FORECASTING MODELS FOR THE U.K. PIG SECTOR

By

Neill W. Calvert B.A.

A thesis submitted to the University of Nottingham

for the degree of Doctor of Philosophy

May 1989

ABSTRACT

In this thesis, forecasting models for the UK pigmeat sector are built using various methodologies with particular interest being paid to the relative forecasting ability of time series models compared with the performance of biological and econometric methodologies. The main determinant of the supply of pigmeat in the UK is the size of the breeding herd, the quantity of meat itself being directly attributable to the number of fat pigs slaughtered and to a lesser extent cullings of older sows and boars from the breeding herd. These three key variables are the ones modelled in this thesis. Prior to building forecasting models an explanation is given of the system underpinning the pig sector, in terms of the biology of the breeding herd pig, the mechanism of how supply responds to prices, and consideration of the well documented 'pig cycle'. Thus, the workings of the biological and economic mechanisms are described in the context of an equilibrium framework before the relevant models are built.

Having built the various models, their relative forecasting performance is measured by consideration of the size of the forecast errors and the ability of the models to forecast the directional movements of the actual series in a specified out-of-sample period. In the concluding chapter, suggestions are made as to how the models might be developed further and how the various approaches might be combined into a single forecasting model.

The availability of data has an important influence on much of the model building methodology and forecasting analysis. Consideration is given at various points in the thesis to circumventing these restrictions.

CONTENTS

ACKNOWLEDGEMENTS

I should like to offer my thanks to my supervisors, Professor Tony Rayner and Mr. John Bates for their continuous guidance and help throughout my time of research at Nottingham, and to Dr. Andrew Jennings who joint supervised with Tony in the first year of my research before leaving his lecturing career. I trust I had nothing to do with his decision to leave!

I also wish to acknowledge the help and interest of the economics department at the M.L.C. for their help in the provision of valuable data and other helpful information and comments.

I thank M.A.F.F. for providing the finance, without which this research would not have been undertaken, and also for providing some of the data.

Last but not least I should like to express my appreciation for the unceasing love, support and encouragement of my parents. You are the best.

I dedicate this thesis to my late Grandparents, Jack and Gladys Menmuir and Jack and Violet Calvert. The memories of the times spent with them in my younger years, their love and their influence on my life will be with me always.

> '....do not throw your pearls to pigs. If you do, they may trample them under their feet and then turn and tear you to pieces.'
>
> Matt 7:6

> 'Because of the Lord's great love we are not consumed, for His compassions never fail. They are new every morning; great is your faithfulness.'
>
> Lam. 3:22-23

CHAPTER ONE

INTRODUCTION

1.1 The Background to and Objectives of This Thesis

One of the most well known phenomenon in agriculture is the presence of the pig cycle, which has long since been of considerable interest to academics and policy makers alike. Academically, the cycle is of interest in that it has considerable implications for policy, and because it is the only significant example of the cobweb theorem. The ups and downs in the size of the pig breeding herd during the course of the cycle in turn produce fluctuations in the supply and, therefore, the price of pigmeat and, subsequently, in the returns to producers. As one of the prime objectives of agricultural policy is to achieve stability, the policy of the EEC's pigmeat regime, for example, is explicitly counter cyclical in the way it operates[1]. Consequently, knowledge of the movements, past and future, of the cycle are of obvious relevance to bodies such as the EEC and the UK's Meat and Livestock Commission, MLC, a statutory body whose job, *inter alia,* is to monitor the UK meat sector, provide information to the sector and improving market efficiency. Because, forecasting the future breeding herd is of such interest, this thesis is concerned primarily with the quantitative aspects of the cycle rather than explicitly addressing the underlying causes of it, in an attempt to build models which, it is hoped, will be of use in forecasting the key variables of the sector.

The prime objective of this thesis is to build forecasting models for the UK pigmeat sector using different methodologies in order to make comparisons of their relative forecasting performances over the short and medium/long term. Specifically, the relative forecasting abilities of econometric, biological and univariate and bivariate Box-Jenkins methodologies will be examined. The prime variables to consider when modelling the pig meat sector are the size of the domestic herd, which then determines the number of pigs able to be produced for a relevant lead period, and the variables which directly affect the supply of pig meat, the number of fat pig slaughterings, and to a lesser extent, sow and boar cullings. Studies in the recent past which have aimed to model the pig meat sectors of the UK and the USA, for example, Ness and Coleman (1976), Savin

1. See The pigmeat Regime MLC's European Handbook Vol 1.

(1978), Burton (1987), and for the US, Westcott and Hull (1985) and Stillman (1985) have all concentrated on modelling the breeding herd. Having determined, from the size of the breeding herd, the production of pigs for the lead time specified the produceable quantity of pigmeat can be inferred.[2]

A common method for modelling the UK pig herd in recent years was initiated by Diane Savin in an article entitled '*Forecasting the Pig Breeding Herd - an Examination of Differential Response to Changes in Profitability*' [3]. In this article she broke from the convention of modelling the breeding herd itself directly as a function of some profit margin, as had been the case in the cited US studies and the UK studies prior to hers. In order to improve the success of modelling the length and amplitude of the well known pig cycle[4] she introduced the idea of modelling the breeding herd as a system of inflows and outflows. Inflows into the breeding herd in the form of boars and gilts-in-pig - or pregnant gilts as they will be referred to in this thesis - which are sows that are in-pig for the first time. Outflow from the breeding herd takes the form of cullings of sows and boars which are no longer considered economically viable to remain in the breeding herd. Both the inflow and the outflow variables are then modelled as functions of a profit variable. Savin was content that she had met with some success in meeting her stated objective to model the breeding herd and its fluctuations over time, her model providing the basis for subsequent work by, for example, Burton (1987) and is the basis of the breeding herd forecasting model currently used by the MLC to provide forecast to the UK pig meat industry and the EEC.

In this thesis, it was decided that a similar inflow/outflow approach to modelling the breeding herd would make a useful basis on which to build a breeding herd forecasting model for the UK. In addition to an econometric approach, however, the decision was taken to analyse the forecasting abilities of a biologically based forecasting model from which the use of profit as an explicit regressor is excluded. Because econometric models of the breeding herd usually include biological features, implicitly or explicitly, the econometric approach to

2. The interested reader can see such methods of determining the number of fat pig and pork marketings in Savin (1978), page 109, and Westcott and Hull (1985), page 37.

3. In '*Supply Response and The World Meat Situation*', Procedings of a Symposium held 13 and 14 April 1978, MLC Economic Information Service, 1978.

4. Work by McClements(1970) suggests that the uk pig cycle varied in the range of 36 to 42 months, though more recent work by Ridgeon given in Green Europe, published by AGRA EUROPE states that the cycle in recent years has increased in length to 5 years.

modelling is, in effect, a more sophisticated type of biological model. Thus, in addition to the insight that the biological model will provide into the biological relationships that exist within and between the breeding and the feeding herds, the comparative forecasting performance of the biological and the econometric approaches will be of obvious interest in the context of this study.

As will become clearer when the breeding herd model is explained more fully, recursive links exist between inflow and the breeding herd and between outflow and the breeding herd lagged for a period appropriate to each case. Consequently, the breeding herd can be expressed as a function of past values of itself as determined by the lags in the inflow,outflow and breeding herd relationships. In view of this, and in view of the fact that I was unaware of the existence of such a study in the UK, it was deemed of interest to examine the comparative forecasting performance of univariate statistical models with those of the biological and econometric models. In this context, Box-Jenkins methodology as proposed by the said authors in their book *'Time Series Analysis- forecasting and control'*[5] will be employed to build appropriate models. Such models are of particular interest for their short term forecasting abilities and have the useful characteristic that a working forecasting model can be developed with no prior knowledge of the variable they are built to forecast. Having said this, they do exhibit the drawbacks of requiring a large amount of homogeneous data and learning the art of model building can be a time consuming affair. An even more sophisticated extension of the univariate Box-Jenkins models are multivariate models. Where it is deemed appropriate, the possibility of building bivariate models will be examined as an alternative to the traditional econometric approach.

Before the methodological approach to modelling and forecasting in this thesis is expounded, a resumé of the pig meat sector in the UK over recent years, a period in which the industry has undergone much change, will be discussed briefly in order to help the reader to understand the context of the period of study.

1.2 A Recent History of the U.K. Pig Meat Sector.[6]

Over the last twenty years or so, the pig meat industry has undergone a considerable number of changes, which have considerably affected methods of

---

5. Box, G.E.P., and Jenkins, G.M. *'Time Series Analysis-forecasting and control'*, Holden-Day, San Francisco, 1970.

6. Much of the information for this section comes from the a special survey published by the MLC in their Market Survey for April 1988.

production and marketing in the sector. These changes have important effects on the biological and economic relationships within and between the breeding and feeding herds and consequently, modelling of key variables of the sector to be considered in this thesis is influenced.

The major effect on the industry over the last 20 years is that it has become much more concentrated and specialised, so that whereas it was quite common 20 years ago to have a large number of small holdings where farmers reared pigs for meat and replenished his breeding herd with gilts from his own stock, today 71% of pig production in England and Wales is concentrated in the hands of producers with herd sizes larger than 100 pigs. The latter figure compares with 54% for such herd sizes as little ago as 1978. In the UK the overall effect of the increased concentration has been to reduce the number of pig farms to 22,000 in 1986, a reduction of 75% of the number in 1968 and there are now fewer holdings with pigs than at any time since the war. In tandem with the increase in concentration, the industry has become increasingly specialised. The number of farms where both breeding and rearing take place is relatively small compared with the late sixties and seventies, 80% of breeding herd boars and 90% of replacement gilts being provided by the specialised breeding companies which have become so common in recent years. Along with increased concentration, the average herd size has increased substantially from 26 in 1978 to 49 in 1986.

Although the number of producers has fallen over the last two decades, the production of pigs, as indicated by the number of slaughterings of fat pigs, has risen. The consequence of this increase in production, added to increased imports of pigmeat and fairly static demand has been to depress the price of pig meat, which has occurred at the same time as increased input prices, principally accounted for by feed costs. Thus, a consequence of the increased efficiency, which has resulted from the economies of scale experienced in the pig sector, has been to squeeze the sector of profits which has accelerated the loss of the less efficient, often smaller farmer.

Like all agricultural sectors, the pig meat sector is provided for under the Common Agricultural Policy - CAP - of the EEC, specifically by the Pig Meat Regime set up on July 1st 1967, the latter being closely linked to the cereal regime because of the high percentage of production costs accounted for by cereal feedstuffs.[7] The UK production of pig meat accounted for 12% of EEC

7. For a detailed exposition of the EEC Pig Meat Regime the reader is directed to section 5 of the MLC European Handbook, *The Common Agricultural Policy*, Volume 1.

production in 1986, the EEC being slightly more than self sufficient in pig meat production and an established net exporter. Although the regime makes provision for intervention buying if the market price falls below a specified 'basic' price, direct intervention has never been used in the pig meat sector. The EEC's main assistance to pig producers has been to prevent imports and aid exports. The Community has also attempted to stabilise the market by providing special storage aid schemes. The latter is preferred to intervention buying on the grounds that it is only a temporary measure, likely to be cheaper than direct EEC intervention, and it is felt that the private sector is better able to judge the local market, knowing when the time is right to release the surpluses back onto the market without depressing prices. The funding of the regime is the responsibility of the European Agricultural Guarantee and Guidance Fund - FEOGA - although the regime accounts for only one percent of the total guarantee fund, of which approximately two-thirds finances export refunds, the remaining third directed to private storage aid. The result of this method of support used by the EEC pigmeat regime is that the UK pig producers, along with their European counterparts, are left susceptible to the ups and, more relevant to the period under discussion, the downward pressures of the market. This was not the case prior to 1973 when the deficiency payments minimum price guarantee scheme was operative in the UK. Having said this the UK government was given permission to give a temporary cleanpig subsidy of 6.7p per kg deadweight in the first half of 1977 to help restore confidence in the sector, characterised by the high culling levels at the end of 1976 and the beginning of 1977.

Pigmeat production is an example of perfect competition and the individual producers, however large they might appear to be, have no way of influencing the price they receive for their pig meat, their prime concern in order to be able to stand these market pressures is to reduce their costs of production. They are able to do this through good husbandry management and technical efficiency, the latter being an area where most has been gained in recent years. Because these technical changes have direct and indirect implications for modelling, the trends in key technical coefficients over the last decade or so are examined below.

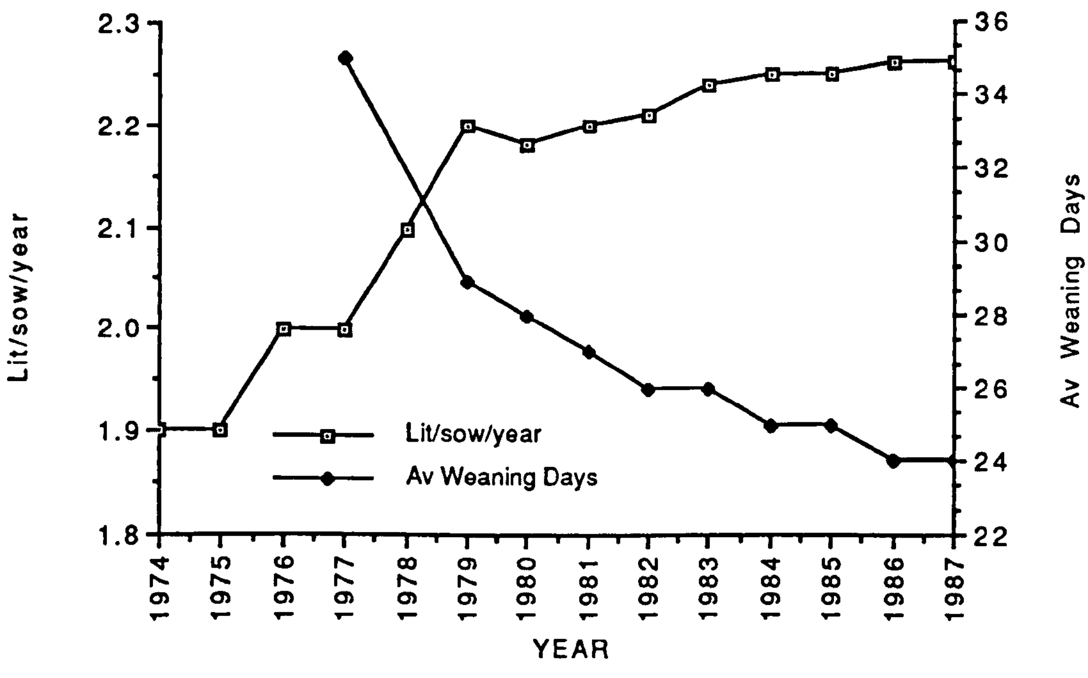
1.3 Recent Trends in the Technical Features of the U.K. Pig Meat Sector.

Basically, the UK pig industry can be split into the breeding sector and the rearing sector, the aims of the producers in both being to produce breeding herds for their reproductive capabilities and feeding herds for their meat as efficiently as possible. The nature of the pig sector is such that the technical features associated with the breeding, rearing and feeding performance of the industry have a significant effect on its performance and, therefore, must be taken into consideration when modelling it. The reason for the importance of such technical features compared with the pig meat sector's two principal red meat rivals is the large number of offspring per litter and a gestation period which although similar to that of sheep is considerably shorter than that of cattle, so that productivity per annum can be affected significantly by improvements in technical factors.

The primary aim of pig farmers is to produce the maximum number of pigs per sow per annum, a figure which is influenced by a number of technical coefficients. The key determinants include the number of litters per sow per annum, which the producer can influence directly by changing the length of the weaning period. The number of litters per year is also affected by the managerial and husbandry abilities of producers, who must aim to ensure that a sow is successfully re-served as quickly as possible after weaning her litters. The second key influence in determining the number of pigs reared per sow per year is the number of piglets successfully reared per litter, a figure determined by the number of pigs born live per sow per litter and the mortality percentage of pigs born live. Figures 1.1 to 1.3 illustrate the trends which have taken place in these technical coefficients for the period 1974 to 1987 as collated by the Pig Plan Survey conducted by the Meat and Livestock Commission - M.L.C.[8]

Figure 1.1

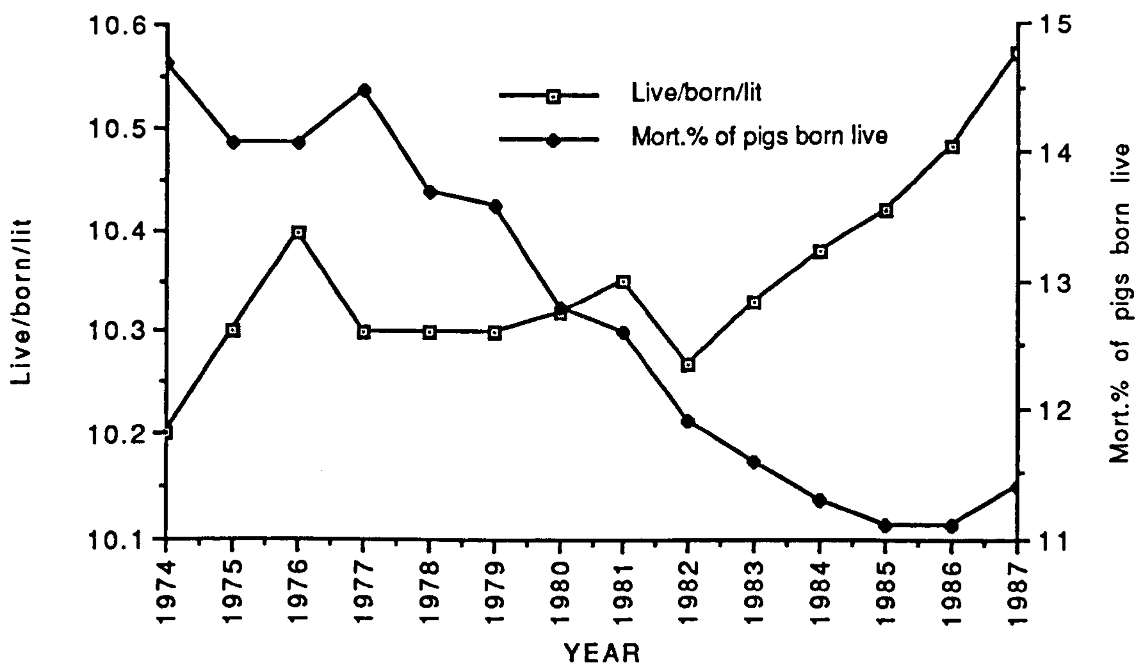
Litters Per Sow Per Year and Average Weaning Age 1974-87



8. All data for figures 1.1 to 1.4 are obtained from M.L.C.'s Pig Yearbooks 1984-88 and are collated by their Pig Plan Survey using a representative sample of the UK breeding herd.

Figure 1.1 indicates that the number of litters per sow per annum has risen steadily from 1.9 in 1974 to 2.26 in 1987, although the rate of increase noticeably slows in the eighties compared with that experienced in the seventies. Unfortunately, the pig plan survey does not provide information on average weaning ages prior to 1979, however, the MLC Market Survey for April 88 quotes the average weaning age for 1977 as 35 days and so this figure is incorporated into figure 1.1. The figure indicates that the slow down in the rate of change in the two trends occurs at about the same time indicating that the reason for the slow down in the increase in the numbers of litters per sow per annum is largely a consequence of the deceleration in the shortening of the weaning period. The alternative explanation for the slow down is a husbandry one, in that the break in the litters per annum series could have occurred as a result of a sudden slowing down in the increased ability on the part of producers to get their sows successfully re-served after weaning. The implausibility of the latter argument coupled with the evidence presented in figure 1.1 however, suggest that it was a change in the rate of decrease in weaning age which was the cause of the decelerated increase in the number of litters per year. Having said this, the MLC Market Survey for April 1988 suggests that the gains in productivity from reductions in weaning age are now less likely and that producers will have to turn their attentions to husbandry and genetic improvements.[9]

Figure 1.2
Live Pigs Born Per Litter and Mortality Percentage of Pigs Born Alive 1974-87

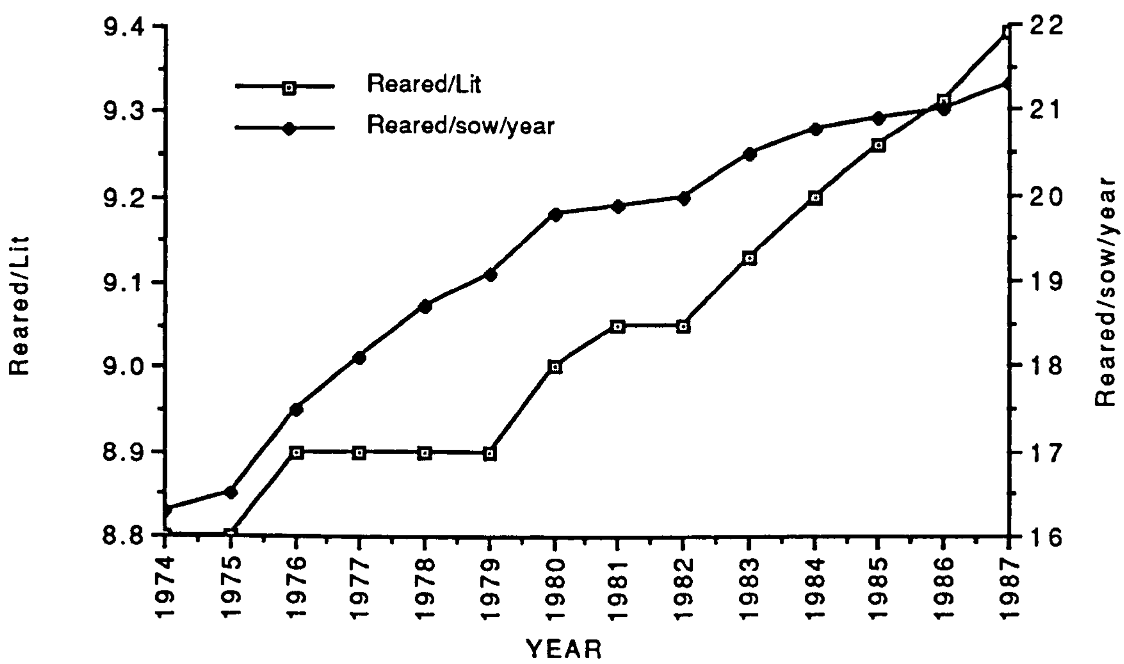


9. MLC Market Survey April 1987, *'The Changing structure of the pig meat marketing chain'*, page 3.

Figure 1.2 shows the trends in the number of pigs born live per litter and the mortality percentage of pigs born live, the latter being the results of losses through disease, illnesses, or more commonly, accidental death, usually through piglets being crushed by the mother. The two trends augur well for the efficiency of the industry, the number of deaths relative to the litter size steadily declining throughout the sample period and the number of live pigs born per litter increasing from around 12 in 1974 to around 15 in 1987. The latter implies an increase in the productivity of the breeding sows, presumably a result of technical and management improvements.

The consequence of the trends illustrated in figure 1.2 is that the number of pigs reared per litter has increased over the period from just above 8.8 in 1974 to almost 9.4 in 1987 as illustrated in figure 1.3 below.

Figure 1.3
Pigs Reared Per Litter and Numbers Reared Per Sow Per Year

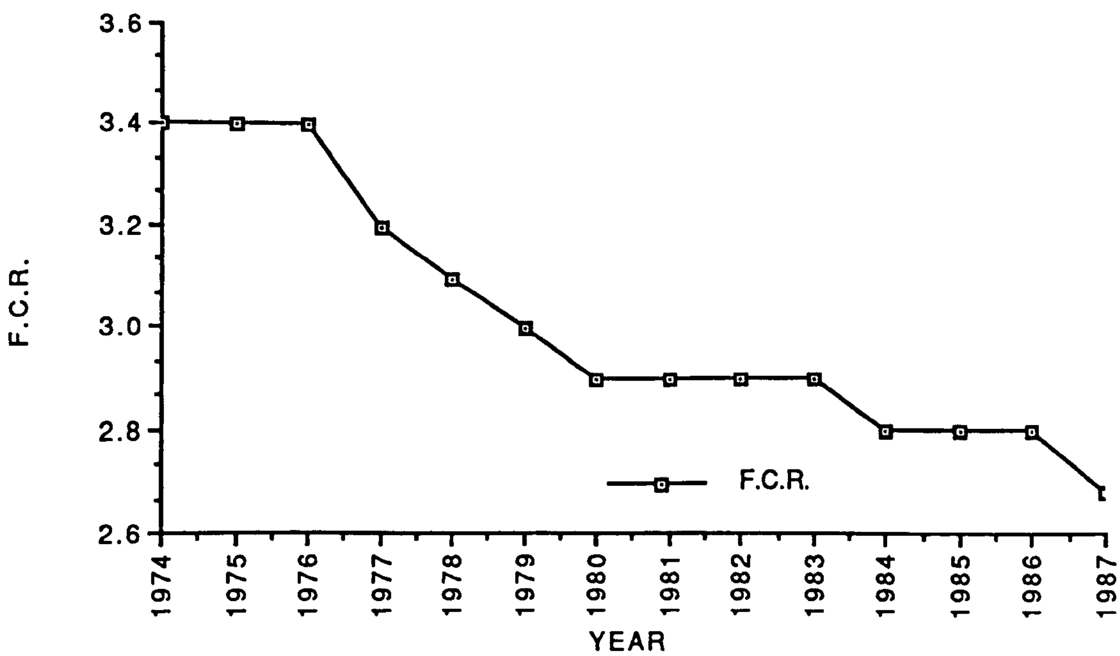


The other series presented in figure 1.3 is the key coefficient of interest, the number of pigs reared per sow per year, whose value is a direct consequence of the technical coefficients analysed above, though ultimately its value is dependent on the number of litters per sow per year and the number of pigs reared per litter. Consequently, the industry has experienced an increase in sow productivity in every year of the sample period, the change in productivity being just over 16

pigs reared per sow per annum in 1974 to over 21 in 1987.
The consequence of these trends in technical coefficients for the modelling of the sector will be seen in later chapters, but the obvious comment to make is that the number of pigs which can be produced by a given size of breeding herd has increased consistently since 1974, or conversely, the herd size required to produce a given number of pigs has decreased.
Moving on from the technical coefficients of the breeding herd, the main concern of the producer given a litter of piglets is to turn them into a saleable commodity as quickly and efficiently as possible. Given that the price of feedstuff is more or less out of the hands of the producer, his basic concern is to make the pig grow as quickly as possible for as little food inputs. Thus, another important technical coefficient in the pig meat industry is the feed conversion ratio - F.C.R. - measured by the quantity of feedstuff required, (usually expressed in Kilo-grammes), per pig in order to achieve one kilo-gramme of liveweight gain. Needless to say, the efficient producer aims to minimise the F.C.R. for each of his pigs. Figure 1.4 below presents the FCR as measured in the rearing herds selected by the MLC Pig Plan survey for the same period as used to analyse the breeding herd above.

Figure 1.4

The Feed Conversion Ratio For the UK Feeding Herd 1974-87



The series shows a feed conversion ratio which has consistently decreased throughout the sample period, illustrating the improved technical efficiency of the industry in the rearing of pigs for their meat.

In both the breeding herd and the feeding herd, therefore, there are clear indications of improved efficiency over recent years. The main consequences of these improvements in technical efficiency is that the industry has become much more concentrated and the increased number of pigs produced has, paradoxically, driven many of the smaller less efficient farmers out of pig meat production as prices have fallen. Future gains in productivity may not be quite so great as they have been in the period of study given natural limits on factors such as the length of the weaning period and animal welfare considerations such as the use farrowing crates and the re-emergence of less intensive open air production.

Having described the aims and the background to the thesis, the methodology employed in the modelling and forecasting process is outlined in the following section.

1.4 Data, Methodology and Outline of the Thesis

Any empirical study such as the one proposed here is dependent to a large extent upon the frequency and quality of data available to the researcher. This fact is very relevant to this study and, as will be explained in further detail at the appropriate points throughout the thesis, the data have a significant influence on the type of analysis employed and the period of estimation and forecasting. The data for the breeding herd come from the June census of agricultural holdings in the UK and the three sample censuses taken at points throughout the year, conducted by the relevant ministries for each member country. Prior to 1974, the three sample censuses took place at quarterly intervals in the months of March, September and December; the accession of the UK to the EEC in 1973 brought a change in the timing of the spring and autumn censuses to April and August respectively. These changes in sample census timing necessitated that the biological and econometric analyses were made post 1973 in order that the intervals between the censuses are equally spaced. One of the consequences of this is that the data for the biological and econometric analyses come only from the three sample censuses, the interval between observations being four months. This time period is referred to as a 'trimestic' time period throughout the thesis. The change in census times meant that Savin's work in 1978, which was concerned with quarterly time intervals, had to be restricted to the English and Welsh herds, for which the March and September sample censuses continued alongside the new April and August censuses up until 1977. Burton's 1987 model of the UK pig sector was conducted with a six-monthly data set, using the information from the June and December censuses. My choice of the trimestic time period, therefore, introduces another new dimension to modelling the breeding herd.

I was particularly interested in the forecasting performance of time series models in comparison with more traditional econometric approaches so that univariate models have been built for all the key variables examined. In the case of the trimestic breeding herd, the forecasts of univariate Box-Jenkins models are compared with those produced by biological and econometric models: the former being expected to perform better in the short term, the latter two in the longer term. The univariate models can be regarded as the simplest form of model in that only one variable is included and the models are purely statistical, no prior knowledge of the series to be modelled being required. The biological models can be viewed as the next most sophisticated, the models having introduced knowledge of the breeding herd system. The econometric models are even more sophisticated in that as well as including the economic phenomenon of prices, biological phenomena are included implicitly and/or explicitly. For the two monthly slaughter series, for which more observations are available, bivariate time series models are built including profit as the second variable rather than taking a more traditional econometric approach. The forecasting performances of the bivariate Box-Jenkins models are then compared with those of the univariate counter-part and a biological model.

The change in census timings will clearly affect the Box-Jenkins analysis of the breeding herd since such methodology requires lengthy time series of equally spaced observations in order to make analysis feasible. How the problem of the census timings affects the Box-Jenkins analysis is dealt with at considerable length in chapter three, in which the Box-Jenkins methodology is applied to build univariate models for the breeding herd series and its component parts. The data for the two slaughter categories modelled are monthly data collated by the MLC by returns from slaughter houses and are less problematic than the census data in that they are not subject to sampling errors and do not suffer from the timing changes experienced by the breeding herd data.

The theory behind the univariate Box-Jenkins methodology is outlined in chapter two for both non-seasonal and seasonal time series. As mentioned above, the methodology described in chapter two is applied to some of the livepig categories recorded by the farm censuses in chapter three, the models being built using the Time Series Package, TSP, available on the mainframe computer here at Nottingham. Using the available package greatly eases the process of model building and forecasting with the Box-Jenkins models although the package is rather inflexible in that it can only deal with monthly, quarterly or annual data and cannot accept the use of intercept dummies with might be useful for dealing with outliers.

The biological models are discussed in chapter four, introduced by considering a steady state equilibrium framework to discuss the theoretic basis around which to build the biological model. The estimated models themselves are proportional models estimated using Ordinary Least Squares, OLS, and Non-Linear Least Squares, LSQ, packages available on TSP. Because the models were built in the context of forecasting, non-biological phenomena such as autocorrelation in residuals are modelled using a Beech-Mackinnon maximum likelihood technique in the case of linear regressions, also available on TSP, and adapting the LSQ models as appropriate when autocorrelation is present in the LSQ estimated models. In order to forecast using the recursive biological model, micro computer software was developed specifically for this purpose. The econometric model, which is discussed in chapter five, is introduced and estimated using similar methodologies to those used in the biological modelling procedure: in addition, a logistic model for a limited dependent variable is also considered. Software, is again developed in order to forecast using the chosen econometric model.

In chapter six univariate Box-Jenkins models are developed for the two monthly slaughter categories and monthly price and profit time series, again using TSP for modelling and forecasting. The theory and practice of Bivariate Box-Jenkins models is the subject of chapter seven, the bivariate models being identified and estimated using a non-linear least squares program developed by an ex-member of the department, a bivariate option not being available on TSP. For forecasting purposes I again developed software for a micro computer. A summary of the models built for each of the principle variables is given in figure 1.5 below.

Figure 1.5
The Models Built for the Principle Variables

Methodology	Trimestic/Quarterly Breeding herd	Monthly Culling	Monthly Slaughter
Univariate	chapter 3	chapter 6	chapter 6
Bivariate		chapter 7	chapter 7
Biological	chapter 4	chapter 4	chapter 4
Econometric	chapter 5		

In the penultimate chapter of the thesis the results of forecasting the trimestic breeding herd, the monthly culling and the monthly fat pig slaughter series using the models developed for each methodology considered, are presented and compared in terms of their ability to forecast the correct level of the variable in question. Because the direction of forecast may be as important to a forecaster as the ability of a model to forecast the correct level, a basic directional analysis is also considered. The short and longer term forecasting performances are analysed by forecasting one-step and two years ahead, and as an intermediate step one year ahead forecasts are also analysed, the latter being the length of time the EEC requires the MLC to forecast the pig sector.

Because of the constraints imposed by the data available, the out-of sample period used for the forecasting analysis is confined to two years, that is, 1986 and 1987. This turns out to be a significant restriction for the trimestic breeding herd analysis, especially as one of the six observations in the out-of-sample period is felt by myself and the MLC[10] to be somewhat suspect in terms of reliability. This particular problem is dealt with at the end of chapter three and in the forecasting analysis chapter. The problem of the lack of good quality out-of-sample data meant that one of my original intentions to combine the various forecasting models to produce one set of forecasts for each of the three key variables studied was not felt to be feasible, but, the methodology behind the idea is discussed in the closing chapter along with other suggestions for future work.

Considering that the models are built from a supply side point of view, forecasting for a period ahead greater than two years was thought to be unwise given the shortness of the pigs life cycle. Forecasts for a longer period would undoubtedly be of benefit to the industry given the cost of investment in pig production equipment, however, forecasting in the longer term would require demand side models to enter into the discussion. Such demand side modelling

10. This information is the result of direct consultation with the MLC.

could become very complex, requiring consideration of factors which influence both the price of pigmeat and the price of feed. The feed costs of pig meat production are largely dependent on the price of cereals and would, therefore, require UK production and trade in grain, to be modelled, incorporating the effects of policy at EEC and probably global level. Modelling the price of pigmeat could also involve a complex system of equations, although the subject has been addressed quite succinctly by Daniels and Savin (1977) in an MLC publication produced after a symposium on meat demand and price forecasting. Modelling demand for pig meat would require the consideration of consumer preferences with respect to different pigmeat and other meat products and, amongst other things, macro variables such as the size of population and disposable income. Were such demand side models to be built, the supply side forecasting system built in this thesis could be closed off and long term forecasts produced with increased confidence in the system.

CHAPTER TWO

BOX-JENKINS UNIVARIATE METHODOLOGY: THE THEORY

2.1 Introduction

The main objective of this thesis is to build forecasting models of the UK pig breeding herd and certain slaughter categories. The aim of this chapter is to give a brief outline of univariate time series model building methodology as proposed by Box and Jenkins (1976) which is a particularly useful approach for the building of short term forecasting models[1]. The first section of the chapter looks at the theory behind non-seasonal time series methodology and is followed by an account of each of the three stages of identification, estimation and diagnostic checking, which are now accepted as the basic steps in the building of univariate time series models. The final section is concerned with the way in which the model building process is adapted to cope with seasonal influences in time series data.

2.2 Time Series Methodology

Time series methodology approaches the subject of model building from an empirical standpoint in that the time series data themselves determine the identification of the appropriate model. Box-Jenkins time series methodology makes use of the fact that all stationary time series data can be represented by a member of the set of general stochastic processes known as Autoregressive Moving-Average, (ARMA) models. The time series variable x_t is said to exhibit weak stationarity if the following conditions hold.

$$
\begin{aligned}
E[x_t] &= \mu && \text{all } t, \\
E[x_t, x_{t-\tau}] &= \sigma^2 < \infty && \text{for } \tau = 0, \\
&= \gamma_\tau && \text{otherwise.}
\end{aligned}
\qquad (2.2.1)^2
$$

where

t = time

$E[x_t]$ = expected value of x_t

τ = magnitude of the lag.

γ_τ = the autocovariance between x_t and $x_{t-\tau}$.

1. For a thorough generalisation of Box-Jenkins methodology see *Granger and Newbold* (1977), chapter one.

2. For a more detailed discussion of weak and strong stationarity see *Nelson* (1973), section 2.1.

The result that all stationary univariate time series can be generated from ARMA models derives from the work of Wold (1954) who proved that any univariate time series could be represented as a realisation of the sum of a self-deterministic component and a moving-average process, possibly of an infinite order. Thus, letting x_t represent the deviation from the mean, x_t can be written as;

$$x_t = \varepsilon_t + \theta_1\varepsilon_{t-1} + \theta_2\,\varepsilon_{t-2}\ldots\ldots\ldots\ldots \tag{2.2.2}$$

where ε_t is a zero-mean white-noise variable such that:-

$$\begin{aligned} E[\varepsilon_t] &= \mu_\varepsilon = 0, \\ E[\varepsilon_t,\varepsilon_{t-\tau}] &= \sigma_\varepsilon^2 \text{ for } \tau = 0, \\ &= 0 \quad \text{otherwise.} \end{aligned} \tag{2.2.3}$$

By making use of the backshift operator, $B^\tau x_t = x_{t-\tau}$, equation (2.2.2) can be rewritten as,

$$x_t = (1 + \theta_1\,B^1 + \theta_2\,B^2 + \ldots\ldots\ldots\ldots\ldots\ldots)\,\varepsilon_t. \tag{2.2.4}$$

The problem with representing all time series in this manner is that many series require the estimation of a large number of parameters in order to adequately describe the data-generating process behind the series. Box and Jenkins suggested that this problem could be overcome by approximating the polynomial in ε_t described in equation (2.2.4.) by the ratio of two lower order and finite polynomials, that is,

$$x_t = \frac{(1 + \theta_1\,B^1 + \theta_2\,B^2 + \ldots\ldots\ldots\ldots\ldots\ldots + \theta_q B^q)}{(1 - \phi_1\,B^1 - \phi_2\,B^2 - \ldots\ldots\ldots\ldots\ldots\ldots\ldots - \phi_p\,B^p)}\,\varepsilon_t \tag{2.2.5}$$

where the numerator is the moving-average component and the denominator is the autoregressive component. By expressing the polynomials in the numerator and denominator as $\theta(B)$ and $\phi(B)$ respectively, equation (2.2.5) can be rewritten as,

$$x_t = \frac{\theta(B)}{\phi(B)}\,\varepsilon_t \tag{2.2.6}$$

Rearranging equation (2.2.6) results in the following expression,

$$\phi(B)\, x_t = \theta(B)\, \varepsilon_t \tag{2.2.7}$$

where $\phi(B)$ is said to be of order p and $\theta(B)$ order q. Thus, Box and Jenkins were able to show that all stationary univariate time series can be approximated by the general ARMA(p,q) model (2.2.7), which accounts for all ARMA processes including the extreme case where p and q both equal zero, in which case equation (2.2.7) reduces to a white noise process,

$$x_t = \varepsilon_t \tag{2.2.8}$$

Other special cases of the general ARMA model arise when one of p or q equals zero, in which case the resulting process is known as a pure moving-average model, {MA(q)} or a pure autoregressive model {AR(p)} respectively.

The polynomial $\phi(B)$ constitutes a p^{th} order difference equation in B. If this is to describe a stationary autoregressive series then it can be shown that the p roots of $\phi(B)$ must all lie outside the unit circle.[3] Given the nature of ε_t, $\theta(B)\varepsilon_t$ will always be stationary provided that q is finite and, therefore, the stationarity of the ARMA model depends solely on the autoregressive component. It is sometimes necessary or desirable to write the ARMA model in a pure AR form and in order that the resulting process be stationary, it can be shown that it is necessary for the q roots of $\theta(B)$ to lie outside the unit circle.[3] This is known as the 'invertibility condition'. Invertibility is also of fundamental importance in respect of the facts that it enables uniqueness of representation for the autocorrelation function of $\theta(B)\varepsilon_t$, and that non-invertible MA processes give rise to inefficient forecasts.[4]

3. See *Granger and Newbold* (1977) section 1.6, p.24

4. Consider an MA(1) model $x_t = \varepsilon_t + \theta_1\varepsilon_{t-1}$. Successive substitutions for lagged values of ε_t gives $x_t = -\Sigma - \theta^j x_{t-j} + \varepsilon_t$. If x_t is not to depend on some value of x_{t-j} in the infinite past then θ must take a value less than one in absolute value. i.e. for the AR representation of x_t to be stationary $|\theta|<1$ must hold so that the root of $(1-\theta B)$ is outside the unit circle. The autocorrelation function is given by,

$$\rho_0 = 1$$
$$\rho_1 = \theta/(1+\theta^2)$$
$$\rho_\tau = 0 \text{ for } \tau > 1$$

It is possible to show that the same autocorrelation function would also be given for the MA (1) model with an absolute parameter value of $(1/\theta)$. Given that $|\theta|<1$, the

Each invertible ARMA process has a distinct pattern of autocorrelations and partial autocorrelations which describe the correlations between values of x_t at various lags in time. The autocorrelation at lag τ, (ρ_τ), can be defined by,

$$\rho_\tau = \frac{\gamma_\tau}{\gamma_0} \qquad (2.2.9)$$

where $\gamma_\tau = E[x_t, x_{t-\tau}]$, and where $\gamma_0 = E[x_t^2]$, which measures the variance of x_t over all values of t. By definition ρ_τ, which measures both the direct and the indirect relationship between x_t and $x_{t-\tau}$ for all values of τ, always takes a value $\leq |1|$. For an MA(q) process it is possible to show that the theoretical set of autocovariances will take values as given by the autocovariance function of equation (2.2.10)

$$\begin{aligned}
\gamma_0 &= (1 + \theta_1^2 + \theta_2^2 + \dots\dots\dots\dots + \theta_q^2)\,\sigma_\varepsilon^2 \\
\gamma_\tau &= (\theta_\tau + \theta_1\theta_{\tau+1} + \dots\dots\dots\dots\dots\dots + \theta_{q-\tau}\,\theta_q)\,\sigma_\varepsilon^2 \quad ; \ \tau = 1 \text{ to } q \\
\gamma_\tau &= 0 \quad ; \ \tau > q
\end{aligned} \qquad (2.2.10)$$

The theoretical autocovariance function for an AR(p) process can be represented by the p^{th} order difference equation as given by:

$$\gamma_\tau = \phi_1\gamma_{\tau-1} + \dots\dots\dots + \phi_p\,\gamma_{\tau-p} \quad \text{for } \tau = 1,2\,\dots\dots \qquad (2.2.11)$$

and hence, the autocorrelation function is given by,

$$\rho_\tau = \phi_1\rho_{\tau-1} + \dots\dots\dots + \phi_p\,\rho_{\tau-p} \quad \text{for } \tau = 1,2,\dots\dots \qquad (2.2.12)$$

The exact pattern taken by the autocorrelations will depend upon the roots and order of the polynomial $\phi(B)$ and also on the magnitude of the parameters (ϕ). For example, if p equals one, the autocorrelations decline in a geometric fashion; smoothly if ϕ lies between zero and plus one, and in an oscillatory manner if ϕ takes a value between negative one and zero. The autocorrelation patterns produced by a mixed ARMA model are much more complex and require a considerable amount of

---

inverse of this is bound to be greater than unity thereby resulting in a non-stationary model. By confining attention to the invertible case the problem of identifying a stationary model is overcome.

identification experience in order to make a successful identification of the model. In theory, the ARMA (p,q) process results in autocorrelations which are dependent on both AR and MA parameters for the first q autocorrelations after which they follow an AR(p) process with a starting value ρ_q rather than ρ_o.

The partial autocorrelation of order k, denoted by a_{kk}, measures only the direct relationship between x_t and x_{t-k} for all values of k. The partial autocorrelations can be obtained by solving the set of k simultaneous linear equations, known as the Yule-Walker equations, which are similar to the autocorrelation function of (2.2.12) expressed as,

$$\rho_\tau = \sum_{j=1}^{k} a_{kj}\, \rho_{\tau-j} \tag{2.2.13}$$

where $a_{kj} = \phi_j$ for $j = 1,2,\ldots\ldots\ldots p$

and $a_{kj} = 0$ for $j > p$

The key point to note here is that the partial autocorrelations from an AR(p) process of any order greater than that of p will be equal to a value of zero and hence the partial autocorrelation will be a crucial tool in the identification of autoregressive processes. Because any invertible MA process can be transformed into a stationary AR process of infinite order, it is also possible to show that the partial autocorrelation function for an MA process will decline as the sum of a set of geometric decay functions similar to the autocorrelation pattern of an AR process as discussed earlier in the section. The partial autocorrelations for a mixed ARMA model will eventually tail-off as well due to the fact that any ARMA process can be transformed into an autoregressive process of infinite order.

2.3 Identification

Identification is the first of the three stages in the Box-Jenkins model building process, the aim of which is to suggest one or more potential time series models to explain the movements in the time series data. Experience appears to suggest that, for non-seasonal time series, at least fifty to sixty data values are required for Box-Jenkins analysis, because of the large number of degrees of freedom that are used up in estimating the autocorrelations of higher order. The first task, having obtained the data, is to generate the auto and partial autocorrelations for the raw data series and for an appropriate number of differences of the data. The rationale for differencing the data is that Box-Jenkins methodology is dependent on the series to be modelled being stationary. Raw economic time series data are rarely stationary in that they are likely to contain, for example, long and short term time trends. Box- and Jenkins recommend that such non-stationary series be made stationary by taking an appropriate number of differences of the raw data, as explained in the next section. The maximum likelihood estimator of ρ_τ, for all values of τ is given by:

$$r_\tau = \frac{\sum_{t=\tau+1}^{n}(x_t - E[x_t])(x_{t-\tau} - E[x_t])}{\sum_{t=1}^{n}(x_t - E[x_t])^2} \qquad (2.3.1)$$

Theoretically, the partial autocorrelations can be estimated from the Yule-Walker equations by substituting the estimated autocorrelation (r_τ) for the theoretical one (ρ_τ) and then solving the set of simultaneous equations in a_{kj}, that is;

$$r_\tau = \sum_{j=1}^{k} \hat{a}_{kj} r_{\tau-1} \qquad \text{for } \tau = 1 \text{ to } k \qquad (2.3.2)$$

where $\hat{a}_{kk}$ is the estimated value of a_{kk}. In practice however, it is much more convenient to use an algorithm similar to that developed by Durbin (1960), which uses ordinary least squares to estimate the k^{th} order partial autocorrelation for an appropriate size of k.[5]

Having obtained these statistics, the analyst needs to ensure that the series he is trying to model is stationary so that the theory in the previous section can be applied. An indication that a series is non-stationary is given by observing that the sample autocorrelations do not die away at higher lags. This is true even if the first few autocorrelations are not large themselves.[6] If this is not the case, so that the ρ_τ's do not die away for a considerable number of lags, Box and Jenkins suggest that differences of the raw data be taken until an indication that stationarity has been achieved is observed in the autocorrelations of the differenced series. If differencing is required then the resulting model is known as an Integrated Autoregressive Moving-Average (ARIMA{p,d,q}) model, where d represents the number of times the raw data have to be differenced in order to achieve stationarity. Equation (2.3.3) represents the general ARIMA model using the backshift operator.

---

5. Durbin,J. (1960), "The Fitting of Time Series", *Rev. Inst. Int. Stats*. 28, pp 233-244.

6. See *Granger and Newbold* (1977) section 3.2, pp. 74.f

$$\phi(B)\,(1-B)^d\,x_t = \theta(B)\,\varepsilon_t \qquad (2.3.3)$$

Economic time series are rarely stationary in their original form, although the need to take anything more than first differences in order to obtain stationarity is unlikely. Having suggested this differencing method for the treatment of non-stationarity, Box and Jenkins warn against the possibility of inducing autocorrelation by over-differencing the data. This problem can be illustrated by the simple case of first differencing the white-noise process given in equation (2.2.8) which results in an MA(1) process with x_t having been first differenced. The latter is commonly referred to as an IMA(1,1) model and is represented by:-

$$(1-B)\,x_t = (1-B)\,\varepsilon_t \qquad (2.3.4)$$

This transformation illustrates the fact that two unnecessary components have been induced into the process. If it is the variance of the time series which is the cause of non-stationarity, then a log transformation of the raw data often ensures that the non-stationary element is removed.

