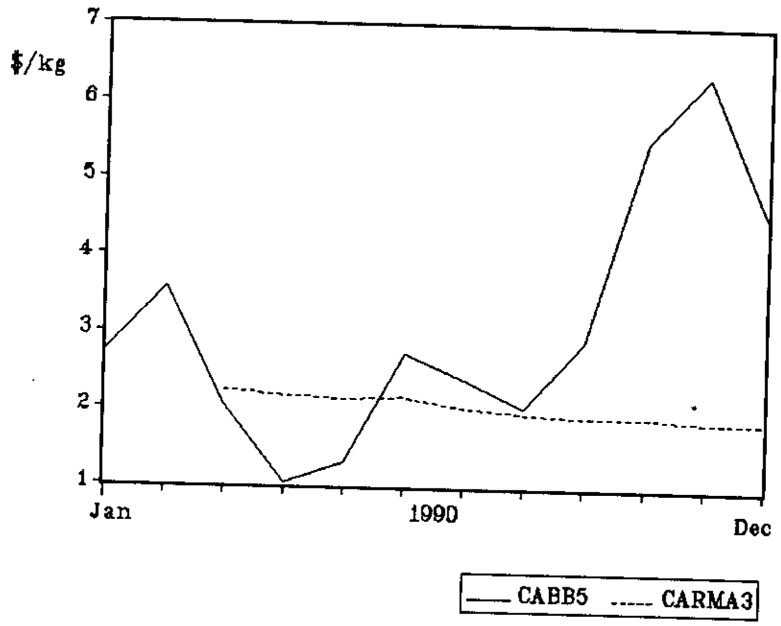

Figure B41 3-steps ahead ARMA forecasts (5 yrs estimation data)

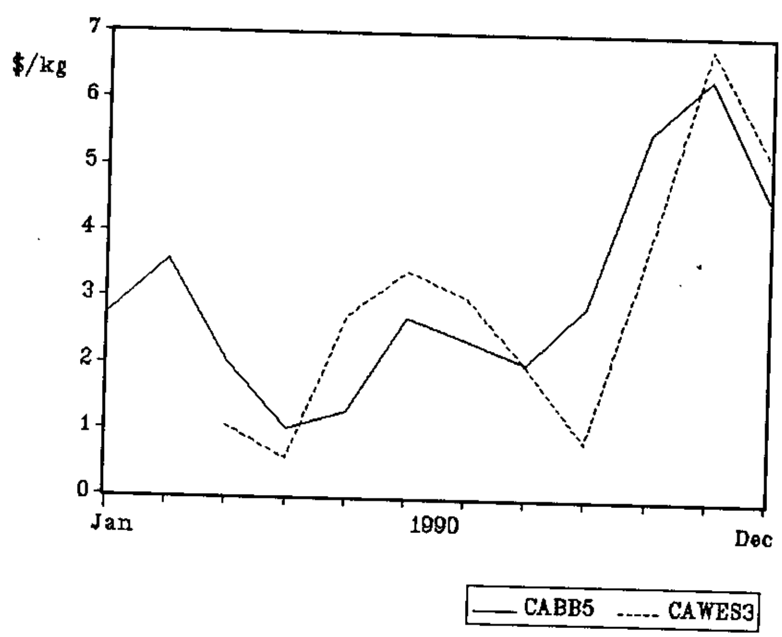

Figure B42 3-steps ahead Winters forecasts (5 yrs estimation data)

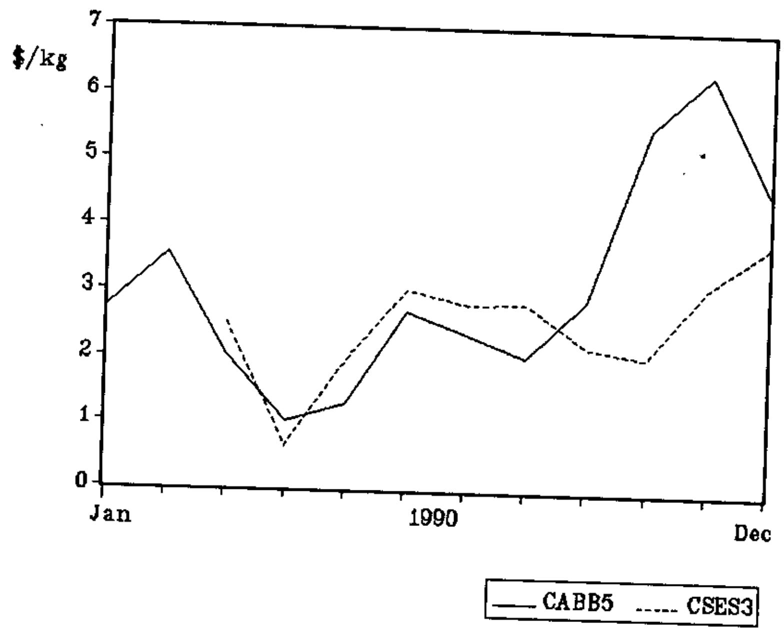

Figure B43 3-steps ahead SES forecasts
(5 yrs estimation data)

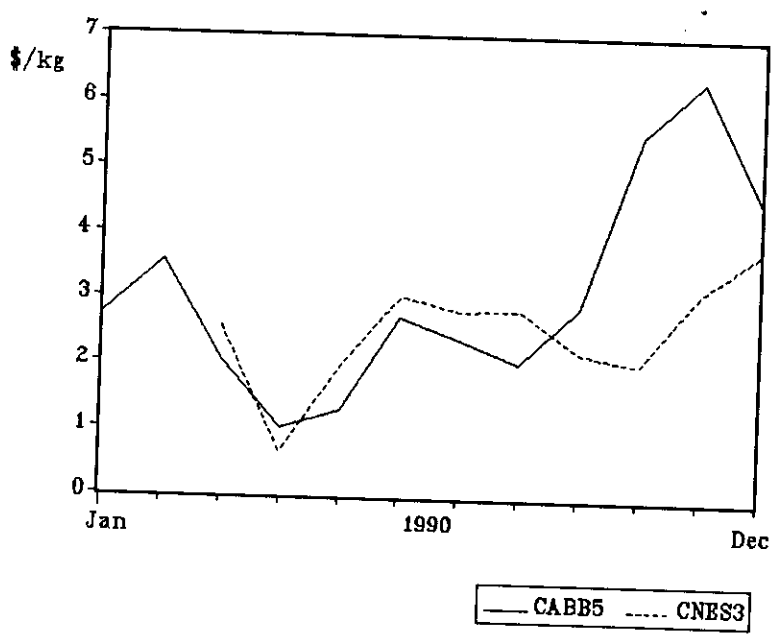

Figure B44 3-steps ahead Naive forecasts
(5 yrs estimation data)

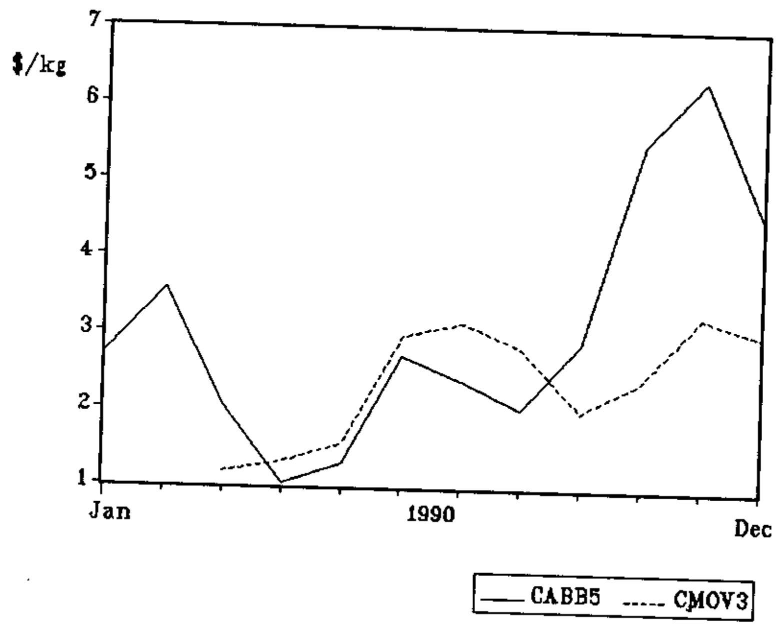

Figure B45 3-steps ahead M avg. forecasts (5 yrs estimation data)

APPENDIX C

PRIMARY ANALYSIS: data and model* forecasts

* The first number in the alpha-numeric model names used in the tables refers to the quantity of estimation data used. The second number highlights the number of steps-ahead being forecasted.

APPENDIX C

Table C1 CABBAGE: 1-step ahead actual and forecasted values (TT$/kg) (3, 4 & 5 years estimation data)

	CABBAGE	ARIMA31	Wint31	SES31	Naive31	M.av31	Factor
1988.01	3.84	2.50	2.82	3.21	3.23	3.02	1.06
1988.02	3.41	4.19	3.12	3.46	3.48	3.18	0.70
1988.03	2.27	4.83	1.38	1.39	1.39	1.42	-1.32
1988.04	1.95	5.36	2.25	2.31	2.34	1.90	-1.25
1988.05	1.48	4.83	2.98	2.65	2.64	2.84	-0.56
1988.06	1.58	6.16	2.79	2.90	2.85	3.43	0.81
1988.07	3.35	6.67	1.45	1.73	1.67	2.31	0.90
1988.08	3.99	6.58	2.28	2.86	2.93	2.09	0.48
1988.09	4.31	4.39	2.29	1.96	2.01	1.48	-1.50
1988.10	4.95	4.44	5.47	5.08	5.18	4.03	-0.63
1988.11	5.99	3.93	6.86	6.60	6.59	6.70	1.01
1988.12	3.26	4.27	6.31	5.33	5.30	5.60	0.32
	Cabbage	ARIMA41	Wint41	SES41	Naive41	M.av41	Factor
1989.01	1.52	2.44	3.22	4.00	4.00	5.02	1.06
1989.02	1.28	3.29	0.87	1.16	1.16	2.40	0.70
1989.03	1.19	3.26	-0.53	-0.74	-0.74	-0.80	-1.32
1989.04	0.95	3.23	1.09	1.26	1.26	0.29	-1.25
1989.05	2.12	3.20	1.61	1.64	1.64	1.80	-0.56
1989.06	1.84	3.17	3.13	3.49	3.49	3.25	0.81
1989.07	2.83	3.14	2.19	1.93	1.93	2.76	0.90
1989.08	2.84	3.12	2.19	2.41	2.41	1.96	0.48
1989.09	1.73	3.09	1.65	0.86	0.86	0.64	-1.50
1989.10	1.97	3.06	2.76	2.60	2.60	2.16	-0.63
1989.11	1.67	3.03	3.67	3.61	3.61	3.92	1.01
1989.12	3.95	3.00	1.23	0.98	0.98	1.95	0.32
	Cabbage	ARIMA51	Wint51	SES51	Naive51	M. av51	
1990.01	2.74	3.65	3.10	4.51	4.54	2.71	0.81
1990.02	3.57	2.29	2.17	2.48	2.47	3.37	0.54
1990.03	2.04	2.24	2.11	1.85	1.86	1.31	-1.17
1990.04	1.04	2.20	1.91	1.93	1.93	1.84	-1.28
1990.05	1.30	2.16	1.80	1.27	1.26	1.71	-1.05
1990.06	2.72	2.12	2.05	2.19	2.19	2.17	-0.16
1990.07	2.39	2.07	3.19	3.39	3.39	3.13	0.51
1990.08	2.04	2.03	1.88	2.39	2.38	2.88	0.50
1990.09	2.90	1.99	0.86	0.87	0.87	1.04	-0.67
1990.10	5.53	1.95	3.77	3.72	3.73	2.71	0.16
1990.11	6.37	1.90	6.83	6.96	6.97	6.07	1.60
1990.12	4.56	1.86	6.47	4.99	4.99	5.29	0.22