Having ensured the stationarity of the series, the patterns in the auto and partial autocorrelations should be examined in order to make an initial identification of the generating process. Because of the influences of disturbances such as data measurement errors, the sample statistics will not be identical to the theoretical values that the ARIMA generating process would imply. In an MA(q) process for example, the sample autocorrelations higher than order q will be small rather than being equal to a value of zero. In order to distinguish whether or not the autocorrelations are large or small, probability theory is employed. The theory implies that an autocorrelation can be said to be significantly large if its value is greater than two standard deviations away from zero. A statistic frequently used to estimate the standard deviation of autocorrelations in time series analysis is one devised by Quenouille, who showed that one standard deviation could be approximated by the reciprocal of the square root of the sample size, that is, $1/\sqrt{n}$.[7] Hence, any auto or partial autocorrelation which is larger than $2/\sqrt{n}$ is said to be significantly different from zero and can be regarded as large. Conversely, a value below this is said to be small, although it should be remembered that this 'rule of thumb' definition is not infallible, and need not be

---

7. Quenouille,M.H. (1949), "Approximate Tests of Correlation in Time Series." *J. Roy. Stat. Soc.* B11. pp 68-84.

interpreted too strictly.

From the theory of the previous section, the key to the identification of an ARIMA processes lies with the patterns present in the sample auto and partial autocorrelations of the stationary series. If a set of data is generated by a pure MA(q) process then one would expect the autocorrelations of the stationary series to lie above two standard errors up to and including a lag of q, at which there should be a cut-off point when the remaining sample autocorrelations will be small. Beyond q, the sample partial autocorrelations should decay as k increases. The key to the identification of an AR(p) process on the other hand, lies in the pattern of the partial autocorrelations, due to the fact that the sample autocorrelations will show a pattern which represents the sum of a geometric decay curve, from which it is virtually impossible to identify p. The sample partial autocorrelations should remain high in value for the first p lags after which there is a cut-off point, so that the higher order partials are below the $2/\sqrt{n}$ mark. The identification of a mixed process however, is considerably less clear cut. The sample partial autocorrelations should tail-off as the value of k increases, whereas the sample autocorrelations should take large values, with an irregular pattern, up to and including lag q, after which they will begin to tail-off as they reflect the autocorrelations of an AR(p). One clue therefore, to the identification of a mixed process as opposed to a pure MA or AR process, is that both of the sample statistics should eventually tail-off, rather than having an abrupt cut-off point.

Having gone through the identification process it is often the case that more than one model appears to be possible. If this is the case, then all the possible models can be taken on to the estimation stage of the model building process.[8]

2.4 Estimation

Having made an initial identification of one or two potential ARIMA models, the objective of the next stage of the Box-Jenkins model building process is the estimation of the parameters of the suggested models. If a model is a pure AR(p) process then all p parameters can be estimated using Ordinary Least Squares (O.L.S) regression techniques. However, as soon as an identified model has an MA term included in it OLS is rendered inadequate and Non-Linear Least Squares (N.L.L.S.) has to be employed instead. Taking an ARMA(1,1) as an example, the process can be written as:

8. Three examples of model identifications using Box-Jenkins methodology can be found in appendix 2.

$$\varepsilon_t = x_t - a_1 x_{t-1} - b_1 \varepsilon_{t-1}. \qquad (2.4.1)$$

where a_1 and b_1 are the sample estimates of ϕ_1 and θ_1 respectively.
The aim of N.L.L.S. is to minimise the sum of the squares of the residuals, $\Sigma\varepsilon_t^2$, which can be estimated by Σe_t^2 , the sum of squares of observed residuals. Having assumed starting values for a_1 and b_1, and by assuming that e_1 takes a value equal to its expected value of zero and inputting this to equation (2.4.1), estimates of lagged values of e_t can be derived. Squaring and summing all values of e_t results in the desired statistic which then needs to be minimised. The minimisation process is done by a search procedure over a range of values of a_1 and b_1 until no significant reduction in Σe_t^2 occurs.[9] Because it is possible that the residual sum of squares contour surface will have multiple minima points, it may be important that good starting values for a_1 and b_1 are given to the computer so that the search will begin in an appropriate part of the contour surface. Starting values for the parameter estimates can be derived from the sample auto and partial autocorrelations. For example, in an MA(1) it can be shown that

$$\rho_1 = \frac{\theta_1}{1 + \theta_1^2} \qquad (2.4.2)$$

By substituting r_1 for ρ_1 and rearranging equation (2.4.2), a quadratic in θ_1 is obtained which can then be solved giving two solutions for θ_1. The starting values of the parameter estimates are obtained by selecting the solutions that will make the resulting model invertible.

2.5 Diagnostic checking

Having identified the model and estimated its parameters, the model should now be checked to see if it is an adequate representation of the data to which it was fitted. If all is well then the model can be used for whatever purpose it was built, otherwise any inadequacies discovered at the checking stage will hopefully indicate the changes which need to be made in order to rectify any model misspecification.
A possible first check is to carry out a t-test on each of the included parameters in order to check their significance. A t-statistic that is greater than or equal to an absolute value of two suggests that the parameter concerned is significantly different from zero, and should therefore be included. If the absolute value of t is

9. See, for example, *Granger and Newbold* (1977) section 3.5,pp 87-89.

less than one then there is a positive reason for its exclusion from the model; a change which should cause a reduction in the value of the estimated residual variance, indicating that the revised model provides a better overall fit to the data. If the absolute value of t lies somewhere between one and two then this suggests that the parameter should perhaps be dropped, although its effect on the residual variance is likely to be adverse when the t-statistic is close to a value of two. If any t-statistics indicate that a model should be respecified, then the estimation and the diagnostic checking procedures must be repeated for the new specification.

A second check which can be made is to test the significance of the model as a whole, by calculating the value of the Mean Square Error (M.S.E.). Because there is no standard of measure to compare the calculated value of the M.S.E., nothing much is done even if it appears to be relatively high. The M.S.E value is much more important when testing whether or not a change in its value, brought about by a change in model specification, is significantly different from what it was under the original specification. The test used to do this is the F-test. R^2 is not usually used to check the significance of time series models due to the fact that any model which picks up a trend in the data will produce a high R^2 value. However, Harvey has suggested that a statistic which he calls R_D^2 , given by;

$$R_d^2 = 1 - \frac{RSS}{\sum_{t=2}^{n} (\Delta X_t - \bar{\Delta X}_t)^2} \qquad (2.5.1)$$

provides a statistic which indicates the relative size of the residual sum of squares, (RSS), of the estimated model.[10] The yardstick used is the residual sum of squares (RSS_0) from having fitted the pure random walk model $x_t - x_{t-1} = \varepsilon_t$. A negative R_D^2 value indicates that the random walk model provides a better fit than the ARIMA model, whereas a small positive value suggests that the improvement gained by fitting the more complex ARIMA model is marginal. A corresponding statistic for seasonal models is discussed in section 2.6.

Having checked to see that the model is not over-identified, by making use of the t-test, it should now be checked to test whether or not there are any parameters missing from the initial identification. This can be done by observing the individual autocorrelations of the estimated residuals, $\rho_\tau(e)$. As an example, consider the case where $|\rho_2(e)|$ is significantly different from zero, (that is, greater than $2/\sqrt{n}$). This would suggest that the error term has the following process;

10. See Harvey, A.C., (1983), "A Unified View of Statistical forecasting Procedures", appendix 2, *L.S.E. Econometrics Programme discussion Paper No. A.40*., L.S.E.

$$e_t = (1 + \beta B^2)\, a_t \qquad (2.5.2)$$

where a_t is white-noise. If these residuals had been the product of an ARMA(1,1) identification, then it is possible to show that the corrected model is given by;

$$(1\text{-}aB)x_t = (1 + bB + \beta B^2 + b\beta B^3)\, a_t. \qquad (2.5.3)$$

Equation (2.5.3) represents an ARMA (1,3) process, although the parameter of B^3 will be small because both b and β will take a value less than unity in order that the process be invertible. If it is believed, therefore, that the true model should contain a given parameter - in this case an MA parameter at lag 3 - whose t-statistic is not quite significant at the 5% level, the analyst may well be justified in including the relevant parameter. This method of correction, although useful sometimes, is often less simple than may appear from the example given, due to the fact that the statistics used are only estimates from a data set which itself will contain measurement errors. A consequence of this is that any cancellation of factors may be masked, so that the augmented model will contain redundant parameters on both sides of the equality. One problem of testing the significance of the residual autocorrelations is that they are small anyway because it is the job of the estimation stage to produce residuals which are as small as possible. Indeed Durbin (1970), has shown that the standard deviation of the residual autocorrelations can be considerably less than $1/\sqrt{n}$.[11] This is particularly true of the autocorrelations at lower orders, that is, where $\tau \leq 6$, otherwise $1/\sqrt{n}$ is still a good approximation to the standard deviation of the residual autocorrelation, as long as it is remembered that $2/\sqrt{n}$ will under-estimate the significance of any deviations from zero.

If a model has been correctly identified then the residuals, as a whole, should exhibit white-noise properties. Attempts have been made to devise a statistic that indicates whether or not the autocorrelations of the estimated residuals deviate from white-noise. Most of the statistics which have been devised are variants of the Box-Pierce statistic, Q, defined as;

$$Q = n \sum \rho_\tau^2(e) \qquad \text{for } \tau = 1 \text{ to } m \qquad (2.5.4)$$

where m represents the m lowest order residual autocorrelations considered for the

11. Durbin, J. (1970), "Testing For Serial Correlation in Least Squares Regression When Some of The Regressors are Lagged Dependent Variables", *Econometrica* 38, pp 410-421.

test.[12] Asymptotically, Q can be shown to have a Chi-Square distribution with m-p-q degrees of freedom. If the value of Q is greater than the tabulated Chi-Square value, then the null hypothesis, that the residuals are white-noise, is rejected and the model should be respecified. Some computer packages, along with the Q-Statistic, print the probability value, (P-value), associated with each of the Q-statistics. Hence, for example, a Q-statistic with a P-value of less than 0.05 is significantly large at the 5% level. Because the Box-Pierce statistic has a distribution that is only asymptotically Chi-Square, there is the problem that the test is not very powerful, and therefore, it is quite likely that a false null hypothesis will be accepted, although Q can still indicate whether or not the residual autocorrelations are, on the whole, too high.[13] For reasons mentioned earlier, the test requires that m is larger than six, and it is preferable that it be greater than or equal to twenty, if this is possible considering the length of the time series. In the event of a rejection of the null hypothesis, the model should be corrected in a manner suggested by the pattern of the residual autocorrelations, as discussed earlier.

A final check to make use of the residuals is an observation of the plots of the estimated residuals to check for homogeneity. Heteroskedasticity, for example, would be indicated by observing that the spread of the residuals' scatter changes over time.

So far, the diagnostic checks which have been employed to test for under-estimation of the model, have made use of the autocorrelations of the estimated residuals; however, because of some of the problems and inadequacies of these methods, it is often much simpler to overfit the initial identification. Although it is important to overfit both sides of the equation, it is even more important that the overfitting be done individually and unidirectionally so that the problem of parameter redundancy is avoided.[14] For each overfitting of the model, t-tests on each of the parameters and F-tests on the adjusted M.S.E. statistics should be made in order to determine whether or not any of the augmented models perform better than the original

12. Box-Pierce (1970), "Distribution of Autocorrelations in ARIMA Time Series Models", *J. Am. Stat. Assoc.* 65, pp1509-1526.

13. See Davies, Triggs and Newbold, (1977), "Significance Levels of the Box-Pierce Portmanteau Statistic in Finite Samples." *Biometrika.* 64, pp 517-522.
The more highly powered Box-Ljung statistic can be used in preference to the Box-Pierce statistic.
i.e. $Q = n(n+2)\sum(n-\tau)^{-1} r_\tau^2(e)$
Ljung, G.M. and Box, G.E.P., (1978), "On a Measure of Lack of Fit in Time Series Models." *Biometrica.* Vol. 65 No. 2 pp297-303.

14. For an example of the problem of overfitting both sides of the equation see example 2 of Appendix 2.

identification. As illustrated in example two of Appendix 2, it is very important to look for the cancellation of factors so that the final model is parsimonious.

All of the tests used so far are in-sample tests, in that they are performed on statistics generated from the fitting of the model to the time series data from which the model itself was identified and estimated. Perhaps a more powerful form of diagnostic checking is out-of-sample testing, which involves testing the forecasting performance of the estimated model. The time series is divided into two, so that the first series is long enough for the Box-Jenkins model building process to be performed. The identified model is then used to make forecasts of the remaining 'out-of-sample' data. The forecasts and the actual out-of-sample data are then compared, and the mean square forecasting errors calculated. These forecasting error statistics are then compared for alternative model identifications, the lowest value indicating the best forecasting model. Obviously, it is imperative that the forecasting method employed is appropriate, considering the type of forecasts for which the model has been developed. For example, if the model is to be used to make one-step-ahead forecasts, then the criteria for choosing the best model must be the lowest mean square forecasting error, resulting from one-step-ahead forecasts of the out-of-sample series. Having obtained a satisfactory forecasting model, the analyst may choose to re-examine the chosen model by re-estimating it from the whole data set, and performing the in-sample diagnostic checks. If none of the models appears to forecast well, then it is important to check, and to allow for events such as structural changes which may have occurred during the period covered by the time series.

Whatever method of diagnostic checking is chosen, it is imperative to check that the final choice of model has parameters which render the process stationary and invertible. This is done by solving the difference equations of the model and checking that the roots lie outside the unit circle.

Having gone through the process of model building for non-seasonal time series, a satisfactory model should have been derived. It is quite possible that two models are almost inseparable at the diagnostic checking stage, in which case the final choice may be made solely on the grounds of parsimony. This idea of parsimonious parameterisation can be very important in shorter time series for releasing degrees of freedom and reducing the chances of multicollinearity problems. Nonetheless, it is possible that two ARIMA models which might appear to be quite different are, in fact very similar when rewritten in a different form, and so the final choice of model may not be too important, especially for the purposes of short term forecasting.

2.6 Seasonal Time Series Models

So far, this chapter has dealt with the building of non-seasonal time series models. The methodology which was employed in this model building process can be adapted, without too much difficulty, to cope with data which contain seasonal components. The seasonal model building process follows the same three stage iterative cycle of identification, estimation and diagnostic checking and, although the identification stage is a little more complex, no new concepts are required.

The most basic seasonal model is the pure seasonal ARMA model, an example of which is the quarterly ARMA(1,1) containing 1 seasonal AR parameter and 1 seasonal MA parameter. The said example can be written as,

$$(X_t - \phi_4 X_{t-4}) = (\varepsilon_t + \theta_4 \varepsilon_{t-4}) \tag{2.6.1}$$

or by making use of the seasonal backshift operator it can be re-written as 2.6.2.

$$(1 - \phi_4 B^4)X_t = (1 + \theta_4 B^4)\varepsilon_t \tag{2.6.2}$$

The latter model is directly comparable with the non-seasonal ARMA (1,1), except that the 1 period lag structure is now replaced with a quarterly lag structure. Were X_t to be non-stationary in levels, one way in which the series can be made stationary is to seasonally difference the raw data series. Assuming the model represented by 2.6.2. needed to be differenced once to meet the stationarity requirements, the resultant model can be represented by equation 2.6.3.

$$(1 - \phi_4 B^4)(1 - B^4)X_t = (1 + \theta_4 B^4)\varepsilon_t \tag{2.6.3}$$

The generalised form of the pure seasonal ARIMA(P,D,Q) is given by

$$\phi_S(B^S)(1-B^S)^D X_t = \theta_S(B^S)\varepsilon_t, \tag{2.6.4}$$

where P is the order of the seasonal autoregressive polynomial, Q is the order of the seasonal moving-average polynomial and D represents the number of times the process has to be seasonally differenced in order to obtain stationarity. B^S is the seasonal backshift operator where s represents the type of seasonal data, so that s = 4 for quarterly data and s = 12 for monthly data.

Although this model caters for the purely seasonal time series, a more general model is required which is able to model seasonal series which also contain non-seasonal components. The simplest and least restrictive of this more general class of model is the model in which the gaps in the seasonal process are filled. Equation (2.6.5) provides an example of a model in which the non-seasonal IMA(1,1) has a quarterly lag added to it.

$$(1-B)\ x_t = (1 + \theta_1 B^1 + \theta_4 B^4)\ \varepsilon_t. \qquad (2.6.5)$$

Again, this model is similar to a non-seasonal ARIMA model, {in this case the MA(4) } except for the fact that there is the seasonal differencing component and there are 'holes' in the lag structure at lags two and three.

The most common alternative type of seasonal model, and the one which is preferred by Box and Jenkins, is the multiplicative ARIMA model given by;

$$\phi(B)\ \phi_s(B^s)\ (1-B)^d\ (1-B^s)^D\ X_t = \theta(B)\ \theta_s(B^s)\ \varepsilon_t, \qquad (2.6.6)$$

where $\phi_s(B^s) = (1 - \phi_{1,s}\ B^s - \phi_{2,s}\ B^{2s} - \ldots\ldots\ldots - \phi_P B^{Ps})$,

and $\theta_s(B^s) = (1 + \theta_{1,s}\ B^s + \theta_{2,s}\ B^{2s} + \ldots\ldots\ldots + \theta_Q B^{Qs})$,

the order of which is (p,P,d,D,q,Q). The multiplicative seasonal model is obtained by replacing the seasonal white noise disturbance term x_t, in the pure seasonal ARIMA(P,D,Q) model,

$$\phi_s(B^s)\ (1-B^s)^D\ X_t = \theta_s(B^s)\ \xi_t, \qquad (2.6.7)$$

with a non-seasonal ARIMA(p,d,q) process,

$$\phi(B)\ (1-B)^d\ u_t = \theta(B)\ \varepsilon_t \qquad (2.6.8)$$

so that the combination of (2.6.7) with (2.6.8) results in the multiplicative seasonal model given by (2.6.4). This type of model is not as general as the model presented in equation (2.6.3) in so much as some of the parameters will be restricted in the values that they can take. Considering the multiplicative MA(1,1),

$$x_t = (1+\theta_1 B)\ (1+\theta_{1,s} B^s)\ \varepsilon_t, \qquad (2.6.9)$$

it is possible to show that $\rho_{s-1} = \rho_{s+1}$. The consequence of this is that the parameter on $B^{s+1}\varepsilon_t$ is restricted and dependent upon θ_1 and $\theta_{1,s}$ as can been seen by expanding the right hand side of equation (2.6.9), as follows.

$$x_t = (1+\theta_1 B + \theta_{1,s} B^s + \theta_1\, \theta_{1,s} B^{s+1})\, \varepsilon_t. \qquad (2.6.10)$$

A third , but less common type of seasonal model is the additive model, whereby a non-seasonal ARMA(p,q) is added to a seasonal ARMA(P,Q) process. The two white noise processes which drive the two series are assumed to be independent of one another. As an example, consider the addition of an AR(1) to a seasonal AR(1) which results in the following:-

$$x_t = \frac{\xi_t}{1 - \theta_4 B^4} + \frac{\varepsilon_t}{1 - \theta_1 B^1} \qquad (2.6.11)$$

This can be rewritten as,

$$(1 - \phi_1 B)\,(1 - \phi_4 B^4)\, x_t = \xi_t + \varepsilon_t - \phi_1 \xi_{t-1} - \phi_4\, \varepsilon_{t-4} \qquad (2.6.12)$$

where the right hand side is an MA(4) with the parameters restricted by ϕ_1 and ϕ_4. Identification of the multiplicative model involves the choosing of values for d,D,p,P,q and Q, which is done by employing the same methodology as was used for non-seasonal models. The first step of the identification stage is to obtain a stationary series. This is achieved by taking first and seasonal differences of the raw data until the autocorrelations begin to die away quickly at higher lags.[15] Determining the number of differences which should be taken is more difficult for seasonal data, especially for quarterly data where the picture is much more cloudy. Granger and Newbold found that models that had been differenced, when differencing was in doubt, were much better forecasters than were the equivalent models which had been left non-differenced.[16] Having obtained a stationary series, the sample auto and partial autocorrelations can be examined, in an attempt to identify the order of the seasonal and non-seasonal polynomials. Again, this methodology follows on directly from that which was used to identify p and q in the non-seasonal models, although the presence of the seasonal component means that the patterns in the auto and partial autocorrelations are more complex and, therefore, more difficult to identify.

15. An alternative to seasonally differencing the data is to subtract seasonal means (i.e. using dummy variables) from the data.

16. See *Granger and Newbold* (1977) p.102.

For a pure multiplicative MA process of order (q,Q), the autocorrelations will obey,

$$\begin{aligned} \rho_\tau = 0 \text{ for } \quad & q < \tau < s-q \\ \text{and} \quad & s+q < \tau < 2s-q \\ & \vdots \\ \text{and} \quad & (Q-1)s+q < \tau < Qs-q \\ \text{and} \quad & Qs+q < \tau . \end{aligned}$$

Thus, for example, a quarterly multiplicative MA (1,1) will have non-zero autocorrelations only at lags 0,1,4 and 5. If the process is a pure multiplicative AR (p,P), the autocorrelations will die out according to the difference equation,

$$\theta(B)\,\theta_s(B^s)\,\rho_\tau = 0 \text{ for all } \tau > 0,$$

and the partial autocorrelations will obey,

$$a_{kk} = 0 \text{ for all } k > Ps+p.$$

If $(1-B)^d\,(1-B^s)^D\,X_t$ follows a multiplicative mixed ARMA process of order (p,q)(P,Q), then the autocorrelations will obey

$$\theta(B)\,\theta_s(B^s)\,\rho_\tau = 0 \quad \text{for all } \tau > q+Qs.$$ [17]

Estimation of the parameters is again a matter of employing non-linear algorithms which minimise the sum of squares of the residual term ε_t. When it comes to the diagnostic checking stage, there are a far greater number of possible models for seasonal time series and therefore, there are a far greater number of possible alternatives to check against. Again, the two main types of check include overfitting the model and inspection of the autocorrelations of the residuals. For the reasons given in the last section the overfitting of the model parameters must only be done on an individual and unidirectional basis, and the clues to how this should be pursued can be obtained from the identification stage. Testing the t-statistic of the additional parameters and the error variance of the augmented model should indicate whether of not the model should be respecified. Again, Harvey has devised a statistic, R_s^2, which indicates the relative size of the residual sum of squares from having fitted the seasonal ARIMA model[10]. The comparative yardstick in this case is the residual sum of squares (RSS_0) from having fitted a model to the first differences of the raw data which contains seasonal dummies. Allowing for degrees of freedom gives the following expression for R_s^2;

---

17. If the reader wishes to examine more generalised autocorrelation patterns belonging to some special cases of multiplicative models, they are refered to *Granger and Newbold* (1977) p. 96-98.

$$R_s^2 = 1 - \frac{\frac{RSS}{(n - sD - d - k)}}{\frac{RSS_o}{(n - 1 - s)}} \tag{2.6.13}$$

where k is the number of deterministic components in the ARIMA model.

Alternatively, the residual autocorrelations can be examined both individually, using the t-statistic, and as a whole, using the Box-Pierce or Box-Ljung statistics, in order to check whether or not they are white noise. If not then, in a similar fashion to the non-seasonal case, they will hopefully indicate ways in which the initial identification could be respecified, although the same reservations on this method hold as well.

A diagnostic check which is not necessary for non-seasonal models but does apply to the seasonal case is a check to see whether or not the multiplicative seasonal model adequately represents the time series being modelled. Thus, for example, the multiplicative quarterly MA given by,

$$(1\text{-}B)\,(1\text{-}B^4)\,x_t = (1+\theta_1 B)\,(1+\theta_{1,4}\,B^4)\,\varepsilon_t, \tag{2.6.14}$$

can be rewritten as,

$$(1\text{-}B)\,(1\text{-}B^4)\,x_t = (1+\theta_1 B^1+\theta_{1,4}\,B^4 + \theta_1\,\theta_{1,4}\,B^5)\,\varepsilon_t. \tag{2.6.15}$$

In order to check the assumption of multiplicity, the data which generated this model could be fitted against the non-multiplicative model given in 2.6.16 below.

$$(1\text{-}B)\,(1\text{-}B^4)\,x_t = (1+\theta_1 B^1+\theta_4\,B^4 + \theta_5 B^5)\,\varepsilon_t. \tag{2.6.16}$$

If the result of the estimation stage is to produce a parameter value for θ_5 which is similar to $\theta_1\theta_{1,4}$ then there is no reason to doubt that the series is multiplicative. If there is a significant difference then the less restrictive non-multiplicative model should be adopted.

2.7 Conclusion

This chapter then has outlined the theory of Box-Jenkins univariate time series methodology for both non-seasonal and seasonal data. The methodology facilitates the construction of models for a time series using only past values of itself and the observed error structure from having fitted the model. The obvious advantages of the methodology is that only one series is required to build forecasting models and arguably a more important advantage is that no prior knowledge of the series to be modelled, or of factors which affect it or are related to it are necessary to enable a workable model to be built. Having described the theoretical basis for the non-seasonal methodology the three stages of identification, estimation and diagnostic checking, suggested as the structured process for the construction of a model, were discussed. A brief outline as to how the theory is applied to seasonal time series was then presented.

As in all walks of life, the application of theory to a 'real world' situation is rarely as straightforward or problem free as the theory suggests. In the following chapter there is a discussion as to how the theory of this chapter has been applied to building quarterly models of the UK pig breeding herd and related census data; a comprehensive discussion of the problems encountered with the data and the model building process itself, and how these problems were resolved.

CHAPTER THREE

BOX-JENKINS UNIVARIATE MODELS FOR THE UK BREEDING HERD

3.1 Introduction

In this Chapter the Box-Jenkins model building methodology described in chapter two is employed to build multiplicative seasonal ARIMA models - hereafter referred to as SARIMA models - for certain key categories of live pigs. The analysis is conducted at the U.K. level of aggregation, so as to be directly comparable with forecasts of the industry produced by the Meat and Livestock Commission - M.L.C. The live pig categories can be broken down into the 'breeding sow herd', ('sows in pig', 'gilts in pig' and 'barren sows'), ' 50kg to 80kg gilts not yet in pig', (unserved gilts), and 'boars'. Census data for each of these categories is available on a quarterly basis. The aggregate of the pregnant sow herd and the pregnant gilt herd will be referred to as the pregnant pig herd. The prime purpose for modelling the chosen live pig categories is to provide forecasts of the U.K. breeding herd which are important for policy making in the sector. A priori, one would expect the univariate models to be particularly useful for the provision of short term forecasts, although their ability to forecasts the medium to long term will also be analysed.
To describe in detail each of the three stages of model building for all of the models produced would be time consuming and laborious, therefore, a detailed description of the methodology employed is given only for the total breeding sows herd model. This is done in order to give the reader some idea as to how the quarterly models were derived using the Box-Jenkins univariate methodology.
Much of the discussion in the chapter revolves around the way in which data problems affected the nature of the analysis and how the problems were resolved. The first such problem was the fact that a civil service strike in 1979 meant that no sample farm census was taken in the first quarter of 1979. To overcome this fundamental problem for time series methodology, initial forecasting models had to be built for each of the breeding herd component series using the data available up to and including the fourth quarter of 1978. A one step forecast could then be made in order to fill in the missing data point. A second major data problem is caused by the shift in the timings of the spring and autumn sample censuses from March and September to April and August respectively. This change, which followed the accession of the UK in to the EEC, had the obvious consequence that the data was

no longer quarterly in the strict sense of the word. For this reason and because of other influences post 1974, such as an Aujezky disease eradication campaign in 1983, and an apparent stabilising in the variability of the pig herd size, the comparison of the forecasting performance of the models built using different sample spaces was felt necessary.

The chapter is rounded off with a discussion of possible actions on the part of the forecaster when he suspects that the sample data for a particular point in time may be suspect in terms of reliability. The latter is included not purely for academic reasons but because it is an actual problem for a particular sample point in the data period set aside for the out-of-sample forecasting analysis. Appendix 3a lists the data used in the analysis presented in this chapter.

3.2 Modelling The UK Breeding Sow Herd

The initial analysis was concerned with modelling the three breeding sow herd categories, sows in pig, gilts in pig - referred to as pregnant sows and pregnant gilts respectively and denoted as 'PS' and 'PG' - and barren sows for breeding not in pig, 'BS'. The two aggregate series, pregnant pigs, 'PP', and total breeding sows, 'H', are also modelled separately. At the time of writing, census data were available from the first quarter of 1957 (1957:1) up to and including the fourth quarter of 1987 (1987:4). Appendix 3b presents a table which outlines the changes which have taken place in the methodology of census data collection for each of the U.K. agricultural ministries over the relevant time period. Up to and including 1973 the data are quarterly -March, June, September and December - and were collected by the Ministry of Agriculture, Fisheries and Foods - M.A.F.F., the Department of Agriculture and Fisheries for Scotland - D.A.F.S., and the Department of Agriculture for Northern Ireland - D.A.N.I. As a result of the U.K.'s accession to the EEC, the sample censuses for March and September from 1974 onwards were moved to April and August respectively. Consequently, the census data beyond this period can only be regarded as pseudo-quarterly and the implications of this will be discussed later.

A feature of all the pig census data is that no data are available for the first quarter of 1979 in which there was civil service strike action. In order that Box-Jenkins analysis could be performed on the complete data set, the first task was to generate data for the gap created by the strike. This was achieved by modelling each of the three component categories of the breeding sow herd on the sample period 1957:1 to 1978:4, and then forecasting the figure for 1979:1. The forecast figure for the two aggregate series were obtained by aggregating the relevant component forecasts.

The structure of the estimated models identified for the 1957-78 sample period, along with the resultant forecasts are given in table 3.1.

At the time the initial models were estimated, data for the breeding sow herd were available from 1957:1-1985:4. Consequently, the models were estimated on the sample up to and including 1981:4 leaving the remainder of the sample available for out-of-sample diagnostic checks. It is the building of the breeding sow herd models for this sample period which are reported in this section. Subsequent biological and econometric models were estimated on data up to and including 1985:4. Therefore, to ensure comparability for the three types of models, the SARIMA models were re-estimated on the sample period 1957:1-1985:4. Only the estimated equations themselves will be presented for the longer sample period time series models.

Table 3.1.

Breeding Sow Herd Forecasting Models and 1979:1 Forecasts.[1]

SERIES	NDIFF	NSDIFF	NAR	NSAR	NMA	NSMA	1979:1
	d	D	p	P	q	Q	FORECAST.*
PS	0	1	2	2	2	0	498
PG	0	1	2	0	1	1	111
BS	0	1	1	2	2	0	241
PP							609
H							850

*. All figures in thousands of pigs

Having identified and estimated Box-Jenkins models on the period 1957:1 to 1981:4, estimated equations were subjected to the usual diagnostic checks, including the calculation of the mean square forecast error, (MSFE), of residuals from having made both one-step ahead conditional forecasts and an unconditional 12-step ahead forecast of the out-of sample period 1983:1-85:4. This was done in order to analyse both the short term and medium-long forecasting abilities of the models. The multiplicative test was to be performed only if the resultant model contained a sufficiently small number of lags to justify such analysis. In the final analysis all but one of the identified univariate models for the breeding sow herd models contained so large a number of lags that any comparison with an unrestricted non-multiplicative model was unlikely to produce any useful and conclusive results. Seasonal dummy models on the first difference series of each of

1. NDIFF = No. of first differences taken. NSDIFF = No. of seasonal differences taken.
NAR = No. of AR parameters. NSAR = No. of seasonal AR, (SAR), parameters.
NMA = No. of MA parameters. NSMA = No. of seasonal MA, (SMA), parameters.

the included categories were also estimated in order that Harvey's R_s^2 value could be calculated.

The plots of the breeding sow herd series' - presented throughout this section of the chapter - appear to suggest that each of the three component series underwent some sort of structural change around the time of the U.K.'s accession to the E.E.C. in 1973. To test this hypothesis, Chow tests were performed on each of the breeding sow herd SARIMA models estimated on the period 1957:1-1985:4.[2] Because of the change in the timings of the Spring and Autumn sample censuses, the start of 1974 was chosen as the dividing point in the sample. The details of the tests and the results are presented in Appendix 3c. Each of the tests proved to be significant at the 1% level, and therefore, the results imply that there are indeed structural changes in the breeding sow herd series, the major effect of which appears to have been to decrease the variability of each of the series. Whether the Chow test results are a consequence of EEC membership or the Aujezky disease eradication campaign of 1983, or a combination of the two, cannot be inferred from the results as given. In an attempt to resolve this question, the Chow test was repeated for the total breeding sow model estimated on the period 1957:1 - 81:4, thereby excluding the period affected by the Aujezky disease eradication campaign. The reason for the choice of the total breeding sow model was that the results from the initial Chow test showed this model to have been least affected by the post 1974 period. The result of the repeated test which implied an even greater effect of the post 1974 period having excluded the Aujezky period for the total breeding sow model. This was, therefore, sufficient to render further Chow tests on the remaining four categories unnecessary. The Chow statistic, again having split the data at 1973/4, for the breeding sow model was 8.27, which is highly significant when compared with a value for $F_{5,82}$ of 3.25 at the 1% level of significance. This result implies the greater significance of the structural change in the post 1974 period when the Aujezky period is removed.

The Chow test results made it apparent that the models as identified and estimated on the period 1957:1-1985:4 may not be appropriate for one or possibly both of the periods pre and post 1974. For forecasting purposes, it was therefore deemed necessary to re-identify and estimate models for the post 1974 period. In order to allow for an EEC entry adjustment period, the subsequently developed biological and econometric models were estimated on the period starting at the first quarter of 1975 and so, for comparison purposes, the decision was taken to start the sample

2. See *Gujarati* pp.297, 305-306 for discussion of Chow test.

space for the later period SARIMA models at 1975:1 also. As for the full sample univariate models, only the estimated equations themselves for the shorter period models are presented.

Having obtained all the relevant Box-Jenkins models, forecasts were produced from the 1957:1-1985:4 and the 1975:1-1985:4 models in order to compare their forecasting abilities. The in-sample forecasts consisted of a 12 step ahead unconditional forecast and 12 one step ahead forecasts for the period 1982:1-1985:4, and an eight step ahead and eight one step ahead forecasts for the out-of sample period, 1986:1-87:4. The MSFE statistic was calculated for each model's forecasts, the exception being for the 12 step ahead in-sample forecasts, which were so adversely affected by the Aujezky eradication campaign of 1983 that any comparison of the MSFE statistics would have been almost meaningless. A table of the mean square forecasting errors is presented in Appendix 3d along with a discussion of the results and the implications for which sample provided the best models for forecasting the relevant periods.

In the light of the results of the analysis of the MSFE statistics from the various types of forecasts which imply that the forecasts from the longer period model are better than those from the later, shorter period, the decision was taken to use the models estimated on the longer sample period, 1957:1-85:4 as the best forecasting models for the UK breeding herd. The implications of the analysis are interesting in that they give a good deal of importance to the length of the time series sample and, therefore, the long run relationships within the time series for the UK pig breeding herd and its components. Thus, despite factors such as the apparent structural change in each of the series after 1974, the change in the sample census timings in and after 1974 and the influence of the Aujezky disease eradication campaign of 1983, the models estimated on the period 1975:1-85:4 are, on the whole, inferior at forecasting both the in-sample and out-of-sample forecasting period.

3.2a A SARIMA Model For The Total Breeding Sow Herd

Because the total breeding sow herd is the primary series for forecasting purposes, the model for this series is presented first and in full. The UK breeding sow herd figures are derived by aggregating the three component series, 'sows in pig', 'gilts in pig', and 'barren sows for breeding'. The forecast figure of 850,000 breeding sows for 1979:1 was obtained by aggregating the forecasts from the previously estimated models for the three component series on the sample period 1957:1-78:4. The plot of the breeding sow series in figure 3.1 differs from the plots of the component series' themselves - presented in subsequent sections - in that it has a much smoother appearance. Although there are signs of a cyclical element, there is less indication of seasonality than in the component series plots. This phenomenon can be explained by the fact that the seasonal pattern of the barren sow series is diametrically opposed to that which is dominant in the 'pregnant sow' series, (see sections 3.2b and 3.2d). The aggregation of these two series' within the breeding sow herd series has resulted, to a certain extent, in the cancelling out of the seasonal effects, confirmed later by the size of the estimated seasonal dummy coefficients of the Harvey model.

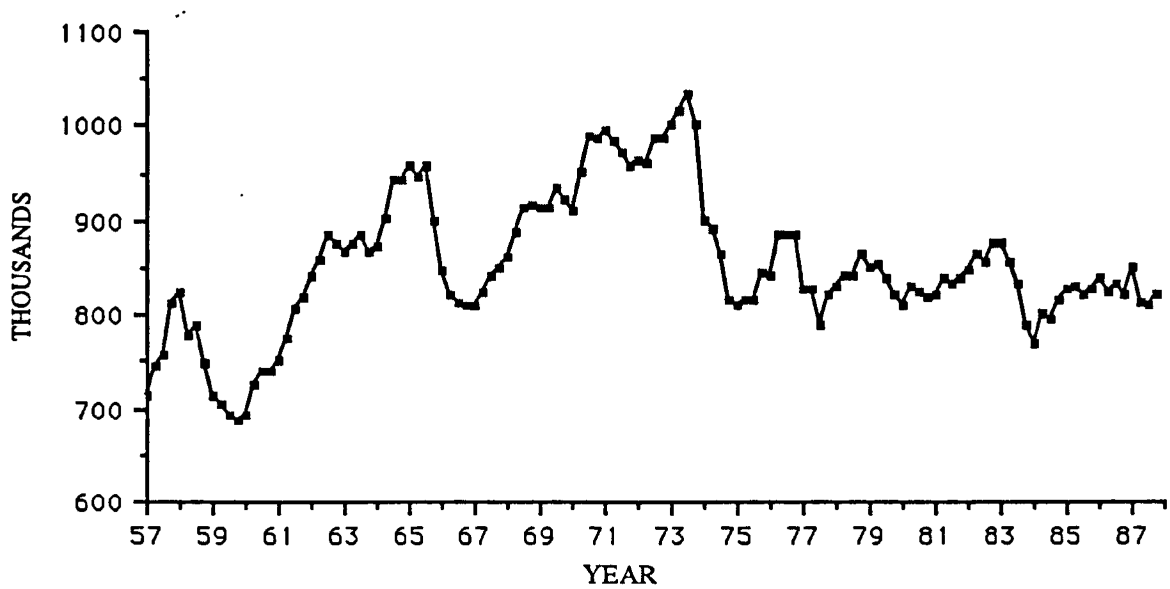
Perhaps the most noticeable feature of the plot is the sharp contrast in the variability of the size of the herd pre and post 1974. Prior to 1974 the herd experiences large variability in size and has a distinct upward trend. The MLC suggest that the sharp decline in the breeding sow herd between 1973 and 1975 was largely the result of lower profitability caused by an upward trend in world feed prices. From 1974 onwards the herd size appears to be far more stable. Although there were no reasons, *a priori*, for taking log transformations of the data, such a transformation was made for the breeding herd series. Because no obvious advantages accrued from this excercise as far as easing the identification process was concerned and because of the forecasting process is eased by not having to make transformations, the decision was made to continue working with the data as given.

The relatively sharp fall in breeding sow herd numbers in 1983 gives the post 1974 plot the appearance of a slight downward trend. It appears that the series from about the end of 1980 to the end of 1983 is behaving somewhat differently from the rest of the series of the post 1974 period. The M.L.C. Market Surveys for this period suggest that the expansion of the breeding sow herd from 1981 to 1982 was the result of high profitability and high gilt numbers, although the continuation of the rise early into 1983 was 'surprising in view of the fact that profits had begun to decline'.[3] The survey also suggested that farmers had started to replace sows at an earlier age.

---

3. See *M.L.C. Market Survey* 1983 No.2

Figure 3.1.
A Plot of the Quarterly Time Series 'Total Breeding Sow Herd', 1957:1-1987:4.



Because of the size of the breeding sow herd at this point in time and because of the fall in profits in 1983 which resulted from weak demand and an increased supply of other meats, a very high culling rate was experienced in 1983. This culling rate was further exaggerated by the Aujezky disease eradication campaign of that year, which at its peak in April and May accounted for up to 1,000 to 2,000 sows a week. The effect of the high culling rate of 1983 was to reduce the size of the breeding sow herd to its lowest since the early 1960's. These apparent changes in the behaviour of the series during the latter part of the sample period are important in that the Box-Jenkins methodology employed is highly dependent on there being no structural changes in the time series concerned. Any forecasts made by the Box-Jenkins models covering 1983 are certain to be adversely affected by the Aujezky factor.

The first step in the identification process of the model on the sample 1957:1-1981:4 was to obtain a stationary series. Although the plot of the breeding sow herd appears to be fairly stationary from the mid 1960's onwards, the autocorrelations of the raw data series in table 3.2 die away only very slowly, and the autocorrelation at lag one is so close to unity that it is evident that the series in levels is non-stationary. Taking a first difference reduces the size of the autocorrelations, however, there is little indication of them or the partials dying away, nor does any clear pattern emerge. The correlograms of the seasonally differenced series die away with acceptable speed and they also indicate the presence of cycles; a well known phenomenon of the pig industry. For these reasons, it was decided for this series, and subsequently for each of the live pig series, that a seasonal difference alone was sufficient to obtain stationary series. The correlograms of the seasonally differenced series presented in table 3.2 show autocorrelations which exhibit a cyclical pattern and which die down by the third lag. The partials at lags 1,2,3, and 5 are the only prominent partials, all being greater than the significant 2 standard errors away from zero. Having identified 'd' as zero and 'D' as 1, the first stage of identification was completed.

The next step in the model building process is to identify the size of p, P, q and Q in order to identify the structure of the SARIMA model as described in section 2.3 of chapter 2. Given the cyclical pattern in the autocorrelations, two AR parameters are included in the initial identification as illustrated in table 3.3. A seasonal autoregressive, (SAR), parameter was included as this would then multiply out with the AR parameter at lag 1, effectively producing a parameter at lag 5 to account for the large partial autocorrelation at this lag. Compared with a critical value for the t-

Table 3.2.
The Autocorrelations and Partial Autocorrelations For The Total Breeding Sow Herd Series 1957:1-1981:4.

SERIES[4]	AUTOCORRELATIONS									
	1	2	3	4	5	6	7	8	9	10
$(1-B)^0(1-B^4)^0$ H	.94	.86	.76	.67	.58	.51	.47	.43	.39	.35
$(1-B)^1(1-B^4)^0$ H	.29	.22	.04	.01	-.29	-.24	-.25	.00	-.03	.09
$(1-B)^0(1-B^4)^1$ H	.84	.57	.22	-.11	-.29	-.36	-.30	-.17	-.01	.11
SERIES	AUTOCORRELATIONS									
	11	12	13	14	15	16	17	18	19	20
$(1-B)^0(1-B^4)^0$ H	.30	.24	.16	.09	.03	-.02	-.06	-.09	-.10	-.10
$(1-B)^1(1-B^4)^0$ H	.05	.22	.00	-.08	-.16	-.02	-.11	-.19	-.01	.04
$(1-B)^0(1-B^4)^1$ H	.15	.12	.02	-.09	-.20	-.25	-.26	-.23	-.15	-.08
SERIES	PARTIAL AUTOCORRELATIONS									
	1	2	3	4	5	6	7	8	9	10
$(1-B)^0(1-B^4)^0$ H	.94	-.19	-.16	.06	-.05	-.07	.13	-.01	-.15	.01
$(1-B)^1(1-B^4)^0$ H	.29	.15	-.07	-.02	-.31	-.12	-.08	.18	.02	.01
$(1-B)^0(1-B^4)^1$ H	.84	-.44	-.38	-.13	.30	-.03	-.03	.01	.13	-.15

statistic of 2.63 for 96 degrees of freedom, each of the three included parameters had significant coefficients at the 1% level. The RSS value is 58,981 and the roots of the parameters indicate that the model is both stationary and invertible. The

4. The powers of 0 and 1 are included to emphasise to the reader that there is either no differencing or only first differencing of the time series respectively.

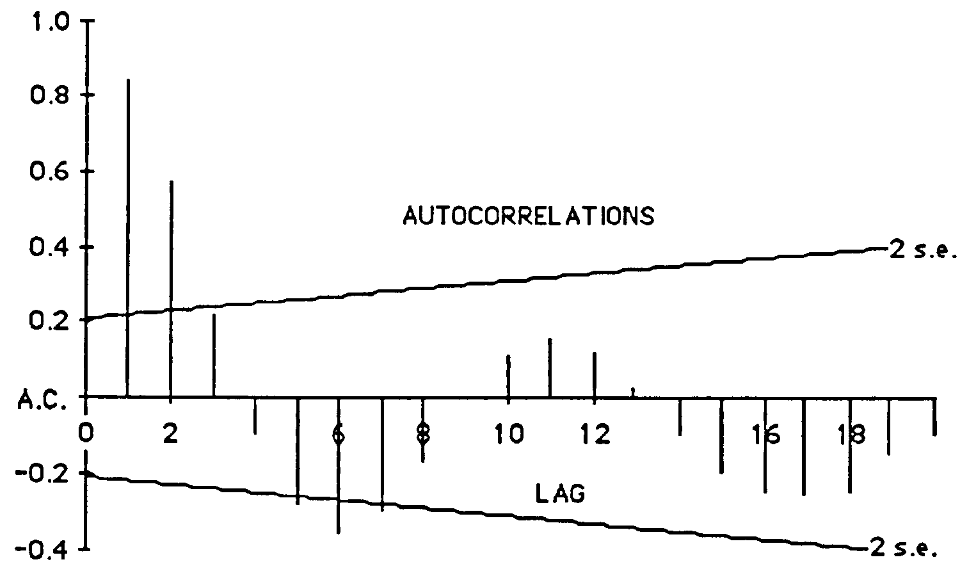
residual autocorrelations indicate that the model is under-parameterised. Compared with a Quenouille statistic of 0.204, the residual autocorrelations at lag 8 is significant and those at lags 1,2, and 3 are large. Also, the Box-Pierce Q-statistic of 39.16 is significantly high at the 1% level as indicated by the associated P-value of 0.0017. The latter statistics indicate that the null hypothesis that the residuals for the first 20 lags are white noise residuals can be rejected at the 1% level.

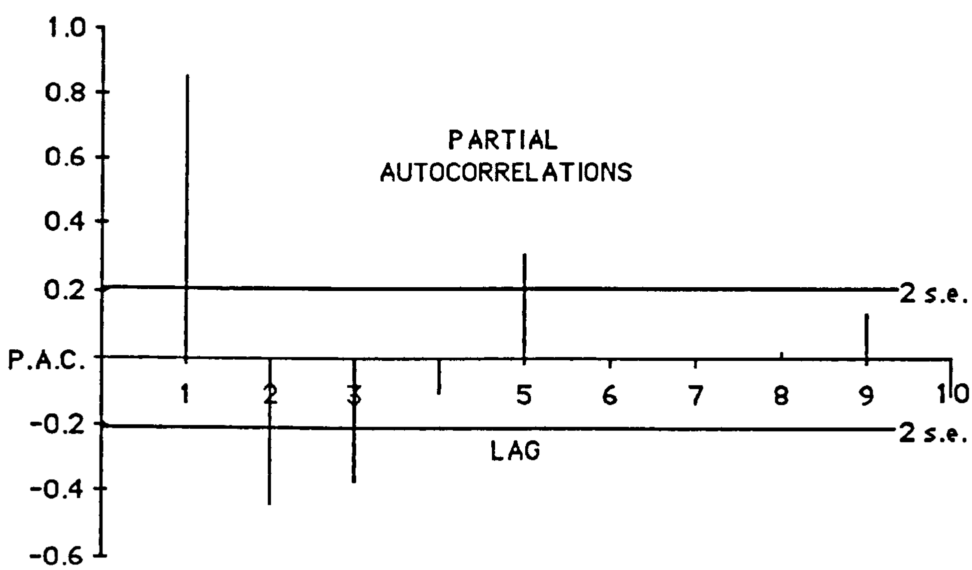
FIGURE 3.2

The Correlograms of the Series $(1\text{-}B)^0(1\text{-}B^4)^1H$ 1957:1-1981:4.

Mean = 3.13

S.E. = 62.4





In an attempt to remove the low lag residual autocorrelation problem, an MA parameter was added to the model. The additional parameter proved to have a highly significant coefficient and had the effect of reducing the RSS value by 7.6% to 54,507. The Q-statistic fell slightly to 32.34 but the P-value of 0.0090 still shows this Q-statistic to be significant at the 1% level. The addition of the MA parameter did have the desired effect of reducing the high residual autocorrelations at lags 1 and 2, although the third is still quite high at 0.177. The remaining outstanding problem is the significant residual autocorrelation of 0.25 at lag 8.

To remove this final obstacle, the model was overfitted with a second SAR parameter. Once again, the additional parameter was highly significant, the RSS being further reduced by 10.5% to 48,787. The most significant effect of the final augmentation was to remove all significant residual autocorrelations, and thereby reduce the Q-statistic to 19.7. The associated P-value of 0.183 indicates that the null hypothesis of white noise residuals is only rejected at the 19% level, and hence the white noise residual requirement appears to have been satisfied.