Table C2 **CABBAGE: 2-steps ahead actual and forecast values (TT$/kg) (3,4 & 5 years estimation data)**

	CABBAGE	ARIMA32	Wint32	SES32	Naive32	M.av32	Factor
1988.01	3.84	N/A	N/A	N/A	N/A	N/A	1.06
1988.02	3.41	4.21	2.10	2.85	2.87	2.77	0.70
1988.03	2.27	4.32	1.09	1.44	1.46	1.31	-1.32
1988.04	1.95	5.41	1.36	1.46	1.46	1.48	-1.25
1988.05	1.48	6.46	3.28	3.00	3.03	2.81	-0.56
1988.06	1.58	7.19	4.29	4.02	4.01	4.11	0.81
1988.07	3.35	7.78	2.66	2.99	2.94	3.23	0.90
1988.08	3.99	8.34	0.38	1.31	1.25	1.57	0.48
1988.09	4.31	5.60	0.58	0.88	0.95	0.53	-1.50
1988.10	4.95	4.76	3.45	2.83	2.88	2.62	-0.63
1988.11	5.99	4.04	7.38	6.72	6.82	6.24	1.01
1988.12	3.26	3.44	7.18	5.91	5.90	5.96	0.32
	Cabbage	ARIMA42	Wint42	SES42	Naive42	M. av42	Factor
1989.01	1.52	N/A	N/A	N/A	N/A	N/A	1.06
1989.02	1.28	3.29	2.57	3.64	3.64	4.15	0.70
1989.03	1.19	3.63	0.94	-0.86	-0.86	-0.24	-1.32
1989.04	0.95	3.77	0.43	-0.67	-0.67	-0.70	-1.25
1989.05	2.12	3.57	1.75	1.95	1.95	1.47	-0.56
1989.06	1.84	3.66	2.62	3.01	3.01	3.09	0.81
1989.07	2.83	2.97	3.48	3.58	3.58	3.46	0.90
1989.08	2.84	3.44	1.55	1.51	1.51	1.92	0.48
1989.09	1.73	2.62	1.00	0.43	0.43	0.20	-1.50
1989.10	1.97	3.10	2.68	1.73	1.73	1.62	-0.63
1989.11	1.67	3.19	4.46	4.24	4.24	4.02	1.01
1989.12	3.95	2.93	3.23	2.92	2.92	3.08	0.32
	Cabbage	ARIMA52	Wint52	SES52	Naive52	M. av52	Factor
1990.01	2.74	N/A	N/A	N/A	N/A	N/A	0.81
1990.02	3.57	2.29	2.53	4.24	4.27	3.35	0.54
1990.03	2.04	2.61	0.71	0.77	0.76	1.21	-1.17
1990.04	1.04	1.39	1.98	1.74	1.75	1.48	-1.28
1990.05	1.30	2.79	2.67	2.16	2.16	2.12	-1.05
1990.06	2.72	2.19	2.55	2.16	2.15	2.38	-0.16
1990.07	2.39	2.33	2.52	2.86	2.86	2.85	0.51
1990.08	2.04	1.53	2.68	3.38	3.38	3.25	0.50
1990.09	2.90	2.13	0.70	1.22	1.21	1.46	-0.67
1990.10	5.53	1.82	1.73	1.70	1.70	1.78	0.16
1990.11	6.37	1.59	5.07	5.16	5.17	4.66	1.60
1990.12	4.56	0.61	6.93	5.58	5.59	5.14	0.22

Table C3 **CABBAGE: 3-steps ahead actual and forecast values** (TT$/kg) (3,4 & 5 years estimation data)

	Cabbage	ARIMA33	Wint33	SES33	Naive33	M.av33	Factor
1988.01	3.84	N/A	N/A	N/A	N/A	N/A	1.06
1988.02	3.41	N/A	N/A	N/A	N/A	N/A	0.70
1988.03	2.27	4.84	0.07	0.83	0.85	0.69	-1.32
1988.04	1.95	4.60	1.07	1.51	1.53	1.30	-1.25
1988.05	1.48	5.54	2.39	2.15	2.15	2.18	-0.56
1988.06	1.58	7.53	4.59	4.37	4.40	4.07	0.81
1988.07	3.35	7.55	4.16	4.11	4.10	4.25	0.90
1988.08	3.99	7.72	1.59	2.57	2.52	2.95	0.48
1988.09	4.31	6.11	-1.32	-0.67	-0.73	-0.25	-1.50
1988.10	4.95	4.99	1.74	1.75	1.82	1.19	-0.63
1988.11	5.99	4.30	5.36	4.47	4.52	4.12	1.01
1988.12	3.26	3.74	7.70	6.03	6.13	5.27	0.32
	Cabbage	ARIMA43	Wint43	SES43	Naive43	M. av43	Factor
1989.01	1.52	N/A	N/A	N/A	N/A	N/A	1.06
1989.02	1.28	N/A	N/A	N/A	N/A	N/A	0.70
1989.03	1.19	3.25	0.77	1.62	1.62	2.39	-1.32
1989.04	0.95	3.20	0.84	-0.79	-0.79	0.14	-1.25
1989.05	2.12	3.10	1.09	0.02	0.02	-0.03	-0.56
1989.06	1.84	2.97	2.76	3.32	3.32	2.60	0.81
1989.07	2.83	2.87	2.97	3.10	3.10	3.22	0.90
1989.08	2.84	2.77	2.84	3.16	3.16	2.98	0.48
1989.09	1.73	2.72	0.36	-0.47	-0.47	0.15	-1.50
1989.10	1.97	2.66	2.03	1.30	1.30	0.96	-0.63
1989.11	1.67	2.66	4.38	3.37	3.37	3.21	1.01
1989.12	3.95	2.60	4.02	3.55	3.55	3.22	0.32
	Cabbage	ARIMA53	Wint53	SES53	Naive53	M. av53	Factor
1990.01	2.74	N/A	N/A	N/A	N/A	N/A	0.81
1990.02	3.57	N/A	N/A	N/A	N/A	N/A	0.54
1990.03	2.04	2.24	1.07	2.53	2.56	1.19	-1.17
1990.04	1.04	2.17	0.58	0.66	0.65	1.32	-1.28
1990.05	1.30	2.14	2.74	1.97	1.98	1.57	-1.05
1990.06	2.72	2.16	3.42	3.05	3.05	2.98	-0.16
1990.07	2.39	2.03	3.02	2.83	2.82	3.16	0.51
1990.08	2.04	1.96	2.01	2.85	2.85	2.84	0.50
1990.09	2.90	1.91	0.86	2.21	2.21	2.01	-0.67
1990.10	5.53	1.92	3.77	2.05	2.04	2.41	0.16
1990.11	6.37	1.87	6.83	3.14	3.14	3.27	1.60
1990.12	4.56	1.86	5.17	3.78	3.79	3.03	0.22