Table 3.3.[5]

The Results of Estimation of the Series $(1-B)^0(1-B^4)^1H$ On The Sample 1957:1 - 1981:4

MODEL									FORECASTS		
d	D	p	P	q	Q	R.S.S.	$Q_{\cdot 20}$	$P_{\cdot 20\text{-}k}$	CMSFE	UMSFE	$\hat{R}^2$
0	1	2	1	0	0	58981	39.2	0.2%	-	-	-
0	1	2	1	1	0	54507	32.3	0.9%	-	-	-
0	1	2	2	1	0	48787	19.7	18.3%	325.6	1001.4	.06

RESIDUAL AUTOCORRELATIONS FOR EACH STAGE OF ESTIMATION

MODEL						RESIDUAL AUTOCORRELATIONS									
d	D	p	P	q	Q	1	2	3	4	5	6	7	8	9	10
0	1	2	1	0	0	-.16	.14	.19	-.08	-.09	-.04	.05	-.31	.10	.10
0	1	2	1	1	0	-.08	.08	.18	-.07	-.07	.03	.08	-.25	.13	.16
0	1	2	2	1	0	-.04	.06	.09	.01	-.12	.05	-.05	-.06	.06	.14

MODEL						RESIDUAL AUTOCORRELATIONS									
d	D	p	P	q	Q	11	12	13	14	15	16	17	18	19	20
0	1	2	1	0	0	-.07	.12	-.03	.01	-.16	-.04	-.01	-.22	.09	-.18
0	1	2	1	1	0	-.02	.11	-.02	-.03	-.20	-.07	-.03	-.21	.06	-.15
0	1	2	2	1	0	.06	-.16	.09	-.01	-.17	-.03	.07	-.20	.04	-.10

RESULTS OF THE ESTIMATION OF THE SEASONAL DUMMY MODEL ON FIRST DIFFERENCES.

SEASONAL DUMMY							DUMMY VARIABLE				FORECASTS.	
d	D	p	P	q	Q	R.S.S.	1	2	3	4	CMSFE	UMSFE
1	0	0	0	0	0	53,513	-7.7	8.8	7.4	-3.8	409.6	1426.7
							(-1.6)	(1.9)	(1.6)	(-.8)		

5. RSS = Residual sum of squares.

$Q_{\cdot 20}$ = Box-Pierce Q-Statistic for residual autocorrelations up to lag 20.

$P_{20\text{-}k}$ = Probability value for Q-statistic at lag 20 in a model containing k parameters.

CFMSE = MSE from conditional forecasts.

UFMSE = MSE from unconditional forecasts.

This model was accepted in view of the fact that it could not be improved significantly by overfitting with additional parameters. The estimated equation, along with the t-statistics of each coefficient is given in equation 3.2.1. The roots of both the seasonal and the non-seasonal AR polynomials are imaginary, thereby indicating a model incorporating two cycles. The length of the cycle produced by the non-seasonal AR polynomial measures 6 years and 10 months, whereas the length of the cycle obtained from the seasonal AR polynomial is approximately 6 years and 6 months long.[6] Both cycles are rather long compared with those obtained by McClements and Ridgeon as referred to in the opening chapter. The presense of two cycles is not easy to interpret though a possible explanation is that the seasonal cycle is modulating the effects of the more dominant non-seasonal cycle.

$$\underset{(-13.6)}{(1 - 1.72B} + \underset{(6.7)}{0.78B^2}) (1 + \underset{(8.6)}{0.80B^4} + \underset{(5.5)}{0.49B^8}) (1 - B^4) H_t = (1 - \underset{(-2.7)}{0.48B}) e_t. \quad (3.2.1.)$$

Because the AR polynomials multiply out to give a polynomial of degree 14, no multiplicative test was carried out. However, in order to check the appropriateness of the identification of D and d, models were also estimated on the raw and first differenced series respectively. The non-stationary parameters resulting from these models vindicated the use of the seasonally differenced series, in that they indicated non-stationarity.

The results of fitting seasonal dummies to the first differenced data - also presented in table 3.3 - confirms all that has been said concerning the lack of seasonality in the total breeding sow herd series resulting from the cancelling out of the opposing seasonal effects in the 'pregnant sows' and 'barren sows for breeding' series. Unlike any of the other census models, none of the individual seasonal dummies is statistically significant at the 5% level, although the summer dummy comes close with a t-statistic of 1.9. The RSS of 53,513 from the seasonal dummy compares with 48,787 for the Box-Jenkins ARIMA model, producing a value for $\hat{R}_s^2$ of 0.06. This figure indicates that the Box-Jenkins model provides a 6% improvement in fit to the breeding sow herd series over that of the seasonal dummy model on first differences.

The total breeding sow herd model was re-estimated on the longer sample period 1957:1-85:4 in order to facilitate comparisons with the forecasts of the subsequently estimated biological and econometric models. Equation 3.2.2 presents the results of

6. The length of cycle is the inverse of the frequency given by $2\pi / \cos^{-1}(\phi_1 / 2 \sqrt{-\phi_2})$ in radians

estimation of this model, for which the same structure was assumed. There is very little change in the coefficients from those presented in equation 3.2.1, and each of the coefficient's t-statistics are larger with the exception of that of the second SAR coefficient.

$$\underset{(-14.0)}{(1 - 1.70B} + \underset{(6.9)}{0.77B^2}) \underset{(9.0)}{(1 + 0.80B^4} + \underset{(5.4)}{0.47B^8}) (1 - B^4) H_t = \underset{(-2.8)}{(1 - 0.48B)} e_t. \quad (3.2.2.)$$

Re-identifying and estimating the model on 1975:1-85:4 resulted in the same structure of model as identified for the 1957:1-81:4 sample. The estimated equation along with the t-statistics - which are to be compared with a 5% significance value of 2.02 for 40 degrees of freedom - is given in equation 3.2.3. below.

$$\underset{(-7.5)}{(1 - 1.53B} + \underset{(4.9)}{0.77B^2}) \underset{(4.9)}{(1 + 0.81B^4} + \underset{(3.2)}{0.51B^8}) (1 - B^4) H_t = \underset{(1.85)}{(1 - 0.53B)} e_t. \quad (3.2.3.)$$

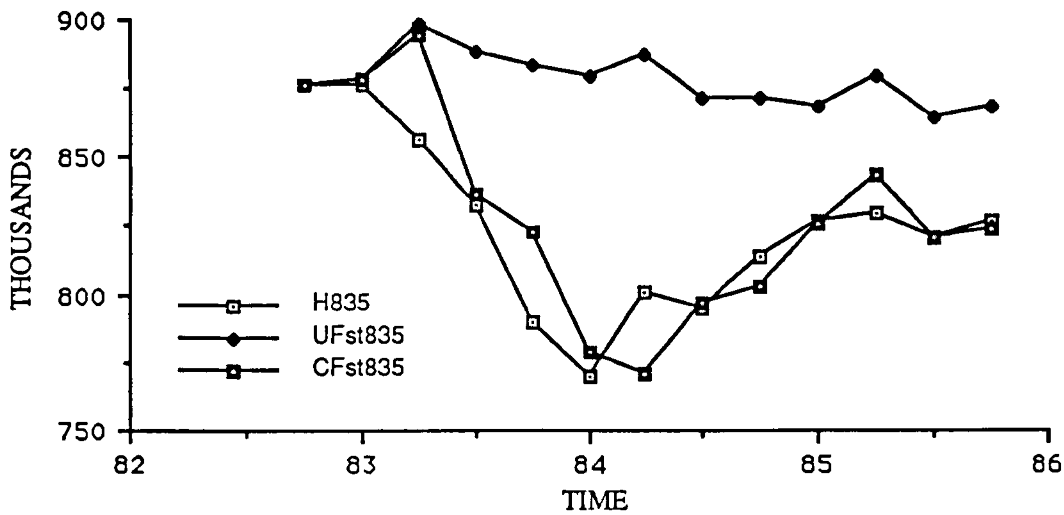
The estimated parameter coefficients of the latter regression are not too different from those estimated in 3.2.1 and 3.2.2, the exception being the first AR parameter, which is lower for the shorter period model. The resulting cycle lengths of the AR and SAR polynomials are 3 years and 1 month and 6 years and 6 months respectively, indicating the sensitivity of the measured cycle length to what is an insignificant change in the value of the first of the AR coefficients.
Figure 3.3a illustrates the in-sample forecasts for the period 1983:1-1985:4. The fall in the breeding sow herd, as a result of the Aujezky disease eradication of 1983, is clearly the cause of the unconditional over-forecasting of the period. Considering the circumstances, the conditional forecasts appear to be reasonable, although over-forecasting of June and December in 1983 is still prevalent. The MSFE of the 1 step forecasts is 323.7. The obvious comment to make about the out-of-sample plot in figure 3.3b is the greater stability of the herd and the forecasts compared with the Aujezky affected in-sample period. Both sets of forecasts miss the relatively sharp increase in the breeding sow herd in the first quarter of 1987, although this figure from the April census does appear somewhat dubious compared with the size of the herd immediately before and after this census point, particularly since the overall trend of the series over the two year period is downwards[7]. The high April census figure also accounts for the relatively large 1 step over-forecast for June of the same year. The MSFE statistics for the 8 step and 1 step forecasts are 150.38 and 185.9

7. Communication with the MLC confirmed the doubt surrounding the reliabiity of the said sample census data.

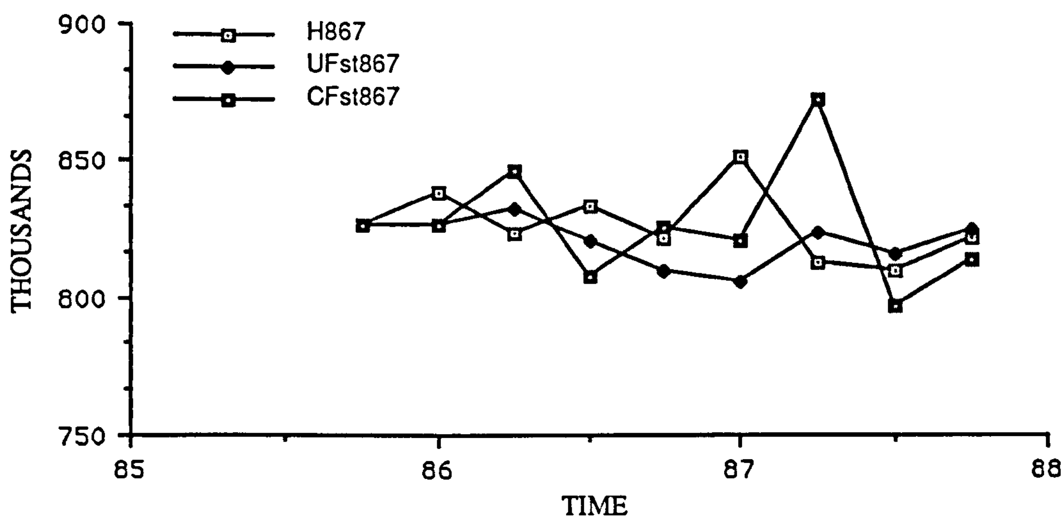
respectively. Not surprisingly, these compare well with the CMSFE of the in-sample period but the statistics are surprising in that the figure for the unconditional forecasts is lower than that of the conditional forecasts. The main reason for this phenomena is the bad 1 step forecast for June 1987 discussed above.

Figure 3.3.

a. The Conditional and Unconditional In-Sample Forecasts For The Breeding Sow Herd Estimated On The Sample 1957:1 to 1985:4



b. The Conditional and Unconditional Out-Of-Sample Forecasts For 1986:1-87:4 Estimated On The Sample 1957:1-85:4.

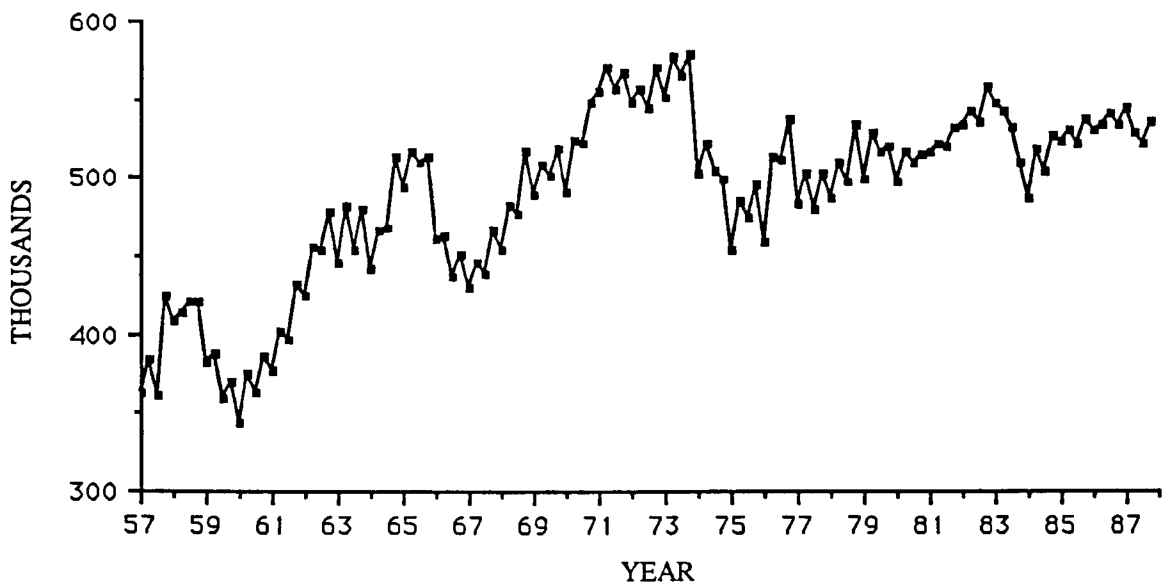


The apparently dubious April sample census data for 1987, and the consequences of this on the one step forecast for the following June - a phenomenon which re-occurs throughout the breeding herd series modelled - prompted a discussion of what actions the the time series forecasted could take faced with such a situation. This discussion, using the breeding herd model built in this section as an example, is presented in section 3.5 towards the end of the chapter.

3.2b A SARIMA Model For The U.K. Sows in Pig Herd.

Having taken a seasonal difference of the series in levels using the sample period 1957:1-78:4, a SARIMA (2,2,0,1,2,0) was identified and estimated and used to produce a forecast for 1979:1 of 498,000 pregnant sows. The plot of the pregnant sow series for the full sample period is given in figure 3.4.

Figure 3.4.
A Plot of the Quarterly Time Series 'Sows in Pig' 1957:1-1987:4.



The plot shows a series that is trending upwards over time with the suggestion of a cyclical element, although the downward slope of the cycle is noticeably steeper than that of the the upward side. This latter phenomenon can be explained by the fact that the herd can be decreased at a faster rate, through slaughtering, than it can be increased through breeding. Seasonal influences are much more apparent than they were in the total breeding sow herd plot, and the post E.E.C. change in behaviour and the Aujezky effect of 1983 are also very apparent. As was the case with the total breeding sow herd, the pregnant sow series has a more stable mean after 1974.

Table 3.4 presents the results of estimating the identified model for the series on the sample 1957:1-81:4. The cyclical element is present once again in the form of 2 AR and 2 SAR parameters whose polynomials produce imaginary roots. Two MA parameters are also included in the model.

Table 3.4.[8]
The Results of Model Estimations of the Series $(1-B)^0(1-B^4)^1 PS$.
1957:1 - 1981:4

MODEL									FORECASTS		
d	D	p	P	q	Q	R.S.S.	Q_{20}	$P_{20\text{-}k}$	CMSFE	UMSFE	$\hat{R}_s^2$
0	1	2	2	2	0	20,623	17.9	20.9%	237.4	287.7	-.06

RESIDUAL AUTOCORRELATIONS

LAG	1	2	3	4	5	6	7	8	9	10
AUTOCORRELATIONS	.03	-.09	.06	.04	-.11	.04	.04	-.03	.09	.08
LAG	11	12	13	14	15	16	17	18	19	20
AUTOCORRELATIONS	-.10	-.15	.05	.01	-.17	-.16	.03	-.12	-.05	-.09

RESULTS OF THE ESTIMATION OF THE SEASONAL DUMMY MODEL ON FIRST DIFFERENCES.

SEASONAL DUMMY							DUMMY VARIABLE				FORECASTS	
d	D	p	P	q	Q	R.S.S.	1	2	3	4	CMSFE	UMSFE
1	0	0	0	0	0	19,264	-26.8	22.4	-10.7	20.9	296.6	472.2
							(-9.2)	(7.9)	(-3.8)	(7.3)		

The Q-statistic and the associated P-value for the first 20 residual autocorrelations suggest that the autocorrelations as a whole are white noise, and the first few residual autocorrelations give no reason to suspect that the model is under-parameterised. Equation 3.2.4. illustrates that each of the model coefficients is significant at the 5% level when compared with the t-statistic of 1.985 for 96 degrees of freedom. The RSS of 20,623 compares with that of 19,264 for the seasonal dummy model on first differences and so the resultant $\hat{R}_s^2$ of -.06 implies that the dummy model gives a 6% better fit to the data than does the Box-Jenkins model. The seasonal dummy model on first differences contains four highly significant dummy variables, indicating that the number of pregnant sows increases in the Summer and in Winter, while falling in the Spring and Autumn. Despite the negative $\hat{R}_s^2$ value, the MSFE statistics imply that the SARIMA model is better than the dummy model for out-of-sample forecasting of the period 1982:1-85:4.

$$\underset{(-7.2)}{(1 - 1.46B} + \underset{(3.6)}{0.61B^2}) (1 + \underset{(3.2)}{0.46B^4} + \underset{(2.5)}{0.26B^8}) (1 - B^4) PS_t = (1 - \underset{(-2.8)}{0.54B} + \underset{(3.4)}{0.43B^2}) e_t \qquad (3.2.4)$$

The AR polynomial cycle has a length of approximately 4 years and 4 months, whereas the SAR polynomial cycle measures 5 years and 8 months. Estimating the

8. See Footnote 5

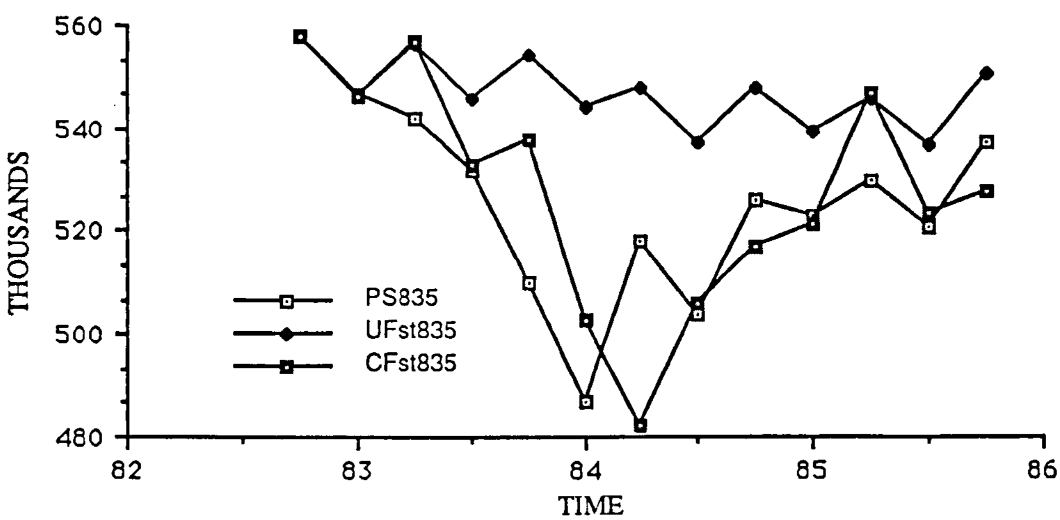
model on the extended sample, results in equation 3.2.5.. The coefficients and the t-statistics for the non-seasonal parameters decrease while those of the seasonal parameters increase, although each change is very small.

$$\underset{(-6.8)}{(1 - 1.45B} + \underset{(3.3)}{0.60B^2}) (1 + \underset{(3.95)}{0.53B^4} + \underset{(2.7)}{0.27B^8}) (1 - B^4) PS_t = (1 - \underset{(-2.6)}{0.54B} + \underset{(3.2)}{0.39B^2}) e_t \quad (3.2.5)$$

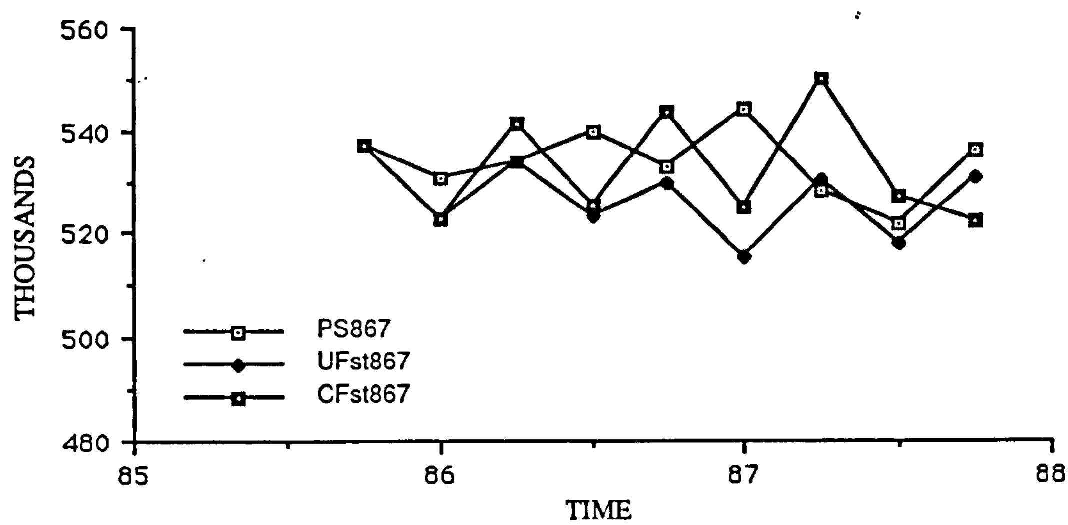
The model re-identified and estimated on 1975-85 - presented in equation 3.2.6 - has a different structure to that of equations 3.2.4 and 3.2.5 in that it has only one seasonal autoregressive parameter, and the first MA parameter is constrained to zero. The imaginary roots of the SAR polynomial produce a cycle length of 5 years and 9 months.

$$(1 - \underset{(-4.1)}{0.62B}) (1 + \underset{(4.6)}{0.72B^4} + \underset{(4.2)}{0.60B^8}) (1 - B^4) PS_t = (1 + \underset{(1.97)}{0.38B^2}) e_t. \quad (3.2.6)$$

Figure 3.5.
a. The Conditional and Unconditional In-Sample Forecasts For The Sows In Pig Herd Estimated On The Sample 1957:1 to 1985:4



b. The Conditional and Unconditional Out-Of-Sample Forecasts For 1986:1-87:4 Estimated On The Sample 1957:1-85:4.

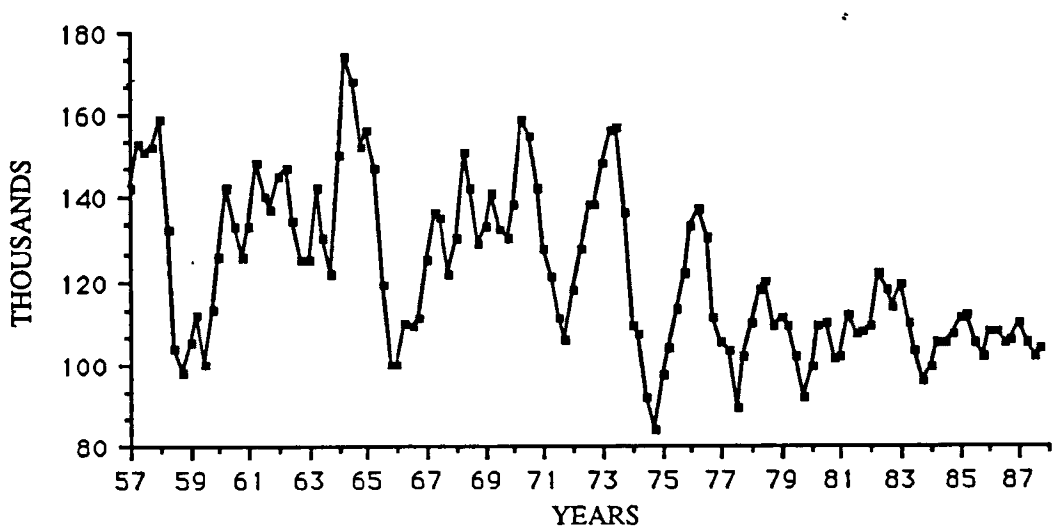


Although affected by the Aujezky factor, the unconditional in-sample forecasts produced by equation 3.2.5. are reasonable, in that they pick up both the seasonal movements and the overall u-shaped trend in the actual data over the forecast period. The conditional one step ahead forecasts for the same period are relatively good for the period beyond the second quarter of 1984, when the effects of the Aujezky factor are diminished. The CMSFE is calculated at 248.8. The comparative statistics for the out-of-sample period are 150.4 for the step unconditional forecasts and 185.9 for the conditional 1 step forecasts. These compare well with the Aujezky affected in-sample period and, as the plot indicates, the forecasts are relatively good. Once again, the magnitude of the MSFE's for the out-of-sample forecasts are the opposite of what one might expect a priori. As with the total breeding sow model, this is a result of the relatively large 1 step over-forecast for June of 1987. Again, the accuracy of the April figure for 1987 might be brought into question and therefore explain the forecast for the following June census figure.

3.2c A SARIMA Model For The U.K. Gilts In Pig Herd

The second of the three breeding sow herd components to be examined is the gilts in pig herd. The initial forecasting model on 1957:1-1978:4, a SARIMA (2,0,0,1,1,1) produced a forecast for 1979:1 of 111,000 gilts. The outstanding features of the plot in figure 3.6 are the slow downward trend, the seasonal fluctuations, and the possibility of a cycle.

Figure 3.6.
A Plot of the Quarterly Time Series 'Gilts in Pig' 1957:1-1987:4.



As was the case with the sows in pig series, the plot illustrates a more stable and less oscillatory series after 1974. The plot suggests a series behaving somewhat differently during 1981 and 1982. This is possibly due to farmers starting to replace sows at an earlier age, thereby increasing the number of gilts at a time when profits in the industry were falling. There is also evidence that the herd was affected by the 1983 slaughter campaign.

Table 3.5.[9]
The Results of Model Estimations of the Series $(1-B)^0(1-B^4)^1PG$.
1957:1 - 1981:4

MODEL								FORECASTS			
d	D	p	P	q	Q	R.S.S.	Q_{20}	P_{20-k}	CMSFE	UMSFE	$\hat{R}^2$
0	1	2	0	1	1	6,549	17.7	34.6%	24.4	34.4	0.30

RESIDUAL AUTOCORRELATIONS

LAG	1	2	3	4	5	6	7	8	9	10
AUTOCORRELATIONS	-.07	-.02	.17	-.06	.06	-.04	.02	.01	.15	-.08
LAG	11	12	13	14	15	16	17	18	19	20
AUTOCORRELATIONS	.08	.07	-.06	.05	-.20	.06	.05	-.11	-.03	-.11

RESULTS OF THE ESTIMATION OF THE SEASONAL DUMMY MODEL ON FIRST DIFFERENCES.

SEASONAL DUMMY							DUMMY VARIABLE				FORECASTS.	
d	D	p	P	q	Q	R.S.S.	1	2	3	4	CMSFE	UMSFE
1	0	0	0	0	0	9,710	5.21	6.84	-6.76	-6.44	30.5	48.7
							(2.5)	(3.4)	(-3.3)	(-3.2)		

9. See Footnote 5

The model identified and estimated for 1957:1-1981:4 on the seasonally differenced data comprised 2 AR , 1 MA, and 1 SMA parameters. The imaginary roots of the AR polynomial produced a cycle length of 3 years and 3 months which is more in line with the length of cycle suggested by McClements (1970). The results of estimation and the model itself are presented in table 3.5 and equation 3.2.7. respectively.

$$\underset{(-18.0)}{(1 - 1.59B} \underset{(10.3)}{+ 0.80B^2)} (1 - B^4) PG_t = \underset{(-2.8)}{(1 - 0.40B)} \underset{(-10.6)}{(1 - 0.74B^4)} e_t. \qquad (3.2.7.)$$

The t-statistics in equation 3.2.7. illustrate that each of the 4 parameters is highly significant and the RSS takes a value of 6,549. The Q-statistic of 17.7 for 20 lags has a corresponding P-value of 34.6% implying a white noise residual process. None of the first few residual autocorrelations is significant, and although that a lag 15 comes very close to being significant, it is not at a seasonal lag and, therefore, it was not thought appropriate to do anything about it.

Fitting the seasonal dummy model on first differences produces dummy coefficients which are all highly significant. The seasonal pattern differs from that of the pregnant sows in that the number of pregnant gilts rises in the Spring and Summer as opposed to rising in Winter and Summer. This phenomenon helps to explain the large oscillations present in figure 3.6. The RSS statistic of 9,710 for the seasonal dummy model combines with that of the Box-Jenkins model to produce an $\hat{R}_s^2$ value of 0.304. This is an encouraging statistic for the SARIMA model, in that it gives a 30.4% better fit to the data over the more naive dummy model.

The expanded form of the model is given in equation 3.2.8. below.

$$(1 - 1.59B + 0.80B^2 - B^4 + 1.59B^5 - 0.80B^6) PG = (1 - 0.40B - 0.74B^4 + 0.30B^5) e_t \qquad (3.2.8)$$

Because the number of parameters in the SARIMA model for pregnant gilts is relatively small, the validity of the multiplicity assumption was tested by fitting an ARMA(6,5) with the coefficients ϕ_3, θ_2 and θ_3 fixed at a value of zero, as presented by equation 3.2.9. Making a comparison of equations 3.2.8 and 3.2.9, it is clear that the multiplicative and the non-multiplicative equations are similar to one another. The critical comparison of the 5th order MA parameters suggests that the multiplicative model does place some restriction upon θ_5, although the restriction does not appear to be great enough to invalidate the use of the multiplicative model.

$$(1 - 1.64B + 0.85B^2 - B^4 + 1.64B^5 - 0.85B^6) PG_t = (1 - 0.43B - 0.85B^4 + 0.53B^5) e_t \qquad (3.2.9)$$

Estimating the model with the complete data set to 1985:4 produces the equation presented in equation 3.2.10 in which each of the the t-statistics are larger than in 3.2.7, implying greater significance for each of the included parameters.

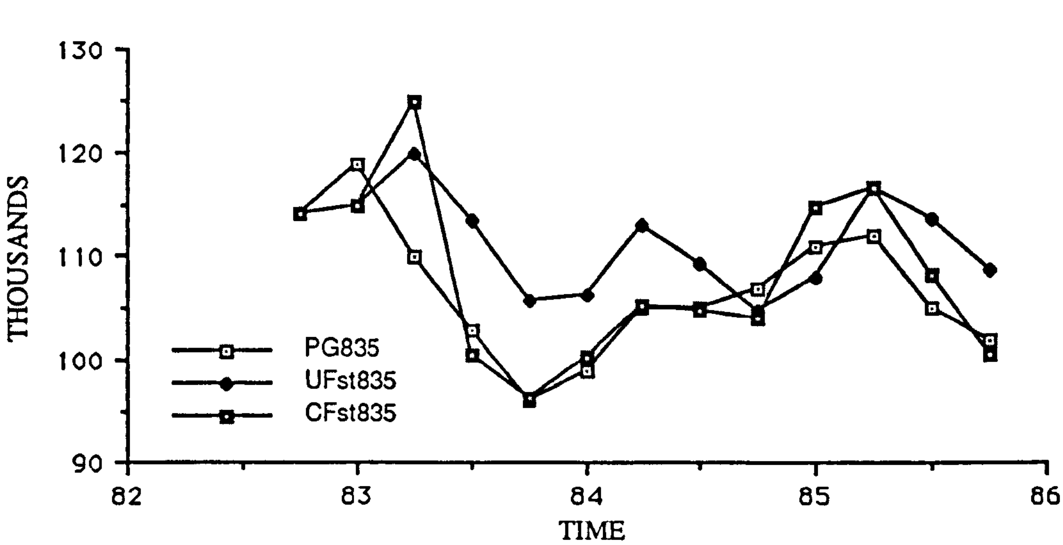
$$\underset{(-20.0)}{(1 - 1.59B} + \underset{(11.5)}{0.81B^2}) (1 - B^4) PG_t = (1 - \underset{(-3.0)}{0.39B}) (1 - \underset{(-10.9)}{0.72B^4}) e_t. \qquad (3.2.10)$$

Identifying and estimating a Box-Jenkins model on the shorter sample period, 1975:1 to 1985:4, produces a model containing more variables, that is, 2 AR, 2 MA and 2 SAR parameters. The estimated model is reproduced in equation 3.2.11. The length of the cycles produced by the AR and the SAR polynomials are 2 years and 3 months and 5 years and 7 months respectively.

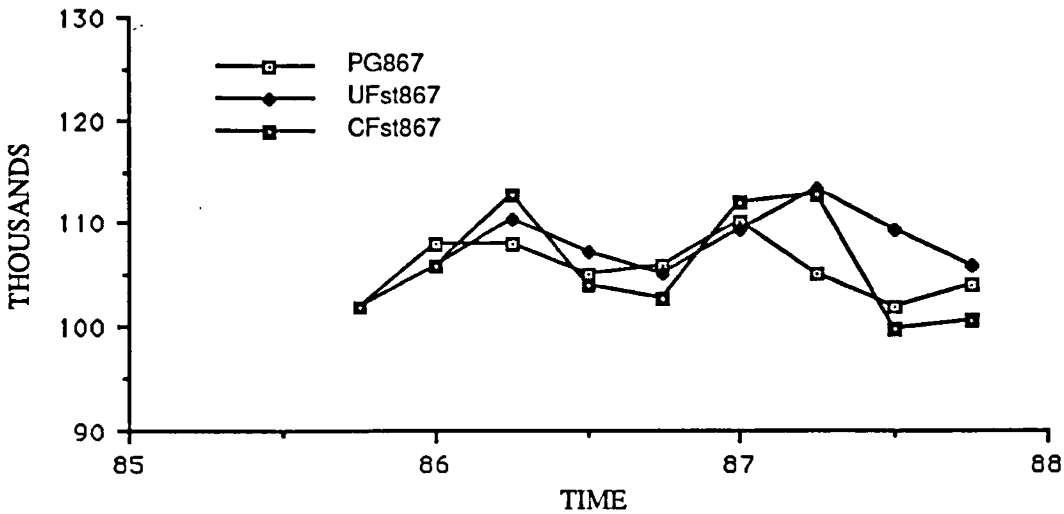
$$(1 - \underset{(-53.2)}{1.52B} + \underset{(73.1)}{0.99B^2})(1 + \underset{(3.4)}{0.56B^4} + \underset{(1.86)}{0.28B^8})(1 - B^4) PG_t = (1 - \underset{(-17.3)}{1.08B} + \underset{(22.5)}{0.94B^2}) e_t. \qquad (3.2.11)$$

The in-sample forecasts for the period 1983:1-85:4 produced by equation 3.2.10. and presented in figure 3.7a show unconditional forecasts which over-forecast for most of the period beyond 1983:2, although they do pick up the seasonal movements in the actual series. The conditional one-step forecasts for the same period follow the actual series very closely, the sole exception being 1983:2 which corresponds with the peak slaughtering period of the Aujezky disease eradication campaign. The MSFE for the conditional forecasts measures 25.22.

Figure 3.7.
a. The Conditional and Unconditional In-Sample Forecasts For The Gilts In Pig Herd Estimated On The Sample 1957:1 to 1985:4



b. The Conditional and Unconditional Out-Of-Sample Forecasts For 1986:1-87:4 Estimated On The Sample 1957:1-85:4.



The unconditional and the conditional MSFE's for the out-of-sample forecasts are 17.5 and 15.0 respectively. These compare well with the CMSFE for the in-sample period. The plot in figure 3.7b illustrates both sets of out-of-sample forecasts picking up the seasonal movements of the pregnant gilt herd, the only exception being for the June figure of 1987 which is over-forecast by both the 1 step and the 8 step forecasts.

3.2d A SARIMA Model For The U.K. Barren Sow Herd

The final component of the breeding sow herd is the Barren sow herd, consisting of those sows which are not in pig at the time of the census. The forecast figure of 241,000 pigs, included in the plot of the series in figure 3.8, was obtained from a model identified as a SARIMA (1,2,0,1,2,0) on the sample 1957:1-1978:4.

The time series plot appears to be non-stationary, especially post 1974 when there is a strong downward trend illustrating the shortening of the weaning time over this period. The cyclical element which is present in the pre 1974 period is not as apparent post 1974. Like the other two breeding sow herd components, the number of barren sows falls during the Aujezky period of 1983.

The results of estimating the seasonal dummy model on the first differences of the sample 1957:1-1981:4 - presented in table 3.6 - reveals a seasonal pattern diametrically opposed to that of the sows in pig series. This result is to be expected if the breeding sow herd is a reasonably constant size over time. All four of the seasonal dummies are significant and the RSS statistic is 15,417.

Figure 3.8.

A Plot of the Quarterly Time Series 'Barren Sows for Breeding' 1957:1-1987:4.

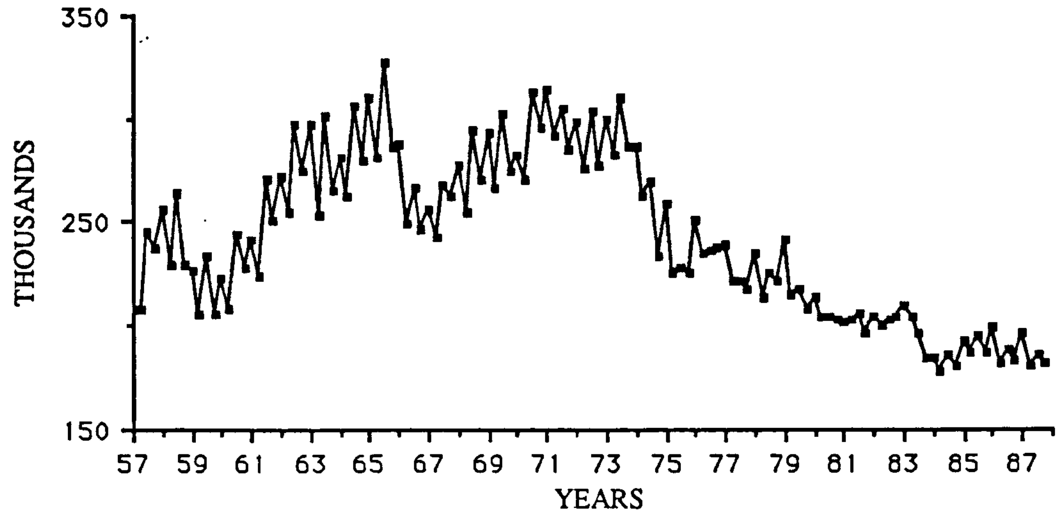


Table 3.6.[10]

The Results of Model Estimations of the Series $(1-B)^0(1-B^4)^1BS$.

1957:1 - 1981:4

MODEL									FORECASTS		
d	D	p	P	q	Q	R.S.S.	$Q_{\cdot 20}$	$P_{\cdot 20\text{-}k}$	CMSFE	UMSFE	$\hat{R}^2$
0	1	1	2	2	0	9,864	12.6	70.2%	25.7	84.6	0.34

RESIDUAL AUTOCORRELATIONS.

LAG	1	2	3	4	5	6	7	8	9	10
AUTOCORRELATIONS	-.05	-.02	.12	.00	-.09	-.07	.03	-.00	-.04	.03

LAG	11	12	13	14	15	16	17	18	19	20
AUTOCORRELATIONS	.20	-.05	-.07	-.01	-.09	-.02	-.10	-.09	-.06	.00

RESULTS OF THE ESTIMATION OF THE SEASONAL DUMMY MODEL ON FIRST DIFFERENCES.

SEASONAL DUMMY							DUMMY VARIABLE				FORECASTS.	
d	D	p	P	q	Q	R.S.S.	1	2	3	4	CMSFE	UMSFE
1	0	0	0	0	0	15,417	14.0	-20.4	24.8	-18.3	248.3	250.2
							(5.4)	(-8.0)	(9.7)	(-7.2)		

Table 3.6 also reveals that the identified Box-Jenkins model on 1957:1-1981:4, for the barren sow series consisted of 1 AR, 2 MA and 2 SAR parameters, the polynomial of the latter having imaginary roots which imply a cycle length of

10. See Footnote 5

approximately 5 years and 8 months. Again, the model was identified on the seasonally differenced data. Because the first of the 2 MA parameters is not significant, it was constrained to a value of zero. The other four estimated parameters were all highly significant, and the RSS of 9,506 compares favourably with that of the more naive dummy model to produce an $\hat{R}_s^2$ value of 0.34. The only significant residual autocorrelation is at lag 11, which is not a seasonal lag. The Box-Pierce Q-statistic at lag 20 is 12.6, which is low enough to have a P-value of 70.2%. The latter indicates that the residual autocorrelations are consistent with a white noise process. The MSFE statistics also show the SARIMA model to be a far superior forecasting model over the seasonal dummy model for the period 1982:1-1985:4. Equation 3.2.12. is the result of estimating the identified Box-Jenkins model on the 1957:1-1981:4 sample.

$$\underset{(-11.6)}{(1-1.82\,B)}(1+\underset{(4.4)}{0.51\,B^4}+\underset{(3.1)}{0.32\,B^8})(1-B^4)\,BS_t = (1+\underset{(3.6)}{0.42\,B^2})\,e_t. \qquad (3.2.12)$$

Once again, the model was re-estimated on the extended sample period up to and including 1985:4, the resultant equation being presented in equation 3.2.13.

$$\underset{(-12.4)}{(1-1.82\,B)}(1+\underset{(4.8)}{0.52\,B^4}+\underset{(3.2)}{0.30\,B^8})(1-B^4)\,BS_t = (1+\underset{(3.9)}{0.42\,B^2})\,e_t. \qquad (3.2.13.)$$

The parameter coefficients of the model on the larger sample are very similar to those of equation 3.2.12., although each of the t-statistics has increased, thereby increasing the significance of each of the included parameters.

Re-identifying and estimating a model on 1975:1-1985:4 produces a model with changes over the previously identified model, which is perhaps not surprising in view of the very different appearance of the plot from 1974 onwards. Identified on the first differenced series, the model includes 2 AR , 2 MA and 2 SAR parameters.

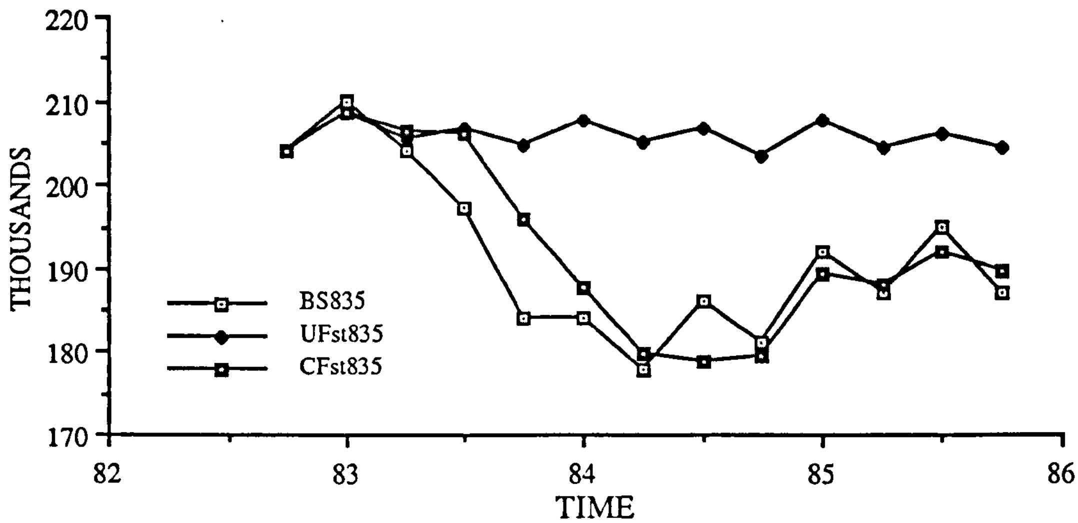
$$(1-\underset{(-9.01)}{1.22B}+\underset{(5.54)}{0.73B^2})(1+\underset{(2.16)}{0.36B^8})(1-B^4)\,BS_t = (1-\underset{(-3.42)}{0.42B}+\underset{(8.05)}{0.82B^2})\,e_t. \qquad (3.2.14)$$

The first of the MA parameters is no longer restricted to zero, but the first of the two SAR parameters is. The results of estimation of this later period model were not as good as those of the full sample model, both in terms of parsimony and the residual autocorrelation checks.

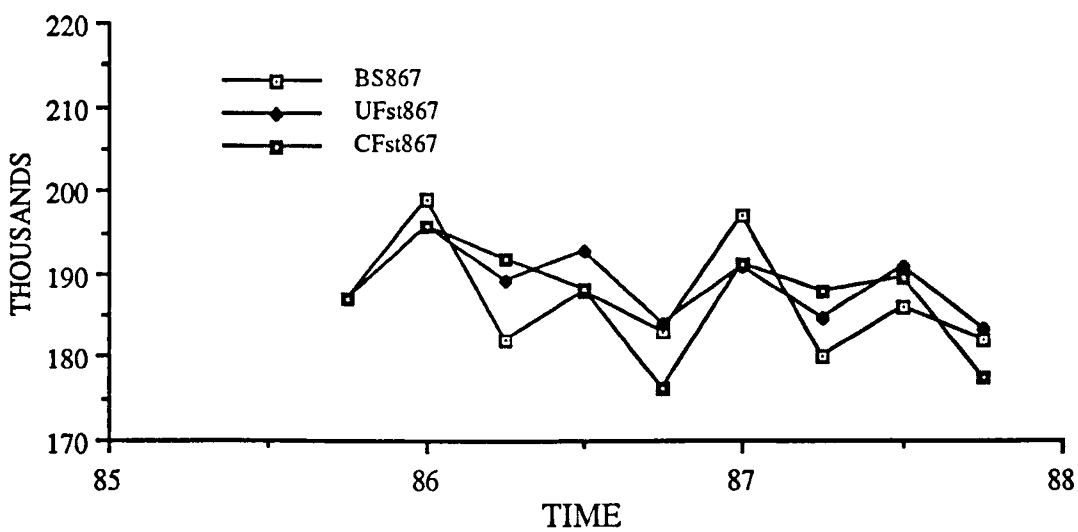
The general results of in-sample and out-of-sample forecasting from the 1957:1-85:4 SARIMA model are very similar to those of the other breeding sow category models. The Aujezky disease eradication campaign has resulted in unconditional over-forecasting during and beyond 1983 and conditional over-forecasting for most of 1983, otherwise both sets of forecasts pick up the seasonal movements quite well. The calculated CMSFE for the in-sample period is 26.6.

Figure 3.9.

a. The Conditional and Unconditional In-Sample Forecasts For The Barren Sow Herd Estimated On The Sample 1957:1 to 1985:4



b. The Conditional and Unconditional Out-Of-Sample Forecasts For 1986:1-87:4 Estimated On The Sample 1957:1-85:4.

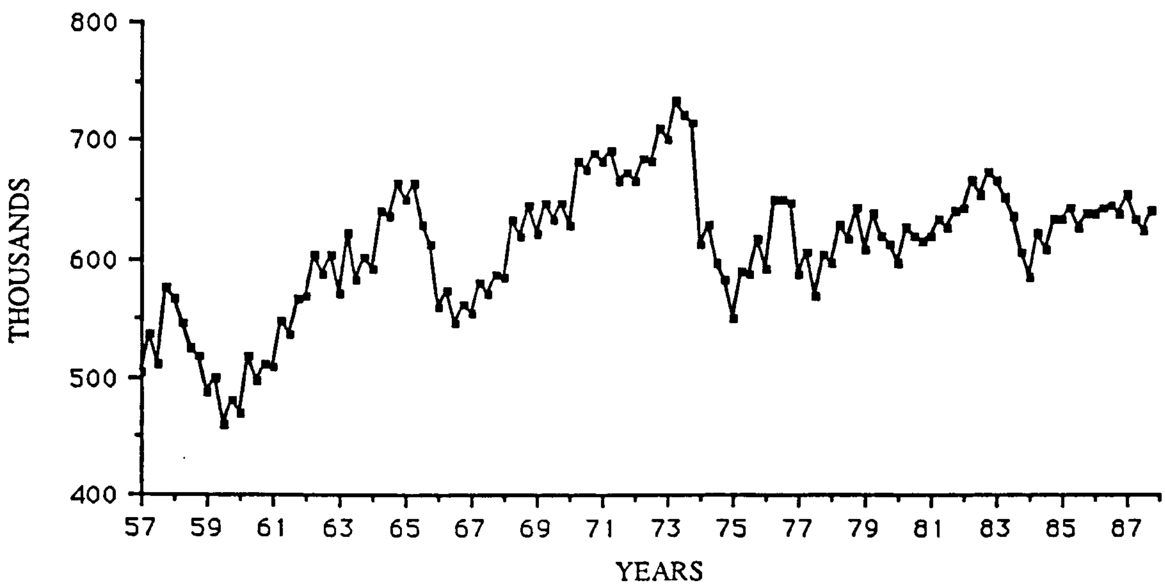


Both sets of forecasts for the out-of-sample period pick up the seasonal movements in the barren sow herd very well but once again, the UMSFE of 21.7 is smaller than the CMSFE of 35.8. There is no obvious reason looking at the plot why this should occur. Furthermore, when translating these statistics into mean absolute errors, the difference between the two is 1.3 thousand pigs only, or approximately 0.7% of the barren sow herd.