Table C4 **TOMATO: 1-step ahead actual and forecasted values (TT$/kg) (3,4 & 5 years estimation data)**

	Tomato	ARIMA31	Wint31	SES31	Naive31	M.av31	Factor
1988.01	5.04	7.90	4.30	5.62	4.26	3.88	0.87
1988.02	2.75	8.07	2.23	3.06	2.48	2.09	-1.69
1988.03	2.45	7.91	1.48	2.46	2.21	2.08	-2.23
1988.04	2.31	7.70	1.60	2.13	2.21	2.09	-2.47
1988.05	1.99	7.56	3.36	4.60	4.78	3.22	-1.51
1988.06	3.88	8.15	3.76	4.02	2.92	3.56	-0.58
1988.07	5.11	8.34	3.67	5.14	5.07	4.59	0.61
1988.08	3.78	9.08	5.03	4.61	4.59	4.39	0.09
1988.09	3.59	8.99	3.76	3.32	2.68	2.99	-1.19
1988.10	6.78	8.99	5.90	5.33	5.60	5.14	0.82
1988.11	7.82	8.48	8.27	8.10	9.48	8.89	3.52
1988.12	6.03	8.91	8.13	8.54	8.24	9.07	3.94
	Tomato	ARIMA41	Wint41	SES41	Naive41	M. av41	Factor
1989.01	6.03	2.53	4.95	5.47	2.96	4.06	0.87
1989.02	3.40	0.60	2.54	2.91	3.47	1.94	-1.69
1989.03	2.55	1.67	1.71	2.37	2.86	2.89	-2.23
1989.04	1.62	3.17	1.39	2.13	2.31	2.47	-2.47
1989.05	1.69	3.61	2.36	3.09	2.58	2.93	-1.51
1989.06	1.93	3.85	3.66	4.02	2.62	3.07	-0.58
1989.07	3.91	3.18	3.85	5.24	3.12	3.46	0.61
1989.08	6.79	1.12	3.99	4.73	3.39	2.82	0.09
1989.09	5.16	1.58	3.47	3.45	5.69	3.90	-1.19
1989.10	7.02	4.34	5.94	5.46	7.17	7.43	0.82
1989.11	7.52	5.89	7.64	8.16	9.72	9.80	3.52
1989.12	7.34	4.20	7.26	8.58	7.94	9.04	3.94
	Tomato	ARIMA51	Wint51	SES51	Naive51	M. av51	Factor
1990.01	3.84	7.57	5.28	5.44	4.93	4.87	0.81
1990.02	3.02	7.05	2.79	3.00	1.40	1.94	-1.63
1990.03	2.90	6.25	1.96	2.59	2.60	1.79	-2.05
1990.04	2.90	4.43	1.54	2.42	2.69	2.54	-2.26
1990.05	4.31	4.71	2.35	2.90	3.38	3.27	-1.78
1990.06	8.11	5.04	3.48	4.15	5.56	5.09	-0.53
1990.07	7.04	6.36	4.12	5.38	9.32	8.05	0.68
1990.08	2.65	7.97	4.86	4.43	6.11	7.25	-0.25
1990.09	5.03	7.05	4.08	3.62	1.84	3.57	-1.06
1990.10	8.27	8.15	6.44	6.02	7.43	5.84	1.34
1990.11	7.45	8.53	7.94	8.14	10.44	10.02	3.51
1990.12	6.63	8.51	7.59	7.83	7.16	8.65	3.22

Table C5 TOMATO: 2-steps ahead actual and forecasted values (TT$/KG) (3,4 & 5 years estimation data)