3.2e A SARIMA Model For The Total In-Pig Herd

The 'in-pig sow' and 'in-pig gilt' herds are aggregated to produce the total in-pig - pregnant pig - herd, (PP), which was then modelled in the usual manner. The series is modelled as an aggregate series due to the importance of the total in pig numbers in the subsequent biological model. The forecasting performance of the aggregate series could also be compared with that of the forecasts obtained from having aggregated the forecasts produced by the two component models. The forecasts from the two component models estimated on the period 1957:1-78:4 were aggregated to produce a forecasts of 609,000 pregnant pigs for the missing census in 1979:1. Because the in-pig sow herd is two and a half times the size of the in-pig gilt herd, it is not surprising that the plot of the pregnant pig series in figure 3.10 is dominated by the pattern that was present in figure 3.4, the plot of the pregnant sow herd.

Figure 3.10.
A Plot of the Series ' Total Pregnant Pigs' 1957:1-87:4



As was the case with the three component models, a seasonal difference of the raw data was required to satisfy the stationarity condition in the correlograms. The model identified and estimated on the 1957:1-1981:4 sample period comprised 2 AR, 2 SAR and 1 MA parameters, which is a similar identification to that of the pregnant sow model except that the second of the MA parameters is not included in the pregnant pig model. The estimated equation and t-statistics are given in equation 3.2.15.

$$(1 - 1.70\,B + 0.78\,B^2)\,(1 + 0.73\,B^4 + 0.42\,B^8)\,(1 - B^4)\,PP_t = (1 - 0.57\,B)\,e_t. \qquad (3.2.15.)$$
$$\quad(-11.8)\quad(6.1)\qquad\qquad(7.1)\qquad(4.3)\qquad\qquad\qquad\qquad(-2.9)$$

The AR and SAR polynomials have imaginary roots, indicating the presence of two superimposed cycles. The measured cycle length of the non-seasonal polynomial is approximately 5 years and 9 months, which compares with 6 years and 5 months for the seasonal cycle. The results of estimating the Box-Jenkins model and the Harvey model are given in Table 3.7.

Table 3.7.[11]

The Results of Model Estimations of the Series $(1\text{-}B)^0(1\text{-}B^4)^1PP$
Estimated on 1957:1 - 1981:4

MODEL									FORECASTS		
d	D	p	P	q	Q	R.S.S.	$Q_{\cdot 20}$	$P_{\cdot 20\text{-}k}$	CMSFE	UMSFE	$\hat{R}_s^2$
0	1	2	2	1	0	32,167	13.8	53.9%	293.3	526.3	-.04

RESIDUAL AUTOCORRELATIONS

LAG	1	2	3	4	5	6	7	8	9	10
AUTOCORRELATIONS	-.05	.07	.06	.03	-.14	-.03	-.02	-.05	.10	.02
LAG	11	12	13	14	15	16	17	18	19	20
AUTOCORRELATIONS	.10	-.15	.06	-.00	-.15	-.09	.02	-.09	-.00	-.07

RESULTS OF THE ESTIMATION OF THE SEASONAL DUMMY MODEL ON FIRST DIFFERENCES.

SEASONAL DUMMY							DUMMY VARIABLE				FORECASTS.	
d	D	p	P	q	Q	R.S.S.	1	2	3	4	CMSFE	UMSFE
1	0	0	0	0	0	32,246	-21.6	29.2	-17.85	14.5	361.5	699.8
							(-5.8)	(7.9)	(-4.7)	(3.9)		

The Q-statistic of 13.8 has an associated probability value of 53.9% indicating the general acceptability of the residuals as a whole. Estimating the parameters of the seasonal dummy on first differences of the raw data produces results similar to those obtained from estimating the Harvey model on the pregnant sow data. The seasonal pattern is the same in direction, although the seasonal pattern of the pregnant gilt series has had the effect of dampening the fluctuations in 'Winter' and 'Spring' while enhancing them in 'Summer' and 'Autumn'. Each of the seasonal dummies is significant as measured by the t-statistic. Comparing the RSS of 32,246 for the dummy model with 32,167 for the SARIMA model results in an $\hat{R}_s^2$ of -0.04 indicating that the seasonal dummy model adds 4% improvement to the fitting

11. See Footnote 5

ability of the SARIMA model. Having said this, the $\hat{R}^2$ value for the pregnant pig model is an improvement on that obtained for the pregnant sow model. The conclusions drawn from comparisons of the MSFE's for the SARIMA and the seasonal dummy models are once again favourable towards the Box-Jenkins derived model.

The model re-estimated on the longer sample period up to and including 1985:4 is given below, and, as has been the case with all the breeding sow models, the extension of the estimation period to 1985:4 produces t-statistics with increased significance.

$$\underset{(-12.6)}{(1 - 1.69\,B} + \underset{(6.5)}{0.77\,B^2}\,)\,(1 + \underset{(7.6)}{0.73\,B^4} + \underset{(4.1)}{0.39\,B^8}\,)\,(1 - B^4\,)\,PP_t = (1 - \underset{(-3.7)}{0.57\,B})\,e_t. \qquad (3.2.16.)$$

The model identified and estimated on the sample period 1975:1-85:4 includes 2 AR, 2 MA and 1 SMA parameters. The model as presented in equation 3.2.17 contains an AR polynomial with unreal roots, indicating the presence of a cycle with a measured length of 2 years and 7 months.

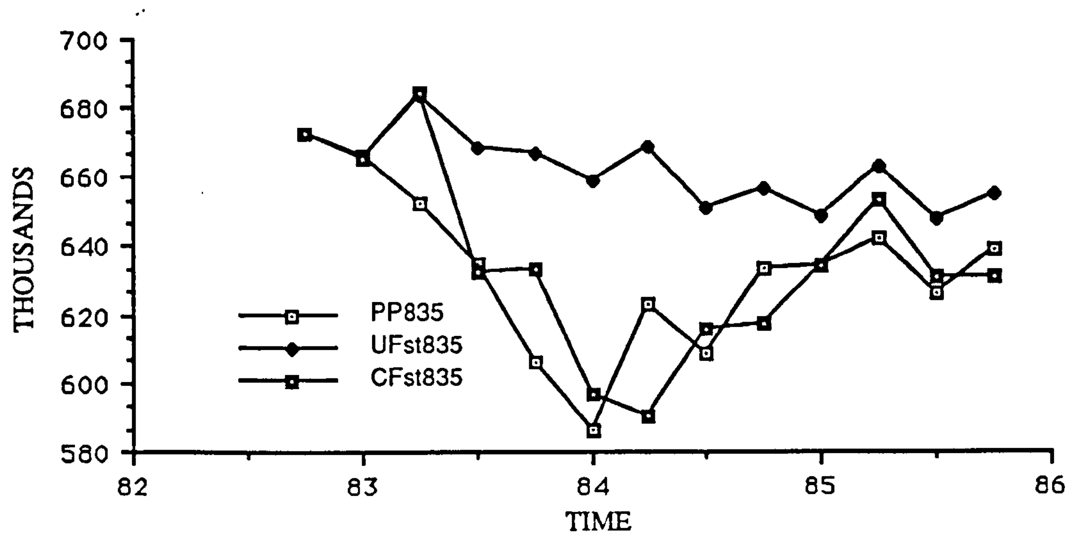
$$(1 - \underset{(-10.6)}{1.48\,B} + \underset{(7.1)}{0.82\,B^2}\,)\,(1 - B^4\,)\,PP_t = (1 - \underset{(-3.2)}{0.67\,B} + \underset{(1.65)}{0.33\,B^2})\,(1 - \underset{(-4.9)}{0.61\,B^4})\,e_t \qquad (3.2.17)$$

The plots of the in-sample and out-of-sample forecasts produced by 3.2.16 are given in figure 3.13. As expected, the unconditional in-sample forecasts fail to pick up the fall in the herd size during the Aujezky period in 1983. The 1 step conditional forecasts for the in-sample period show a very similar pattern of performance to those of the pregnant sow model presented in section 3.2b. The CMSFE measures 287.5.

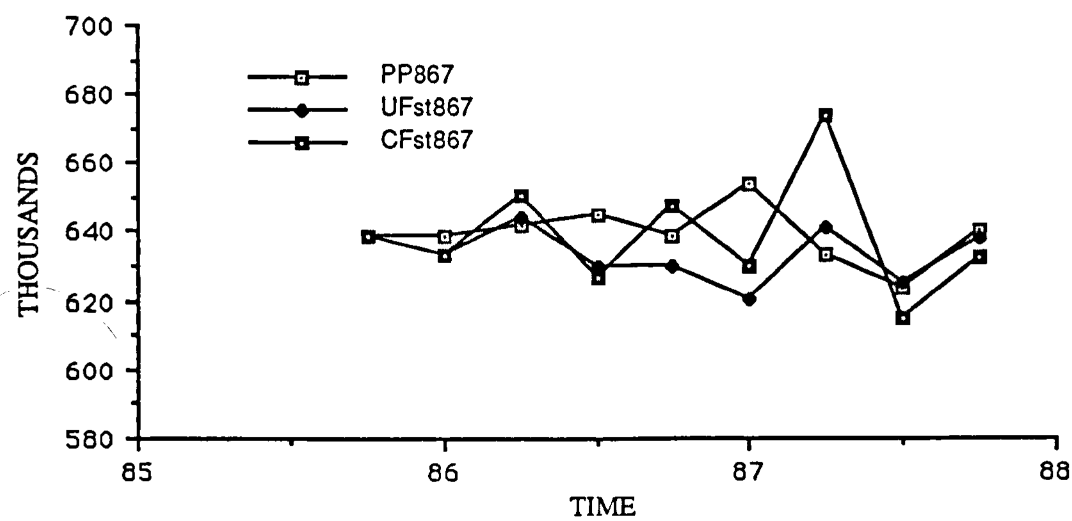
The out-of-sample forecast results are also similar to those of the pregnant sow model. The 8 step unconditional forecasts are particularly good at the start and the end of the period, and the UMSFE of 188.7 compares very well with the equivalent statistic from the 1 step conditional forecasts of the in-sample period. The out-of-sample 1 step forecasts have a CMSFE of 358.7 due once again to the relatively large over-forecast of the June 1987 figure. Again, the responsibility for this must lay at the hands of the rather questionable April figure in that year.

Figure 3.11.

a. The Conditional and Unconditional In-Sample Forecasts For The Pregnant Pig Herd Estimated On The Sample 1957:1 to 1985:4



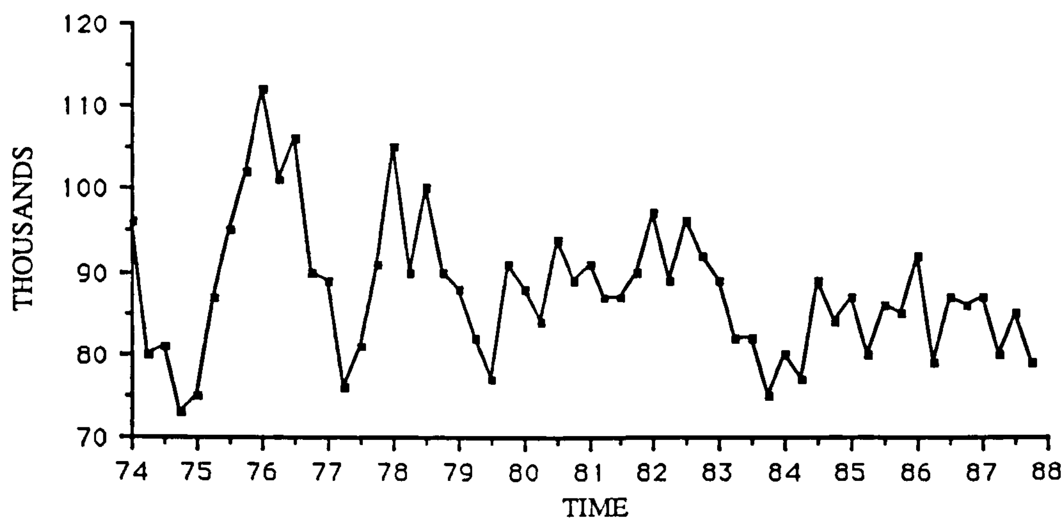
b. The Conditional and Unconditional Out-Of-Sample Forecasts For 1986:1-87:4 Estimated On The Sample 1957:1-85:4.



3.3 A SARIMA Model For The U.K. Unserved Gilt Herd

Collation of the unserved gilt herd only began in 1974 as a result of the U.K.'s accession to the E.E.C. 'Unserved gilts' is the term which is used to describe those gilts over 50Kg which are to enter the breeding sow herd, but are not yet in pig. The series is an important leading indicator of the number of gilts in pig, itself a leading indicator of the size of the breeding sow herd. The data are collected along with breeding sow herd data in the four pseudo-quarterly censuses carried out by M.A.F.F. Along with each of the other census categories, the datum for 1979:1 is missing as a result of the civil service strike action of that quarter. As the sample size was not considered large enough for the building of a Box-Jenkins model with which to forecast the missing figure, a three point moving-average was used to provide an interpolated figure of 88,000 pigs. The appropriateness of this figure was later checked by forecasting the 1979:1 figure again using the Box-Jenkins model subsequently built using the interpolated value of 88,000. This was done to ensure that seasonality was taken in to consideration when interpolating the missing observation. The subsequent forecast of 87,800 suggested that the initial interpolated value was indeed a reasonable value for 1979:1. The plot of the complete series from 1974:1-1985:4, including the interpolated figure of 88,000, is presented in figure 3.12.

Figure 3.12.
Plot of the Unserved Gilts Series 1974:1-1987:4



The plot gives the impression of a slight downward trend over the given period. Apart from the Aujezky period, (1983), the plot appears much more stable in the 1980's than it does in the 1970's.

For compatibility with the subsequently developed biological and economic models, the SARIMA model was estimated on the period 1975:1 to 1985:4 inclusive, the data for 1986:1-87:4 being reserved for out of sample diagnostic checks. At the identification stage, it was difficult to interpret whether or not the correlograms of the raw data series were stationary. Although a model identification on the raw series was attempted, the estimated parameters for the resultant model implied that the series was indeed non-stationary. The autocorrelations of the first differences series gave no indication of stationarity and so, as with the breeding sow herd series, the unserved gilt series had to be seasonally differenced to achieve stationarity.

The identified model consisted of 1 AR, 1 SAR and 2 SMA parameters and, therefore, the model does not contain a cyclical element. Estimation produces an RSS value of 1,166 and, as equation 3.3.1. illustrates, each of the 4 parameter coefficients have t-statistics which are greater than the critical value of 2.02 for 39 degrees of freedom.

$$\underset{(-6.3)}{(1-0.74B)}\underset{(48.0)}{(1+0.99B^4)}(1-B^4)\,UG_t = (1-\underset{(-2.1)}{0.27B^4}-\underset{(-5.1)}{0.60B^8})\,e_t. \tag{3.3.1}$$

None of the first 20 residual autocorrelations were as large as the Quenouille statistic of 0.30 for 44 degrees of freedom, and the P-value of 53%, associated with the Q-statistic of 14.92, gives no reason to suspect that the first 20 residual autocorrelations show anything other than a white noise residual process.

Table 3.8.[12]
The Results of Estimation of the Unserved Gilt Series
On the Sample 1975:1 - 1985:4

MODEL									IN-SAMPLE		
d	D	p	P	q	Q	R.S.S.	Q_{20}	P_{20-k}	CMSFE	UMSFE	$\hat{R}^2$
0	1	1	1	0	2	1,166	14.9	53.0%	21.9	74.2	0.327

RESIDUAL AUTOCORRELATIONS.

LAG	1	2	3	4	5	6	7	8	9	10
AUTOCORRELATIONS	-.06	.13	-.13	-.04	-.20	-.19	.07	.04	.17	-.08
LAG	11	12	13	14	15	16	17	18	19	20
AUTOCORRELATIONS	.12	-.14	-.11	-.13	.06	-.02	.15	.09	.08	-.09

RESULTS OF THE ESTIMATION OF THE SEASONAL DUMMY MODEL ON FIRST DIFFERENCES.

SEASONAL DUMMY							DUMMY VARIABLE				IN-SAMPLE.	
d	D	p	P	q	Q	R.S.S.	1	2	3	4	CMSFE	UMSFE
1	0	0	0	0	0	1,876	3.2	-6.0	5.3	-1.3	14.4	138.1
							(1.5)	(-2.9)	(2.5)	(-0.6)		

Table 3.8 also shows the results of having estimated the seasonal dummy model on the first differenced series for the 1974-1985 sample. The RSS value of 1,876, which is higher than that of the SARIMA model, is a result of only the Spring and

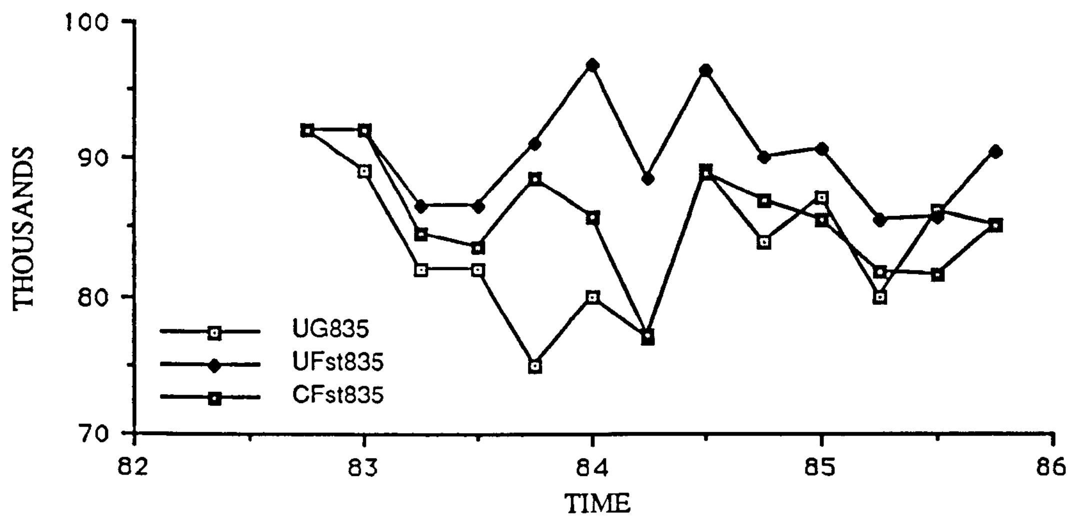
12. See Footnote 5

Summer dummies being significant at the 5% level. The resultant $\hat{R}_s^2$ value of 0.327 indicates that the SARIMA model provides a 32.7% better fit to the sample than does the seasonal dummy model on first differences.

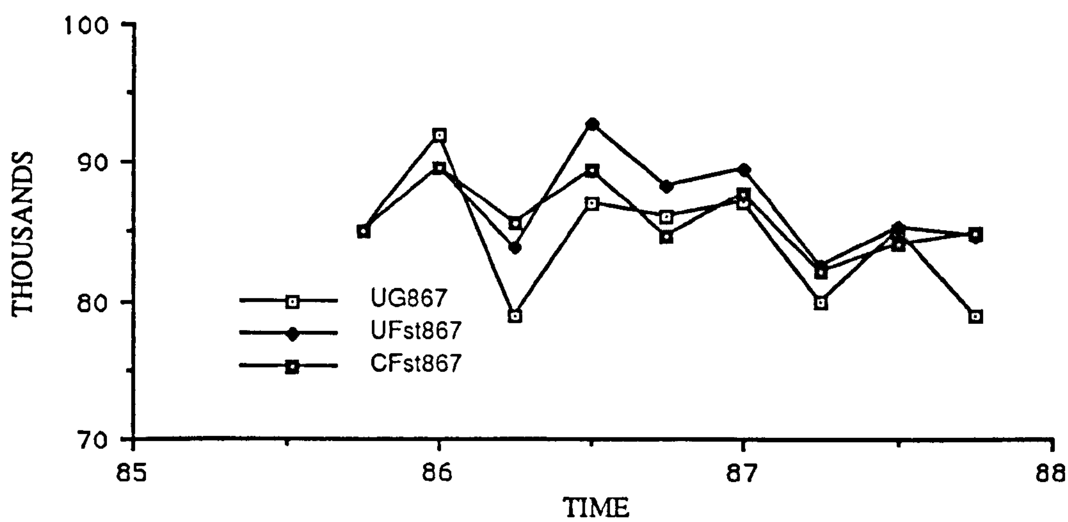
The in-sample and out of sample forecasting periods are the same as those used for the breeding sow herd models, and as was the case with these models, the in-sample forecasts are characterised by over-forecasting of the herd numbers for 1983, and is particularly true of the December forecast. Consequently, the unconditional forecasts over-forecast for the remainder of the in-sample period although they do pick up the seasonal movements in the series very well. Beyond the second quarter of 1984, the conditional forecasts are never worse than a 5.2% error, three of the forecasts being very accurate.

Figure 3.13.

a. The Conditional and Unconditional In-Sample Forecasts For The Unserved Gilt Herd Estimated On The Sample 1975:1 to 1985:4



b. The Conditional and Unconditional Out-Of-Sample Forecasts For 1986:1-87:4 Estimated On The Sample 1974:1-85:4.

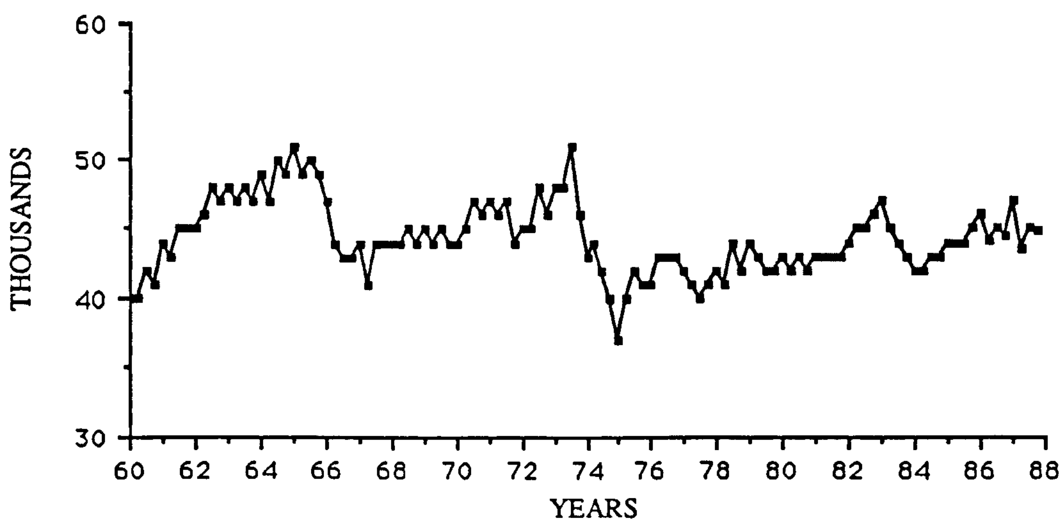


The MSFE statistics in table 3.8 indicate that the SARIMA model is far superior to the seasonal dummy model at unconditionally forecasting the in-sample period, although the equivalent statistic for the conditional forecasts implies that the forecasts from the dummy model are better than the SARIMA model forecasts, this result, however, is due solely to the relatively large over-forecast for the December figure in 1983 as referred to above. The UMSFE and the CMSFE for the out-of-sample period are calculated at 14.1 and 12.2 respectively, which compare very well with all of the in-sample equivalents. Both sets of forecasts pick up the seasonal movements in the unserved gilt herd very well.

3.4 A SARIMA Model For The U.K. Boar Herd

The data for the U.K. boar herd have been collected on exactly the same basis as that of the breeding sow herd except that the starting date for the available data is 1960:1. The forecast figure for 1979:1 was 44,000 boars, forecast using a SARIMA (2,1,0,1,2,0). Having learned from the building of the breeding sow herd models that the models built on the sample 1957:1-1981:4 were not going to be of explicit use and that the models estimated on the longer time period were better at forecasting, the model for the boar herd was estimated only for the sample period up to and including 1985:4.

Figure 3.14.
A Plot Of The U.K. Boar Herd 1960:1-1987:4.



The plot of the series gives the impression of a series with both seasonal and cyclical components. The herd experiences a relatively large fall in numbers in 1965 and around the time of the UK's entry into the EEC during the latter half of 1972 and 1973. Although the herd shows signs of increasing during 1982, the growth turns into a decline in 1983, partially a result of the eradication campaign of that year, after which the herd size is relatively stable.

Table 3.9.[13]
The Results of Estimation of the Boar Series On the Sample 1960:1 - 1985:4

MODEL									IN-SAMPLE		
d	D	p	P	q	Q	R.S.S.	Q_{20}	P_{20-k}	CMSFE	UMSFE	$\hat{R}_s^2$
0	1	2	1	2	0	179.8	16.2	37.1%	1.81	3.36	0.06

RESIDUAL AUTOCORRELATIONS.

LAG	1	2	3	4	5	6	7	8	9	10
AUTOCORRELATIONS	.02	-.01	.00	-.02	.12	-.04	-.04	-.05	.14	-.01
LAG	11	12	13	14	15	16	17	18	19	20
AUTOCORRELATIONS	-.00	-.05	.02	-.06	-.03	-.29	-.04	.03	-.10	-.03

RESULTS OF THE ESTIMATION OF THE SEASONAL DUMMY MODEL ON FIRST DIFFERENCES.

SEASONAL DUMMY							DUMMY VARIABLE				IN-SAMPLE	
d	D	p	P	q	Q	R.S.S.	1	2	3	4	CMSFE	UMSFE
1	0	0	0	0	0	200.0	0.48	-.42	1.0	-0.85	1.21	10.9
							(1.7)	(-1.5)	(3.6)	(-3.0)		

Table 3.9 gives the results of having estimated both a SARIMA model, and a seasonal dummy model on first differences, for the period 1960:1-1985:4. Only the third and the fourth seasonal dummy parameters of the Harvey model are significant at the 5% level, and the resultant RSS statistic is 200.0.

Although attempts at model estimation were made using the non-differenced data, the estimated coefficients implied non-stationary models and, therefore, as was the case with each of the live pig categories studied, it was the autocorrelations of the seasonally differenced series which indicated stationarity. The identified model consisted of 2 AR, 1 SAR and 2 MA parameters, all of which were highly significant and the estimated parameters of the Autoregressive terms implied a cycle of 3 years and 4 months. With the exception of lag 16, each of the residual autocorrelations is small, producing a Q-statistic of 16.2 for the first 20 lags, producing an associated P-value of 0.371 implying that the residuals as a whole are consistent with white noise. The RSS of the SARIMA model is 179.8 which, when compared with the equivalent statistic of the seasonal dummy model gives an $\hat{R}_s^2$

13. See Footnote 5

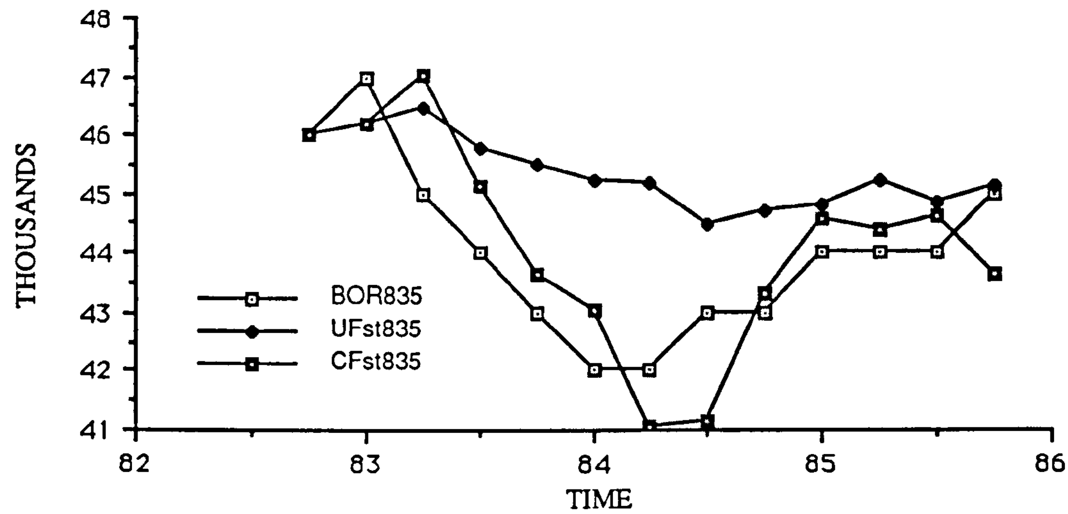
value of 0.06. Although the SARIMA model is better than the seasonal dummy model in terms of the in-sample UMSFE, the UMSFE statistics imply a slight superiority to the Harvey model. The estimated Box-Jenkins model, along with the t-statistics of the estimated coefficients, is given in equation 3.4.1.

$$\underset{(-13.6)}{(1 - 1.55B} + \underset{(7.7)}{0.76B^2})(1 + \underset{(3.6)}{0.43B^4})(1 - B^4)\,B_t = (1 - \underset{(-6.9)}{0.87\,B} + \underset{(6.4)}{0.60B^2})\,e_t. \qquad (3.4.1)$$

The 1 step conditional and the unconditional forecasts for both the in-sample and out-of sample periods 1983:1-85:4 and 1986:1-87:4 respectively, are presented in figure 3.15.

Figure 3.15.

a. The Conditional and Unconditional In-Sample Forecasts For The Boar Herd Estimated On The Sample 1960:1 to 1985:4



b. The Conditional and Unconditional Out-Of-Sample Forecasts For 1986:1-1987:4 Estimated On The Sample 1960:1-85:4.

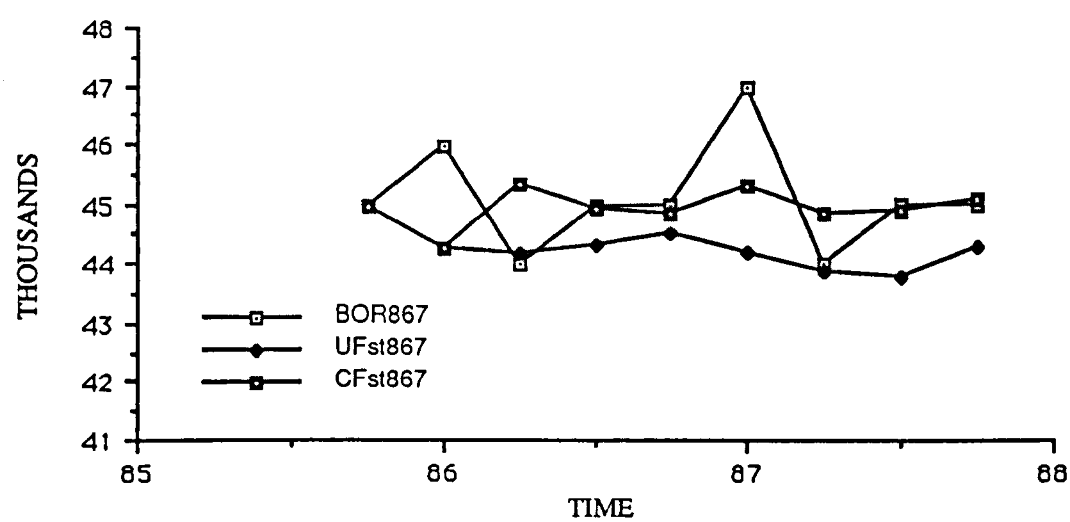


Figure 3.15a illustrates a boar herd which decreases up to 1984 then recovers over the latter half of the period. Not surprisingly, the unconditional forecasts are too high for the period affected by the Aujezky eradication campaign, although they forecast 1985 relatively well. The story is a similar one for the in-sample one-step forecasts except that the errors are much smaller as is expected. It appears to be the case that the model is a little slow in forecasting the turning points in the actual series over the given period. As for the out-of-sample forecasts, the UMSFE and CMSFE measure 1.13 and 1.06 respectively which reflect the improvement in the model's ability to forecast the out-of-sample period, the obvious exception being the forecast for the April figure of 1987. As was the case with many of the the breeding sow herd series, the reliability of the census figure for April 1987 looks suspect as the plot in figure 3.15b illustrates. The high April figure is the reason for the the 1 step under-forecast of the April 87 figure, and furthermore, the one-step over-forecast of the following June figure. There is an obvious tendency for the 8-step unconditional forecast to under-forecast the out-of-sample period.

3.5 Overcoming the Problem of Suspect Sample Data

To round off this discussion of forecasting the breeding herd using time series models it was thought appropriate to examine the problem of suspect sample data which appears to be a recurrent problem in April 1987 for the breeding herd models examined. The nature of the problem is that the sample data for the breeding sow and boar components appear to augment the size of the individual and, therefore, the aggregate herds. The immediate consequence of this is that the models appear to badly under-forecast the herd sizes for the said sample census. A second, and more important problem from the point of view of forecasting, is that the one-step forecast for sample points following the suspect period is clearly affected, as will be a number of forecasts into the future depending on the number of autoregressive and moving-average terms in the identified model. These problems are true for many of the models studied in this chapter, and no more so than in the case of the first model concerned with forecasting the size of the total breeding sow herd.

The aim of this section is to describe a method for overcoming the problems created by the presence of a suspect value in the forecasting period, using the example of the total breeding sow model discussed in section 3.2a. The specific nature of the problem in the breeding sow model is that the model, as presented in equation 3.2.2, contains 2AR, 2SAR and 1 MA parameters, together with a seasonal difference. Consequently, the forecasts for up to 14 periods ahead of the suspect sample data period are derived using the suspect datum itself through one of the AR terms. The worst effect in the breeding sow model, as it was in the case of the other breeding herd models, is that the forecast for June 1987 is considerably above the actual census figure of 813,000. The reason for this is not only the fact that the April census figure of 851,000 gives a high base value from which the June figure is forecasted, but the large positive error of the April forecast causes a further increase in the forecast for June through the MA term present in the identified model. The presence of the MA terms means that all such forecasting errors will have a bearing on all future one-step forecasts. This effect is, of course, in addition to the effect of the suspect April figure itself through the AR terms.

Faced with such a situation, the forecaster may wish to consider the following actions in order to forecast the period beyond the suspect value. Firstly, he could use the one-step forecast for April 1987, and substitute this forecast value for the actual suspect value. This action has two main consequences. The first is that the AR effects of the suspect value on the forecasts for up to 14 steps ahead has been revised, hopefully to a value which is itself consistent with the rest of the data, and secondly, the April forecast error has been eliminated, therefore, removing the effect on future one-step forecasts through the MA term. If the forecaster was interested only in forecasting from a period prior to the suspect period, then he may consider an n-step ahead unconditional forecast, however, the latter is obviously of limited use. A two-step forecast for June 1987 with December 1986 as the base point using the latter method will yield the same forecast as the more general method of replacing the suspect datum and forecasting one-step ahead.

The general method was employed in order to revise the one-step forecasts for the period beyond April 1987 so that these revised forecasts could be compared with the results of the original forecasts presented in figure 3.3b. The results of the revised forecasts are presented along with the original forecasts and the actual data in table 3.10 below.

Table 3.10

The Original and Revised Forecasts for June, August and December 1987

Time	Breeding Sows	Original	Revised
1987:2	813	870.7	833.8
1987:3	810	797.8	803.4
1987:4	822	813.2	815.9

The one-step forecast for April 1987 is 820.9 thousand, which when substituted for the actual dubious value for April yields a one-step forecast for June of 833.8

thousand, that is, 20.8 thousand above the true census figure. This compares with the original forecast of 870.7 thousand; an over-forecast of 57.7 thousand. The revised forecasts for August and December also provide an improvement to the original forecasted values giving further support to the action of replacing the suspect April figure.

Although it would appear that the revision method has led to an improvement in the three forecasts after the suspect period, the above analysis is obviously dependent upon the forecaster's knowledge that the April figure is indeed suspect and needs to be replaced. More than likely, this knowledge is only going to become available in the light of future data - in this case the figure from the June census - and as such, is very much an ex-post method. Having said this, there is nothing to stop the forecaster using the method if the forecast error for his latest available figure is out by an amount which he regards as unacceptable. He is then free to compare forecasts of points beyond the latest figure and choose which he feels is likely to be the more reliable. In the light of subsequent data he may of course reverse his original decision.

Although the method of revision discussed above has been with reference to a particular example of a univariate model, the reader will hopefully appreciate that the method can easily be generalised in order to apply it to other models, univariate or multivariate.

3.6 Conclusion

In this chapter, univariate statistical models have been built for the key components of the breeding herd using methodology advocated by Box and Jenkins as outlined in chapter two. Such models are of interest in that they are expected to be particularly useful for forecasting in the short run and have the advantage over more traditional econometric methods that they require only the data for the variable of interest to be collected and require no prior knowledge of the variable concerned in order to build a workable forecasting model. The disadvantages include the fact that large quantities of uniformally spaced time series are required if the methodology is to be strictly adhered to.

The data available indeed proved to be somewhat problematic and much of the discussion in the chapter was concerned with the nature of the data problems and how they were overcome. The missing data problem for April 1979 was overcome by building initial forecasting models using the data available up to the end of 1978 and making a one-step forecast for April 1979. The second major problem with the breeding herd data was the apparent stabilising of the component variables after about 1974, presumed to be a consequence of the UK's accession to the EEC, this also being the reason for the change in the timings of the spring and autumn sample censuses. The latter problem meant that data were no longer quarterly in the true sense of the term after 1974, and the 'rules' of the methodology were broken to some extent as they were continued to be treated as quarterly. Chow tests were used to confirm the notion that there were structural changes in the breeding herd component and aggregate data post 1974, necessitating a re-identification and estimation of the models on the post 1974 sample period. Interestingly enough, despite the changes in the post 1974 period including the stabilisation, the census timing changes and the obvious effects of the 1983 Aujezky disease eradication campaign, it is the models built on the 1957 to 1985 sample space which were deemed to be better at forecasting the in-sample and out-of-sample periods. Whether this situation continues in the future when more data become available which will benefit the identification and estimation of the post 1974 models, diminishing the effects of the 1983 eradication campaign, remains to be seen. Given that the larger sample models are the best for the period analysed, it is they which shall be used in the forecasting analysis of chapter eight.

Although no prior knowledge of the variables to be modelled is necessary to build Box-Jenkins models, any knowledge available can be made taken into consideration at the identification stage. The historical knowledge of a pig cycle was made use of by encouraging the use of second order AR terms where appropriate which might yield cycles through having imaginary roots. All but the unserved gilt model - which was estimated over a shorter period due to non-collation of the series prior to 1974 - contained cycles, but of variable lengths, ranging from 6 years and 10 months in the case of the total breeding sow model to 3 years and 3 months for pregnant gilts, with evidence of sensitivity to the values of the estimated polynomial coefficients. All the models were estimated after taking a seasonal difference of the raw data to achieve stationarity, and all identifications included both AR and MA terms of various orders. The fact that there were only four observations per annum made identification quite difficult, in that it was often hard to separate seasonal and non-seasonal effects from one another. The estimation of the seasonal dummy models on the first differenced data as suggested by Harvey illustrated the presence of seasonality in the component series, and although the computed $\hat{R}_s^2$ statistic often implied that the SARIMA models were not much better, if at all, at modelling the variables concerned over the given period, the SARIMA models were superior when it came to forecasting.

The chapter is completed with a discussion of how the Box-Jenkins models could be used to overcome the final data problem concerned with the problem of suspect sample data in April 1987. The analysis indicated how the suspect sample data can affect the forecasting ability of the Box-Jenkins models, and how adjustment of the data, using forecasts from the derived models, can improve the forecasts beyond the observation in question. The full implications of the April 1987 data will be discussed in the forecasting chapter along with the analysis concerned with investigating the relative forecasting abilities of the total breeding sow univariate model and the aggregate forecasts from the three breeding sow herd component models.

In the following two chapters comparative forecasting models for the breeding herd will be built using a biological approach and an econometric approach. In chapter six, the methodology employed to build the univariate quarterly models of this chapter will be applied to the monthly time series for culling and fat pig slaughter and various price and profit series.

CHAPTER FOUR

A BIOLOGICAL MODEL OF THE UK BREEDING HERD

4.1 INTRODUCTION

The biological model comprises a system of equations based on the biological relationships within the breeding sow herd and factors such as the number of days from birth to slaughter of fat pigs. The main reason for building the biological model is to derive a forecasting model for the breeding herd and the two slaughter categories, 'sows and boar cullings', (M), and fat pig slaughter, (FP). Whereas the time series models of the breeding herd presented in chapter 3 are expected to be of greater use in the short term, it is anticipated that models based upon the biological relationships within the UK pig herd will yield more useful forecasts for the medium term.[1] The biological analysis will also help our understanding of the system in which the producers operate and should lead to a greater awareness of the implications of policy on the industry.

The first section of the chapter is concerned with a theoretical approach to an overall model for the breeding herd, modelling it as a system of inflows and outflows under conditions of an assumed steady state equilibrium, (S.S.E.). Having shown how the steady state equilibrium model can be used to build a recursive forecasting model, the results of estimating the relationships for the biological model are presented in section 4.4. The equations are initially estimated using ordinary least squares, (O.L.S.), regression techniques, other techniques being used and modifications being made as thought necessary. The data are four monthly - hereafter referred to as trimestic - for the live pig categories, the source being the three sample censuses carried out by M.A.F.F. at the beginning of April, August and December since 1974. The slaughter data, on the other hand, are monthly. The models are estimated on the sample period 1975-1985 inclusive, in order to allow for a period of adjustment for the sector following the entry of the U.K. into the E.E.C., and to avoid the effects on the sector produced by the abnormal behaviour of the world's agricultural markets in 1974; a result of the sharp increases in commodity prices of that year. The inclusion of the 1974 data had a noticeable effect on the parameter estimates of the models due to the relative shortness of the sample period, thereby justifying their exclusion. Indeed, the size of the first few residuals in a number of the models justifies the need for dummy variables in order to remove any influence on seasonal or time trend parameters.

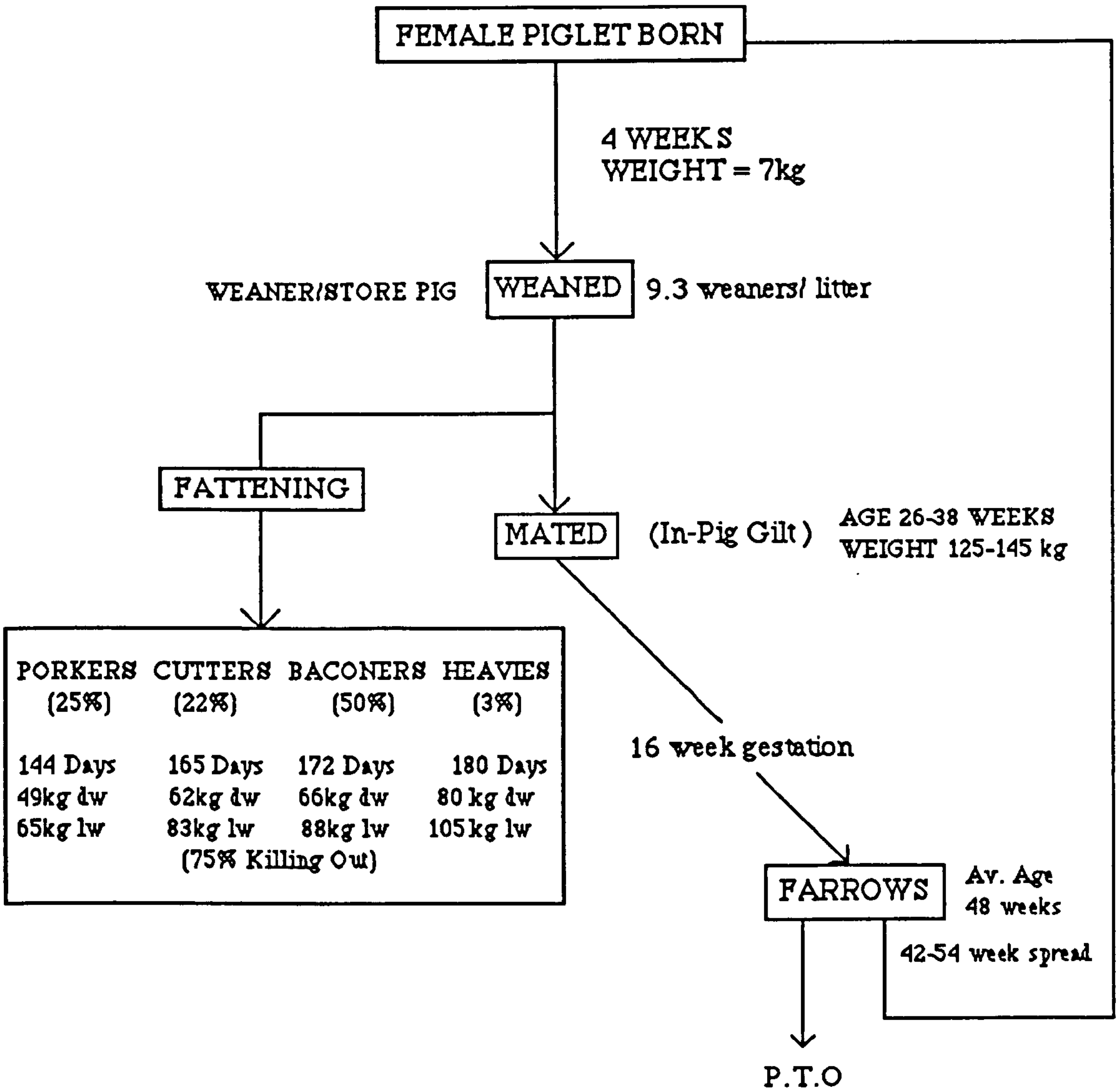
---

1. These expectations of the usefulness of biological models for forecasting in the medium term are partly a result of work comparing the ability of statistical, economic and biological models in forecasting key variables of the English and Welsh dairy sector discussed in:- Rayner, A.J. and Young, R. J., " Information, Hierarchical Model Structures and Forecasting." *European Review of Agricultural Economics*, No. 7, 1980.

4.2 The Biological System of the Pig

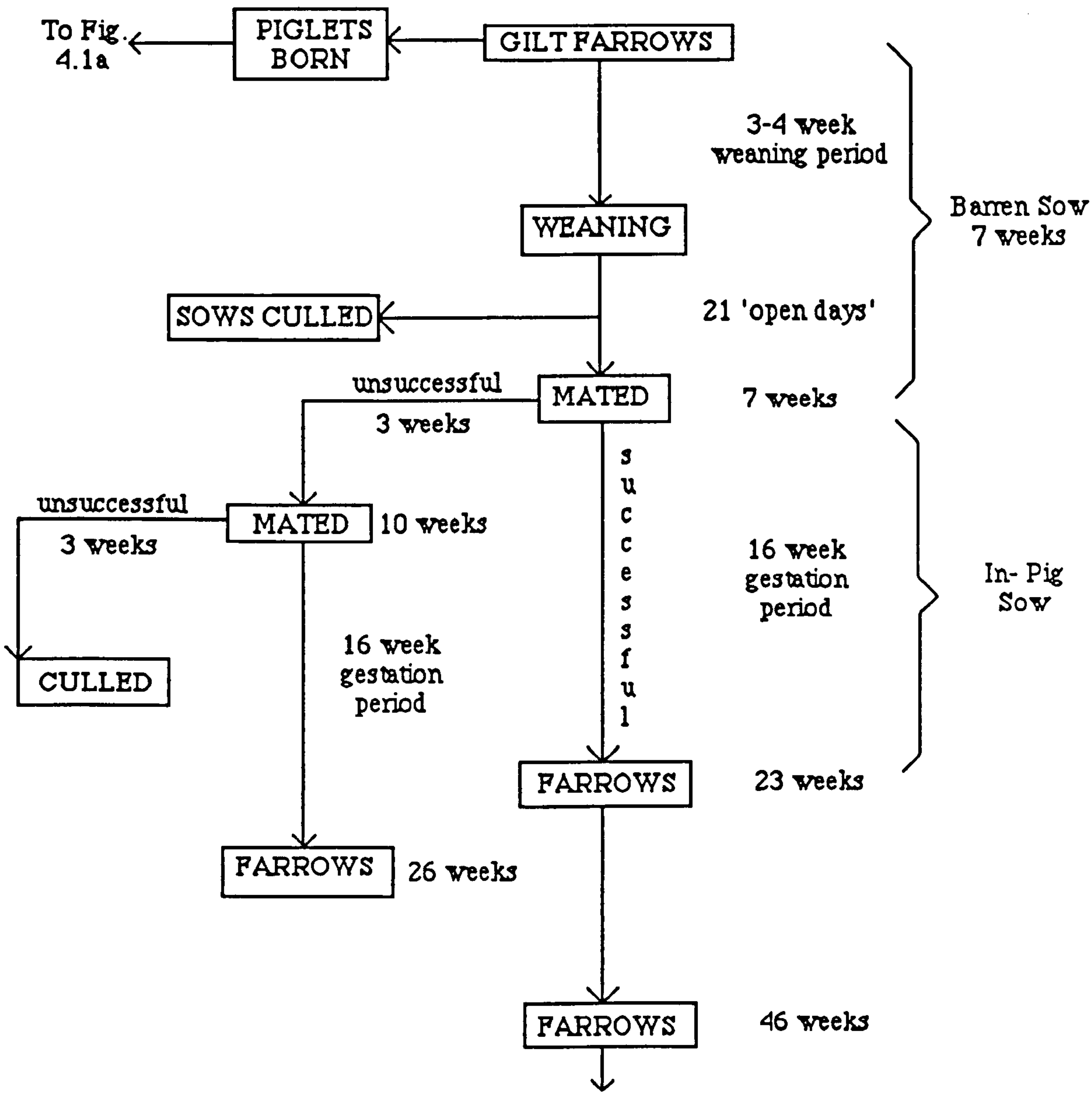
In order to build the biological model, it was necessary to make certain assumptions and generalisations concerning the nature of the biological lags within the system.