	Tomato	ARIMA32	Wint32	SES32	Naive32	M. av32	Factor
1988.01	5.04	N/A	N/A	N/A	N/A	N/A	0.87
1988.02	2.75	4.52	1.85	3.06	1.70	1.51	-1.69
1988.03	2.45	5.26	1.22	2.46	1.94	1.74	-2.23
1988.04	2.31	6.57	1.11	2.13	1.97	1.91	-2.47
1988.05	1.99	5.88	3.00	4.60	4.68	3.11	-1.51
1988.06	3.88	7.59	4.46	4.02	4.20	4.17	-0.58
1988.07	5.11	7.93	3.61	5.14	4.11	4.43	0.61
1988.08	3.78	7.45	4.30	4.61	4.55	4.13	0.09
1988.09	3.59	6.10	4.40	3.32	3.31	3.30	-1.19
1988.10	6.78	10.02	5.99	5.33	4.69	4.84	0.82
1988.11	7.82	9.48	7.82	8.10	8.30	8.07	3.52
1988.12	6.03	7.70	8.36	8.54	9.90	9.61	3.94
	Tomato	ARIMA42	Wint42	SES42	Naive42	M. av42	Factor
1989.01	6.03	N/A	N/A	N/A	N/A	N/A	0.87
1989.02	3.40	1.39	2.54	2.91	0.40	0.95	-1.69
1989.03	2.55	-0.74	1.71	2.37	2.93	2.16	-2.23
1989.04	1.62	1.27	1.39	2.13	2.62	2.64	-2.47
1989.05	1.69	3.41	2.36	3.09	3.27	3.35	-1.51
1989.06	1.93	4.92	3.66	4.02	3.51	3.68	-0.58
1989.07	3.91	4.27	3.85	5.23	3.81	4.03	0.61
1989.08	6.79	4.02	3.99	4.72	2.60	2.59	0.09
1989.09	5.16	1.43	3.47	3.44	2.11	1.91	-1.19
1989.10	7.02	0.35	5.94	5.46	7.70	6.81	0.82
1989.11	7.52	4.83	7.64	8.16	9.87	10.00	3.52
1989.12	7.34	3.95	7.26	8.58	10.14	10.18	3.94
	Tomato	ARIMA52	Wint52	SES52	Naive52	M. av52	Factor
1990.01	3.84	N/A	N/A	N/A	N/A	N/A	0.81
1990.02	3.02	6.76	2.77	3.00	2.49	2.46	-1.63
1990.03	2.90	6.43	1.93	2.59	0.98	1.25	-2.05
1990.04	2.90	4.79	1.49	2.42	2.39	1.99	-2.26
1990.05	4.31	4.30	2.28	2.90	3.17	3.09	-1.78
1990.06	8.11	3.84	3.36	4.15	4.63	4.58	-0.53
1990.07	7.04	4.70	3.91	5.38	6.77	6.54	0.68
1990.08	2.65	5.31	4.60	4.43	8.39	7.75	-0.25
1990.09	5.03	7.67	3.86	3.62	5.30	5.87	-1.06
1990.10	8.27	9.49	6.20	6.02	4.24	5.11	1.34
1990.11	7.45	8.54	7.67	8.14	9.60	8.80	3.51
1990.12	6.63	6.66	7.33	7.83	10.15	9.94	3.22

Table C6 **TOMATO: 3-steps ahead actual and forecasted values** (TT$/kg) (3,4 & 5 years estimation data)

	Tomato	ARIMA33	Wint33	SES33	Naive33	M. av33	Factor
1988.01	5.04	N/A	N/A	N/A	N/A	N/A	0.87
1988.02	2.75	N/A	N/A	N/A	N/A	N/A	-1.69
1988.03	2.45	2.06	0.84	2.46	1.16	0.88	-2.23
1988.04	2.31	1.88	0.84	2.13	1.70	1.41	-2.47
1988.05	1.99	2.19	2.51	4.60	4.44	2.83	-1.51
1988.06	3.88	3.55	4.41	4.02	4.10	4.01	-0.58
1988.07	5.11	4.60	4.31	5.14	5.39	5.35	0.61
1988.08	3.78	4.90	4.24	4.61	3.59	3.89	0.09
1988.09	3.59	3.23	3.67	3.32	3.27	2.91	-1.19
1988.10	6.78	5.16	6.63	5.33	5.32	5.30	0.82
1988.11	7.82	7.76	7.91	8.10	7.39	7.62	3.52
1988.12	6.03	7.24	7.91	8.54	8.72	8.38	3.94
	Tomato	ARIMA43	Wint43	SES43	Naive43	M. av43	Factor
1989.01	6.03	N/A	N/A	N/A	N/A	N/A	0.87
1989.02	3.40	N/A	N/A	N/A	N/A	N/A	-1.69
1989.03	2.55	2.66	1.71	2.37	-0.14	0.69	-2.23
1989.04	1.62	2.20	1.39	2.13	2.69	1.54	-2.47
1989.05	1.69	2.28	2.36	3.09	3.58	3.61	-1.51
1989.06	1.93	4.31	3.66	4.02	4.20	4.32	-0.58
1989.07	3.91	4.02	3.85	5.23	4.70	4.96	0.61
1989.08	6.79	2.15	3.99	4.73	3.29	3.44	0.09
1989.09	5.16	4.82	3.47	3.45	1.32	1.58	-1.19
1989.10	7.02	4.56	5.94	5.46	4.12	3.82	0.82
1989.11	7.52	2.64	7.64	8.16	10.40	9.06	3.52
1989.12	7.34	4.24	7.26	8.58	10.29	10.49	3.94
	Tomato	ARIMA53	Wint53	SES53	Naive53	M. av53	Factor
1990.01	3.84	N/A	N/A	N/A	N/A	N/A	0.81
1990.02	3.02	N/A	N/A	N/A	N/A	N/A	-1.63
1990.03	2.90	5.93	1.93	2.59	2.07	2.03	-2.05
1990.04	2.90	4.35	1.49	2.42	0.77	1.18	-2.26
1990.05	4.31	4.11	2.27	2.90	2.87	2.26	-1.78
1990.06	8.11	4.20	3.36	4.15	4.42	4.31	-0.53
1990.07	7.04	4.46	3.90	5.38	5.84	5.76	0.68
1990.08	2.65	4.66	4.58	4.43	5.84	5.49	-0.25
1990.09	5.03	6.03	3.84	3.62	7.58	6.62	-1.06
1990.10	8.27	9.78	6.18	6.02	7.70	8.56	1.34
1990.11	7.45	6.54	7.64	8.14	6.41	7.71	3.51
1990.12	6.63	6.87	7.30	7.83	9.31	8.11	3.22

APPENDIX D

Secondary Analysis: data and model* forecasts

* The first number in the alpha-numeric model names used in the tables refers to the quantity of estimation data used. The second highlights the number of steps ahead being forecasted.