Figure 4.1a:- The Biological System For The Rearing of Pigs.



SOURCES:- Cambridge Survey (various)
Exeter Survey (various)
Pig Improvement Company
M.A.F.F.
M.L.C.
M.L.C. Pig Year Book (various)

Figure 4.1b:- The Biological system within the Sow Herd.



The above implies 2 litters in 46 weeks.

Therefore implies 2.26 litters in 52 weeks.

Theseassumptions and generalisations are illustrated in the flow diagrams of figures 4.1a and 4.1b and are outlined below. Starting from the birth of a female piglet, the piglet will be weaned at approximately 4 weeks, at which point the store pig - those pigs recorded as weighing 0-20kg (recorded as 0 to 2 months of age prior to 1974) - will be fattened for fat pig slaughter ready for the meat market. Fat pigs fall in to one of four categories; 'porkers', 'cutters', 'baconers' and to a lesser extent 'heavy hogs'. Figure 4.1a illustrates the approximate percentage of pigs going into each category along with the number of days from birth to slaughter and the average dead weights for each category. The associated live weights have been derived using the formula employed by the M.L.C.[2] A proportion of the young females are not fattened but are retained for entry into the breeding herd. From the weight of 50kg, such a female is recorded as a 'gilt for breeding not yet in pig' and will be referred to as an unserved gilt, (UG). The approximate average weight of the gilt at first service is 125kg to 145kg at which stage the gilt will be about 28 to 34 weeks old, (6.5 - 8 months old). Assuming the service to be a successful one, the gilt will then remain in pig for an approximate 16 week gestation period, so that the average gilt will farrow between the ages of 42 and 54 weeks of age.

Having farrowed, the gilt will then be recorded as a 'barren sow for breeding' for approximately 7 weeks. Four of these will be spent weaning the litter, followed by an average 21 day anaestrone or 'open day' period, after which she will be served again. The cycle is then repeated so that having farrowed once, a sow can expect to farrow a further two litters in the space of 46 weeks. The latter implies 2.26 litters will be born per sow per annum, (52 weeks). With an average of 9.3 weaners per litter, 21.02 weaners per sow per annum can be expected. If a service is unsuccessful, a sow can expect to be served a second time, although a further failure is likely to result in the sow being culled.

4.3 A Steady State Equilibrium Model For The U.K. Breeding Herd

The steady state equilibrium, (S.S.E.), model expresses the breeding herd as a system of inflows and outflows in which it is assumed that the herd is in a state of equilibrium. The herd is thus treated as a capital flow system in which there is investment in the form of inflow into the breeding herd and scrapping, (outflow), in the form of culling. The steady state assumption means that the inflow and outflow variables are treated in such a way as to prevent the herd from increasing or decreasing consistently over time. Seasonality is temporarily ignored. The total breeding herd in period t is defined in 4.3.1 as the aggregate of the breeding sow herd and the boar herd.[3]

---

2. See MLC's Pig Year Book 1984 p.49.

3. Throughout the text t refers to time as represented by the model, so that it represents 4 month periods in the semestic models and 1 month periods in the monthly models.

$$HB_t \equiv H_t + B_t \tag{4.3.1}$$

Considering the breeding sow herd component individually, equation 4.3.2 defines it as the aggregate of pregnant sows, (PS), pregnant gilts, (PG), and barren sows for breeding, (BS).

$$H_t \equiv PS_t + PG_t + BS_t \tag{4.3.2}$$

Equation 4.3.3 represents an identity for the breeding sow herd, expressing it as a system of inflows and outflows.

$$H_t \equiv H_{t-1} + IG_{t-1,t} - MS_{t-1,t} - LS_{t-1,t} \tag{4.3.3}$$

Thus, the breeding sow herd is defined as comprising the sow herd in the previous time period, plus the inflow of gilts which have become pregnant between t-1 and t, ($IG_{t-1,t}$), minus the number of sows culled from t-1 to t, ($MS_{t-1,t}$), and less the number of sows lost through disease, injury, etc. between t-1 and t, ($LS_{t-1,t}$), which are not recorded as cullings. The problem with 4.3.3 as it stands is that none of the variables, with the exception of breeding sows, are recorded as distinct categories in the census and slaughter data and are, therefore, unobserved. Having said this, it is possible to derive an estimate of the gilt inflow between t-1 and t figure by calculating 17/16ths of the pregnant gilt figure at time t.

$$\hat{IG}_{t-1,t} = 17/16\ PG_t \tag{4.3.4}$$

The reasoning for this relationship is that there are approximately 17 weeks between census timings but only 16 weeks in the gestation period, hence, 1/17 th of the true gilt inflow figure from t-1 to t miss being recorded as pregnant gilts at time t, and are recorded instead as barren sows. This occurs because an average of 1/17 th of the pigs recorded as unserved gilts at t-1 will conceive in the week immediately following the census at t-1 and will, therefore, farrow in the week immediately preceding the census at time t.

As sow cullings and losses are not directly observed, it is more appropriate to re-define the problem in terms of the total breeding herd, that is, the 'breeding sows' plus 'boars for service'. The consequences for the sow herd can then be derived utilising the relationship between the boar herd and the sow herd as expressed in equations 4.3.5 and 4.3.6.

$$B_t = \alpha H_t \tag{4.3.5}$$

and therefore,

$$H_t = 1 / (1+ \alpha) HB_t \tag{4.3.6}$$

The identity in equation 4.3.7 is an expression for the total breeding herd equivalent to that for the sow herd given in 4.3.3, where M is the observed 'culling of sows and boars', I, the actual inflow of sows and boars, and L, the losses of sows and boars from the breeding herd.

$$HB_t \equiv HB_{t-1} + I_{t-1,t} - M_{t-1,t} - L_{t-1,t} \tag{4.3.7}$$

The problem now presented is that 'losses' and total inflow of gilts and boars are unobserved and as a consequence, the inflow figure - estimated by $\hat{I}_{t-1,t}$ using expression 4.3.8, which is itself obtained by rearranging 4.3.7 - captures the unobserved losses and the unobserved inflow of boars.

$$\hat{I}_{t-1,t} \equiv I_{t-1,t} - L_{t-1,t} \equiv HB_t - HB_{t-1} + M_{t-1,t} \tag{4.3.8}$$

Because of the variability of the estimated inflow series, and because of the fact that it is the aggregate of three unobserved components, inflow will be proxied by a transformation of the pregnant gilt series, based on the relationship described in equation 4.3.4. If it is assumed that the length of breeding herd life of sows and boars is the same, referring to 4.3.5, the unobserved inflow of boars can be estimated by $\alpha \times IG_t$. The consequence of this is that the inflow of sows and boars can be estimated by the expression given in equation 4.3.9, and can then be substituted into 4.3.7 in order to derive the breeding herd at time t.

$$\hat{I}_{t-1,t} = 17/16 (1 + \alpha) PG_t \tag{4.3.9}$$

Having defined the primary variables, and having distinguished between the observed and the unobserved variables, a recursive forecasting model for the breeding herd can now be built by consideration of the biological relationships for the inflow and outflow variables. Under the steady state equilibrium -SSE - assumption, the outflow variable at time t can be regarded as a constant proportion, θ, of the breeding herd at time t-1.

$$M_{t-1,t} = \theta HB_{t-1} \tag{4.3.10}$$

From this it can be shown that:-

$$\theta = 1/L \tag{4.3.11}$$

where L is the average lifetime of sows and boars in the breeding herd under the

SSE assumption.

On the inflow side of the equation, the estimated inflow is expressed as a function of the herd size lagged 12 months - three periods. The reason for this lag is that unserved gilts are approximately 8 months old by the time they are first served. Added to the four months spent in pig, a lag of one year can be expected from the time of a sow being in pig to the time that one of her offspring will themselves become in-pig gilts.

$$I_{t-1,t} = \phi\, HB_{t-3} \tag{4.3.12}$$

Under SSE conditions, ϕ is assumed to be equal to θ, otherwise the herd size will either increase or decrease consistently over time.

Using 4.3.10 and 4.3.12, forecasts of the breeding herd for any required lead time can be made. For example, a one step ahead forecast for the total breeding herd, $HB_{t-1}{}^{*}$, is obtained by employing expression 4.3.13.

$$HB^{*}_{t+1} = (1 - \theta)\, HB_t + \phi\, HB_{t-2} \tag{4.3.13}$$

Alternatively, a one year - three period - ahead forecast can be generated using expression 4.3.14.

$$HB^{*}_{t+3} = \{(1-\theta)^3 + \phi\}\, HB_t + (1-\theta)\phi\, HB_{t-1} + (1-\theta)^2\phi\, HB_{t-2} \tag{4.3.14}$$

Having shown how a recursive forecasting model can be built based on the steady state equilibrium framework, biologically based relationships will be estimated introducing phenomena such as seasonality in order to build trimestic and monthly forecasting models of the breeding herd, cullings and fat pig slaughterings. The models will also make a correction for autocorrelation in the residuals.

4.4 The Methodology of Estimation for the Trimestic Models

The biological models developed in this chapter are built in order to forecast the trimestic breeding herd and the monthly culling and fat pig slaughter categories, and to explain some of the biological relationships within the breeding herd system. In terms of the recursive breeding herd forecasting model, explained within a steady state equilibrium framework, in the previous section, the prime relationships to model will be those between the inflow and its proxy variable, pregnant gilts, and between culling and the breeding herd lagged 1 period and pregnant gilts and the breeding herd lagged 3 periods. Having said this, other models will be built for these variables for comparison and interest purposes. For the two monthly slaughter categories, models based upon trimestic models for the same variables will be developed.

The models are concerned with the breeding sow herd, (H), and its components, the boar herd, (B), unserved gilts between 50kg and 80kg, (UG), culled sows and boars, (M), and fat pig slaughter, (FP). The models for which the dependent parameter is a livepig category are trimestic models based on the data collected from the sample censuses in April, August and December. The data for the two slaughter categories are monthly data and, therefore, aggregation to four-monthly periods is required for use in the trimestic models. The slaughter data are based on the weekly estimates of slaughterings by M.A.F.F., and are collected in such a way that weekly periods are aggregated to arrive at the published monthly figures. The weekly estimates are aggregated in such a way that the data for January, April, July and October represent five, rather than four week periods of accounting. Furthermore, there is the additional problem that in certain '53 week' years December is also counted as a five week period. For the sample period 1975:1 to 1985:12, on which the models are built, the latter is true for the Decembers of 1976 and 1981. All the data used in the estimation of the trimestic models are presented in Appendix 4a.

The trimestic flow periods are defined so that 'period 1' comprises December to March inclusive, 'period 2', April to July inclusive and 'period 3', August to November inclusive. Because of the four and five week accounting months , period 2 is always a 17 week period, period 3 an 18 week period and period 1 will be either a 17 or an 18 week period; depending on whether or not the accounting year is 52 or 53 weeks long respectively. In order that all three periods represent 17.33 weeks - the average length of time between each of the April, August and December census dates - the trimestic totals for the culling and the fat pig slaughter series are adjusted by multiplying by 52/51 and 52/54 for 17 and 18 week periods respectively.

Although the models presented in this chapter are called biological, the models estimated are more sophisticated than the term biological might infer. Because the models are built primarily in a forecasting context, the decision was made to include in the regressions all non-economic factors which might affect the simple biological relationships. Thus, factors such as seasonality and time trends are modelled, as are 'shocks' to the system, such as the influence of the Aujezky disease eradication campaign of 1983, and the possibility of outliers at the beginning of the estimation sample period. The latter are a possible 'carry-over' effect of the UK's accession to the EEC and the influence of commodity price increases in 1974 upon the agricultural sector.

The models are estimated initially using OLS methodology, modelling seasonality, Aujezky and early outliers using intervention dummy variables.[4] The presence of a time trend is modelled using a simple linear time trend variable, T, which enters the model as a multiplicative term given that the models are proportional and do not contain an intercept. Consequently, the model parameters are estimated using non-linear least squares methodology - LSQ. Where the Durbin-Watson statistic indicates the presence of residual autocorrelation, the residuals are put through the Box-Jenkins identification procedure to analyse the nature of the autocorrelation.[5] Where first order autocorrelation is indicated, the Beech-Mackinnon Maximum likelihood estimation procedure for first-order autocorrelation correction is employed to estimate the model.[6] Where the initial model has been estimated using LSQ methodology, the first order autocorrelation problem is corrected by the use of a first order rho-transformation of the model, again estimating using LSQ. Where the autocorrelation problem is of an order greater than one, the correction is made using the appropriate rho-transformation, estimating using LSQ methodology. In the case of regressions including the dependent variable lagged one period, the DW statistic is replaced by Durbin's h-statistic, a large sample statistic asymptotically distributed as a N(0,1) random variable under the null hypothesis that the autocorrelation coefficient, ρ, equals zero.[7] Although the critical values of the DW and h-statistics are calculated on the assumption that the models are estimated using OLS they are nevertheless used as an indicator of the absence, or otherwise, of first order residual autocorrelation.

Although it has been said that economic variables were not to be included in the models, a common feature of many of the relationships modelled is that they are affected by a government pigmeat subsidy of 5.5p per kg deadweight, available from Jan 31 to June 11 of 1977. In addition, the residual plots of a few models indicated the presence of other outlier observations for which no explanation could be found. Being a biological model, it was deemed inappropriate to present the results of regressions including such subsidy and outlier dummies in the main text. But because the exclusion of these effects often has a significant effect on the magnitude, and even the signs of seasonal and time trend parameters, the

---

4. Notation for the dummy variables used in the semestic models is by the letter D - or A in the case of the Aujezky dummies - followed by the value for the time period in which the dummy is used. Thus, for example, a dummy representing the first period in 1975 is labeled D75:1.

5. The DW statistics throughout this chapter are compared with Farebrother's tables of significance levels presented in his paper; "The Durbin-Watson Test for Serial Correlation When There is No Intercept in The Regression", *Econometrica* Vol. 48 No. 6 Sept 1980 pp. 1553ff

6. Beech, Charles M and James G. Mackinnon, " A Maximum Likelihood Procedure for Regression Containing Autocorrelated Errors", *Econometrica* 46, 1978, pp.32-61.

7. For reference to the Durbin H-test see *Pindyck & Rubinfeld* pp. 194-5

decision was taken to allow for these factors in some way. In order to do this, all the models were initially estimated including dummies to represent the subsidy and outlier effects, thus enabling an unbiased analysis of whether or not the seasonal and time trend variables should be included in the models. The decision as to whether or not variables should be included in the final model is made using the t-statistics of the estimated parameters concerned or, where the analysis of t-statistics did not produce a clear picture, the analysis of variance F-test is employed.

The resultant trimestic models, including the subsidy and outlier dummies are presented in Appendix 4c[8]. The estimated regressions having excluded these dummies are then presented in the text as the chosen 'biological' model for each particular relationship. Section 4.5 gives a comprehensive discussion of how models for each of the individual relationships were built.

4.5 The Trimestic Models Estimated

The following sub-sections discuss the modelling of the various components of the breeding herd necessary to achieve an overall forecasting model for the breeding herd itself. The section is concluded with a model for the slaughter of fat pigs.

4.5a The Boar Herd and the Breeding Sow Herd

Model 4.5a is concerned with the relationship between the number of boars for service, (B), and the breeding sow herd, (H), at time t as expressed in equation 4.5a.1 in which e_t is a white noise error term.

$$B_t = \alpha H_t + e_t \tag{4.5a.1}$$

The use of this model was discussed in section 4.3. The initial OLS regression model indicates that the boar herd is the equivalent of 5.15% of the breeding sow herd. The residuals of the regression indicate a clear positive time trend which shows no signs of having ended, suggesting that the number of boars in the breeding herd as a percentage of the number of sows is increasing over time. The latter phenomenon is a possible consequence of the increased productivity of sows over the given period. In order to model this, a simple linear time trend variable was added to the model structure and the regression estimated using non-linear least squares, LSQ, methodology. To model potential seasonal effects, two seasonal intercept dummies were included representing the August and December censuses. The residuals from having estimated the latter regression

[8]. The equations presented in the appendices have the same number as the equivalent equation presented in the main text except that the number is suffixed with the letter c

suggested two sets of excluded variables. The first two residuals appeared to be somewhat out of line with the rest, hence, the decision was taken to include two intercept dummies for the observations for 1975:1 and 1975:2. This was done in order to remove any potential effect of their exclusion from the estimation procedure on the estimates of the seasonal dummy parameters, and is justified on the grounds that the sector may not have fully recovered from the effects of an unstable world market in 1974 and the UK's accession to the EEC. The second set of missing variables are dummy variables to model the effects of the Aujezky disease eradication campaign of 1983 which decreased the proportion of sows to boars. The results of estimating the regression having modelled all apparent effects on the boar-sow relationship are presented in equation 4.5a.2 below.

$$
\begin{aligned}
B_t = &\underset{(117.2)}{(0.0485\, H_t} - \underset{(-0.68)}{0.0002\, AugH_t} - \underset{(-2.81)}{0.0009\, DecH_t})(1 + \underset{(9.49)}{0.0036\, T}) - \underset{(-3.8)}{2.77\, D75{:}1} + \underset{(2.69)}{1.94\, D75{:}2} - \\
&\underset{(-0.38)}{0.27\, A83{:}2} + \underset{(2.07)}{1.45\, A83{:}3} + \underset{(0.80)}{0.56\, A84{:}1} \qquad (4.5a.2)
\end{aligned}
$$

Obs = 33; RSS = 10.4; $\hat{R}^2 = 0.87$; DW = 2.14

The results imply that the size of the boar herd is of the order of 5% of the breeding sow herd, although there is a relatively large fall in the proportion of boars in relation to breeding sows at the time of the December census. The time trend parameter and the two outlier dummies for the first two observations are all highly significant. Of the Aujezky dummies, only the second is significant and indicates that the sow herd was reduced by a larger relative percentage than was the boar herd by the eradication campaign of 1983.

The adjusted R-squared value indicates that 87% of variation in the size of the boar herd is explained by the regression. The Durbin-Watson statistic takes a value of 2.14 which confirms that the residuals are consistent with white noise.

4.5b Inflow

Model 4.5b relates the derived inflow figure for the four months preceding the census dates, to the number of pregnant gilts at time t. The rationale for looking at this relationship is that pregnant gilts are often used as a proxy for inflow in biological and econometric models of the pig breeding herd. In section 4.3 it was noted that the derived inflow figure not only included the inflow of pregnant gilts but also the inflow of boars for service and a negative element for losses from disease of sows and boars not recorded as cullings. The inflow estimate, ($\hat{I}_{t-1,t}$), is derived using equation 4.5b.1,

$$\hat{I}_{t-1,t} = HB_t - HB_{t-1} + M_{t-1,t} \tag{4.5b.1}$$

where $M_{t-1,t}$ is the adjusted culling of sows and boars figure for the period t-1 to t, so that if t represents April, $M_{t-1,t}$ is an estimate of cullings for the period between the December and the April census dates.

Initial OLS regression of inflow on pregnant gilts including none of the dummy variables produced a residual plot with highly variable residuals pre-1978. In an attempt to ease the modelling process, the decision was taken to model this relationship on the post 1977 period initially, later applying the resulting model structure to the normal 1975-85 estimation period. The result of estimation post 1977 including seasonal dummies produces residuals which gave no indication of excluded variables. The possibility of seasonality was entertained because of the implicit inclusion of losses and boar inflow in the estimated inflow variable. Although neither of the estimated seasonal dummy parameters were significant, because they were of opposite signs - August negative and December positive - an F-test was performed to test their overall significance.

The analysis of variance F-test is a standard procedure for testing the overall explanatory power of one or more regression variables. Using the case in question as an example, the null hypothesis states that the regression excluding the seasonal dummies is the best for explaining the given relationship. The alternative states that the regression including the seasonal dummies has the greater explanatory power. The explanatory power of the two regressions is measured by the ratio of their unexplained variances, or to be more precise, their sums of squares of residuals having taken account of the differing degrees of freedom of the two regressions. Thus, the F-ratio measures whether or not the RSS of the regression including the seasonal dummies is significantly lower than that of the regression excluding the said dummies. If the resultant F-statistic is significant when compared with the F-statistic tables for v_1, $n-v_1-k$ degrees of freedom, where n is the number of observations, k is the number of estimated parameters in the initial regression and v_1 is the difference in the number of estimated parameters in the two regressions in question, then the null hypothesis can be rejected and the additional variables remain in the regression model on the basis that they have significant statistical explanatory power.[9]

The results of the F-test for model 4.5b produced an F-statistic of 2.561, which compared with the F-tables for 2,21 degrees of freedom is not significant even at the 10% level. Consequently, the null hypothesis of no seasonality in the

9. For a more comprehensive discussion of the F-test the reader is refered to Gujarati, D., *Basic Econometrics*, (1978) p 87. & p. 130.

relationship could not be rejected and the seasonal dummies were removed from the estimation procedure. The results of the chosen model estimated on the post 1977 period are presented in equation 4.5b.2.

$$\hat{I}_{t-1,t} = \underset{(58.7)}{1.0957}\, PG_t \qquad (4.5b.2)$$

Obs = 24   RSS = 2199.2 $\hat{R}^2 = 0.35$ DW = 2.695

The estimated parameter suggests that estimated inflow can be approximated by the equivalent of approximately 110% of the pregnant gilt herd. Having said this only 35% of the relationship is explained by the regression. Re-estimating on the whole sample period produces the model given in equation 4.5b.3.

$$\hat{I}_{t-1,t} = \underset{(44.0)}{1.0863}\, PG_t \qquad (4.5b.3)$$

Obs = 33   RSS = 7549.0 $\hat{R}^2 = .0581$ DW = 2.22

The estimated parameter is slightly lower than that for the regression excluding the period prior to 1978 and the explanatory power of the model has fallen to a mere 6%, the greater variability of the relationship pre-1978 being very evident in the residual plot.

Using the results from model 4.5a which indicates that the boar herd is approximately 5% of the size of the breeding sow herd, equation 4.3.9 implies that a proxy for actual inflow of sows and boars, ($I_{t-1,t}$), is $1.116PG_t$, that is, 17/16 multiplied by 1.05 PG_t. When compared with this expected coefficient of 1.116 the results of models 4.5b.2 and 4.5b.3 appear to support the idea of using the pregnant gilt figure as a proxy for inflow, the deficit in the estimated parameters being explained by the inclusion of losses in the inflow estimate. The estimated parameter from equation 4.5b.2, the better fitting of the two models, suggests that losses in the breeding sow herd from t-1 to t can be accounted for by a figure approximating 2% of the pregnant gilt herd at time t.

4.5c The Breeding Sow Component Proportions

Like the models already presented in this chapter, the three models presented under model 4.5c are proportional in that they do not contain an intercept term and are static in that they do not concern a lagged independent variable. The models are concerned with the breeding sow herd and how it breaks down into its three component parts; pregnant sows, pregnant gilts and barren sows for breeding, that is:-

$$PS_t = f_a(H_t) \qquad (4.5c.1)$$
$$BS_t = f_b(H_t) \qquad (4.5c.2)$$
$$PG_t = f_c(H_t) \qquad (4.5c.3)$$

The relationships are estimated for descriptive purposes and to provide estimates of the future composition of the breeding herd. For example, having forecast the future breeding sow herd, the estimated models for the three expressions above could be utilised to provide an estimate of the composition of the sow herd of that same period. Were any of these components themselves independent variables in other models, these forecasts could then be used to forecasts other variables in the system.

Taking each component in turn, the first to be modelled is the pregnant sow percentage. The result of OLS estimation of the model containing only the seasonal dummies produced a residual plot which implied a positive time trend was missing from the model, although the plot also suggests that the trend has slowed, if not ended, from about 1983 onwards. To allow for the time trend, which is almost certainly a result of the shortening of weaning length over the sample period, the simple linear time trend variable was added to the regression which was then estimated using non-linear least squares - LSQ - regression. The residuals of the latter included a residual at observation 1976:1 which was more than three standard errors below the fitted line. The results of model estimation having included a dummy to remove the effects of this unexplained outlier on the seasonal dummy parameters is presented in appendix 4c. It was clear from re-estimation that there was no Aujezky effect and no 1977 subsidy influence on the relationship. The seasonal dummy parameters have associated t-ratios which are highly significant for December and high, though not quite significant at the 5% level, for August. The regression has an $\hat{R}^2$ value of 0.92 and a Durbin-Watson of 1.46. Although this DW is not significant at the 5% level when compared with the Farebrother tables, the residuals were put through the Box-Jenkins identification procedure. The result of the latter exercise was to produce correlograms with no obvious pattern to them, and no autocorrelations or partials much greater than one standard error away from zero.

The estimated equation and the diagnostics of the regression having excluded the dummy for the possible outlier are presented in equation 4.5c.4.

$$PS_t = \underset{(131.9)}{(0.5666\, H_t} + \underset{(2.13)}{0.0090\, AugH_t} + \underset{(4.54)}{0.0191\, DecH_t})(1 + \underset{(11.0)}{0.0039\, T}) \qquad (4.5c.4)$$

Obs. = 33 RSS = 2193.5 $\hat{R}^2$ = 0.88 DW = 1.39

The results show that all but the August dummy are significant at the 1% level. The removal of the outlier dummy has resulted in the August dummy parameter being significant at about the 5% level. Both seasonal dummies indicate an increase in the proportion of pregnant sows in August and December compared with the proportion in April. The linear time trend variable is positive and highly significant, illustrating the fact that the pregnant sow proportion of the breeding sow herd has increased over the given period, as a result of the shortening of the weaning period. The $\hat{R}^2$ figure indicates that 88% of the variation in the proportion is explained by equation 4.5c.4, and compares with a value of 92% for the regression including the outlier - presented in Appendix 4c.

The second of the breeding sow proportion models is that relating the percentage of barren sows for breeding to the breeding sow herd. As one would expect, the results are in many ways the opposite of the those obtained for the pregnant sow proportion. The residual plots of the initial OLS model illustrated the presence of a negative time trend which appears to have stopped around 1982, and there is evidence that the first observation in the sample period is rather large compared with the rest of the sample. As with the previous model, there is no evidence of any Aujezky or subsidy effect. To model each of the possible influences, the time trend along with the seasonal and the 1975:1 dummy are included in the initial non-linear least squares regression. The results of LSQ estimation are presented in equation 4.5c.5 below.

$$BS_t = (\underset{(90.2)}{0.2970}\, H_t - \underset{(-3.1)}{0.0102}\, AugH_t - \underset{(-5.0)}{0.0167}\, DecH_t)(1 - \underset{(-17.1)}{0.0068}\, T) + \underset{(3.3)}{20.34}\, D75{:}1 \qquad (4.5c.5)$$

Obs. = 33   RSS = 875.7 $\hat{R}^2 = 0.93$ DW = 1.62

It is clear that as the proportion of pregnant sows to breeding sows increases throughout the year, the proportion of barren sows for breeding moves in the opposite direction. The overall time trend is also in the opposite direction to that of the pregnant sow proportion, also caused by the shortening of the weaning period. The DW statistic lies in the region of uncertainty at both the 1% and 5% levels of significance. Putting the residuals through the Box-Jenkins identification procedure produced no significant autocorrelations or partial autocorrelations.

The final proportional relationship to be estimated is the pregnant gilt to breeding sow proportion. A priori, one expects the results for this model to be the complement of the previous two models. OLS estimation of the regression including the seasonal dummies produced residuals and a DW statistic which clearly indicated the presence of first order serial correlation. This was confirmed having put the residuals through the Box-Jenkins identification procedure. Because there was no time trend in the residuals, the regression could be estimated using a Beech-Mackinnon AR1 procedure. The initial estimation including the two seasonal dummies indicated that the proportion of the sow herd comprising pregnant gilts decreases significantly in December. The residuals indicated that observation 1975:1 was a possible outlier and so the appropriate intervention dummy was included in the regression along with three subsidy dummies and Aujezky dummy A83:2. The parameters on the subsidy dummies illustrate a significant fall in the proportion of pregnant gilts in 1977:2. This can be explained by the fact that many of the potential gilts for that period had been hived off into the feeding herd for slaughter at the time of the subsidy. It is interesting to note that the fall in gilt numbers as a result of the 1977 subsidy and the Aujezky disease eradication campaign are rectified by the next census, hence the lack of need for Aujezky dummies A83:3 and A84:1 and hence the insignificance of the dummy parameter for 1977:3 presented in Appendix 4c. These results illustrate that the size of the gilt herd can be changed by production decisions faster than the other breeding sow herd components for the reason that it is not constrained by the same biological lags. The parameter on the 1975:1 dummy is significant and negative and the December parameter's significance is increased by the inclusion of these other dummies. The regression presented in equation 4.5c.6 is the result of re-estimating the latter equation having dropped the subsidy dummies.

$$\underset{}{PG_t} = \underset{(38.0)}{0.1329\, H_t} + \underset{(0.01)}{0.00003\, AUGH_t} - \underset{(-2.53)}{0.0059\, DECH_t} - \underset{(-2.05)}{13.26\, D75{:}1} - \underset{(-1.23)}{6.82\, A83{:}2} + U_t$$

$$U_t = \underset{(4.21)}{0.6053\, U_{t-1}} \qquad (4.5c.6)$$

Obs. = 33 RSS = 1065.2 $\hat{R}^2 = 0.98$

The removal of the subsidy dummies has had the effect of changing the August dummy parameter from positive to negative. A possible explanation for the fall in the pregnant gilt percentage at the December census is the fact that more young pigs than normal are diverted into the feeding herd as opposed to the breeding herd prior to the Christmas period, thus depriving the December census of its quota of in-pig gilts. The DW statistic of 1.62 is not presented because the regression has been estimated using Beech-Mackinnon: however, there is no sign of any autoregressive problem in the plot of residuals, nor with the statistic itself. The exclusion of the three subsidy dummies in the above regression increases the value of the RSS statistic from 797.0 and the significance of the Aujezky dummy variable is much reduced.

It should be said that the three models presented in this section were all derived independently of one another which explains why, for example, the pregnant sow model considered a dummy for an outlier for the 1976:1 observation, a dummy not required by the other two models. Because the three models are proportional, one would expect an extreme movement in one direction in one of the components to be compensated by movements in the opposite direction in one or both of the other components. Obviously in the case of 1976:1, the compensating movements in the pregnant gilt and barren sow components were not large enough individually for them to appear as outliers in their respective residual plots. Were these three models important in the context of a breeding herd forecasting model it could be argued that it would be necessary to model the three components in a constrained manner, thereby producing an identical structure for each of the three models

A summary of the estimated proportions for three components of the breeding herd is presented in Table 4.3 below, using the estimated proportions for the end of the estimation period, at which point the time trends in the pregnant sow and pregnant pig models appear to have finished. The estimates of the proportions for pregnant sows and pregnant gilts have been obtained using the regressions which appear in appendix 4c: these regressions include outlier and subsidy dummies respectively, thereby removing these effects from the estimated seasonal parameters.

Table 4.3.
Seasonal Proportions for Each Component of the Breeding Sow Herd

HERD	APRIL	AUGUST	DECEMBER
PS	0.6376	0.6383	0.6564
BS	0.2343	0.2197	0.2139
PG	0.1331	0.1347	0.1270
TOTAL	1.0050	0.9927	0.9973

4.5d Culling and the Total Breeding Herd

Model 4.5d is concerned with the relationship between culled sows and boars and the total breeding herd at the previous census. Thus, for example, the cullings from April to July inclusive, are modelled as being dependent upon the size of the total breeding herd in April. This is the model which is included as the representative model for inflow in the recursive forecasting model for the trimestic breeding herd developed in section 4.3 and given in equation 4.3.10.

Assuming a weaning period of four weeks, few, if any, of the sows recorded as being pregnant in the census - taken at the beginning of April - will be culled in April itself. Consequently, a model was estimated relating the cullings in the relevant four months to the size of the breeding herd lagged both one and two periods. However, spurious parameters resulted, presumably the result of multicollinearity, and therefore the less satisfactory model based solely on the one period lag had to suffice.

An initial OLS regression including the seasonal, Aujezky and subsidy dummies produced significant parameters on the Aujezky and subsidy dummies, both of which indicated increases in cullings. The August dummy indicated a significant decrease in culling percentage with respect to period one. The residuals gave no indication of a time trend in the relationship. The outstanding problem with the regression is that the DW statistic of 1.26 although not significant at the 5% level, was thought low enough to justify an investigation of the residuals using the Box-Jenkins identification procedure. The latter process produced a correlogram in which there was evidence of first order autocorrelation.

To deal with this problem the relationship was re-estimated using the Beech-Mackinnon first order autocorrelation methodology. Estimation including all the usual dummies produced a regression in which the August seasonal dummy shows a significant drop in culling proportions and all six of the Aujezky and subsidy dummies indicate significant increases in culling. The RSS is 876.8 and there are no obvious problems with the residual plots. The regression results presented in equation 4.5d.1 are those from having estimated the latter equation after having dropped the subsidy dummies, the fuller model appearing in Appendix 4c. The exclusion of the subsidy dummies has substantially increased the value of the RSS statistic. The results indicate that approximately 13% of the breeding herd is culled between each census and - using expressions 4.3.9 and 4.3.10 - implies that the average life of a member of the breeding herd is approximately 2 years and 6 months.

$$M_{t-1,t} = \underset{(33.8)}{0.1327}\, HB_{t-1} - \underset{(-2.55)}{0.0069}\, AugHB_{t-1} + \underset{(0.24)}{0.0006}\, DecHB_{t-1}$$

$$+ \underset{(5.43)}{39.07}\, A83{:}2 + \underset{(3.07)}{24.55}\, A83{:}3 + \underset{(1.83)}{13.08}\, A84{:}1 + U_t$$

$$U_t = \underset{(4.22)}{0.62}\, U_{t-1} \qquad (4.5d.1)$$

Obs. = 32 RSS = 1285.0 $\hat{R}^2 = 0.98$

4.5e Culling and the Pregnant Sow Herd

Model 4.5e is similar to 4.5d except that culling is now estimated as a proportion of the number of pregnant sows at time t-1. The reason for investigating this relationship is that one would expect very few pregnant gilts to be slaughtered having farrowed once only, and by definition neither will barren sows for breeding be slaughtered, leaving pregnant sows and boars as the only plausible source of cullings in the next period. Although models including the boar herd as a second independent regressor lagged one period were estimated, the coefficient on boars was anomalous. Aggregating the pregnant sow and the boar herd and using the lagged variable as the independent regressor also proved fruitless in that the model did not perform as well as the model presented below in terms of RSS and adjusted R-squares and hence the chosen model does not contain the boar variable.

The regression was estimated using the Beech-Mackinnon procedure which produced the same structure of model as that in model 4.5d. Again, the three subsidy dummies indicated large increases in cullings in 1977, and the three Aujezky dummies showed a similar effect for 1983/4. The difference between this and the previous model is that the seasonal dummies were no longer significant. Because the parameters on the seasonal dummies were of the opposite signs, an F-test was performed before the decision was taken to remove them. The resulting F-value of 2.42 was not significant even at the 10% level and so the regression was re-estimated having removed the seasonal dummies. The resulting model and the diagnostic checks are given in appendix 4c, the regression having removed the three subsidy dummies being presented in equation 4.5e.1.

$$M_{t-1,t} = \underset{(31.7)}{0.2240}\, PS_{t-1} + \underset{(4.64)}{35.24}\, A83{:}2 + \underset{(2.98)}{25.42}\, A83{:}3 + \underset{(1.85)}{14.00}\, A84{:}1 + U_t \qquad (4.5e.1)$$

$$U_t = \underset{(4.59)}{0.6373}\, U_{t-1}$$

Obs. = 32 RSS = 1656.5 $\hat{R}^2 = 0.972$

As expected, the removal of the subsidy dummies has increased the RSS of the estimated regression and decreased the adjusted R-square value. The H-statistic of -0.098 implies almost no autocorrelation in the residuals; a very different picture from that presented by the H-statistic for the model including the four subsidy dummy variables. The latter phenomenon illustrates the volatility of the test statistics over the given period for a model containing a small number of degrees of freedom.

Comparing the sums of squares of residuals and the adjusted R-square values of models 4.5d and 4.5e, it would appear that the model using the total breeding herd as the independent regressor is the superior model. This is true for the models both including and excluding the 1977 subsidy dummies, and so model 4.5d appears to be the better culling model of the two.

4.5f Outflow as a Function of Inflow

Using the results from model 4.5d which imply that the breeding sows remain in the breeding herd for approximately two and a half years, cullings in time period t-1 to t should have derived from inflow lagged 7 and 8 periods. Under steady state assumptions, and with no deaths of sows and boars once they have entered the breeding herd, one would expect the coefficients on the Inflow variables to sum to unity. Using pregnant gilts as the proxy for inflow and assuming that boar inflow is equivalent to 5% of sow inflow, it is expected that the coefficients on the lagged pregnant gilt variables should sum to 1.116. Model 4.5f attempts to estimate this relationship between outflow and lagged inflow.

Initial OLS regression including two seasonal dummy variables produced residual plots and a DW statistic which clearly indicated serial correlation. Putting the residuals through the Box-Jenkins identification procedure produced an autocorrelation correlogram with an obvious cyclical pattern and a partial autocorrelation correlogram in which the first two partials were significantly different from zero implying an AR(2) process. In order to allow for this, LSQ was used to make a second order autocorrelation correction. The residuals of the initial regression also indicated the need for the first two of the Aujezky dummies. The residuals at observations 1979:2-1980:2 inclusive also appeared to be out of line with the remainder of the residuals. With a subsidy effect on inflow at observations 1977:1-77:3, one would expect, with a lag of 7 and 8 periods that the effect would show through on cullings in 1979:2-80:3 and, therefore, four subsidy dummies representing the four potential outliers were also included in the model.

Estimation of the regression produces highly significant parameters on the autoregressive and pregnant gilt variables, and the August dummy has a significant negative parameter. The roots of the AR polynomial suggest a cycle in the model with a measured length of approximately 3 years and 10 months. The first Aujezky dummy parameter is highly significant and positive followed by a significant and negative parameter on the second Aujezky dummy implying a greater proportion of culling at the time of the campaign, followed by a lull presumably caused by the reduced availability of potential sow and boar culls. The first of the subsidy dummies has a highly significant positive parameter, whereas the parameters on the third and fourth subsidy dummies show a drop in cullings, though not a significant one. With inflow low in 1977:1 and high in 1977:3, the estimated coefficients on the four subsidy dummies are those that one would expect. The coefficients on the pregnant pig variables sum to 1.110 with a slightly higher weighting on the longer of the two lags, the difference between this and the expected value of 1.116 being explained by losses from the breeding herd. The only potential problem with the model is the H-statistic of -3.16, which is highly significant. Given that there are only 11 degrees of freedom in the model, nothing was done to rectify the situation; a decision somewhat justified by the fact that the H-statistic for the model having excluded the four subsidy dummies falls to a value of -0.098. The resulting model and diagnostic check statistics are given in Appendix 4c and the re-estimated regression, having dropped the four subsidy dummies is presented in equation 4.5f.1.

$$
\begin{aligned}
M_{t-1,t} = {} & \underset{(7.5)}{1.1738}\ M_{t-2,t-1} - \underset{(-5.0)}{0.6093}\ M_{t-3,t-2} + REG_t(\underset{(3.88)}{0.5306}, \underset{(4.33)}{0.5512}, \underset{(-2.47)}{-0.0481}, \underset{(-0.52)}{-0.0097}) - \\
& 1.1738\ REG_{t-1}(0.5306, 0.5512, -0.0481, -0.0097) - \\
& 0.6093\ REG_{t-2}(0.5306, 0.5512, -0.0481, -0.0097) + \\
& \underset{(5.1)}{43.68}\ A83{:}2 - \underset{(-2.02)}{23.05}\ A2
\end{aligned}
\tag{4.5f.1}
$$

Where $REG_t(a_1,a_2,b_1,b_2) = (a_1\ PG_{t-7} + a_2\ PG_{t-8})(1 + b1\ AUG_t + b2DEC_t)$,

and $REG_{t-1}(a_1,a_2,b_1,b_2) = (a_1\ PG_{t-8} + a_2\ PG_{t-9})(1 + b1\ AUG_{t-1} + b2DEC_{t-1})$,

and $REG_{t-2}(a_1,a_2,b_1,b_2) = (a_1\ PG_{t-9} + a_2\ PG_{t-10})(1 + b1\ AUG_{t-2} + b2DEC_{t-2})$,

Obs. = 23 RSS = 872.1 $\hat{R}^2 = 0.72$ H = -0.098

The exclusion of the subsidy dummies has caused a significant increase in the size of the RSS and slightly reduced the size of the coefficients on the pregnant gilt variables. The roots of the AR polynomial now produce a shorter cycle length of approximately 2 years and 11 months. The difference in length of the cycles produced by the regressions including and excluding the subsidy dummies illustrates the volatility of the parameter estimates over the short sample period on which the model is estimated. The removal of the subsidy dummies has still produced similar parameters on the two lagged pregnant gilt variables which one would expect with an average life of 30 months in the breeding herd indicated by the results from model 4.5d.

Comparing the mean square errors - MSE - of models 4.5d and 4.5f, that is, 40.16 and 37.91 respectively, it would appear that the latter model provides the better fit to the cull data over the estimation period. However, as was seen in model 4.5b, due to effects such as the 1977 subsidy and world market effects at the start of the estimation sample period, the data are noticeably more volatile in the first half of the sample period. Because model 4.5f includes much longer lags on the right hand side, it was estimated over the shorter more stable period than model 4.5d, which contains only a one period lag. Consequently, it was deemed necessary to re-calculate the residual sums of squares for model 4.5d using the same estimation period as that for model 4.5f, in order to make a fair comparison. Doing the latter produces a RSS of 652.1 for the latest 23 observations, which compares with a figure of 872.1 for model 4.5f and converts to a MSE of 28.4. Over a comparable estimation period, therefore, the estimated regressions suggest that the best trimestic culling model is that containing the total breeding herd lagged one period as the independent variable.

4.5g Pregnant Gilts and the Total Breeding Herd

This equation attempts to model the pregnant gilt herd - the inflow proxy - as a function of the total breeding herd lagged 12 months - three census periods and is the inflow side of the recursive breeding herd forecasting model discussed in section 4.3 and expressed in equation 4.3.12. The reasoning for this lag is that we expect gilts entering the breeding herd to have come from pregnant pigs which were in pig three periods earlier. Initial estimation including seasonal and Aujezky dummies produced a residual plot with large residuals for the first two observations and indicated a subsidy effect in 1977. The outlier and subsidy effects were modelled by use of the appropriate dummy variables. OLS estimation including the usual set of dummies implied no seasonal effect and so the seasonal dummies were removed from the estimation procedure. The resulting re-estimation was satisfactory as far as the t-statistics were concerned but the DW of 1.07 was low enough to render it advisable to run the residuals through the Box-Jenkins identification process. The resultant correlograms clearly indicated first order serial correlation, thereby making estimation by Beech-Mackinnon appropriate.

Once again, the seasonal dummy parameters were not significant at the 5% level and were, therefore, dropped. The resulting re-estimation produced significant negative parameters on the Aujezky dummies and on the second and third of the subsidy dummies. The RSS value was 1028.4, the $\hat{R}^2$ value 0.99 and there was no apparent problem with the plot of the residuals. This model was accepted as the best and the results of estimation are reproduced in Appendix 4c; the results of re-estimation without the subsidy dummies are presented in equation 4.5g.1.

$$PG_t = 0.1222\,HB_{t-3} - 11.77\,A83{:}2 - 18.36\,A83{:}3 - 13.66\,A84{:}1 + 28.99\,D76{:}1 + 23.82\,D76{:}2 + U_t$$
$$\quad(36.9) \qquad (-1.41) \qquad (-2.02) \qquad (-1.63) \qquad (3.1) \qquad (2.84)$$

$$U_t = 0.4625\,U_{t-1} \qquad (4.5g.1)$$
$$(2.65)$$

Obs. = 30 RSS = 1634.5 $\hat{R}^2 = 0.982$

The removal of the subsidy dummies has had little effect on the estimates of the included parameters though the value of the RSS had increased significantly.

4.5h Pregnant Gilts and the Pregnant Pig Herd

This model is similar to the previous model except that it considers pregnant gilts as a proportion of the pregnant pig herd, rather than the total breeding herd, lagged three periods. The process of estimation and the results of diagnostic checking are virtually identical to those of the previous model and, therefore, only the results of the estimation of the final regression, which has an identical structure to that of the previous model, are discussed. The results of the regression including the subsidy dummies are presented in Appendix 4c. There is no obvious problem with the residuals plots of either estimation.

$$PG_t = 0.1728\,PP_{t-3} - 12.47\,A83{:}2 - 20.45\,A83{:}3 - 14.78\,A84{:}1 + 37.53\,D76{:}1 + 27.04\,D76{:}2 + U_t$$
$$\quad(20.4) \qquad (-1.40) \qquad (-2.18) \qquad (-1.66) \qquad (4.0) \qquad (3.1)$$

$$U_t = 0.3251\,U_{t-1} \qquad (4.5h.1)$$
$$(1.71)$$

Obs. = 30 RSS = 1828.3 $\hat{R}^2 = 0.987$

As was the case with model 4.5g the removal of the subsidy dummies affects only the RSS statistic although the increase is highly significant. Once again, it is the model which includes the total breeding herd as the independent parameter which is the better model if comparing the RSS statistics, and it is for this reason that model 4.5g is preferred to model 4.5h.

4.5i Pregnant Gilts and the Unserved Gilt Herd

Model 4.5i. relates pregnant gilts to the number of unserved gilts in the previous time period. Initial OLS estimation excluding all dummies indicated the presence of a positive time trend in the relationship and the first three residuals were positive and somewhat larger than the other residuals. The positive time trend could be explained by the fact that management techniques had improved over the given time period, thereby increasing the proportion of unserved gilts conceiving.To allow for these factors and to model the possibility of seasonality and an Aujezky effect LSQ was employed for estimation. The results of estimation indicated no Aujezky effect and so the three Aujezky dummies were removed from the estimation procedure. Re-estimation minus the Aujezky dummies produced a satisfactory model. The results as given in Appendix 4c imply that a figure the equivalent of 112.6% of pigs recorded as unserved gilts are recorded as pregnant gilts at the next census. The reason for this is that unserved gilts are only recorded as such between the weights of 50kg and 80kg. The unserved gilts will be a weight within this range for a period substantially less than four months and hence the figure greater than 100%. The estimated percentage would of course be slightly higher were the gestation period not one week short of the interval between censuses. The positive time trend is significant and the subsidy dummies illustrate that the number of pregnant gilts fell significantly in the August of 1977, presumably because of unserved gilts being diverted into the feeding herd at the time of the subsidy. In the December of 1977, however, the pregnant gilt herd was significantly increased, presumably because producers were attempting to replenish the breeding herd stock. All three of the dummies representing the observations for 1975:2 to 1976:1 have significantly positive parameters. There is nothing obviously wrong with the residual plots and the DW of 2.60 is not significant at either the 5% or 1% levels. The adjusted R-squared value is 0.75 and the RSS value 529.4. Re-estimating the regression minus the subsidy dummies produces the results presented in equation 4.5i.1

$$\begin{aligned} PG_t = {} & \underset{(34.4)}{1.1280}\, UG_{t-1} - \underset{(-1.25)}{0.0342}\, AugUG_{t-1} - \underset{(-2.13)}{0.0589}\, DecUG_{t-1})(1 + \underset{(3.23)}{0.0045}\, T) + \\ & \underset{(4.9)}{30.60}\, D75{:}2 + \underset{(3.05)}{19.53}\, D75{:}3 + \underset{(2.51)}{16.40}\, D76{:}1 \end{aligned} \tag{4.5i.1}$$

Obs. = 32 RSS = 835.4 $\hat{R}^2 = 0.65$ DW = 2.67

The removal of the subsidy dummies has an influence on the seasonal dummy parameter estimates and significantly worsens the values of the RSS and adjusted R-squared statistics.