APPENDIX D

Table D1 **CABBAGE: one-step ahead actual and forecasted values (TT$/kg) (3, 4 & 5 yrs estimation data)**

	Cabbage	ARIMA31	Wint31	SES31	Naive31	Factor
1990.01	2.74	3.90	1.74	4.64	4.64	1.04
1990.02	3.57	3.70	2.37	2.95	2.95	1.03
1990.03	2.04	3.67	2.68	2.83	2.83	0.77
1990.04	1.04	3.54	2.14	1.21	1.21	0.77
1990.05	1.30	3.53	1.41	0.50	0.50	0.62
1990.06	2.72	3.45	1.56	0.76	0.76	0.94
1990.07	2.39	3.43	3.47	3.38	3.38	1.32
1990.08	2.04	3.38	2.66	4.50	4.50	1.43
1990.09	2.90	3.38	1.36	2.91	2.91	1.00
1990.10	5.53	3.35	2.92	3.42	3.42	1.18
1990.11	6.37	3.34	5.23	8.85	8.85	1.36
1990.12	4.56	3.32	7.16	7.73	7.73	0.89

	Cabbage	ARIMA41	Wint41	SES41	Naive41	Factor
1990.01	2.74	4.02	2.73	4.41	4.64	1.04
1990.02	3.57	3.80	2.51	3.05	2.95	1.03
1990.03	2.04	3.09	2.33	2.80	2.83	0.77
1990.04	1.04	2.57	1.92	1.32	1.21	0.77
1990.05	1.30	2.31	1.44	0.53	0.50	0.62
1990.06	2.72	1.70	1.45	0.77	0.76	0.94
1990.07	2.39	1.75	3.27	3.21	3.38	1.32
1990.08	2.04	1.67	2.41	4.43	4.50	1.43
1990.09	2.90	2.07	1.47	2.93	2.91	1.00
1990.10	5.53	2.15	3.43	3.42	3.42	1.18
1990.11	6.37	1.31	6.71	8.51	8.85	1.36
1990.12	4.56	1.17	6.46	7.57	7.73	0.89

	Cabbage	ARIMA51	Wint51	SES51	Naive51	Factor
1990.01	2.74	3.65	2.86	4.55	4.61	1.25
1990.02	3.57	2.29	2.34	4.07	4.07	1.19
1990.03	2.04	2.24	2.20	3.05	3.06	0.72
1990.04	1.04	2.20	1.87	1.02	0.99	0.67
1990.05	1.30	2.16	1.49	0.53	0.52	0.74
1990.06	2.72	2.12	1.53	0.87	0.88	0.92
1990.07	2.39	2.07	3.19	2.75	2.79	1.11
1990.08	2.04	2.03	2.12	2.96	2.97	1.12
1990.09	2.90	1.99	1.41	2.05	2.04	0.89
1990.10	5.53	1.95	3.88	2.93	2.93	1.13
1990.11	6.37	1.90	7.46	9.41	9.50	1.52
1990.12	4.56	1.86	7.03	10.30	10.37	1.07

Table D2 **CABBAGE: 2-steps ahead actual and forecasted values**
(TT$/kg) (3, 4 & 5 yrs estimation data)

	Cabbage	ARIMA32	Wint32	SES32	Naive32	Factor
1990.01	2.74	2.91	0.83	1.28	1.28	1.04
1990.02	3.57	3.64	1.61	4.59	4.59	1.03
1990.03	2.04	3.91	1.90	2.20	2.20	0.77
1990.04	1.04	3.12	2.65	2.83	2.83	0.77
1990.05	1.30	4.13	2.41	0.97	0.97	0.62
1990.06	2.72	3.35	1.67	0.75	0.75	0.94
1990.07	2.39	3.96	2.14	1.07	1.07	1.32
1990.08	2.04	2.88	3.56	3.66	3.66	1.43
1990.09	2.90	3.95	1.68	3.15	3.15	1.00
1990.10	5.53	3.04	1.49	3.43	3.43	1.18
1990.11	6.37	3.71	3.03	3.94	3.94	1.36
1990.12	4.56	2.28	6.10	5.79	5.79	0.89

	Cabbage	ARIMA42	Wint42	SES42	Naive42	Factor
1990.01	2.74	2.51	1.18	1.31	1.28	1.04
1990.02	3.57	3.78	2.50	4.37	4.59	1.03
1990.03	2.04	2.89	1.67	2.28	2.20	0.77
1990.04	1.04	2.23	2.17	2.80	2.83	0.77
1990.05	1.30	2.12	2.52	1.06	0.97	0.62
1990.06	2.72	1.85	1.59	0.81	0.75	0.94
1990.07	2.39	1.91	1.78	1.08	1.07	1.32
1990.08	2.04	1.73	3.22	3.47	3.66	1.43
1990.09	2.90	2.20	1.71	3.10	3.15	1.00
1990.10	5.53	2.63	1.78	3.46	3.43	1.18
1990.11	6.37	2.43	4.27	3.94	3.94	1.36
1990.12	4.56	1.38	6.78	5.57	5.79	0.89

	Cabbage	ARIMA52	Wint52	SES52	Naive52	Factor
1990.01	2.74	2.64	1.27	1.39	1.38	1.25
1990.02	3.57	2.29	2.45	4.33	4.39	1.19
1990.03	2.04	2.61	1.43	2.46	2.46	0.72
1990.04	1.04	1.39	2.02	2.83	2.85	0.67
1990.05	1.30	2.79	2.70	1.12	1.10	0.74
1990.06	2.72	2.19	1.76	0.65	0.64	0.92
1990.07	2.39	2.33	1.78	1.05	1.07	1.11
1990.08	2.04	1.53	2.84	2.78	2.81	1.12
1990.09	2.90	2.13	1.47	2.35	2.36	0.89
1990.10	5.53	1.82	1.87	2.60	2.59	1.13
1990.11	6.37	1.59	5.22	3.94	3.94	1.52
1990.12	4.56	0.61	8.24	6.62	6.69	1.07

Table D3 **CABBAGE: 3-steps ahead actual and forecasted values**
(TT$/kg) (3, 4 & 5 yrs estimation data)

	Cabbage	ARIMA33	Wint33	SES33	Naive33	Factor
1990.01	2.74	2.89	1.00	1.74	1.74	1.04
1990.02	3.57	2.94	0.72	1.27	1.27	1.03
1990.03	2.04	3.55	1.27	3.43	3.43	0.77
1990.04	1.04	3.32	1.86	2.20	2.20	0.77
1990.05	1.30	3.29	3.00	2.28	2.28	0.62
1990.06	2.72	3.27	2.92	1.48	1.48	0.94
1990.07	2.39	3.26	2.30	1.06	1.06	1.32
1990.08	2.04	3.19	2.14	1.16	1.16	1.43
1990.09	2.90	3.18	2.28	2.56	2.56	1.00
1990.10	5.53	3.23	1.88	3.72	3.72	1.18
1990.11	6.37	3.27	1.48	3.96	3.96	1.36
1990.12	4.56	3.25	3.46	2.58	2.58	0.89

	Cabbage	ARIMA43	Wint43	SES43	Naive43	Factor
1990.01	2.74	3.66	1.73	1.75	1.74	1.04
1990.02	3.57	2.40	1.05	1.30	1.27	1.03
1990.03	2.04	2.79	1.66	3.26	3.43	0.77
1990.04	1.04	2.01	1.55	2.28	2.20	0.77
1990.05	1.30	1.73	2.85	2.25	2.28	0.62
1990.06	2.72	1.59	2.82	1.61	1.48	0.94
1990.07	2.39	2.01	1.96	1.14	1.06	1.32
1990.08	2.04	1.94	1.73	1.17	1.16	1.43
1990.09	2.90	2.30	2.31	2.43	2.56	1.00
1990.10	5.53	2.78	2.08	3.66	3.72	1.18
1990.11	6.37	2.16	2.18	3.98	3.96	1.36
1990.12	4.56	2.64	4.29	2.58	2.58	0.89

	Cabbage	ARIMA53	Wint53	SES53	Naive53	Factor
1990.01	2.74	2.76	2.08	2.19	2.18	1.25
1990.02	3.57	2.64	1.06	1.32	1.31	1.19
1990.03	2.04	2.24	1.50	2.62	2.66	0.72
1990.04	1.04	2.17	1.30	2.29	2.29	0.67
1990.05	1.30	2.14	2.92	3.13	3.15	0.74
1990.06	2.72	2.16	3.22	1.40	1.36	0.92
1990.07	2.39	2.03	2.05	0.79	0.78	1.11
1990.08	2.04	1.96	1.57	1.06	1.08	1.12
1990.09	2.90	1.91	1.97	2.21	2.23	0.89
1990.10	5.53	1.92	1.95	2.98	2.99	1.13
1990.11	6.37	1.87	2.49	3.50	3.48	1.52
1990.12	4.56	1.86	5.75	2.77	2.77	1.07

Table D4 TOMATO: 1-step ahead actual and forecasted values
(TT$/kg) (3, 4 & 5 yrs estimation data)

	Tomato	Wint31	SES31	Naive31	Factor
1990.01	3.84	5.55	5.58	5.58	1.26
1990.02	3.02	3.12	3.38	3.39	0.70
1990.03	2.90	2.63	1.36	1.31	0.62
1990.04	2.90	2.25	1.07	1.06	0.59
1990.05	4.31	2.20	0.95	0.95	0.56
1990.06	8.11	3.41	2.33	2.35	0.97
1990.07	7.04	4.99	10.10	10.31	1.31
1990.08	2.65	5.44	10.34	10.40	1.13
1990.09	5.03	4.39	2.97	2.81	0.94
1990.10	8.27	6.45	6.86	6.94	1.46
1990.11	7.45	7.29	20.23	20.57	1.71
1990.12	6.63	6.98	21.08	21.13	1.66
	Tomato	Wint41	SES41	Naive41	Factor
1990.01	3.84	5.55	5.14	5.58	1.26
1990.02	3.02	2.11	2.86	3.39	0.70
1990.03	2.90	2.44	2.52	1.31	0.62
1990.04	2.90	2.38	2.40	1.06	0.59
1990.05	4.31	3.43	2.27	0.95	0.56
1990.06	8.11	6.04	3.94	2.35	0.97
1990.07	7.04	11.43	5.32	10.31	1.31
1990.08	2.65	7.01	4.59	10.40	1.13
1990.09	5.03	2.10	3.82	2.81	0.94
1990.10	8.27	8.17	5.93	6.94	1.46
1990.11	7.45	10.72	6.94	20.57	1.71
1990.12	6.63	7.15	6.74	21.13	1.66
	Tomato	Wint51	SES51	Naive51	Factor
1990.01	3.84	5.06	5.37	4.97	1.29
1990.02	3.02	2.80	2.91	3.47	0.70
1990.03	2.90	2.06	2.46	1.25	0.59
1990.04	2.90	1.67	2.23	0.91	0.53
1990.05	4.31	2.38	2.74	0.99	0.65
1990.06	8.11	3.40	4.00	2.67	0.95
1990.07	7.04	4.04	5.34	9.75	1.26
1990.08	2.65	4.64	4.31	9.05	1.02
1990.09	5.03	3.92	3.47	2.21	0.82
1990.10	8.27	6.22	6.05	5.88	1.43
1990.11	7.45	7.74	8.32	23.42	1.98
1990.01	6.63	7.42	8.04	28.19	1.91

Table D5 TOMATO: 2-steps ahead actual and forecasted values (TT$/kg) (3, 4 & 5 yrs estimation data)