4.5j Unserved Gilts and the Total Breeding Herd

This model relates the number of unserved gilts to the breeding herd lagged two periods. The OLS model with no intercept dummies produced a residual plot which indicated the possibility of a negative time trend, a possibility because of the increased productivity of sows over time having the consequence that fewer unserved gilts are required to produce the same number of piglets over time. A linear time trend variable was included as were seasonal dummies, the first three Aujezky dummies and four dummies to model the effects of the 1977 subsidy. Having estimated the latter regression by LSQ, it was apparent from inspection of the residuals and the DW statistic that there was a first order autocorrelation problem. To model this, a first order rho-transformation of the model was estimated using LSQ. The residuals of the latter regression indicated a couple of outliers at observations 76:1 and 79:2. These were corrected for by the addition of intervention dummies to remove their effects on the estimated parameters of other variables. Although the time trend parameter was no longer significant, it remains in the model as it was felt biologically justifiable. Because the seasonal dummy parameters were not significant they were dropped from the regression and the resulting model accepted as the best for this relationship. Appendix 3c gives the results of estimating the full model.

The Aujezky dummies indicate that the eradication campaign significantly reduced the number of unserved gilts at the time of the campaign. The effect of the 1977 subsidy is similar to that of the Aujezky effect in that the number of maiden gilts is significantly reduced at the time of the subsidy, almost certainly due to the fact that potential gilts were diverted into the feeding herd, although numbers increase significantly as a proportion of the breeding herd in the second quarter of 1978, presumably a result of producers trying to replenish the depleted breeding herd.

The diagnostics of the model give a RSS value of 316.91 and an adjusted R-squared of 0.73, and the plot of the residuals show no sign of an autocorrelation problem confirmed by an H-statistic of -0.94. The results of estimating the regression minus the subsidy dummies and the outlier for 1979:2 is presented in equation 4.5j.1 below.

$$\begin{aligned} UG_t = {} & \underset{(2.23)}{0.3662}\, UG_{t-1} + REG_t(\underset{(17.5)}{0.1023}, \underset{(0.11)}{0.0003}) - 0.3662\, REG_{t-1}(0.102\dot{3}, 0.0003) - \\ & \underset{(-1.56)}{11.61}\, A83{:}2 - \underset{(-2.00)}{15.42}\, A83{:}3 - \underset{(-0.37)}{2.99}\, A84{:}1 + \underset{(2.39)}{18.53}\, D76{:}1 \end{aligned} \qquad (4.5j.1)$$

Where $REG_t(\alpha,\gamma) = \alpha\, HB_{t-2}(1 + \gamma T)$,
and $REG_{t-1}(\alpha,\gamma) = \alpha\, HB_{t-3}(1 + \gamma\, \{T-1\})$

Obs. = 30 RSS = 1161.2 $\hat{R}^2 = 0.24$ H = 1.70

The removal of the subsidy and 1979:2 outlier dummies has a significant effect on the results of the diagnostic statistics. The RSS statistic and the adjusted R-square values have both been adversely affected and the h-statistic of 1.70 is significant at the 5% level. The estimated time trend parameter has actually changed sign from negative to positive.

4.5k Unserved Gilts and the Pregnant Pig Herd

This model is similar to the latter except that the independent variable is replaced by pregnant pigs, they being the direct source of the unserved gilts two periods later. The process of estimation followed that of the previous model, and the structure of the final regressions are identical. The time trend is significantly negative and the effects of the subsidy and the eradication campaign are almost identical to what they were in the model using the total breeding herd as the regressor. The results of estimation, both including and excluding the subsidy dummies are presented in Appendix 3c and equation 4.5k.1 respectively.

$$\underset{}{UG_t} = \underset{(1.97)}{0.3339}\, UG_{t-1} + REG_t(\underset{(16.7)}{0.1517}, \underset{(-0.82)}{-0.0023}) - 0.3339\, REG_{t-1}(0.1517, -0.0023) -$$

$$\underset{(-1.60)}{12.40}\, A83{:}2 - \underset{(-1.88)}{15.19}\, A83{:}3 - \underset{(-0.43)}{3.58}\, A84{:}1 + \underset{(2.09)}{17.23}\, D76{:}1 \qquad (4.5k.1)$$

Where $REG_t(\alpha,\gamma) = \alpha\, PP_{t-2}(1 + \gamma T)$,

and $REG_{t-1}(\alpha,\gamma) = \alpha\, PP_{t-3}(1 + \gamma\,\{T-1\})$

Obs. = 30 RSS = 1264.7 $\hat{R}^2$ = 0.17 H = 1.84

As with the previous model, the h-statistic for the model having removed the subsidy and the outlier dummies is significant at the 5% level, but the value of -0.86 for h in the regression including the said dummies implies that there is no serial correlation in the residuals. The RSS and the adjusted R-square statistics are again much worse having removed the subsidy and outlier dummies. The time trend variable is still negative in sign but is no longer significant at the 5% level. Comparing the RSS and adjusted R-Square statistics of models 4.5j and 4.5k, it would appear that model 4.5j provides a better fit to the data when the subsidy and the outlier dummies are excluded; the situation is, however, the reverse when the subsidy and 1979:2 outlier dummies are included. If a choice between the two were to be made, it could be argued that 4.5j is the better of the two for three main reasons. In no particular order these are; firstly for convenience, that is, it is easier to forecast the total breeding herd - the independent variable in 4.5j - than it is to forecast the pregnant pig herd, which first requires a forecast of the breeding sow herd any way before 4.5c.4 and 4.5c.6 can be used to forecast the two components of the pregnant pig herd. A second reason for using 4.5j is for consistency, that is, all comparisons of the other biological relationships have produced the result that the model including the total breeding herd as the independent regressor provides the better fit. A third and final reason for choosing 4.5j in preference to 4.5k is that it is almost certainly the greater significance of the time trend variable in model 4.5k which has contributed to the equality, if not superiority of fit of this model compared with that of 4.5j. Given that the residuals of model 4.5c.4 and 4.5c.5 indicate that these time trends - which were the consequence of producers reducing the weaning period during the late seventies and early eighties - have ended, it suggests that using 4.5k, with its significant time trend to forecasts the future unserved gilt herd may not be wise. One these grounds then, 4.5k should be rejected in favour of 4.5j for modelling the unserved gilt herd.

4.5l Fat pigs and the Total Breeding Herd

This model is concerned with the relationship between fat pigs and the total breeding herd lagged 2 and 3 periods, the lags expected when the age of fat pig slaughter is considered. Modelling using OLS methodology produces a residual plot in which there is an obvious positive time trend, which one would expect if sows and gilts become more productive over time. To model this, a non-linear least squares regression was estimated incorporating 2 seasonal dummies and the simple linear time trend variable T. The residual plot illustrated the possibility of outliers for the first two observations in the estimation period, a subsidy effect from 1977:2 -1978:1 inclusive, and the need for the Aujezky dummies A83:3 and A84:1. Dummies to represent these effects were subsequently added to the regression and their parameters estimated.

The parameters on the two lagged breeding herd variables were both significant at the 5% level and had a weight on the shorter of the two lags which was twice the magnitude of that on the variable lagged three periods. The coefficients on the seasonal dummies are highly significant indicating that slaughterings as a percentage of the breeding herd decreases in August and increases in December. As expected, the coefficient on the time trend variable is highly significant. Although A83:3 and A84:1 are not significant at the 5% level, the residuals of the previous regression illustrated the obvious effect of the eradication campaign on slaughtering numbers at the turn of 1983, and so the dummies were kept in the regression. The coefficients on the dummies for the first two observations are both large and negative. The coefficients on the subsidy dummies indicate that slaughterings significantly decrease in the period from April to July inclusive, after which they show signs of recovery in the following two periods. A possible explanation for the decrease in the number of slaughterings in the second period of 1977 is that slaughterings had been relatively high at the end of 1976 and the start of 1977, due primarily to a pessimism on the part of producers for expected profits in 1977 and the introduction of the subsidy at the end of January 1977. The former factor had encouraged producers to start reducing the size of their breeding herds, hence the shortage of fat pigs towards the middle of 1977, while the subsidy may have encouraged producers to slaughter slightly earlier than they had planned anticipating that the subsidy would not last. Alternatively, and possibly a more plausible explanation in the light of other models, is that the subsidy may have encouraged producers to start increasing the size of the depleted breeding herd in the second period of 1977, hence the increase in the number of unserved gilts towards the end of 1977. The latter increase in unserved gilts would have reduced the number of fat pigs available for slaughter in the second period of 1977.

As there are no residual or DW problems with the regression, confirmed by the use of Box-Jenkins identification, the latter regression was accepted as the best for this relationship - the results of which are presented in Appendix 3c - and so the re-estimated equation minus the subsidy dummies is presented in equation 4.51.1 below.

$$\begin{aligned} FP_{t-1,t} = {} & (\underset{(5.72)}{3.4702}\, HB_{t-2} + \underset{(2.52)}{1.5354}\, HB_{t-3})\,(1 - \underset{(-4.96)}{0.0371}\, AUG + \underset{(3.02)}{0.0235}\, DEC)\,(1 + \underset{(13.7)}{0.0062}\, T) - \\ & \underset{(-2.42)}{194.04}\, D76{:}1 - \underset{(-2.60)}{213.66}\, D76{:}2 + \underset{(0.84)}{66.89}\, A83{:}2 - \underset{(-1.52)}{120.26}\, A83{:}3 - \underset{(-1.95)}{174.13}\, A84{:}1 \end{aligned} \tag{4.51.1}$$

Obs. = 30 RSS = 104779 $\hat{R}^2 = 0.94$ DW = 1.61

The overall effect of having removed the three subsidy dummies is to adversely effect the RSS and the $\hat{R}^2$ statistic.

4.5m Fat pigs and the Pregnant Pig Herd

This model is identical to the previous one except that the breeding herd is replaced by pregnant pigs as the independent regressor. The estimation procedure followed that of the previous model and the coefficients of the included variables were of similar magnitude and significance, although it is noticeable that the time trend is smaller due to the fact that barren sows have accounted for an decreasing percentage of the breeding sow herd over the given period as illustrated in model 4.5c. This effect is not present on the right hand side of the current model, hence the smaller parameter on the time trend variable. The only potential problem with the diagnostics is the DW statistic of 1.07. This value lies within the range of uncertainty at the 1% level when compared with the Farebrother tables, but although a model incorporating a rho-transformation was estimated, the first order autoregressive parameter proved to be insignificant at the 5% level. Putting the residuals of the initial regression through the Box-Jenkins procedure illustrated the fact that there was no other possible order for any potential serial correlation and so the structure of the models remains identical to that of model 4.5l. The results of estimation with and without the subsidy dummies are given in Appendix 4c and equation 4.5m.1 respectively.

$$FP_{t-1,t} = \underset{(5.42)}{(4.5972}\, PP_{t-2} + \underset{(3.28)}{2.8332}\, PP_{t-3})\,(1 - \underset{(-5.29)}{0.0463}\, AUG + \underset{(2.05)}{0.0222}\, DEC)\,(1 + \underset{(6.36)}{0.0031}\, T) -$$

$$\underset{(-1.73)}{165.68}\, D76{:}1 - \underset{(-2.71)}{260.48}\, D76{:}2 + \underset{(0.70)}{64.53}\, A83{:}2 - \underset{(-1.50)}{139.22}\, A83{:}3 - \underset{(-1.92)}{201.04}\, A84{:}1 \qquad (4.5m.1)$$

Obs. = 30 RSS = 143,190 $\hat{R}^2 = 0.92$ DW = 1.24

The diagnostic results of the latter two models in terms of RSS and adjusted R-squared value indicate that the first of the two models, which has the total breeding herd as the independent regressor, is the better of the two. This is true for the models both including and excluding the subsidy dummies and so model 4.5l would be used in preference to 4.5m as the best trimestic model for fat pig slaughter.

4.6 Monthly Models For Cullings and Slaughterings

Sow and Boar Culling models similar to 4.5d and 4.5e, and fat pig slaughter models similar to 4.5l and 4.5m are developed here with the dependent variables measured in monthly rather than trimestic terms. As with the trimestic biological models, the regressions are estimated on the sample period 1975-85 inclusive. Both the said monthly series are adjusted so that each of the months represents a four week period in order to allow for the fact that the data for some of the months are aggregated using five rather than four week accounting periods. The monthly culling and fat pig slaughter data are listed in Appendix 3b. Seasonality is accounted for by 11 seasonal dummies which measure deviations from the base month of January. Once again, the simple linear time trend is used to model trending relationships, and the appropriate Aujezky, subsidy and outlier dummies are included where necessary to remove any effects they might otherwise have on the estimated seasonal and time trend coefficients. As was the case with the trimestic models, the results of estimation having included the subsidy and outlier dummies are presented in an appendix - Appendix 4d - rather than in the main text. Serial correlation is remedied using the appropriate order autocorrelation correction.

4.6a Culling and the Total Breeding Herd

The first model relates the culling of sows and boars to the total breeding herd. The monthly culling series is regressed as a proportion of the breeding herd at the previous census provided that the census does not take place in the same month as the culling. Allowing for an average weaning period of three to four weeks, it is not expected that many of the sows classed as pregnant in the census - taken at the beginning of the month - will be culled in that same census month. Consequently, the cullings in April, August and December are regressed on the breeding herd as measured at the previous December, April and August censuses respectively. In order to model the above specification using TSP the census data are arranged so that the first four observations for each year are the observation from the April census; the second four, the August census observation and the last four, the December observation. Consequently, the specifications of the two monthly culling models presented in this chapter are such that cullings in any single month are regressed on the breeding herd with an apparent lag of four months.

Initial OLS estimation including the 11 seasonal dummies resulted in a regression in which it was clear that there was serial correlation. The residuals were put through the Box-Jenkins identification procedure, the resulting correlograms clearly illustrating first order serial correlation. As there was no evidence of a time trend in the residuals, the Beech-Mackinnon maximum likelihood procedure could be employed to estimate the model. It was also evident from the residuals of the initial regression that there was a potential outlier observation in December of 1975 and that culling visibly increased in the March to June of 1983 - a period corresponding to the Aujezky disease eradication campaign of that year. To model these effects five individual intervention dummies were included in the regression even though no reason could be found for the discrepancy in December 1975. Finally, any effect of the 1977 subsidy dummy is removed by the inclusion of five individual dummies for the months of February to June of the said year - the months when the subsidy was operative. The results of estimation including all the said dummies is presented in Appendix 4d along with the results of estimating similar regressions for the other monthly models discussed in this chapter.

The RSS has a value of 225.1 and the regression has an adjusted R-squared value of 0.91. The parameter on lagged breeding herd variable implies that an average of 3% of the breeding herd at any single census is culled in each of the following four months. The seasonal dummies indicate significant decreases in cullings with respect to the January figure in the months of April, July, August and December and significant increases in February and November. The most significant parameters are those on the November and December dummies, illustrating the increase in cullings in November in order to meet the additional Christmas demand. This is followed in December by a large fall in cullings, presumably the result of the November increase depleting the number of sows and boars available for culling in December. The outlier dummy has a highly significant positive parameter and the four Aujezky dummies are all highly significant. None of the 5 estimated subsidy dummy parameters is anywhere near the accepted significance levels and there is no sign of any serial correlation problem in the plot of the residuals, confirmed by the Box-Jenkins identification correlograms of the residuals.

Dropping the subsidy and the outlier dummy produces the regression presented in equation 4.6.1 below. The estimated coefficients and t-statistics appear in table 4.6a.1.

$$M_{t-1,t} = a\,HB_{t-4} + b2\,febHB_{t-4} + b3\,marHB_{t-4} + b4\,aprHB_{t-4} + b5\,mayHB_{t-4} + b6\,junHB_{t-4} +$$

$$b7\,julHB_{t-4} + b8\,augHB_{t-4} + b9\,sepHB_{t-4} + b10\,octHB_{t-4} + b11\,novHB_{t-4} + b12\,decHB_{t-4} +$$

$$D1\,A83{:}3 + D2\,A83{:}4 + D3\,A83{:}5 + D4\,A83{:}6 + U_t$$

$$U_t = R1\,U_{t-1} + e_t \qquad (4.6a.1)$$

Obs. = 128 RSS = 313.54 $\hat{R}^2 = 0.93$

Table 4.6a.1.
The Results of Estimating the Monthly Culling of Sows and Boars as a Proportion of the Total Breeding Herd.

VARIABLE	COEFFICIENT	ESTIMATE	t-RATIO
U_{t-1}	R1	0.7819	13.6
HB_{t-4}	a	0.0306	31.4
febHB	b2	0.0016	2.62
marHB	b3	0.0011	1.25
aprHB	b4	-0.0023	-2.39
mayHB	b5	-0.0008	-0.81
junHB	b6	-0.0005	-0.43
julHB	b7	-0.0019	-1.83
augHB	b8	-0.0021	-1.99
sepHB	b9	0.0010	1.00
octHB	b10	0.0009	0.99
novHB	b11	0.0026	3.17
decHB	b12	-0.0023	-3.71
A83:3	D1	3.36	1.96
A83:4	D2	8.10	3.92
A83:5	D3	7.08	3.44
A83:6	D4	5.85	3.42

Compared with the model which includes the subsidy and the outlier dummies, the regression presented in equation 4.6a.1 has a higher $\hat{R}^2$ value, presumably the result of having dropped the insignificant subsidy dummies, but a higher RSS value from having dropped the significant outlier dummy for 1975:12.

4.6b Culling and the Pregnant Sow Herd

The second monthly biological model is the same as model 4.6a except that the total breeding herd is replaced by the pregnant sow herd as was done for the trimestic culling model presented in section 4.5e. Initial OLS estimation of the model including the seasonal dummies produced residuals and test statistics very similar to those for the monthly culling model with the total breeding herd as the independent regressor. Consequently, a model with a specification equivalent to that of model 4.6a was estimated by Beech-Mackinnon. The resultant model appeared totally satisfactory in terms of the diagnostics and is therefore accepted as the best model for the relationship between culling and lagged pregnant sow herd.

The structure of the regression and the results of estimation are presented in equation 4.6.2 and table 4.6b.1 respectively. The equivalent model including the subsidy and outlier dummies appears in Appendix 4d.

$$M_{t-1,t} = a1\ PS_{t-4} + b2\ febPS_{t-4} + b3\ marPS_{t-4} + b4\ aprPS_{t-4} + b5\ mayPS_{t-4} + b6\ junPS_{t-4} +$$

$$b7\ julPS_{t-4} + b8\ augPS_{t-4} + b9\ sepPS_{t-4} + b10\ octPS_{t-4} + b11\ novPS_{t-4} + b12\ decPS_{t-4} +$$

$$D1\ A83{:}3 + D2\ A83{:}4 + D3\ A83{:}5 + D4\ A83{:}6 + U_t$$

$$U_t = R1\ U_{t-1} + e_t \qquad (4.6b.1)$$

Obs. = 128 RSS = 318.18 $\hat{R}^2 = 0.92$

Table 4.6b.1.
The Results of Estimating the Monthly Culling of Sows and Boars as a Proportion of Pregnant Sows.

VARIABLE	COEFFICIENT	ESTIMATE	t-RATIO
U_{t-1}	R1	0.7979	14.3
PS_{t-4}	a	0.0514	29.7
febPS	b2	0.0028	2.65
marPS	b3	0.0019	1.32
aprPS	b4	-0.0038	-2.28
mayPS	b5	0.0004	0.24
junPS	b6	0.0009	0.47
julPS	b7	-0.0017	-0.93
augPS	b8	-0.0019	-1.05
sepPS	b9	0.0028	1.66
octPS	b10	0.0027	1.73
novPS	b11	0.0056	4.01
decPS	b12	-0.0031	-2.88
A83:3	D1	3.25	1.88
A83:4	D2	8.01	3.85
A83:5	D3	6.66	3.20
A83:6	D4	5.66	3.27

The coefficients on the pregnant sow variables indicate that the equivalent of approximately 5% of the herd are culled in each month following the censuses, and again November has the highest culling ratio. The results of comparing the diagnostics of the models including and excluding the subsidy and outlier dummies are similar to those of model 4.6b.

Comparing the $\hat{R}^2$ and the RSS values of the two monthly models for culling presented in sections 4.6a and 4.6b, indicates that little or nothing is gained by changing the independent regressor from lagged total breeding herd to lagged pregnant sow herd. As a consequence model 4.6a.1 is accepted as the better of the two monthly sow and boar culling models.

4.6c Fat pigs and the Total Breeding Herd

Model 4.6c is concerned with the relationship between the monthly slaughterings of fat pigs as a proportion of the total breeding herd lagged two and/or three census periods. The information on the number of days from birth to slaughter - given in figure 4.1a - for the various types of fat pigs indicates that the range is approximately 144 days for porkers to over 200 days for heavy hogs. In terms of months, these figures approximate four and a half to six and a half months. Working on this basis, slaughterings in any given month can be expected to have come from sows and gilts recorded as being in pig 6 to 8 months earlier. For the majority of fat pigs, therefore, this implies a lag of two census periods, however, for slaughterings in January, May and September some fat pigs, born shortly before the census 2 periods earlier will require a lag of three census periods. Consequently, the initial specification for the model is that presented in equation 4.6c.1 which does not allow for seasonality or any other potential factor.

$$FP_{t-1,t} = a1\ HB_{t-8} - a2\ DumX * HB_{t-8} + a2\ DumX * HB_{t-12} + e_t \quad (4.6c.1)$$

where DumX = 1 for January, May and September, that is:-

$$FP_{t-1,t} = (a1 - a2)\ HB_{t-8} + a2\ HB_{t-12} + e_t \quad (4.6c.2)$$

and DumX = 0 otherwise, that is:-

$$FP_{t-1,t} = a1\ HB_{t-8} + e_t \quad (4.6c.3)$$

Equation 4.6c.3 has been specified so that parameters a1 and a2 are constrained in such a way that the sum of parameters on the total breeding herd lagged both 8 and 12 months for January, May and September fat pigs is constrained to a value equal to that taken by the parameter on HB lagged 8 months in the other 9 months of the year. Estimation of 4.6c.3 having included 11 seasonal dummies produced residual plots in which there was an obvious positive time trend, illustrating the increased productivity of sows over the sample period. The simple linear time trend was included in the regression to model the trend. The plot of residuals from having estimated the latter regression and a low DW statistic indicated the possibility of first order autocorrelation. This was supported having put the residuals through the Box-Jenkins identification procedure and observing the resulting correlograms. To model the first order autocorrelation, a rho-transformation of the equation was estimated. An obvious problem indicated by the plot of residuals of the rho-transformed model was the possibility of an Aujezky eradication campaign influence effective from April of 1983 through to the April of the following year. The effect was to substantially increase slaughterings in April 1983 and thereafter to increasingly reduce slaughterings over the following 12 months. Because the inclusion and estimation of 13 separate dummies seemed somewhat cumbersome and extravagant in terms of usage of degrees of freedom, the decision was taken to model the Aujezky effect by a single dummy covering the whole of the relevant period. The estimated Aujezky dummy parameter therefore, measures an average effect of the eradication campaign over the given period: though not ideal, it does diminish the effect of the campaign on the magnitude and direction of the seasonal dummy parameter estimates.

The sole remaining alteration required was to add the five individual dummies for the subsidy months of February to June of 1977. The structure of the regression and the results of estimation are presented in appendix 4b. The subsidy effect is similar to the Aujezky effect in that the immediate effect is for slaughterings to be larger than expected after which they fall. Having said this, the results of estimation - presented in Appendix 4d - show only the June 1977 dummy to be significant at the 5% level.

Removing the subsidy dummies produces an equation with a structure represented by equation 4.6c.4 and estimated coefficients given in table 4.6c.1.

$$FP_{t-1,t} = R1\ FP_{t-2,t-1} + REG_t - R1\ REG_{t-1} + D1\ DUMA + e_t \qquad (4.6c.4)$$

Where $REG_t = (a1\ HB_{t-8} - a2\ DumXHB_{t-8} + a2\ DumXHB_{t-12})(1 + c\,T)(1 + b2\ Feb + b3\ Mar + b4\ Apr + b5\ May + b6\ Jun + b7\ Jul + b8\ Aug + b9\ Sep + b10\ Oct + b11\ Nov + b12\ Dec)$
and $REG_{t-1} = (a1\ HB_{t-9} - a2\ DumXHB_{t-9} + a2\ DumXHB_{t-13})(1 + c\,(T-1))(1 + b2\ Feb(-1) + b3\ Mar(-1) + b4\ Apr(-1) + b5\ May(-1) + b6\ Jun(-1) + b7\ Jul(-1) + b8\ Aug(-1) + b9\ Sep(-1) + b10\ Oct(-1) + b11\ Nov(-1) + b12\ Dec(-1))$

Obs. = 119 RSS = 62325.5 $\hat{R}^2$ = 0.88 H = -1.78

Table 4.6c.1.
The Results of Estimating the Monthly Slaughterings of Fat pigs as a Proportion of the Total Breeding Herd.

VARIABLE	COEFFICIENT	ESTIMATE	t-RATIO
FP_{t-1}	R1	0.5095	6.00
HB_{t-8}	a1	1.0517	58.1
HB_{t-12}	a2	0.6620	4.36
T(ime)	c	0.0018	10.1
Feb	b2	0.0345	4.02
Mar	b3	0.0331	3.13
Apr	b4	-0.0036	-0.32
May	b5	-0.0010	-0.09
Jun	b6	-0.0055	-0.47
Jul	b7	-0.0270	-2.32
Aug	b8	-0.0169	-1.45
Sep	b9	0.0537	4.48
Oct	b10	0.0601	5.18
Nov	b11	0.0997	9.07
Dec	b12	0.0298	3.41
ADUM	D1	-16.28	-2.05

The parameters on the lagged dependent variable, the lagged breeding herd variables and the time trend variable are all highly significant. The estimated parameters on the lagged breeding herd variables indicate an average slaughter figure at the start of the estimation period equivalent to 105% of the breeding herd 2 censuses earlier. Because of the presence of the time trend, the latter figure obviously increases over time. The parameter estimates also indicate a higher weighting on the longer of the two lags for January, May and September. The seasonal dummy parameters clearly indicate increased slaughter - above the

January figure - from September through to March inclusive and reduced slaughter for the months of April to August. Differences in the parameters to those estimated having included the 1977 subsidy dummies are that the February dummy parameter has become slightly larger and the June parameter has become negative. The Aujezky dummy parameter indicates an average drop in slaughterings of 15,740 in each of the months from April 1983 to April 1984.

The RSS of 62,325 compares with that of 55,681 for the regression including the subsidy dummies, while the respective adjusted R-squared values are 0.88 and 0.89. The Durbin H-statistic from the regressions both including and excluding the subsidy dummies take values of -1.42 and -1.78 respectively and provide little or no evidence of residual autocorrelation.

4.6d Fat pigs and the Pregnant Pig Herd

The last of the monthly models is a fat pig slaughter model, built using identical methodology to the previous regression discussed under model 4.6c, except that the regressor is now the pregnant pig herd rather than the total breeding herd as was done for the equivalent trimestic model in section 4.5m. The results of estimation of equation 4.6d.1 are presented in table 4.6d.1 below, and those for the model including the subsidy dummies of 1977 are presented in Appendix 4d.

$$FP_{t-1,t} = R1\ FP_{t-2,t-1} + REG_t - R1\ REG_{t-1} + D1\ DUMA + e_t \qquad (4.6d.1)$$

Where $REG_t = (a1\ PP_{t-8} - a2\ DumXPP_{t-8} + a2\ DumXPP_{t-12})\ (1 + c\ T)\ (1 + b2\ Feb + b3\ Mar + b4\ Apr + b5\ May + b6\ Jun + b7\ Jul + b8\ Aug + b9\ Sep + b10\ Oct + b11\ Nov + b12\ Dec)$

and $REG_{t-1} = (a1\ PP_{t-9} - a2\ DumXPP_{t-9} + a2\ DumXPP_{t-13})\ (1 + c\ (T-1))\ (1 + b2\ Feb(-1) + b3\ Mar(-1) + b4\ Apr(-1) + b5\ May(-1) + b6\ Jun(-1) + b7\ Jul(-1) + b8\ Aug(-1) + b9\ Sep(-1) + b10\ Oct(-1) + b11\ Nov(-1) + b12\ Dec(-1))$

Obs. = 119 RSS = 68401.4 $\hat{R}^2 = 0.87$ H = -1.16

Table 4.6d.1.
The Results of Estimating the Monthly Slaughterings of Fat pigs as a Proportion of the Pregnant Pig Herd.

VARIABLE	COEFFICIENT	ESTIMATE	t-RATIO
FP_{t-1}	R1	0.5598	6.80
PP_{t-8}	a1	1.6058	52.8
PP_{t-12}	a2	0.9596	4.96
T(ime)	c	0.0010	5.31
Feb	b2	0.0249	2.76
Mar	b3	0.0247	2.21
Apr	b4	-0.0110	-0.92
May	b5	-0.0119	-0.97
Jun	b6	-0.0217	-1.74
Jul	b7	-0.0422	-3.39
Aug	b8	-0.0315	-2.54
Sep	b9	0.0464	3.40
Oct	b10	0.0645	5.17
Nov	b11	0.1051	8.94
Dec	b12	0.0355	3.83
ADUM	D1	-15.71	-1.88

The results of estimation of equation 4.6d.1 and the same model including the subsidy dummies are very similar to those for the equivalent regressions presented under model 4.6c. The estimated value for a1 indicates that the equivalent of about 161% of the pregnant pig herd two censuses previous were slaughtered at the start of the estimation period, and again the weighting on the longer lag is larger than that on the shorter lag in the expressions for slaughterings in January, May and September. The estimated time trend trend coefficient and its t-statistic are approximately half the value they take in model 4.6c. The latter is the result of the absence of the negatively trending barren sow series from the lagged independent variable. The H-statistic of the models including and excluding the subsidy dummies are not significant at the 5% level measured at -1.30 and -1.16 respectively.

The RSS and the adjusted R-square statistics for the regressions presented in model 4.6d are not as good as the equivalent statistics for the regressions in model 4.6c and for this reason the monthly fat pig model using the total breeding herd as the independent regressor is chosen as the better model for the adjusted monthly fat pig slaughterings.

4.7 Forecasting With the Biological Models

As was stated in the introduction to this chapter, the main reason for estimating the biological models was to build forecasting models for the breeding herd and the monthly culling and fat pig slaughter series, which could then be compared with the forecasting performance of the equivalent time series and economic models. In this section it is proposed to outline the method by which forecasts for the trimestic breeding herd and the monthly cull and fat pig slaughter series can be made using the trimestic and the monthly culling models built in this chapter.

The expression from which the trimestic forecasts of the total breeding herd will be produced is derived from that presented in equation 4.7.1, which is the equality 4.3.7 presented earlier in section 4.3 during the discussion of the steady state equilibrium model.

$$HB_t \equiv HB_{t-1} + I_{t-1,t} - M_{t-1,t} - L_{t-1,t} \tag{4.7.1}$$

Because losses of sows and boars are not observed, 4.7.1 reduces to 4.7.2 in which $\hat{I}_{t-1,t}$ is the derived estimated inflow variable which measures actual inflow $I_{t-1,t}$ minus sow and boar losses, $L_{t-1,t}$.

$$HB_t \equiv HB_{t-1} + \hat{I}_{t-1,t} - M_{t-1,t} \tag{4.7.2}$$

Because estimated inflow is a derived variable, it has been suggested that it be replaced by the more readily available proxy variable, pregnant gilts at time t, PG_t. Using the results of the regression estimated in model 4.5b, the total breeding herd generating function is given by equation 4.7.3 below.

$$HB_t \equiv HB_{t-1} + 1.0863\, PG_t - M_{t-1,t} \tag{4.7.2}$$

In order to produce forecasts for the breeding herd, therefore, it is necessary to forecast pregnant gilts and culling.

The pregnant gilt forecasts are derived using model 4.5g which relates pregnant gilts to the total breeding herd lagged three periods. An alternative route would have been to use a two step procedure employing models 4.5i, which relates pregnant gilts to the unserved gilt herd at the previous census, and 4.5j, relating the number of unserved gilts to the total breeding herd lagged two censuses. Because the latter route is more cumbersome, and because 4.5i and 4.5j contain time trends which may not continue in the future, this route is not considered as a viable alternative to using 4.5g. On the culling side of the equation model 4.5d, which models culling as a function of the total breeding herd lagged one period is the chosen forecasting model. For all biological forecasting models, whether trimestic or monthly, the models including the subsidy and other outlier dummies as presented in appendices 4c and 4d are used, in the expectation that better estimates of the seasonal and other affected parameters are obtained.

Having made the choice of best forecasting models for the inflow and outflow variables, the two models can be combined, using the recursive model given in 4.7.3, to produce one-step conditional forecasts and n-step unconditional forecasts of culling, pregnant gilts and the key variable, the breeding herd.

For the two monthly series analysed, the models developed in sections 4.6a and 4.6c, relating culling and fat pig slaughter to the breeding herd lagged an appropriate number of months respectively, will be used to make monthly forecasts of the two slaughter series. The unconditional forecasts beyond four months ahead will require forecasts of the breeding herd to be made and so the trimestic model described above will be employed for this purpose, giving further importance to the trimestic breeding herd model. A 24-month ahead forecast of the monthly culling figure, for example, would require a 6-trimester step forecast of the breeding herd to be made.

4.8 Conclusion

In this chapter, trimestic and monthly models have been built illustrating the biological relationships which exist within and between the breeding and feeding herds. A theoretical framework was used to explain how some of the models built could be used to derive a recursive forecasting model of the breeding herd; the key variable under examination. Because the models were built primarily for the purpose of forecasting, the methodology used was somewhat more sophisticated than a purely biological approach, Beech-Mackinnon maximum likelihood and non-linear least squares being employed to model phenomena such as autocorrelation and time trends. Potential outliers primerily the result of the commodity price rises of 1973, the 1977 pigmeat subsidy and the Aujezky eradication campaign of 1983, were all modelled by using the relevant intercept dummies to remove their effects on seasonal and time trend parameters. Technical features of the sector and seasonal influences such as increased demand for meat at Christmas meant that variables representing these two features are frequently included through having significant estimated parameters.

It is an interesting feature of the estimated models, that where two models are directly comparable, the model including the total breeding herd as a regressor, rather than one or more components of it, is always the better model of the two. It is also an interesting feature of the trimestic models that the only variables which had an autocorrelation problem after least squares estimation were all variables which could be regarded as decision variables for the producer, that is, sow and boar cullings, pregnant gilts and the unserved gilt herd. The presence of autocorrelation in these variables may imply that producers are unable to implement decisions to increase or decrease these variables in the space of one trimestic period, possibly due to the effect of adjustment cost in the system. It is possible that such effects may be picked up when the economic model is estimated in the following chapter.

Having estimated and presented all the models section 4.7 indicated how the key models could be used to produce forecasts for any specified period of time ahead for both the trimestic and monthly variables of concern. In chapter eight, the results of forecasting in the short and medium to long term using the biologically based models will be discussed and compared with similar forecasts using Box-Jenkins and econometric models.

CHAPTER FIVE

AN ECONOMETRIC MODEL FOR THE BREEDING HERD

5.1 Introduction

In Chapters three and four models for the breeding herd have been developed on a statistical and a biological basis respectively. In this chapter a model is built introducing, as an explicit explanatory variable the hitherto absent economic phenomenon of profit. It is hoped that the model will be of use for forecasting the medium term, and that some of the unexplained variations in the biological model can be accounted for by economic phenomena. Because response times of production effects from given price changes are often dependent on the biological lags in the system, the econometric model implicitly includes a biological element.

The economic variable employed is an indicator of the profitability of production defined as the ratio of the Average All Pig Price, AAPP, to an index of feed costs. Two feed cost indices are considered: the first is simply the index of compound feed, CF, and the second is the arithmetic mean of the latter and the index of barley feed, BF; these are the two profit ratios employed by Savin and the M.L.C. in their respective models.[1] The AAPP is usually quoted in pence per kg deadweight - p/kg dw - and is derived from representative deadweight and liveweight quotations for fat pigs in the U.K. Because the AAPP in its present form has only been available since June 1975, figures for the first five months of 1975 were derived using the series which predated the AAPP using the following methodology. This earlier series, which indicated the returns to pig producers, was quoted for the remainder of 1975 alongside the AAPP. Using the arithmetic means of the AAPP and the original series for June to December of 1975, the data for the original series for January to May could be converted into a comparable AAPP. The completed series from January 1975 to 1987 was then converted into an index of prices. At the time of writing, all indices quoted by M.A.F.F. and M.L.C. had a base year of 1980, and as the prime purpose of modelling is to forecast, the decision was taken to base the AAPP and all other price indices on 1980. It should be noted that the AAPP series for February to June of 1977 inclusive includes a subsidy of 5.5p per kg dw paid to farmers from 31 January to 11 June. The pig compound feed price index has been obtained from M.A.F.F.

---

1. See Savin (1978). The MLC kindly gave private access to their model but requested that it was not published in detail.

through the M.L.C. already in index form. The barley feed price usually quoted in pounds sterling per tonne was also converted into an index with a base year of 1980. All price data are collated on a weekly basis and are aggregated and converted into monthly averages. The census data, being the same as that used in the biological model of the previous chapter, are presented in appendix 4c. The average monthly price indices used in this chapter are presented in appendix 6.
In the following section the breeding herd is viewed as a capital flow system involving investment - inflow into the herd - and scrapping in the form of culling - outflow from the herd - in the context of an equilibrium framework. This is done in order to give the reader an insight into how the pig sector is expected to respond - in an equilibrium context - to economic phenomena such as profits and interest rate. Discussion is centered on the number of breeding sows, as they far outnumber the size of the breeding boar herd, and much of the discussion applies equally well to boars as to sows anyway. Having expounded the capital theory, the models for investment and scrapping are presented, examining a variety of econometric approaches.

5.2 The Breeding Herd As A Capital Flow System.

In the biological model of chapter four, the breeding herd was viewed as a system of inflows and outflows expressed by equation 5.2.1 in which $\hat{I}_{t-1,t}$ is an estimate of the true inflow figure minus losses from the system through disease, accidents etc. As was the case with the biological model, the time subscripts represent the trimesters measured between the three sample censuses of April, August and December.

$$HB_t = HB_{t-1} + \hat{I}_{t-1,t} - M_{t-1,t} \tag{5.2.1}$$

For the reasons discussed in chapter four, the derived inflow variable is to be proxied by the pregnant gilt herd variable. In section 4.5b of chapter four, the best estimate of $\hat{I}_{t-1,t}$ from the biological model was deemed to be 1.0863 PG_t so that the function used to generate forecasts of the breeding herd is given by 5.2.2 below.

$$HB_t = HB_{t-1} + 1.0863\ PG_t - M_{t-1,t} \tag{5.2.2}$$

The expression given by 5.2.2 will also be used to generate the econometric model forecasts of the breeding herd and so the task, as it was for the biological model, is to model the inflow of pregnant gilts and the culling of sows and boars. The difference from the biological model is that rather than viewing the system from a biological point of view, the inflow of gilts is now seen as an investment decision, the life of the capital being represented by the lifetime of sows and boars in the breeding herd. At some point in the future, it is deemed more profitable to scrap - cull - the pig rather than maintain it in the breeding herd for its productive capacity. The objective in this section is to describe the conditions for maintaining an equilibrium level of investment and scrapping, the implications of this for culling age and the age distribution of the herd and also to examine the comparative statics of the steady state equilibrium. Finally, reference is made to the role of adjustment costs of investing and scrapping.

We start with the equilibrium conditions to obtain the optimal level of investment and scrapping and, therefore, the optimal culling age. A gilt is added to the breeding herd as an alternative to placing it into the feeding herd for fattening. The young pig will go into the breeding herd rather than the feeding herd as long as the expected present value of future net revenues from the sale of the capital item's production and its scrapping value is greater than the expected revenue from having sold the pig as a consumer good itself. In other words, if the expected future net revenues from selling the progeny of the young gilt over its expected optimal lifetime, plus its own expected cull value are greater than the current fat pig slaughter value net of fattening costs, investment in the young pig will take place. On the scrapping side of the equation, a sow will not be culled if delaying the culling age by one parity - one farrowing period - gives rise to expected discounted future net revenues which exceed the current value of culled sows and boars.

Under conditions of a steady state equilibrium, which are expressed by equation 5.2.3 below, net revenues are such that investment in the breeding herd, that is inflow, is matched by culling from it, so that the breeding herd has no long term tendency to increase or decrease over time.

$$I = M = \theta\, HB = 1/L\, HB \qquad (5.2.3)^{2}$$

The consequences of this are that there will be an optimal age for the retention of pigs in the breeding herd, hence, an optimal culling age and a uniform age distribution throughout the breeding herd. The parameter ' θ ', the proportion of the breeding herd accounted for by culling, is the reciprocal of the optimal culling age, L.

---

2. The steady state assumption negates the need for time subscripts.

Given a situation of steady state equilibrium so that the breeding herd is at its optimal size with an optimal lifetime for the capital, the comparative static effects of changes in a given number of exogenous shocks to the system can be examined. Firstly, consider a permanent rise in the margin of fat pig receipts over costs of maintaining the capital good, *cetertis paribus,* primarily in the form of feed costs.[3] This provides an inducement for producers to increase the optimal size of the breeding herd. They respond in the short run by increasing inflow into, and decreasing outflow out of the breeding herd. Over the longer term inflow increases at a diminishing rate but the culling rate will also gradually increase as the size of the breeding herd increases. Eventually, a new equilibrium will be reached with a larger herd size and a higher level of matched inflow and outflow than under the old equilibrium. If we also abstract from any induced rise in culling price, then there would be a permanent rise in the optimal life of the breeding sow.

Now consider a permanent rise in the scrapping value on the other hand, that is, an increase in the cull price of breeding sows and boars *ceteris paribus*. This will also have the effect of increasing the long run optimal size of the breeding herd through increased inflow. This occurs because pig production is now more profitable, breeding sows and boars now having greater worth as an end product in themselves. In the short term the herd is likely to undergo a temporary reduction in size as barren sows which were about to be served for a final time are now more profitable as culls. The increase in the cull value will shorten the optimal lifetime of the breeding herd because net returns at the margin of culling are now increased. Moreover, because it is almost certain that an increase in fat pig prices is the prime cause of an increase in the price of culls, the position regarding the overall effect of an increase in fat pig price on optimal lifetime is not clear. The certain effect is that the optimal size for the breeding herd will increase.

Investment decisions are invariably tied up with the real rate of interest, so let us consider the case of an exogenous increase in real interest rates. Such a rise reduces the net present value of pig production. The likely consequence of this is to reduce the optimal life of a breeding sow, L, thereby increasing the value of theta due to the fact that the breeding sow is now less profitable as productive capital relative to her scrap value as a culled sow.

As discussed in chapter one, one of the most striking features of the pig sector in the last 20 years or so has been the sharp increase in sow productivity, mainly a result of shortened average weaning periods. An increase in the productivity of

---

3. We abstract from any feedback effects that changing the herd size might have on profits via changes in the future availablility of fat pigs.

sows due to improved technology in the industry reduces costs per pig reared and, therefore, increases profitability in the industry. As discussed above with increased profitability, the optimal herd size will increase, the position with respect to optimal life of the breeding herd being uncertain.

It is the opinion of some in the industry that future productivity gains are to come from gains in genetic engineering and better husbandry.[4] Such gains are likely to come through increasing the productivity of gilts, which are considerably less productive in terms of litter size and length of time taken after weaning to be returned to first service.[5] If there are such increases in the productivity of gilts relative to older sows, a tendency to reduce L because of the improved efficiency of replacement of the less productive sows by the more productive gilts is likely.

Finally, in this theoretical assessment of the breeding herd as a capital flow, the role of adjustment cost is considered. On the inflow side, a farmer wishing to increase the size of his breeding herd quickly will need to hold a large number of young gilts. These gilts will compete with the feeding herd and the rest of the breeding herd for the production unit's resources. These adjustment costs will increase with the speed of increase required and can be associated with the number of in-pig gilts present in the breeding herd. At the scrapping end of the system, there are no obvious adjustment costs associated with the act of culling; except that as culling is reduced, again competition for resources can be considered. A disincentive and, therefore, a barrier to excessively high culling rates is rising productivity of sows at the culling margin with each farrowing - parity - up to about the optimal fifth parity, after which productivity levels off and starts to fall again. Gilts are the least productive members of the breeding herd. The consequences of such adjustment cost for modelling are that decisions to expand or contract the breeding herd are likely to be seen over more than one census period, introducing the possibility of autocorrelation. This possibility is discussed in the modelling sections of the chapter.

Having considered the breeding herd as a capital flow system, and having briefly examined some of the characteristics of the system under equilibrium, the following two sections are concerned with the results of building forecasting models for the proxy variable for investment, pregnant gilts, and the scrapping variable, sow and boar culls.

5.3 The Methodology of Modelling

The pregnant gilt - inflow[6] - model and the culling models are built using a base

---

4. See MLC market survey April 1988, page 3f.
5. See pages 56 to 58 of MLC pig yearbook April 1988.

methodology in which it is intended to incorporate the models developed by Savin and the MLC derivative of the Savin model, both of which are discussed within the chapter. As was the case with the biological models, the sampling period for estimation is 1975 to 1985 inclusive, 1986 and 1987 being made available for out-of-sample forecasting tests. Initially, the inflow and outflow models are built within a dynamic framework incorporating a profit variable and the dependent variable lagged a number of periods. The lags extend in order that they cover a full calendar year, systematically removing from the model specification the longer lagged variables which prove to be insignificant. The possibility of augmenting the models by the inclusion of explicit biological variables is also examined.

In the process of modelling, a number of problems need to be resolved. The first, which has already been referred to, is determining the number of lagged dependent and independent variables which are to be included in the model specification. The usual t-tests for testing the specific significance of estimated coefficients and F-tests on the overall significance of included variables will be employed for this purpose. For each of the inflow and outflow variables, the appropriate profit variable needs to be obtained. Biological knowledge and the lags included in the profit variables in the existing Savin and MLC models are both used to determine the chosen lag. A similar though perhaps less important question regarding the nature of the profit variable is to decide upon the appropriate ratio to use. Two alternatives are considered. The first is what I refer to as the simple profit ratio employed by Savin, and is simply the ratio of the Average All Pig Price, AAPP, and the compound feed price for pigs for any given month. The ratio which will be referred to as the complex profit ratio, is the same as the latter except that the denominator is the arithmetic mean of the compound feed price and the barley feed price, as employed in the MLC model. All prices are in index form, having 1980 as a base year.

As was the case with the biological models, various intervention dummies will be employed as necessary in each case to model any possible outliers, the exclusion of which may have significant effects on the size and sign of seasonal dummy parameter estimates included to model seasonal influences. These intervention dummies primarily explain the Aujezky disease eradication campaign of 1983, the temporary subsidy on pig meat in the first half of 1977 and possible effects on observations at the very start of the sample period resulting from the influences on the industry of the 1974 commodity price increases and the UK's accession to the

6. Because pregnant gilts are being used as a proxy for inflow, the two terms will be used interchangeably.

EEC.

Having obtained satisfactory solutions to the above questions and having obtained a satisfactory model, the specification of the model is investigated with particular interest paid to the possibility of autocorrelation in the given model which may arise through the effects of adjustment costs as discussed towards the end of section 5.2. The derived models and the methodology employed are compared with those of Savin and the MLC.

5.4 The Inflow Model Estimated.

As previously explained and as was the case with the biological model, inflow is to be proxied by the pregnant gilt herd. The model for the pregnant gilt herd is estimated on the sample period 1975 to 1985 inclusive and, like the biological model is a four-monthly - trimestic - model. The initial specification is dynamic in that it includes lagged dependent variables, lagged values of the profit regressor also being included. The lags for the two said variables initially extends to cover one calendar year so that the initial model therefore takes the form of the expression given in equation 5.4.1,

$$PG_t = c_0 + \alpha_0 P_t + \beta_1 PG_{t-1} + \alpha_1 P_{t-1} + \beta_2 PG_{t-2} + \alpha_2 P_{t-2} + \beta_3 PG_{t-3} + \alpha_3 P_{t-3} + \varepsilon_t \quad (5.4.1)$$

where P_t is the appropriate profit variable associated with the pregnant gilt herd at time t and ε_t is a zero-mean white-noise error term. Seasonality has temporarily been ignored.

Putting aside for the moment the question of which of the simple or the more complex profit ratio should be used, it is necessary to obtain an appropriate profit variable before estimation of the model can commence. Specifically, biological knowledge can be used in order to determine the appropriate lag between the level of profit and its subsequent effect on the pregnant gilt herd.

Consider a pregnant gilt at the December census. These gilts will have been put in pig from the middle of August to the middle of November. Assuming a 30 week average age for the gilt to be put in pig, the oldest of the December pregnant gilts would have been born in the middle of January. Assuming that the average age for slaughter of the fat pigs is six months, the decision to put the young pig into the breeding herd must be made before the middle of July. As the decision is unlikely to be made at either extreme, the decision was taken to use the midpoint, that is, three months of age. On this basis, the decision for the gilt put in pig in August is deemed to have taken place in April. Assuming a uniform distribution for the time at which the December gilts were put into pig, the decision period would span April through to July. This happens to be a lag of 5 to 8 months, the same lag derived by Savin using empirical methods, observing the correlation between the pregnant gilt and the profit ratio time series data.

To allow for the fact that gilts can be put in pig from the age of 5 to 8 months, however, a weighted average for the profit ratio lagged 3 to 9 months was derived assuming a uniform distribution for the age at which the gilts are served. The expression for the chosen lag structure is given by 5.4.2 below in which the subscript i is time measured in months.

$$P_t = 1/16\,(P_{i-3} + 2P_{i-4} + 3P_{i-5} + 4P_{i-6} + 3P_{i-7} + 2P_{i-8} + P_{i-9}) \qquad (5.4.2)$$

In other words, the appropriate profit variable for the pregnant gilt herd at time t is a weighted average of the average profit ratio three to nine months earlier inclusive. For example, the gilts recorded as being in-pig in the December census are deemed to be a consequence of level of profit in the previous March through to the previous September inclusive. This lag structure incorporates the lags used by Savin and the MLC in their models for pregnant gilts both of which were arithmetic means of profit in the months i-5 to i-8 and i-3 to i-6 respectively.

Equation 5.4.1 was estimated with the addition of the two seasonal dummy variables, using the lag structure for the profit variable described by the expression in 5.4.2: the resulting residual plot indicated the presence of possible outliers. In order to remove the effect of such outliers on the estimated parameters of the seasonal dummies, the decision was taken to include appropriate dummy variables with which to model the various effects. The dummies concerned are the first three Aujezky disease eradication campaign dummies A83:2, A83:3 and A84:1 and a dummy for August 1977. The residuals at the time of the eradication campaign clearly indicate a fall in the numbers of pregnant gilts which is also the case in August 1977. The reason for the former needs no explanation, and the latter is the result of the temporary subsidy on pig meat from the end of January to the beginning of June, which appears to have resulted in potential gilts having been transferred instead into the feeding herd in order for producers to take full advantage of the temporary subsidy. The addition of these 'outlier' dummies to the regression produced the results presented in equation 5.4.3 below, in which the t-ratios are presented in parentheses.

$$\begin{aligned} PG_t = {} & \underset{(0.94)}{27.34} + \underset{(2.09)}{0.651}\,P_t + \underset{(3.20)}{0.643}\,PG_{t-1} - \underset{(-1.77)}{0.702}\,P_{t-1} + \underset{(1.41)}{0.314}\,PG_{t-2} - \underset{(-1.41)}{0.513}\,P_{t-2} - \underset{(-0.01)}{0.003}\,PG_{t-3} + \underset{(1.43)}{0.335}\,P_{t-3} \\ & + \underset{(1.28)}{5.11}\,AUG + \underset{(0.28)}{0.99}\,DEC - \underset{(-2.71)}{18.50}\,A83{:}2 - \underset{(-1.63)}{10.60}\,A83{:}3 - \underset{(-1.08)}{6.88}\,A84{:}1 - \underset{(-3.56)}{25.13}\,D77{:}2 \end{aligned} \qquad (5.4.3)$$

$$\text{Nobs.} = 28 \quad RSS = 319.66 \quad \hat{R}^2 = 0.586 \quad DT^7 = 0.78$$

The coefficient on the profit ratio variable P_t is significant and has the correct positive sign indicating that an increase in profit produces a lagged increase in inflow into the breeding herd which the producers are aiming to expand. The Durbin T-statistic for regressions including lagged dependent variables indicates no autocorrelation problems with the residuals. The residual plot illustrated that there are no other outstanding outliers missing from the model specification.
The outstanding problem with the estimated regression 5.4.3 is the insignificance of the t-statistics of the estimated parameters for the profit variable and the pregnant gilt variable lagged 2 and 3 trimesters. To test their overall significance, the variables for pregnant gilts and profit lagged three trimesters were dropped from regression 5.4.3 and an F-test on the residual sum of squares was performed. The resulting F-statistic of 1.875 is not significant even at the 10% level when compared with the critical value from the table of F-statistics for 2,14 degrees of freedom. Consequently, the alternative hypothesis of significance for the said lagged variables could be rejected and PG_{t-3} and P_{t-3} removed from the model. The estimated model excluding pregnant gilts and profit lagged three trimesters also produced t-statistics for the estimated parameters of PG_{t-2} and P_{t-2} which were not significant at accepted levels. Consequently, the process for testing the significance of the said variables lagged three periods was repeated for the variables lagged two periods. The resultant F-statistic of 0.73 is clearly not significant, once again leading to a rejection of the null-hypothesis and an acceptance of the model minus the pregnant gilt and profit variables lagged two trimesters. The residual for the first observation available for estimation in the latter regression, that is, 1976:1, was greater than two standard errors above zero. Because this observation is at an extreme of the estimation period and, therefore, has the potential to be misleading in terms of whether or not the residual plot contains a time trend, as well as the possible effects on the estimated seasonal parameters, the decision to include the intervention dummy D76:1, in order to model its effect, was made. The estimated regression for the latter model is given in equation 5.4.4 below.

---

7. The Durbin t-statistic, see Durbin(1970), is given by TSP output for OLS regressions including a lagged dependent variable, and is a large sample statistic similar, though more general statistic, to that of the Durbin H-statistic employed in chapter 3. A significant Durbin t-statistic indicates the presence of first order autocorrelation

$$PG_t = 38.95 + 0.532\,P_t + 0.672\,PG_{t-1} - 0.564\,P_{t-1} + 3.88\,AUG - 2.08\,DEC -$$
(2.93) (2.69) (4.30) (-3.03) (1.54) (-.79)

$$13.64\,A83{:}2 - 4.48\,A83{:}3 - 4.15\,A84{:}1 - 19.56\,D77{:}2 + 16.19\,D76{:}1 \qquad (5.4.4)$$
(-2.20) (-.82) (-.76) (-3.75) (2.55)

Nobs. = 30 RSS = 440.2 $\hat{R}^2 = 0.752$ DT = 0.44

The estimated parameters for the included pregnant gilt and profit variables are all significant in terms of their t-statistics and have the correct signs. Although the RSS statistic is higher than it is in 5.4.3 it is primarily the result of the regression being estimated over a longer time period, and moreover, the adjusted R-squared value of 0.752 is considerably larger than that obtained in the initial regression, implying that 75%, as opposed to 59% of the variation in the pregnant gilt herd is explained by the model presented in 5.4.4. The Durbin t-statistic for regressions incorporating lagged dependent variables indicates no autocorrelation problem in the residuals.

The structure of the dynamic model presented in equation 5.4.4 can be viewed as an unrestricted form of autoregressive model. Again ignoring seasonality and other dummy variables, the basic form of equation 5.4.4, for example, can be written in the following way.

$$PG_t = c_0 + \alpha_0 P_t + \beta_1 PG_{t-1} + \alpha_1 P_{t-1} + \varepsilon_t \qquad (5.4.5)$$

Now assume that the static model given by equation 5.4.6 is AR1 in the error v_t.

$$PG_t = c + \alpha P_t + v_t \qquad (5.4.6)$$

In other words,

$$PG_t = c(1 - \rho) + \alpha P_t + \rho\, PG_{t-1} - \alpha\rho\, P_{t-1} + \varepsilon_t, \qquad (5.4.7)$$

in which ρ is the AR1 parameter and ε_t is the white-noise error term. The expression given in equation 5.4.7 is a restricted form of equation 5.4.5.

If the model estimated in 5.4.4 does indeed have an autocorrelation structure, we would expect the product of the coefficients on P_t and PG_{t-1} to be the negative of the value of the coefficient on P_{t-1}. In 5.4.4, therefore, 0.532 multiplied by 0.67, which equals 0.357, and compares with 0.564. Given that the standard error on the two profit variables are approximately 0.19, the product and the coefficient on lagged profit differ by slightly more than one standard error indicating the possibility of equality. As a further check on the possibility of a first order autocorrelation structure, 5.4.4 was re-estimated using non-linear least squares, LSQ, methodology constraining the coefficients of the profit and lagged pregnant gilt variables to take the first order autocorrelation structure. In other words, the coefficients concerned will be forced to take values such that the product of the estimated coefficients on P_t and PG_{t-1} will be the negative of the value of the estimated coefficient on P_{t-1}. Having estimated the LSQ model, an F-test, in which the null-hypothesis that the constrained and the unconstrained regressions are the same will be refuted if the resultant F-statistic, which compares the RSS statistics of the two models, is larger than the critical value for F. The RSS of 535.12 for the constrained regression is naturally larger than the equivalent statistic for the unconstrained regression, but the resultant F-statistic of 4.05 is not significant at the 5% level compared with the critical value for 1,19 degrees of freedom. The test, therefore, supports the use of the constrained regression and, moreover, the presence of a first order autocorrelation model structure.

This conclusion is similar to the model derived by Savin which was estimated employing the Durbin 2-stage procedure. However, she made the methodological error of excluding seasonal dummies which she appeared to think would fall out of the model as the first order autocorrelation structure was imposed. In addition, Savin made a methodological error in removing from her OLS models, any seasonal dummies which did not have significant t-statistics. This is erroneous because the significance of estimated seasonal dummy parameters is dependent on the season chosen as the base from which to measure the deviation by all other seasons. It is therefore argued that all or none of the seasonal dummies should be included. Rather than using the Durbin 2-stage procedure, which is similar in methodology to that of the aforementioned LSQ model, for estimation of a model imposing the first order autocorrelation structure the Beech-Mackinnon maximum likelihood procedure can be employed instead. The latter offers the advantage of using less degrees of freedom in estimating the regression because no lagged variables are estimated. This gives the dual advantage of releasing an extra observation as well as reducing the number of coefficients to estimate. With the relatively short sample period available for estimation of the trimestic models, increasing the number of degrees of freedom available is considered to be an important factor, and hence the choice of the Beech-Mackinnon estimation procedure over the Durbin two-stage method. The model as estimated using Beech-Mackinnon maximum likelihood is given in equation 5.4.8 and represents the most restrictive form of the general model, where the coefficients of all the variables including the dummies are restricted by the AR term.

$$PG_t = 38.46 + 0.691\,P_t + 3.42\,AUG - 0.48\,DEC -$$
$$\quad (2.60) \quad (4.98) \quad (2.32) \quad (-.32)$$
$$11.29\,A83{:}2 - 7.32\,A83{:}3 - 4.82\,A84{:}1 - 17.67\,D77{:}2 + 3.58\,D76{:}1 + U_t\,,$$
$$(-2.64) \quad (-1.51) \quad (-1.14) \quad (-4.89) \quad (0.76)$$
$$U_t = 0.794\,U_{t-1} \qquad (5.4.8)$$
$$(6.9)$$

Nobs. = 30 RSS = 388.38 $\hat{R}^2 = 0.89$

Again, the key variables have significant coefficients and the correct signs and the estimated coefficient on the August dummy is significantly greater than zero. The RSS statistic is smaller than that reported for the previous OLS regression including lagged dependent and profit variables, and although the DW statistic is not presented in the table of diagnostic statistics because the model was not estimated using OLS, the value of 1.85 and the plot of the residuals give no cause for concern about autocorrelated residuals. As a confirmatory test of the AR1 nature of the model structure, the latter regression was re-estimated using OLS and the residuals put through the Box-Jenkins identification procedure. The resulting correlograms suggested that the autocorrelation in the model was nothing other than first order AR.

One final restriction examined in the general form of the model 5.4.5 was to force α_1 to equal the negative of α_0 so that the two profit level variables are replaced by a single profit change variable. Because one less parameter is estimated, this form of restriction also releases one degree of freedom which is good on parsimony grounds. The results of estimating this model are given in equation 5.4.9.

$$PG_t = 37.90 + 0.652\,PG_{t-1} + 0.550\,(P_t - P_{t-1}) + 3.84\,AUG - 1.98\,DEC -$$
$$\quad (3.25) \quad (4.30) \quad (3.03) \quad (1.57) \quad (-.79)$$
$$13.03\,A83{:}2 - 4.14\,A83{:}3 - 3.85\,A84{:}1 - 19.41\,D77{:}2 + 15.57\,D76{:}1 \qquad (5.4.9)$$
$$(-2.56) \quad (-.83) \quad (-.76) \quad (-3.87) \quad (2.97)$$

Nobs. = 30 RSS = 441.0 $\hat{R}^2 = 0.76$ DT = 0.54

Again, comparing these results with those of the unrestricted form, the restrictions appear to be plausible. The estimated coefficients and the RSS statistic have changed little so that the adjusted R-squared value has increased marginally to 0.76, and the Durbin t-statistic indicates no problem with residual autocorrelations.

A fundamental difference of the latter model compared with the previous models examined is that it is short term rather than long term in respect of the reaction of gilt numbers to a change in the level of profits. The previous models are all specified in terms of the profit level so, for example, an increase in the profit level would produce a permanent increase in the level of gilts. In 5.4.9 however, an increase in the profit level at time t has a partial effect on gilt numbers at time t, but were profits to remain at their new level, because there is no change in the level of profits, they would have no direct effect on the size of the gilt herd. In terms of what we would expect in terms of our initial equilibrium framework, and given that we expect the econometric model to perform better in the medium to longer term rather than the short term, this latter model is not as appealing as the others examined.

The model as represented in 5.4.8 was accepted as the best model for the inflow proxy, the pregnant gilt herd, from having followed the given methodology. Having allowed for the fact that Savin's original model was built at the time when pure quarterly data were available, my model incorporates that developed by Savin. The fundamental differences include a wider range of model specifications having been considered here, eliminating other possible autocorrelation structures. Also, Savin's methodological errors have been eliminated and, having determined that the autocorrelation in the model is of the first order, the model has been estimated using the Beech-Mackinnon rather than the Durbin two-stage procedure which is more cumbersome to estimate and slightly less efficient in terms of use of degrees of freedom.

The trimestic model developed by the MLC, supposedly based upon the Savin model is, however, different in a number of ways: three main differences can be identified. Firstly, the lag structure of the profit variable is shorter by two months, although there is no indication of how the given lag structure was derived, and the profit variable, is the same as the simple profit ratio used by myself, but replacing the compound feed price index denominator by the arithmetic mean of the compound feed price index and the index of barley feed. Secondly, the model does not account for seasonality. The final and most fundamental difference in the model structure is that it attempts to model the autocorrelation in a less general manner than the Savin-derived approach, simply including a lagged dependent variable as a regressor, implying that the autocorrelation is the result of stickiness in adjusting the size of the pregnant gilt herd. If my model is correct then it implies that the MLC model is misspecified in terms of its dealing with autocorrelation. Estimating the MLC model adding the seasonal and other intervention dummies included in 5.4.4 produces a DT statistic of 2.50, which is significant at the 5% level giving strong indication of the presence of first order residual autocorrelation. This result backs up the argument that the MLC model is indeed misspecified.

Many of the above regressions were re-estimated exchanging the simple and the MLC profit ratios, with results which indicated little or no improvement in using the more complex of the two ratios. Consequently, and in view of the fact that the simpler profit ratios are easier to calculate and require less data, use of the MLC ratio will no longer be considered in this thesis. Exchanging my profit variable lag structure for those chosen by Savin and the MLC, showed my lag structure to be superior in terms of residual sums of squares.

Having arrived at a satisfactory but fundamentally short run model, it was decided to investigate the possibility of explicitly including a long run element by adding the breeding herd variable lagged three periods to the unrestricted model 5.4.4. This is the lag derived in the biological model for pregnant gilts discussed in section 4.5g. The effect of adding this variable to the estimated model was to eliminate the significance of the intercept term. Hence, as an alternative to the model specification represented by 5.4.5, a proportional model in which the intercept term is replaced by the lagged breeding herd variable was deemed appropriate, so that the general form of the unrestricted model is that given in 5.4.10.

$$PG_t = \gamma_1 HB_{t-3} + \alpha_0 P_t + \beta_1 PG_{t-1} + \alpha_1 P_{t-1} + \varepsilon_t \qquad (5.4.10)$$

The re-estimated model is presented in equation 5.4.11 below in which the seasonal dummies are now tied to the lagged breeding herd term as they were in the biological proportional models presented in chapter four.

$$\begin{aligned} PG_t = {} & \underset{(3.37)}{0.042\, HB_{t-3}} + \underset{(3.39)}{0.595\, P_t} + \underset{(5.15)}{0.736\, PG_{t-1}} - \underset{(-3.84)}{0.672\, P_{t-1}} + \underset{(1.80)}{0.005\, AugHB_{t-3}} - \underset{(-.81)}{0.002\, DecHB_{t-3}} - \\ & \underset{(-2.69)}{16.26\, A83{:}2} - \underset{(-1.22)}{6.55\, A83{:}3} - \underset{(-1.24)}{6.75\, A84{:}1} - \underset{(-4.31)}{22.13\, D77{:}2} + \underset{(2.88)}{17.66\, D76{:}1} \qquad (5.4.11) \end{aligned}$$

Nobs. = 30 RSS = 399.0 $\hat{R}^2 = 0.776$ DT = -0.075

The estimated unrestricted regression presented in 5.4.11 is directly comparable with that of 5.4.4. The estimated coefficients of the two models are quite similar for the variables included in both, although the t-statistics in 5.4.5 are all larger than the those presented in 5.4.4. The diagnostic test statistics of 5.4.5 are all better than those of 5.4.4, the adjusted R-squared value indicating that 77.6% rather than 75.2% of the variability in the pregnant gilt series is explained by the latter regression. In view of this, the proportional model was re-estimated imposing the same three restrictions to check for autocorrelation as were imposed on the unrestricted intercept model 4.5.4. The results of estimation were similar to the results obtained for the intercept model in that the restrictions appeared to be valid, none of the estimated coefficients changing by significant amounts. The RSS statistics for the restricted and unrestricted models estimated using the intercept and the proportional approaches are presented in table 5.1 below in which the definitions of the restrictions refer to the parameters of the unrestricted general equations presented in equations 5.4.5 and 5.4.10 above.

Table 5.1
The RSS Statistics From the Estimated Restricted and Unrestricted Forms of Intercept and Proportional Econometric Models for Pregnant Gilt.

RESTRICTIONS	MODEL SPECIFICATION INTERCEPT	PROPORTIONAL
Unrestricted	440.20	398.97
AR1 (Beech-Mac)	388.38	407.53
$\alpha_1 = -(\alpha_0 \beta_1)$	535.12	521.92
$\alpha_1 = -\alpha_0$	440.96	403.37

As discussed above in reference to the results of modelling using the intercept approach, there is little to choose between the models in terms of estimated coefficients and all the models show no problems of residual autocorrelation. It therefore seems reasonable to analyse the models in terms of the RSS statistics presented in the table above. The obvious thing to notice is the superiority of the proportional models relative to their intercept equivalents, the exception being the Beech-Mackinnon AR1 model, where the RSS for the intercept model rather curiously is actually lower than for the unrestricted model. This could only be put down to the way in which TSP calculates the RSS for AR1 models. Given that the proportional approach is accepted as the better of the two, the choice of best model lies between the Beech-Mackinnon pure AR1 model and the profit difference model. On the grounds that the latter is less desirable from a theoretical stance, the Beech-Mackinnon proportional model is chosen as the best econometric forecasting model for the pregnant gilt herd, the estimated equation being presented in 5.4.12 below.

$$PG_t = \underset{(2.25)}{0.031}\, HB_{t-3} + \underset{(7.11)}{0.798}\, P_t + \underset{(2.22)}{0.0039}\, AugHB_{t-3} - \underset{(-.25)}{0.0004}\, DecHB_{t-3} -$$

$$\underset{(-2.62)}{11.43}\, A83{:}2 - \underset{(-1.56)}{7.80}\, A83{:}3 - \underset{(-1.35)}{5.98}\, A84{:}1 - \underset{(-4.88)}{18.15}\, D77{:}2 + \underset{(0.63)}{3.03}\, D76{:}1 + U_t$$

$$U_t = \underset{(7.4)}{0.8105}\, U_{t-1} \qquad (5.4.12)$$

Nobs. = 30 RSS = 407.5 $\hat{R}^2 = 0.89$

5.5 The Outflow Model Estimated

Similar methodology to that employed above is now applied to build an equivalent outflow model. Because the methodology for investigating the pregnant gilt model was discussed at considerable length in the previous section, the methodological discussion here will be considerably briefer. As referred to in the previous section, use of the more complex of the two profit ratios considered has been rejected. Also, because of the apparent superiority of the proportional models over the intercept approach, and the appeal they have by explicitly incorporating biological as well as economic phenomena, only the proportional approach will be considered.

The first consideration when attempting to model culling is to derive an appropriate lag for the profit variable. It seems reasonable to assume that the producer will make the decision to cull shortly after a sow has farrowed. If the sow is to be culled, she can be fattened ready for culling during the 3 to 4 week weaning period, the act of culling following shortly after weaning has taken place. In view of this, a lag of one month between the decision to cull based on a given level of profits and the act of culling is chosen so that culling in February, for example, is deemed to be a results of profit levels in January. This lag compares with a zero lag used by the MLC, and a lag of two months which Savin derived purely on statistical grounds. To allow for variability in the decision process and weaning period, and to incorporate the lags used by the MLC and Savin, the decision was taken to use a weighted average profit ratio similar to that used for the profit ratio variable in the pregnant gilt models presented earlier. The weighted average which is constructed giving equal weight to MLC's zero lag, my one month lag and Savin's two month empirical lag produced the following profit ratio variable,

$$\Pi_t = 1/12\,(P_i + 2\,P_{i-1} + 3\,P_{i-2} + 3\,P_{i-3} + 2\,P_{i-4} + P_{i-5}), \qquad (5.5.1)$$

where i is time measured in months, t is time measured in trimesters and P_i is the profit ratio in a given month. For example, therefore, cullings in December through to March, are deemed to be a function of profits in October to March inclusive.

Temporarily ignoring seasonality and 'outliers', the initial model including lags on the appropriate variables to cover one calendar year, is that given in equation 5.5.2 in which the appropriate breeding herd variable lagged one trimester is included as the biological variable.

$$M_{t-1,t} = c_1 HB_{t-1} + a_0 \Pi_t + b_1 M_{t-2,t-1} + a_1 \Pi_{t-1} + b_2 M_{t-3,t-2} + a_2 \Pi_{t-2} + b_3 M_{t-4,t-3} + a_3 \Pi_{t-3} + \varepsilon_t \quad (5.5.2)$$

The latter model was estimated including seasonal dummy variables and an appropriate number of intercept dummies were added to the model, determined by the observations with large residuals in the residual plot. The intervention dummies included were four Aujezky dummies A83:2 to A84:2 and a dummy for the April 1977. Having estimated the regression, the longer lagged variables were systematically dropped from the regression, F-tests being performed as they had been in the pregnant gilt model, to ascertain the overall significance of the variables removed. As occurred in the case of pregnant gilts, the variables lagged two and three periods were deemed to be insignificant and were, therefore, removed from the model. The resulting estimated regression is presented in equation 5.5.3.

$$\begin{aligned} M_{t-1,t} = {} & \underset{(3.50)}{0.0694\, HB_{t-1}} - \underset{(-4.7)}{0.827\, \Pi_t} + \underset{(6.56)}{0.677\, M_{t-2,t-1}} + \underset{(3.17)}{0.613\, \Pi_{t-1}} - \underset{(-4.1)}{0.013\, AugHB_{t-1}} + \underset{(1.79)}{0.004\, DecHB_{t-1}} + \\ & \underset{(6.19)}{31.78\, A83{:}2} - \underset{(-1.63)}{11.07\, A83{:}3} - \underset{(-2.1)}{11.97\, A84{:}1} - \underset{(-1.53)}{7.66\, A84{:}2} + \underset{(2.13)}{10.61\, D77{:}1} \qquad (5.5.3) \end{aligned}$$

Nobs. = 31 RSS = 394.0 $\hat{R}^2 = 0.90$ DT = -0.01

The results of estimation appear to be satisfactory in term of the parameters estimated and the overall diagnostics of the model. The sign on the profit variable is negative as expected, indicating that farmers respond to an increase in profits in the short run by decreasing culling and increasing the size of the breeding herd in order to increase future production of fat pigs. The August seasonal dummy parameter is significantly negative and that for the December positive and almost significant at the 5% level. The Aujezky dummies illustrate how culling increased at the time of the eradication campaign but was low in the following months, presumably as farmers were concerned with replenishing the breeding herd. The subsidy dummy measuring the effect of the 1977 temporary subsidy on pig meat indicates that the level of culling was still abnormally high at the time of the subsidy. The temporary subsidy was used to give relief to a depressed pig sector typified by a high culling level. The adjusted R-squared statistic indicates that the regression explains 90% of the variation in cullings over the estimation period and the Durbin t-statistic is almost zero indicating no residual autocorrelation problem.

Again, the regression presented in 5.5.3 can be regarded as an unrestricted autoregressive specification and so the restricted forms of autocorrelation analysed in the pregnant gilt model are imposed on 5.5.3 to investigate the form of any autocorrelation which might be present. The results of estimating the various restricted models are summarised by the RSS statistics as presented in table 5.2 below.

Table 5.2
The RSS Statistics From the Estimated Restricted and Unrestricted Forms of Proportional Econometric Models for Culling.

RESTRICTIONS [8]	RSS
Unrestricted	394.0
AR1 (Beech-Mac)	687.9
$a_1 = -(a_0 b_1)$	400.7
$a_1 = -a_0$	500.7

The results contrast with those of the pregnant gilt models in that the Beech-Mackinnon AR1 specification, in which all the included variables are subject to the autoregressive constraint, appears to be too great a restriction, the RSS increasing from 394.0 to 687.9. The clear choice of restriction in this case lies with the non-linear estimation presented in 5.5.4.

$$M_{t-1,t} = \underset{(3.67)}{0.071}\, HB_{t-1} - \underset{(-4.7)}{0.818}\, \Pi_t + \underset{(7.24)}{0.696}\, M_{t-2,t-1} + 0.818{*}0.696\, \Pi_{t-1} - \underset{(-4.08)}{0.013}\, AugHB_{t-1} + \underset{(1.90)}{0.005}\, DecHB_{t-1} +$$

$$\underset{(6.41)}{30.86}\, A83{:}2 - \underset{(-2.12)}{12.79}\, A83{:}3 - \underset{(-2.45)}{13.12}\, A84{:}1 - \underset{(-1.81)}{8.50}\, A84{:}2 + \underset{(2.10)}{10.21}\, D77{:}1 \qquad (5.5.4)$$

Nobs. = 31 RSS = 394.0 $\hat{R}^2 = 0.91$ DT = 0.20

8. The parameters used to define the restrictions come from the following model in which seasonality and other dummy variables are ignored.

$$M_{t-1,t} = c_1 HB_{t-1} + a_0 \Pi_t + b_1 M_{t-2,t-1} + a_1 \Pi_{t-1} + \varepsilon_t$$

The estimated parameters are very similar to those of the unrestricted regression and the small increase in the RSS brought about by the restriction is more than offset by the small increase in degrees of freedom through having to estimate one less parameter, so that the adjusted R-squared statistic increases to 91%. The Durbin t-statistic indicates that all is well in terms of autocorrelation in the residuals.

The MLC model which is not presented for confidentiality reasons, but which again models autocorrelation by using a lagged dependent regressor, was estimated adding seasonal dummies and the relevant outlier dummies. The results of estimation were satisfactory in terms of autocorrelation, the residuals producing a Durbin t-statistic of 0.87, so unlike the MLC's pregnant gilt model, their culling model is not misspecified, but the RSS of 965.9 and the adjusted R-squared statistic of 0.764 are not as good as those obtained by the proportional models discussed above.

Savin modelled culling as a percentage of the breeding herd at the previous census. This approach, which implicitly includes the biological element incorporated into my proportional model was chosen because, as was explained in the biological model for culling, the vast majority of sows culled between census dates will have been classed as sows in-pig at the previous census. In the light of this approach, the decision was taken to estimate the unrestricted and the restricted forms analysed for actual cullings, using the specification given in 5.5.5, in which seasonality and other intercept dummies are temporarily ignored.

$$MHB_{t-1,t} = C_1 + A_0 \Pi_t + B_1 MHB_{t-2,t-1} + A_1 \Pi_{t-1} + \varepsilon_t \;, \qquad (5.5.5)$$

where $MHB_{t-1,t} = M_{t-1,t} / HB_{t-1}$.

Estimation of the unrestricted form given in 5.5.5, including the intercept dummies and two seasonal dummies produces the regression presented in 5.5.6.[9]

$$\underset{}{MHB_{t-1,t}} = \underset{(1.84)}{0.0435} - \underset{(-5.1)}{0.0011\,\Pi_t} + \underset{(6.51)}{0.867\,MHB_{t-2,t-1}} + \underset{(3.96)}{0.0009\,\Pi_{t-1}} - \underset{(-3.9)}{0.013\,AUG} + \underset{(2.01)}{0.0054\,DEC} +$$

$$\underset{(5.27)}{0.031\,A83{:}2} - \underset{(-1.75)}{0.014\,A83{:}3} - \underset{(-2.15)}{0.015\,A84{:}1} - \underset{(-1.75)}{0.011\,A84{:}2} + \underset{(1.68)}{0.01\,D77{:}1} \qquad (5.5.6)$$

Nobs. = 31 RSS = 0.000545 $\hat{R}^2 = 0.86$ DT = -0.45

9. Because of the presence of the intercept, the seasonal dummies are now the usual intercept seasonal dummies, unlike the proportional models where the seasonal dummy is multiplied by the breeding herd lagged the relevant period.

The estimated coefficients of all the prime variables take the correct signs and are significant as measures by their t-statistics. The Durbin t-statistic indicates that all is well in terms of the residuals and the lack of autocorrelation, though the adjusted R-square value of 0.86 is not as high as the equivalent 0.90 from the model for actual culling presented in equation 5.5.3. The three restricted forms of model looked at in this chapter were estimated for the cull percentage variable and the fitted values converted into equivalent actual culling numbers. These fitted actual cullings were then compared with the true culling figures and an RSS statistic calculated as presented in table 5.3 below. These RSS statistics are directly comparable with the RSS statistics presented for the actual cull models as given in table 5.2 above.

Table 5.3
The RSS Statistics For Actual Culling Derived From the Estimated Restricted and Unrestricted Forms of Econometric Models For the Cull percentage Variable.

RESTRICTIONS [10]	RSS
Unrestricted	417.4
AR1 (Beech-Mac)	583.9
$A_1 = -(A_0 * B_1)$	427.6
$A_1 = -A_0$	525.1

Comparing the RSS statistics of tables 5.2 and 5.3, the results are similar in that it is the $A_1 = -(A_0 * B_1)$ restriction which is the most valid and the pure AR1 specification which is the least valid as measured by the increase in the RSS from that of the unrestricted model. Three of the four comparisons with the equivalent statistics in table 5.2 indicate that the fitted values from the actual culling models are a better fit to actual cullings than the converted fitted values from the cull percentage models. The exception to this is the Beech-Mackinnon AR1 specification, for which the RSS statistic of 583.9 is 15% lower for the fitted values from the cull percentage models than those produced by the AR1 model for actual culling. This AR1 model is the equivalent of the model derived by Savin, except that it does not contain her methodological errors and the model has been estimated using Beech-Mackinnon rather than the two-stage Durbin-Watson procedure. The results of the analysis imply that Savin may well have been able to improve her model had she not restricted her model fully by use of the two-stage procedure.

10. The parameters used to define the restrictions come from equation 5.5.5

Because the culling percentage dependent variable takes values between zero and unity by definition, the decision was taken to estimate the unrestricted model expressed in its general form in equation 5.5.5, using a logistic transformation of the dependent variable. The theory of the logistic model and the estimation of the model is outlined in the following section.

5.6 A Logit Model For Cull Percentage

5.6a The Theory of Logit Modelling

Savin's approach was to model culling as a percentage of the breeding herd at the previous census. Because the dependent variable is a percentage it can only take values within the unit range and this creates some problems for OLS estimation as discussed below. Thus it was decided that it would be appropriate to attempt a Logit approach to model the cull percentage.[11] This technique allows explicitly for a dependent variable which is defined only on the unit interval. Let F represent the vector of n sample proportions f_i, that is, the number culled divided by the total number in the one period lagged breeding herd. The model to be estimated is given as follows.

$$
\begin{aligned}
f_i &= x_i' \beta + u_i \qquad\qquad \text{for } i = 1,2,\ldots\ldots,n; \\
&= \phi_i + u_i \text{, say} \qquad\qquad (5.5.7)
\end{aligned}
$$

where x_i is a $(k\times 1)$ vector of explanatory variables, β being a $(k\times 1)$ vector of unknown parameters and u_i is a disturbance term. Ordinary least squares estimation of the above expression yields the linear probability estimator of β, denoted, b. The problem with using OLS estimation is twofold: firstly, although it is appropriate to assume that the disturbances have zero mean, the variance of u, $E[u^2] = W = \phi(1-\phi)/n$, dropping observational subscripts, and thus the model is heteroskedastic resulting in inefficient OLS estimates of β. The heteroskedasticity problem is overcome by use of feasible generalised least squares (GLS) estimation, where each of the dependent and independent variables are weighted, that is, multiplied by the reciprocal of the square root of w, where unknown parameters in w are replaced by b. OLS can then be performed on the transformed variables producing asymptotically efficient estimates of β. Secondly, there is no guarantee that OLS, or indeed GLS estimation will produce

11. For a discussion of the reasoning and theory of logit models see Judge et al. (1982).

fitted values which lie within the unit interval. Not only is this unfortunate, but GLS may not even be operational since there is no reason why OLS estimation should not produce negative estimates of the variance of u.

The way that the logit model imposes the required constraint on the estimated values of ϕ is to equate ϕ with the following expression.

$$\phi_i = 1/\{1 + \exp(-x_i'\beta)\} \tag{5.5.8}$$

Because exp(-z) is always greater than zero, the denominator in the above expression is always greater than unity and, therefore, ϕ_i always lies within the unit interval.

An estimate of f can be obtained by employing non-linear least squares on the original regression, however, it is easily shown that the logistic transformation $\mathrm{Log}\{\phi/(1-\phi)\}$ is equal to $x_i'\beta$ and hence, OLS regression of the logistic transformation can be used. Letting $y = \mathrm{Log}\{f/(1-f)\}$,

$$y_i = \mathrm{Log}\{\phi_i/(1-\phi_i)\} + e_i \quad = \quad x_i'\beta + e_i; \tag{5.5.9}$$

which is valid as long as $f \neq 0$ or 1.

Zellner and Lee (1965) show that Var(e) is given by $1/n\,(\phi(1-\phi))$, - denoted by $1/V$ - and so the steps to modelling a limited dependent variable can be described as follows:-

i. Generate values for the logistically transformed variable y_i using the observed sample proportions f_i.

ii. fit equation 5.5.9 by OLS.

iii. generate V by firstly generating ϕ using the fitted values obtained from the OLS regression in step 2.

iv. do weighted, (generalised), least squares regression using the square root of V as weights.

5.6b The Logit Cull percentage Model Estimated

The procedure outlined above was applied to the cull percentage variable using the unrestricted model specification generalised in equation 5.5.5. The model was estimated for the logistically transformed variable y_t defined as;

$$y_t = \mathrm{Log}\{MHB_t/(1-MHB_t)\}. \tag{5.5.10}$$

The model fitted has an adjusted R-square value equal to 0.85 and a Durbin-

Watson statistic of 2.16, and all the estimated parameters were significant or almost significant at the 5% level. The fitted values were transformed, firstly into fitted values for the cull percentage variable MHB_t and then into fitted values for actual cullings. The latter were compared with the actual recorded culling numbers and an RSS statistic of 405.5 resulted. This statistic compares directly with the 417.4 for the RSS derived from the fitted values of the non-logistic MHB_t model as presented in table 5.3, and indicates, therefore, that the logistic transformation model gives a slight improvement to the explanatory power of the unconstrained model. On the basis that the actual culling models are on the whole better at modelling culling than the equivalent cull percentage models, and given that the logistic approach would appear not to improve the explanatory power of the cull percentage approach sufficiently enough to improve on the performance of the model for actual culling, the latter are deemed to represent the best class of forecast models for the trimestic culling data and the decision not to pursue the logistic approach further was taken. In addition to the above reasons, the actual culling models are less cumbersome to use as forecasting models in that they do not require transformations of variables in order to model and derive the required forecasts. Of the models for actual culling, the best restricted form of model is the non-linear least squares model as given in equation 5.5.4, in which the parameter on the profit variable lagged one period is restricted to equal the negative of the product of the parameters on the lagged dependent variable and profit at time t. This model is therefore chosen as the best econometric model for forecasting trimestic culling.

5.7 Conclusions

In this chapter, having taken a theoretical look at an economic system for the breeding herd in an equilibrium framework, a recursive forecasting model of the breeding herd has been developed by modelling investment - inflow - in the form of the pregnant gilt herd and outflow - scrapping - in the form of culling. Various approaches to modelling the two key variables have been considered and compared with the approaches taken by the MLC and Savin, on whose model the MLC model is based. For both the derived pregnant gilt and culling models, the chosen specification was viewed as an unconstrained form of autoregressive model, and as such, various autocorrelation restrictions were imposed on the models to ascertain the validity of the restrictions. Not only did such restrictions help to identify the nature of the autocorrelation in the models but the parameter restriction released degrees of freedom, which could be helpful, given that the number of observations is not large.

The various approaches and restrictions were compared in terms of the estimated models' RSS statistics, which were directly comparable given that the estimation period and included dummy variables were the same for all compared models. Both the culling and the pregnant gilt models explicitly included a biological element in the form of the breeding herd lagged an appropriate period. The model chosen for the pregnant gilt herd is the most restrictive AR1 model estimated using the Beech-Mackinnon maximum likelihood technique. On the other hand, this appears to be the least satisfactory restriction for the cull model. The best model for the latter variable is the one in which only the parameter on the lagged profit variable is constrained, as estimated by non-linear least squares. A logistic modelling approach for limited dependent variables proved to give little or no improvement and was therefore rejected in favour of the existing models. Before proceeding to forecasting, a brief summary is given of some of the economic implications of these gilt and culling models. In particular, ignoring seasonality and assuming the system to be in equilibrium with a profit ratio of 100, a one percent sustained one unit increase in the profit ratio implies a short run increase in gilts of 0.798 and a decrease in culling of 0.818. These two effects cause a 1.65 increase in the breeding herd which implies a short run supply elasticity of 0.2. This figure, though not directly comparable, is in line with equivalent short run elasticities produced by other studies, for example, Westcott (1985), and McClements (1971). The long run elasticity is constrained to take a value of unity because of the proportional specification of the model. There is convergence to a new equilibrium, half the change being made within one year of the initial change in profit. The new eqilibrium position is obtained through both the short term effects of profits increasing inflow and decreasing culling resectively, followed by the longer term effect of the increase in the breeding herd increasing both inflow and culling. The net effect on culling becomes positive at a point between the fifth and sixth period after the initial change in profit. Of course I have abstracted here from the fact that increase in the breeding herd will eventually increase the number of clean pigs available which will then produce a fedback effect on profits. Long run elasticities calculated by Jones (1958) and McClements (1971) measure between 2 and 2.5.

The two chosen models can now be used in combination with the expression for the breeding herd, given in equation 5.2.2, to produce trimestic forecasts of the breeding herd which can then be compared with those of the univariate statistical models developed in chapter three and the biological model of chapter four.

CHAPTER SIX

MONTHLY UNIVARIATE BOX-JENKINS MODELS FOR CULLING, FAT PIG SLAUGHTER AND PRICES

6.1 Introduction

In this chapter , the Box-Jenkins univariate time series methodology outlined in Chapter Two and used in Chapter Three to build models for the quarterly breeding herd data is applied again to build monthly models. Seasonality is now a monthly phenomenon so that a seasonal difference, for example, is achieved by taking a twelfth and not a fourth difference. Five monthly series are modelled; the two slaughter categories, sow and boar cullings and fat pig slaughter; the two price indices, real average all pig price, AAPP, and the real compound pig feed price, which are used to derive a profit ratio for the industry throughout the thesis, and the final series is the profit ratio itself.

Because of the need to look at real rather than nominal AAPP and Compound feed prices, a deflator was required. The first deflator considered was an index of agricultural input prices, 'prices of the means of agricultural production - goods and services currently consumed'. Although the data were obtained and models built using this deflator, the results of model building provided no obvious improvements over similar models built using the RPI deflator. Since forecasts of the RPI are readily available from external sources, the decision was made giving preference to the use of the RPI deflator, thereby reducing the number of variables to be forecast by one. Thus, whenever real prices are referred to, the deflator is always the RPI with a base year of 1980.

The rationale for building the slaughter models is to derive forecasting models, the results of which can be compared with the equivalent biological models built in chapter four, and bivariate Box-Jenkins models including profit as the explanatory variable to be built in the following chapter. For the time series models it is their short term forecasting abilities which are expected to be particularly useful. The price index and profit models are of interest because they will provide a simple technique for forecasting the said series, the forecasts then being available to provide data which can be used by the models including profit as an explanatory variable to produce longer term forecasts of inflow and outflow to and from the breeding herd. The penultimate section in the chapter compares the forecasting ability of the profit model compared with the forecasts of the profit ratio derived by forecasting the AAPP and the compound feed price and dividing the former by the latter.

The univariate slaughter models and the profit model will play an integral part in the identification and estimation of the bivariate models for the two slaughter categories discussed in chapter seven.

The models are presented in a similar fashion to the quarterly univariate models of chapter three. Only the first of the models, that for culling, is described in detail in order to give the reader an insight into how the Box-Jenkins methodology is applied to monthly rather than quarterly data. The models are built on the sample period 1975-1985 in order to make them compatible with the biological and econometric models. The data for 1985 are used as an in-sample forecast period for diagnostic checks and the following two years provide an out-of-sample period. A list of all the data used in the analysis in this chapter is presented in appendix 6.

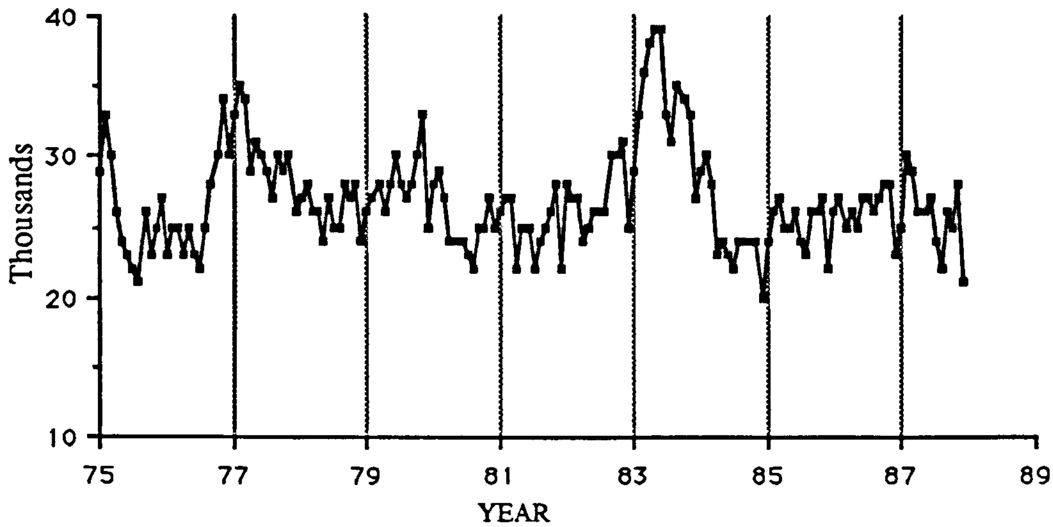
6.2 A SARIMA Model For The Culling Of Sows and Boars

The series 'cullings' refers to the culling of sows and boars from the breeding herd. In the previously developed biological model, this series is an important factor in determining the future size of the breeding herd in that it is a measure of outflow from it. The culling data are collected on a monthly basis by M.A.F.F., and are available from 1968:1 to 1987:12.

In order to compare the results of the chosen Box-Jenkins SARIMA model with those of the biological model, the model is to be identified and estimated on the sample 1975:1-1985:12. Because of the way the data are collated, some of the figures represent cullings from a 5 rather than a 4 week month, giving the impression of a larger number of cullings in certain months. To remove any problems that this might cause, the data are adjusted by taking 4/5 of the culling figure for a 5 week month so that all months represent 4 week periods. Since the data are monthly there are no degrees of freedom problems and the data for 1986-87 provide a comparison period for the out-of-sample forecasts.

The SARIMA model is built using the same methodology as that used to build the quarterly models, except that autocorrelations up to lag 40 are examined, and the Q-statistic represents the first 25, as opposed to the first 20, residual autocorrelations in order to catch any omitted seasonal effect likely to show itself at and around lag 24.

Figure 6.1
A Plot Of The Monthly U.K. Sows and Boar Culling Series 1975:1-1987:12



The plot of the sows and boars culling series from 1975:1-1987:12 looks reasonably stationary, although there is evidence of a slow downward trend, halted only by a relatively large increase in cullings peaking in May/June of 1983. This period corresponds closely with the peak culling period of the Aujezky disease eradication campaign. This heavy culling period is followed by a pronounced fall in cullings, the figures falling to their lowest point in December of 1984 when only 20,000 sows and boars were culled. The cullings in 1985 and 1986 show a much more stable series. It is difficult to tell from the plot whether or not seasonality is present in sow and boar culling.The autocorrelations resulting from the Box-jenkins identification procedure are reproduced in table 3.10.

The autocorrelations of the series in levels indicate non-stationarity in that they die away very slowly, most of the autocorrelations up to and including lag 26 being significantly large. Those of the first differenced series indicate stationarity at the non-seasonal lags, however, the autocorrelations at the seasonal lags 12, 24 36 etc. die away only slowly, implying seasonal non-stationarity. To remove this problem a seasonal difference was taken in addition to the first difference already used.

Table 6.1
Autocorrelations For The Sows and Boars Culling Series.

SERIES	AUTOCORRELATIONS									
	1	2	3	4	5	6	7	8	9	10
$(1-B)^0(1-B^{12})^0$ M	.74	.62	.53	.48	.36	.23	.19	.17	.02	-.08
$(1-B)^1(1-B^{12})^0$ M	-.24	-.10	-.04	.12	.01	-.15	-.06	.27	-.11	-.14
$(1-B)^1(1-B^{12})^1$ M	-.12	.11	.14	-.20	.17	.01	-.08	.19	-.10	-.05
SERIES	AUTOCORRELATIONS									
	11	12	13	14	15	16	17	18	19	20
$(1-B)^0(1-B^{12})^0$ M	-.10	-.05	-.23	-.30	-.32	-.29	-.35	-.38	-.33	-.26
$(1-B)^1(1-B^{12})^0$ M	-.17	.44	-.18	-.11	-.11	.21	-.07	-.16	-.05	.20
$(1-B)^1(1-B^{12})^1$ M	.03	-.53	.04	-.07	-.04	.01	-.08	-.06	-.02	-.07
SERIES	AUTOCORRELATIONS									
	21	22	23	24	25	26	27	28	29	30
$(1-B)^0(1-B^{12})^0$ M	-.28	-.28	-.21	-.06	-.18	-.20	-.16	-.06	-.10	-.12
$(1-B)^1(1-B^{12})^0$ M	-.04	-.14	-.13	.51	-.19	-.09	-.11	.25	-.02	-.10
$(1-B)^1(1-B^{12})^1$ M	-.03	.05	.00	.07	.04	-.02	.01	.08	.02	.03
SERIES	AUTOCORRELATIONS									
	31	32	33	34	35	36	37	38	39	40
$(1-B)^0(1-B^{12})^0$ M	-.07	-.03	-.08	-.11	-.06	.05	-.08	-.07	-.06	.00
$(1-B)^1(1-B^{12})^0$ M	.02	.15	-.02	-.18	-.14	.44	-.22	-.03	-.09	.16
$(1-B)^1(1-B^{12})^1$ M	.01	.00	.10	-.09	-.09	.09	-.12	.06	-.01	-.11

The correlograms of the differenced series illustrate the need for seasonal parameters, although there are also relatively large auto and partial autocorrelations at lags which are multiples of 4. As the autocorrelation at lag 12 is so dominant, and the partials at lags 12 and 24 are also large, the initial identification was that of an SMA model of order one.Although the parameter coefficient was highly significant, the Q-statistic at lag 25 indicated that the residuals were not from a white noise process, and consequently the model required augmentation. The residual autocorrelations at lags 4 and 12 were significantly large when compared with a Quenouille statistic of 0.183. This residual autocorrelation pattern, and the correlograms of figure 6.2 suggested that the appropriate overfit was an SAR parameter.