	Tomato	Wint32	SES32	Naive32	Factor
1990.01	3.84	5.37	5.56	5.54	1.26
1990.02	3.02	3.12	3.10	3.10	0.70
1990.03	2.90	2.63	2.99	3.00	0.62
1990.04	2.90	2.25	1.29	1.24	0.59
1990.05	4.31	2.20	1.01	1.01	0.56
1990.06	8.11	3.41	1.65	1.65	0.97
1990.07	7.04	4.99	3.14	3.17	1.31
1990.08	2.65	5.44	8.71	8.89	1.13
1990.09	5.03	4.39	8.60	8.65	0.94
1990.10	8.27	6.45	4.61	4.37	1.46
1990.11	7.45	7.29	8.04	8.12	1.71
1990.12	6.63	6.98	19.64	19.97	1.66
	Tomato	**Wint42**	**SES42**	**Naive42**	**Factor**
1990.01	3.84	5.45	5.14	5.54	1.26
1990.02	3.02	3.05	2.86	3.10	0.70
1990.03	2.90	1.71	2.52	3.00	0.62
1990.04	2.90	2.01	2.40	1.24	0.59
1990.05	4.31	2.81	2.27	1.01	0.56
1990.06	8.11	4.80	3.94	1.65	0.97
1990.07	7.04	8.51	5.32	3.17	1.31
1990.08	2.65	11.37	4.59	8.89	1.13
1990.09	5.03	5.53	3.82	8.65	0.94
1990.10	8.27	3.41	5.93	4.37	1.46
1990.11	7.45	10.59	6.94	8.12	1.71
1990.12	6.63	10.28	6.74	19.97	1.66
	Tomato	**Wint52**	**SES52**	**Naive52**	**Factor**
1990.01	3.84	5.03	5.37	4.90	1.29
1990.02	3.02	2.81	2.91	2.70	0.70
1990.03	2.90	2.06	2.47	2.92	0.59
1990.04	2.90	1.66	2.23	1.12	0.53
1990.05	4.31	2.36	2.74	1.12	0.65
1990.06	8.11	3.37	4.00	1.45	0.95
1990.07	7.04	3.98	5.34	3.54	1.26
1990.08	2.65	4.61	4.31	7.89	1.02
1990.09	5.03	3.94	3.47	7.27	0.82
1990.10	8.27	6.20	6.05	3.86	1.43
1990.11	7.45	7.71	8.32	8.14	1.98
1990.01	6.63	7.41	8.02	22.60	1.91

TABLE D6 **TOMATO: 3-steps ahead actual and forecasted values** (TT$/kg) (3, 4 & 5 yrs estimation data)

	Tomato	Wint33	SES33	Naive33	Factor
1990.01	3.84	5.22	6.10	6.07	1.26
1990.02	3.02	3.02	3.09	3.08	0.70
1990.03	2.90	2.63	2.75	2.75	0.62
1990.04	2.90	2.25	2.85	2.86	0.59
1990.05	4.31	2.20	1.23	1.18	0.56
1990.06	8.11	3.41	1.76	1.75	0.97
1990.07	7.04	4.99	2.23	2.23	1.31
1990.08	2.65	5.44	2.71	2.73	1.13
1990.09	5.03	4.39	7.25	7.40	0.94
1990.10	8.27	6.45	13.36	13.43	1.46
1990.11	7.45	7.29	5.40	5.11	1.71
1990.12	6.63	6.98	7.80	7.89	1.66
	Tomato	Wint43	SES43	Naive43	Factor
1990.01	3.84	6.77	5.14	6.07	1.26
1990.02	3.02	2.99	2.86	3.08	0.70
1990.03	2.90	2.47	2.52	2.75	0.62
1990.04	2.90	1.41	2.40	2.86	0.59
1990.05	4.31	2.38	2.27	1.18	0.56
1990.06	8.11	3.94	3.94	1.75	0.97
1990.07	7.04	6.77	5.32	2.23	1.31
1990.08	2.65	8.47	4.59	2.73	1.13
1990.09	5.03	8.97	3.82	7.40	0.94
1990.10	8.27	8.98	5.93	13.43	1.46
1990.11	7.45	4.43	6.94	5.11	1.71
1990.12	6.63	10.15	6.74	7.89	1.66
	Tomato	Wint53	SES53	Naive53	Factor
1990.01	3.84	5.02	5.37	6.33	1.29
1990.02	3.02	2.80	2.91	2.66	0.70
1990.03	2.90	2.06	2.46	2.27	0.59
1990.04	2.90	1.67	2.23	2.62	0.53
1990.05	4.31	2.35	2.74	1.38	0.65
1990.06	8.11	3.34	4.00	1.63	0.95
1990.07	7.04	3.94	5.34	1.93	1.26
1990.08	2.65	4.53	4.31	2.87	1.02
1990.09	5.03	3.90	3.47	6.35	0.82
1990.10	8.27	6.23	6.05	12.68	1.43
1990.11	7.45	7.68	8.32	5.35	1.98
1990.01	6.63	7.39	8.02	7.85	1.91

APPENDIX E

APPENDIX E

ARIMA MODEL SPECIFICATION WITH MicroTSP

The aim in ARIMA modelling is to specify a model that allows us to accounts for all of the systematic patterns within the data set. Conversely the presence of random residuals (white noise) after model specification indicates a well-specified model. Two tests can be used to check for the presence of white noise.

One is visual check of the ACF and PACF of the residuals. With very few exceptions, statistically these should not be significantly different from zero (eg. they should all lie within the 95% confidence bands for a random series).

The Ljung-Box Q statistic is another test for white noise. The Ljung-Box Q statistic estimates the probability that autocorrelations as large or larger than those observed could have been the result of random variation. Statistically this should not be significant. It should thus assume a value greater than say 5%. Therefore a probability of 70% indicates that white noise could have generated autocorrelations as large or larger than those observed 70% of the time, (SPSS Inc., 1990).

The actual model-fitting procedure involves first determining stationarity of the data. This is accomplished through use of the Adjusted Dickey-Fuller(ADF) test. As described in chapter 4 of this report, non-stationarity exists when the ADF test t-statistic value is lower in absolute value than the critical MacKinnon t-statistic. The opposite case exists when the data is stationary. If the original data series are non-stationary then the first difference of that series is tested. If this differenced series is stationary then the series is integrated of order one. Once the data is stationary it can be used in the specification process.

Model-specification using the Micro-TSP software involves a procedure in which the model is initially "overfit" with AR and MA terms, followed by a process wherein those with insignificant t-statistics are dropped from the equation. Thus each statistically significant spike in the AC and PAC functions are included first as an ARI regression equation and then as an IMA regression equation. In both equations the terms with insignificant t-statistics are dropped. The significant terms in the two equations are then placed in one ARIMA regression equation. Again the terms with insignificant t-statistics are dropped from this ARIMA equation. The terms with significant t-statistics are those in the well-specified ARIMA equations.

The Ljung-Box statistic and/or visual test is used to check the residuals for white noise.