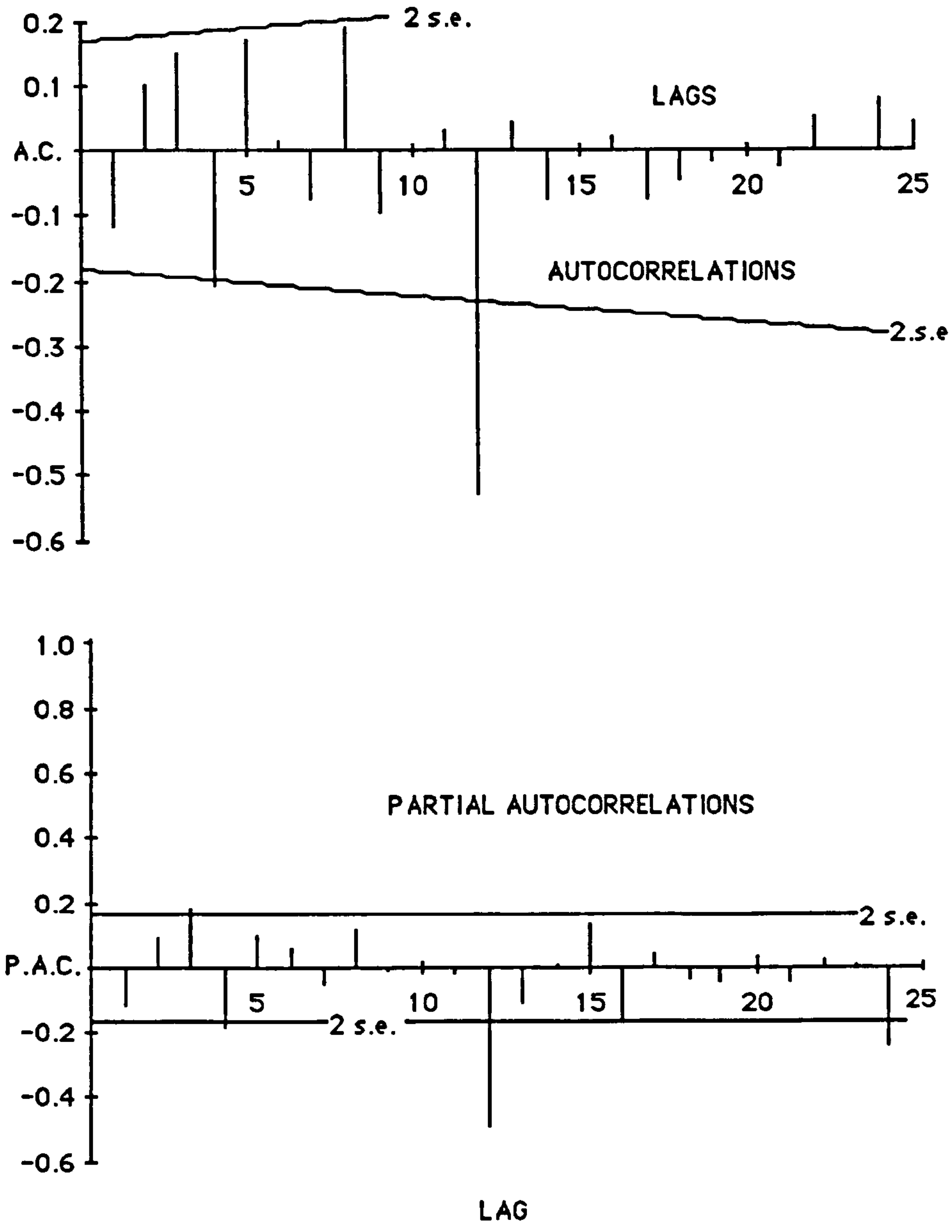
Estimation of the augmented model shows the additional parameter to be significant and its inclusion has the desired effect of removing all significant residual autocorrelations at seasonal lags.

FIGURE 6.2

The Correlograms of the Series $(1-B)^1(1-B^{12})^1M$.

Mean = 0.067

S.E. = 2.68



The sole remaining problem was the residual autocorrelation at lag 4 which verges on significance. Although not a seasonal lag, the fact that four months is the approximate length of the gestation period in pigs justified the decision to further overfit the model with an AR parameter at lag 4. The latter augmentation proved to be the final adjustment to the model.

Table 6.2[1]
The Results of Model Estimation of The Culling Series.

MODEL									IN-SAMPLE		
d	D	p	P	q	Q	R.S.S.	$Q_{.25}$	$P_{.25\text{-}k}$	CMSFE	UMSFE	$\hat{R}_s^2$
1	1	4	1	0	1	386.1	12.95	93.5%	0.25	0.42	-0.00

RESIDUAL AUTOCORRELATIONS.

LAG	1	2	3	4	5	6	7	8	9	10
AUTOCORRELATIONS	-.02	.09	.02	.03	.08	-.10	.06	-.01	-.08	-.01
LAG	11	12	13	14	15	16	17	18	19	20
AUTOCORRELATIONS	.01	.08	-.03	-.06	-.06	.05	-.07	-.11	.02	-.10
LAG	21	22	23	24	25	26	27	28	29	30
AUTOCORRELATIONS	-.02	-.05	-.01	-.02	-.10	-.02	-.01	.06	.01	-.02

RESULTS OF THE ESTIMATION OF THE SEASONAL DUMMY MODEL ON FIRST DIFFERENCES.

DUMMY VARIABLE	1	2	3	4	5	6	7	8	9	10	11	12
COEFFICIENT	2.2	1.64	-.45	-2.64	.73	.18	-1.82	-.09	2.73	-.09	1.45	-4.27
T-STATISTICS	3.8	3.0	-.83	-4.8	1.3	0.3	-3.31	-.17	4.96	-.17	2.65	-7.8

RSS = 395.8

IN-SAMPLE FORECAST RESULTS
CMSFE = 0.73
UMSFE = 8.07

The estimated model in equation 3.5.1. illustrates the significance of each of the included parameters at the 5% level, ($t_{120}{}^{.975} = \pm 1.98$, $t_{120}{}^{.995} = \pm 2.62$). The residual sums of squares of the estimated model is 386.1 and none of the residual autocorrelations is significant. The Q-statistic at lag 25 has an associated P-value of 93.5% providing a good indication that the residuals as a whole are white noise.

$$\underset{(2.5)}{(1 + 0.22\,B^4)}\underset{(4.5)}{(1 + 0.36\,B^{12})}(1 - B)(1 - B^{12})M_t = \underset{(-22.3)}{(1 - 0.85\,B^{12})}e_t. \qquad (6.2.1)$$

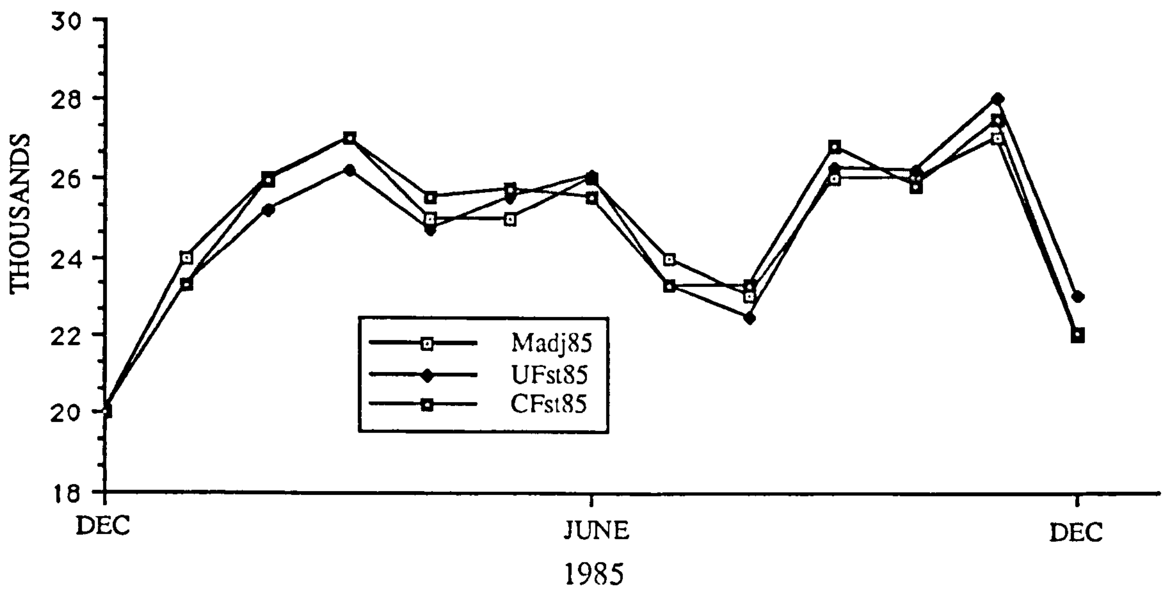
Table 6.2 also gives the results of having fitted Harvey's seasonal dummy model on the first differences of the culling series. Harvey's model has an RSS of 395.8, which is slightly higher than that of the SARIMA model and the comparison of the two yields a Harvey $\hat{R}_s^2$ value of -0.0007. The latter statistic is negative because the SARIMA model uses more degrees of freedom. Seven of the twelve seasonal dummies are significant at the 1% level, ($t_{119}{}^{.995} = \pm 2.62$). Although

1. RSS = Residual sums of squares.
Q.25 = Box-Pierce Q-Statistic for residual auto-correlations up to lag 25.
$P_{20\text{-}k}$ = Probability value for Q-statistic at lag 20 in a model containing k parameters.
CMSFE = mean square error of conditional forecasts.
UMSFE = mean square error of unconditional forecasts.

the $\hat{R}_s^2$ statistic shows that the SARIMA model is marginally worse at fitting the data than the seasonal dummy model, the MSFE statistics for the 12 in-sample forecasts for 1985 suggest that the SARIMA model is better at forecasting the in-sample period. The latter comparison of the in-sample forecasting abilities is of interest because forecasting evaluation is the primary concern of this thesis.

Figure 6.3

a. The Conditional and Unconditional In-Sample Culling Forecasts For 1985 From The SARIMA Model Estimated On 1975:1 to 1985:12



b. The Conditional and Unconditional Out-of-Sample Culling Forecasts For 1986-87 From The SARIMA Model Estimated on 1975:1-85:12.

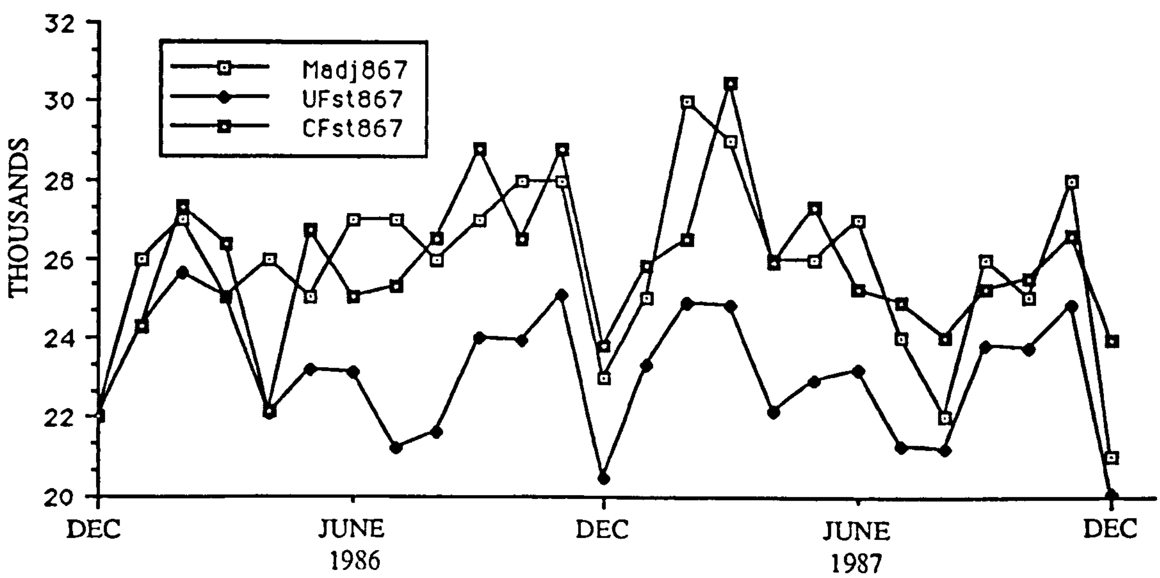


Figure 6.3a illustrates good conditional and unconditional forecasts for the in-

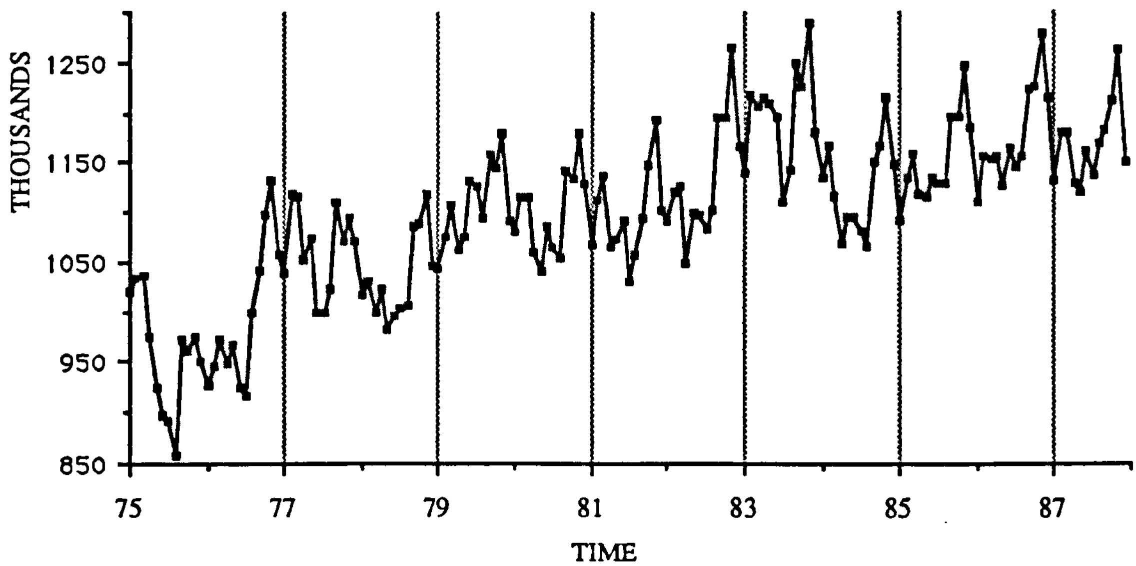
sample period produced by the Box-Jenkins model. The superiority of the SARIMA model's in-sample forecasting over that of the seasonal dummy model, however, is not repeated in the out-of-sample forecasts for 1986/7. The unconditional forecasts have a MSE of 10.1 which compares with 8.07 for the seasonal dummy model. Both models unconditionally under-forecast culling in 1986/7, a fact which is almost certainly a result of the depressed state of the market in the latter part of the sample space. Having said this, the figure shows that the SARIMA model forecasts the underlying seasonality very well. There is little difference in the MSE statistics for the conditional forecasts from the SARIMA model and the Harvey model for the out-of-sample period, having respective values of 3.01 and 2.79.

6.3 A SARIMA Model For Fat Pig Slaughter

Data for the slaughter of fat pigs are also collated on a monthly basis, and again the data have been adjusted to iron out the effects of 5 week months. Although data are available from 1968 onwards, the model is again estimated on the sample 1975:1-1985:12. The plot of the series shows an upward trend over the given period, with obvious seasonal influences, such as the increase in slaughterings in the month of November, presumably to meet the increase in demand for pig meat over the Christmas period.

Figure 6.4

A Plot Of The U.K. Monthly Series: 'Fat Pig Slaughter', 1975:1-1987:12



As with the Culling series, both a first and a seasonal difference had to be taken in order to arrive at correlograms which indicated stationarity. The identified model consisted of an MA and an SMA parameter, both of which were significant at the 1% level, producing an RSS value of 95,205.

Table 6.3[2]
The Results of Model Estimation of The Fat Pig Series:- 1975:1 - 1985:12.

MODEL									IN-SAMPLE		
d	D	p	P	q	Q	R.S.S.	$Q_{\cdot 25}$	$P_{\cdot 25\text{-}k}$	CMSFE	UMSFE	$\hat{R}_s^2$
1	1	0	0	1	1	95,205	19.2	68.8%	245.1	557.3	-.05

RESIDUAL AUTOCORRELATIONS.

LAG	1	2	3	4	5	6	7	8	9	10
AUTOCORRELATIONS	-.04	-.06	.06	.01	.06	-.02	.08	-.07	.09	-.15
LAG	11	12	13	14	15	16	17	18	19	20
AUTOCORRELATIONS	.03	-.05	-.07	-.10	-.06	-.08	-.05	.04	-.16	.03
LAG	21	22	23	24	25	26	27	28	29	30
AUTOCORRELATIONS	-.09	-.02	.13	-.05	-.05	.05	.05	.04	.09	.00

RESULTS OF THE ESTIMATION OF THE SEASONAL DUMMY MODEL ON FIRST DIFFERENCES.

DUMMY VARIABLE	1	2	3	4	5	6	7	8	9	10	11	12
COEFFICIENT	-30.8	38.0	1.0	-41.0	2.4	-.9	-19.6	8.8	78.9	2.8	41.5	-68.8
T-STATISTICS	-3.5	4.5	0.1	-4.9	0.3	-0.2	-2.3	1.1	9.4	0.3	4.9	-8.2

RSS = 91,980

IN-SAMPLE FORECAST RESULTS
CMSFE = 181.7
UMSFE = 357.9

Of the first 12 residual autocorrelations, only that at lag 10 is greater than 1 standard error away from zero as measured by the Quenouille statistic. The Box-Pierce Q-statistic at lag 25 is 19.22 which, having a P-value of 68.8%, gives a good indication of white noise residuals. Although models containing AR and SAR parameters were estimated, they provided no improvement to the model presented in equation 3.6.1, either in terms of parsimony or forecasting ability.

$$(1-B)(1-B^{12})FP_t = \underset{(-3.33)}{(1-0.295B)}\underset{(-28.0)}{(1-0.87B^{12})}e_t. \qquad (6.3.1)$$

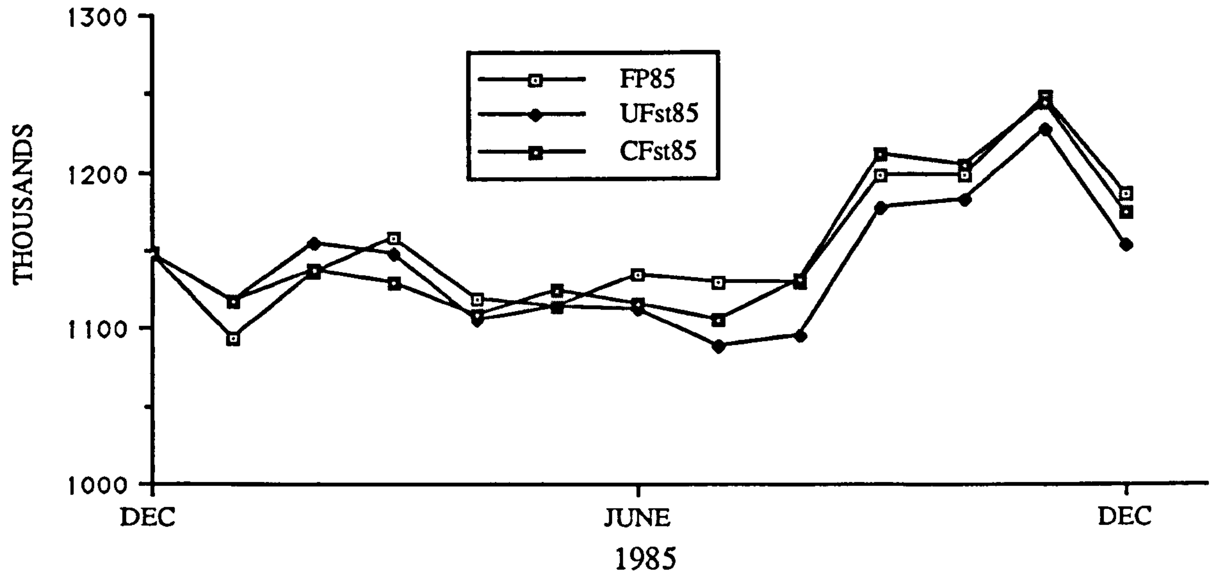
The seasonal dummy model on first differences provided a model with an RSS of 91,980 which is lower than that of the SARIMA model, resulting in an $\hat{R}_s^2$ value of -0.05. Seven of the twelve seasonal dummies are significant at the 5% level, the December dummy indicating a fall in slaughterings of 68,818 from the high

2. See footnote 1

level of November. Unlike the Culling model, the SARIMA model for fat pig slaughterings was not as good as the seasonal dummy model when comparing the MSE statistics from the in-sample forecasts.

Figure 6.5.

a. The Conditional and Unconditional In-Sample Forecasts of Slaughterings For 1985 From The SARIMA Model Estimated On The Sample 1975:1 to 1985:12



b. The Conditional and Unconditional Out-of-Sample Forecasts of Slaughterings For 1986-7 From The SARIMA Model Estimated on 1975:1-1985:12.

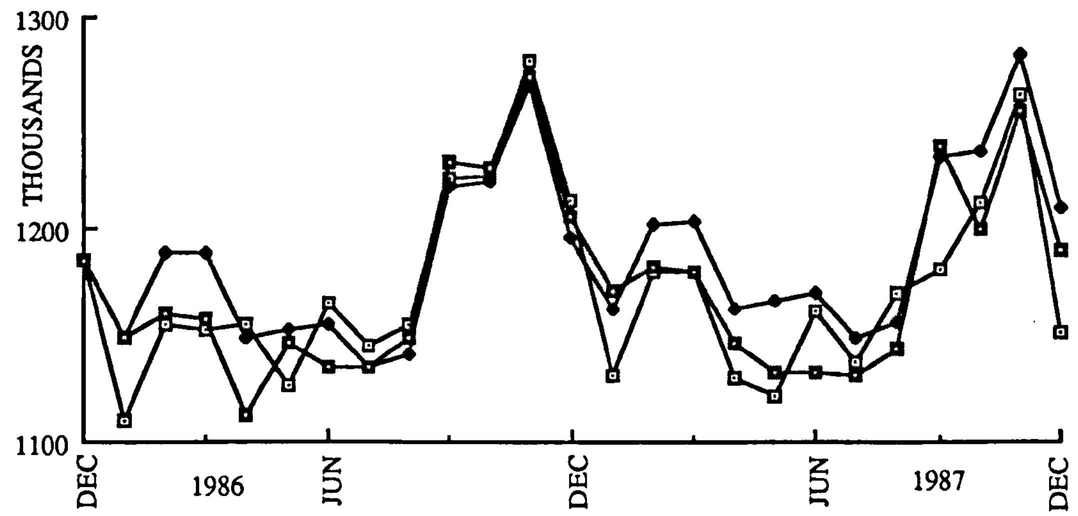


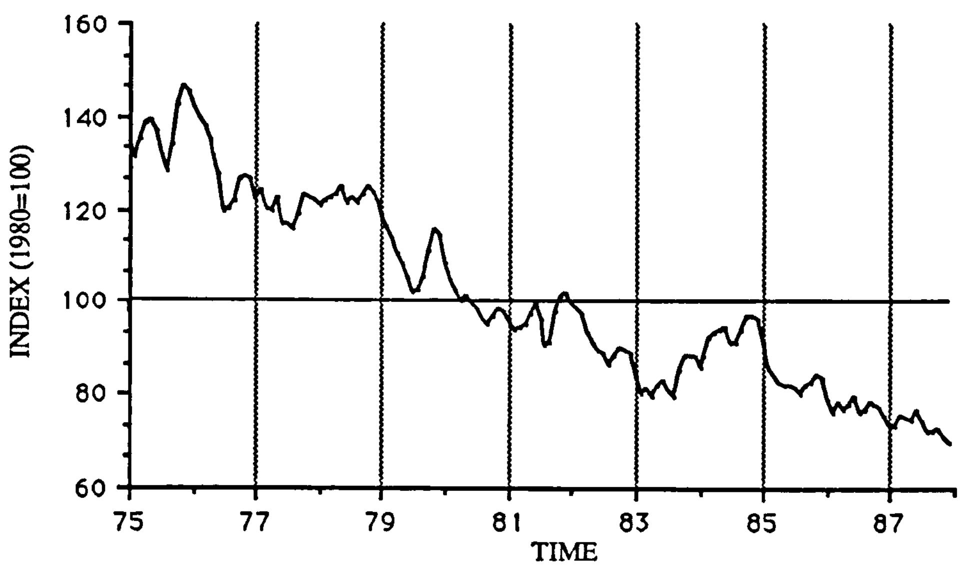
Figure 6.5a illustrates that both the conditional and the unconditional forecasts pick up the seasonal trends in the slaughtering series for 1985 very well. The plot in figure b shows a very similar seasonal pattern in slaughterings in 1986 and 1987 to that seen in 1985. Again, the model forecasts the seasonality in the slaughtering

reasonably well for the out-of-sample period, the conditional and the unconditional forecast MSE statistics being 573.9 and 767.4 respectively. The equivalent statistics for the seasonal dummy model on first differences are 806.6 and 799.2 respectively, so that the Unconditional forecasts have an MSE statistic which is slightly lower than that of the conditional forecasts. These figures imply that the SARIMA model is the better forecasting model for the out-of-sample period, the reverse of what was the case in the in-sample period of analysis.

6.4 A SARIMA Model for the Real AAPP Index Deflated by the RPI

The publication of the AAPP began in June 1975. Figures for the first five months of 1975 are derived using the original series 'Monthly Average Returns For All Pigs' - (£ per score dw), as described earlier. Having said this, it could be argued that the first 10 observations in both the feed price and the AAPP series' appear to be somewhat out of line with the subsequent data. On the grounds that the market was still being affected by the world commodity price increases of 1974, the decision was made to identify and estimate the three univariate price models on the sample space 1976:1 to 1985:12 inclusive. Consequently, the Harvey seasonal dummy model is also estimated on the shortened sample period.

FIGURE 6.6
Plot of the Real AAPP Index 1975:1-1987:12



The plot of the index shows that the real AAPP has been declining steadily since 1975. There is a relatively large fall in 1976 which is almost certainly one of the contributing factors for the imposition of the 5.5p per kg dw subsidy on pig meat in the first half of 1977. Although there were signs of a recovery in the index in

1983 and 1984, there was another minor slump at the turn of 1984 and 1985, after which the index continues to decline. Although it is difficult to see any recurring general seasonal pattern in the plot, there are signs that the AAPP falls at the turn of most years.

The identification stage of the Box-Jenkins procedure indicated that both a seasonal and a first difference were required to satisfy the conditions of stationarity in the correlograms. The model identified and estimated was one containing 1 AR and 1 SMA parameter. The results of estimation are presented in equation 4.2.5.

$$\underset{(3.06)}{(1 - 0.2856\,B)}\,(1 - B)\,(1 - B^{12})\,AAPP_t = \underset{(22.0)}{(1 - 0.8559\,B^{12})}\,e_t. \qquad (6.4.1)$$

The size of the residual autocorrelations are not as satisfactory as they might be, in particular the value of -0.21 at lag 13 which is significant when measured against the Quenouille statistic. Much searching and overfitting failed to provide a model which performed better in terms of the diagnostic checks. The Box-Pierce Q-statistic of 22.87 at lag 25 has an associated P-value of 46.8 so that the null hypothesis of white noise residuals cannot be rejected until the 47% level, thereby implying residuals consistent with white-noise. The Harvey model confirms the notion that prices fall in December, and fall significantly in January and February, April June, July and August. Significant increases in prices occur in September and October. A comparison of the RSS of the SARIMA and the Harvey dummy models produces an $\hat{R}_s^2$ value of -0.075, implying a 7.5% better fit for the Harvey model.

The in-sample forecasting results, as expressed by the conditional and unconditional MSFE are slightly better for the Harvey model than they are for the SARIMA model.

Table 6.4[3]

The Results of Model Estimation For The Real AAPP Index

Estimated on 1976:1 - 1985:12.

MODEL									IN-SAMPLE		
d	D	p	P	q	Q	R.S.S.	$Q_{.25}$	$P_{.25\text{-}k}$	CMSFE	UMSFE	$\hat{R}^2_s$
1	1	1	0	0	1	484.5	22.87	46.8	2.95	61.19	-.075

RESIDUAL AUTOCORRELATIONS.

LAG	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
AUTOCORRELATIONS	-.01	-.00	-.02	-.14	-.00	.09	-.04	.06	.06	-.06	-.08	-.01	-.21	-.07	.05

LAG	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
AUTOCORRELATIONS	-.12	.02	.10	.06	-.01	-.13	-.19	.03	.01	-.02	.08	.11	.09	-.05	-.02

RESULTS OF THE ESTIMATION OF THE SEASONAL DUMMY MODEL ON FIRST DIFFERENCES.

DUMMY VARIABLE	1	2	3	4	5	6	7	8	9	10	11	12
COEFFICIENT	-3.42	-1.58	-.65	-1.53	.27	-1.51	-2.2	-1.33	2.03	3.29	0.97	-.56
T-STATISTICS	-5.0	-2.41	-.99	-2.34	.41	-2.3	-3.36	-2.03	3.10	5.02	1.48	-.85

RSS = 459.2

IN-SAMPLE FORECAST RESULTS

CMSFE = 2.59

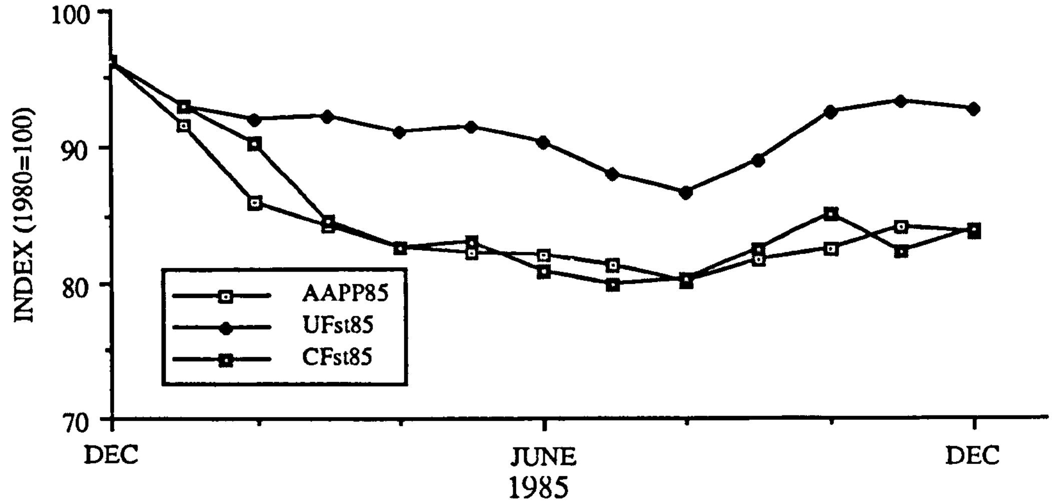
UMSFE = 33.5

The unconditional SARIMA forecasts over-forecast the actual figures although they pick up the seasonal movements in the actual series very well. As one would expect, the conditional MSFE is smaller than the equivalent unconditional statistic for both the Harvey and the SARIMA models, although the conditional SARIMA forecasts also over-forecast in the first quarter.

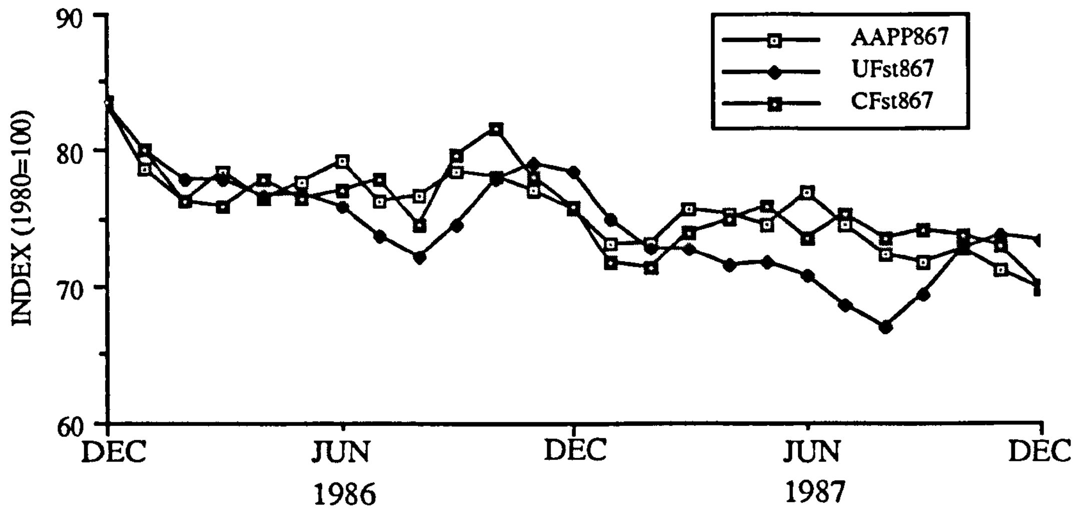
3. See footnote 1

Figure 6.7

a. The Conditional and Unconditional In-Sample Forecasts Of The Real AAPP For 1985 From The SARIMA Model Estimated On 1976:1 to 1985:12



b. The Conditional and Unconditional Out-of-Sample Forecasts of The AAPP For 1986-87 From The SARIMA Model Estimated on 1976:1-1985:12

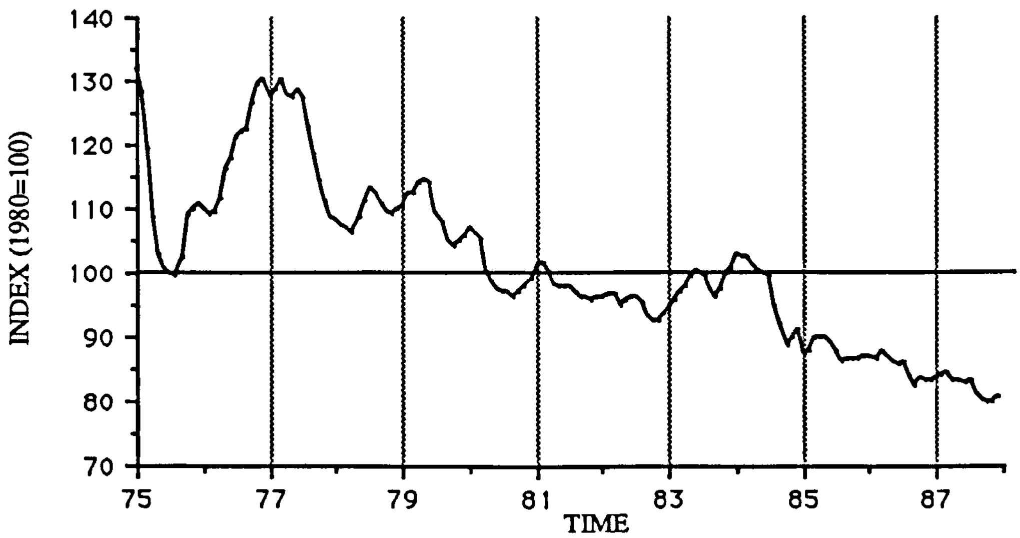


The relative forecasting performances of the Harvey and the SARIMA models are reversed when it comes to the more stringent test of forecasting the out-of-sample period. The UMSFE of 9.53 for the SARIMA model compares very well, not only with the equivalent statistic for the in-sample period, but also with the out-of sample equivalent for the Harvey dummy model which takes a value of 14.31. Having said this, there is a tendency for the unconditional forecasts to under-forecast much of 1987. Apart from the odd hiccup, the one-step conditional forecasts from the SARIMA model forecast very favourably in the out-of sample period. The CMSFE of 2.945 is as good, if not slightly better than the same statistic for the in-sample period, but even more significantly, it is an improvement on the equivalent statistic of 3.51 produced by the Harvey model.

6.5 A SARIMA Model for the Real Compound Feed Price Index Deflated by RPI

The plot of the series shows a real price which has fallen steadily over the period of estimation with no obvious seasonal pattern observed. There is a sharp drop in real prices at the very start of the sample period in 1975 and another sharp fall starting in the July of 1977, which is the period immediately following the end of the 1977 subsidy on fat pig prices. The real price of feed increases during the Aujezky disease eradication campaign. One other noticeable decline in the index is experienced in the latter half of 1984, after which the index continues to decline but at a decreased rate.

Figure 6.8
A Plot of the Compound Feed Price Index;- 1975-87 (1980=100)



The index was put through the Box-Jenkins identification procedure, the autocorrelations of the raw series showing obvious signs of non-stationarity. The autocorrelations died away only very slowly, and the partial autocorrelation at lag one was virtually equal to unity. Taking first differences of the raw series produced correlograms which indicated that stationarity had been obtained without the need for a seasonal difference. The resultant model, identified on the first differenced series, was one containing an AR and a second order SAR parameter - the SAR parameter at lag one being constrained to a value of zero.

$$\underset{}{(1-B)}\,\underset{(6.38)}{(1-0.5082\,B)}\,\underset{(4.38)}{(1-0.3635\,B^{24})}\,CF_t = e_t. \qquad (6.5.1)$$

Both the included estimated parameters are highly significant as measured by the t-statistics; the regression having a RSS statistic of 262.3. The Q-statistic of 21.71 for the first 25 residual autocorrelations has an associated P-value of 53.8 giving almost conclusive evidence for accepting the null hypothesis of white noise residuals. The results of the diagnostic checks are presented in the table below along with the results of having estimated the Harvey seasonal dummy model.

Table 6.5[4]

The Results of Model Estimation of The Index for Compound Feed Prices Estimated on the Sample Period 1976:1 - 1985:12.

MODEL									IN-SAMPLE		
d	D	p	P	q	Q	R.S.S.	$Q_{\cdot 25}$	$P_{\cdot 25\text{-}k}$	CMSFE	UMSFE	$\hat{R}_s^2$
1	0	1	2	0	0	262.3	21.71	53.8%	2.78	34.8	0.22

RESIDUAL AUTOCORRELATIONS.

LAG	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
AUTOCORRELATIONS	.02	-.04	-.04	-.01	-.06	-.10	.05	-.04	.01	.10	.12	-.03	-.01	-.12	-.15

LAG	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
AUTOCORRELATIONS	-.15	-.07	-.02	.07	.04	.09	-.06	.07	-.14	-.09	.01	-.06	.00	.05	.20

RESULTS OF THE ESTIMATION OF THE SEASONAL DUMMY MODEL ON FIRST DIFFERENCES.

DUMMY VARIABLE	1	2	3	4	5	6	7	8	9	10	11	12
COEFFICIENT	-0.16	0.40	0.13	-.77	0.67	0.36	-0.36	-1.62	-1.5	-0.49	0.42	0.59
T-STATISTICS	-.28	.75	.24	-1.44	1.25	0.67	-0.67	-3.02	-2.80	-.91	.78	1.12

RSS = 307.0

IN-SAMPLE FORECAST RESULTS

CMSFE = 2.07

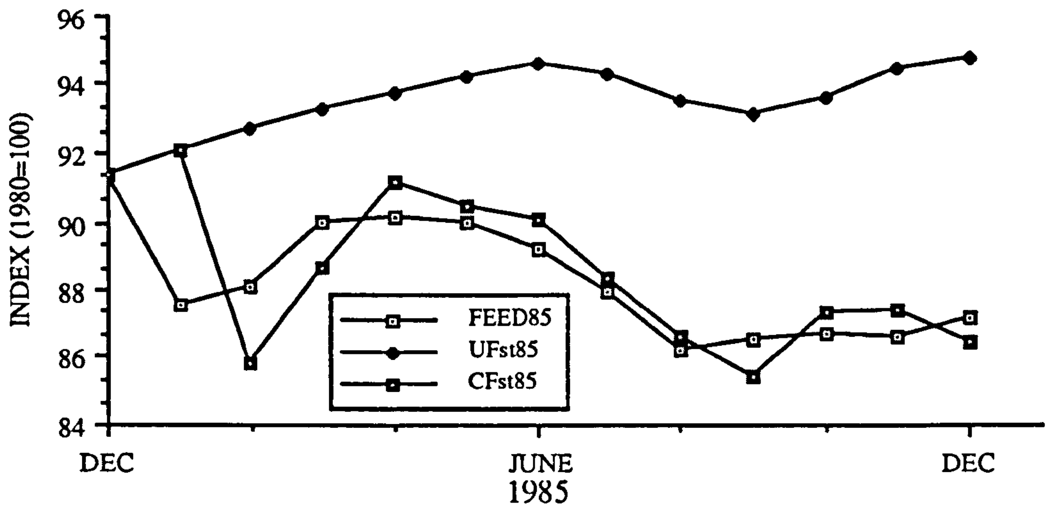
UMSFE = 7.04

The SARIMA model compares very well with the Harvey seasonal dummy model, producing an $\hat{R}_s^2$ value of 0.22. This indicates that the SARIMA model gives a 22% improvement in fit to the sample data over that of the Harvey model. The negative coefficients for August and September are the only significant coefficients in the Harvey model, suggesting that the only significant seasonal effect on the compound feed price for pigs is the fall in price around harvest time when the cost of inputs fall.

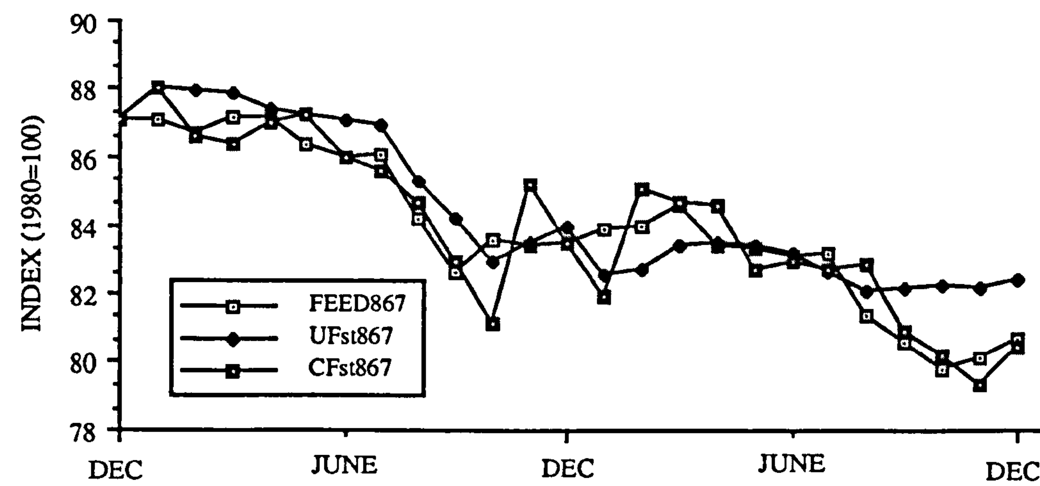
4. See footnote 1

Figure 6.9

a. The Conditional and Unconditional In-Sample Forecasts for Compound Feed Prices For 1985 From The Model Estimated On The Period 1976:1 to 1985:12



b. The Conditional and Unconditional Out-of-Sample Forecasts for Compound Feed Prices for 1986-7 From The Model Estimated on 1976:1-1985:12.



The conditional and in particular the unconditional in-sample MSFE's of the Harvey model compare favourably with those of the SARIMA model. The unconditional SARIMA forecasts continually over-forecast the actual 1985 figures, although the forecasts do pick up the seasonal movements in the index in the latter half of the year quite well. It is low index figures for the first three months of 1985, the result of the relatively sharp drop in the actual index for January which is the feature not picked up by either of the SARIMA model forecasts in the in-sample period.

As was the case with the AAPP model forecasts the MSFE's suggest that the

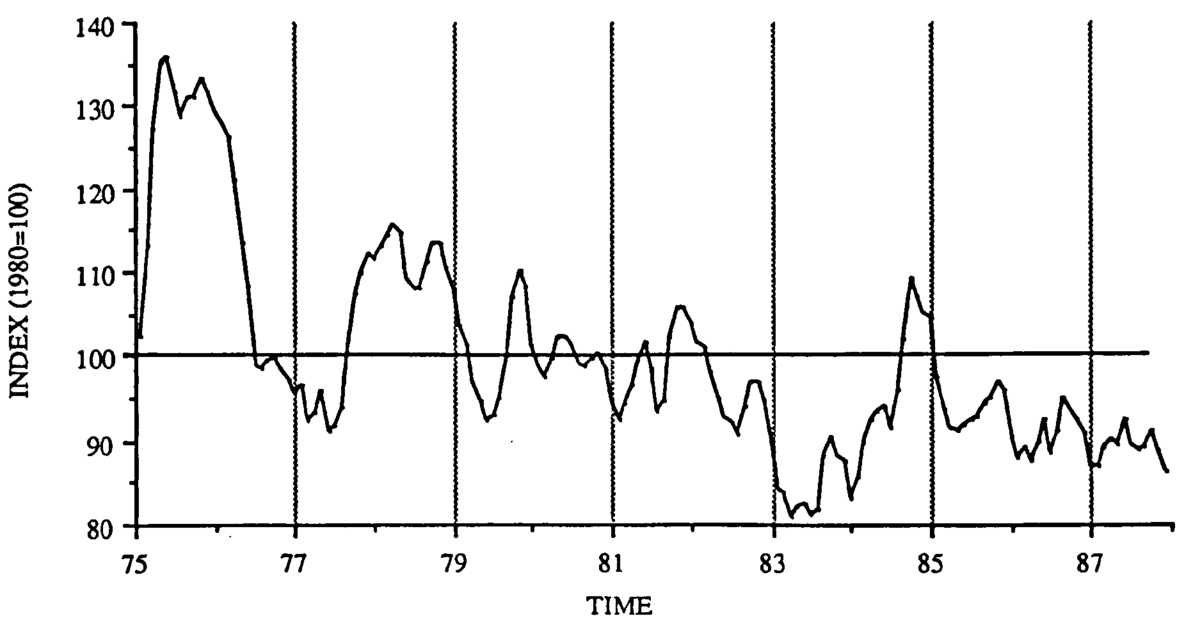
SARIMA model forecasts considerably better in the out-of-sample period than it does in the in-sample period, and also out-performs the Harvey dummy model. The unconditional 24-step forecast from the SARIMA model is very good at picking up both the general and the seasonal trends in the index, although they show signs of over-forecasting in the last five months of 1987. The ability to forecast the out-of-sample period well is reflected in the UMSFE of 1.29, an average absolute error of slightly more than one percent of the average for the index for the given forecast period. The latter figure compares favourably with the equivalent 1.85 of the Harvey model. As one would expect, the conditional one-step forecasts are an improvement over the unconditional forecasts, the CMSFE of the SARIMA model forecasts taking a value of 0.94. This figure is matched by the same statistic for the Harvey model forecasts.

6.6 A SARIMA Model For The Ratio of AAPP and Pig Compound Feed Price

The plot of the profit ratio - 'PR' - index illustrates that the series exhibits a general downward trend from 1975 onwards and no obvious seasonal pattern is discernible. The ratio is very volatile during the first 18 months of the sample period and specifically so during the first five months of the sample space, a time period which includes derived figures for AAPP. In order to remove the possibility of the derived data affecting the size of the estimated coefficients, the decision was taken to identify and estimate the model for the profit ratio on the period 1976:1-85:12 as was the case with the AAPP and compound feed price index models.

Figure 6.10

Plot of the Ratio of AAPP to the Compound feed price 1975:1-87:12



The raw data required both a first and a seasonal difference before the correlograms indicated stationarity. The identified model contained one AR parameter and and one SMA parameter, the results of estimation being given below.

$$(1-B)(1-B^{12})(1-\underset{(1.93)}{0.1854}B)PR_t = (1-\underset{(21.9)}{0.8600}B^{12})e_t. \qquad (6.6.1)$$

The AR coefficient is on the verge of significance at the 5% level and the SMA coefficient is very highly significant. None of the residual autocorrelations is close to being significant, and the Q-value of 17.4 at lag 25, which has an associated p-value of 79.0, provides positive evidence that the residuals are nothing other than white-noise. The RSS of 1132.7 does not compare very favourably with the Harvey seasonal dummy on first differences model, the $\hat{R}_s^2$ value implying a 13% superior fit for the Harvey model. The negative January and February coefficients and the positive September and October coefficients are the only significant seasonal dummy coefficients in the Harvey model.

Table 6.6[5]
The Results of Model Estimation For The Ratio of AAPP and Pig Compound Feed Price Estimated On The Sample 1976:1 - 1985:12.

MODEL									IN-SAMPLE		
d	D	p	P	q	Q	R.S.S.	$Q_{\cdot 25}$	$P_{\cdot 25-k}$	CMSFE	UMSFE	$\hat{R}_s^2$
1	1	1	0	0	1	1132.7	17.4	79.0	8.9	54.9	-0.13

RESIDUAL AUTOCORRELATIONS.

LAG	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
AUTOCORRELATIONS	-.11	.06	-.00	-.07	-.12	.12	.00	.03	.02	-.09	.04	-.07	-.09	-.15	-.05

LAG	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
AUTOCORRELATIONS	-.13	.03	-.05	.11	.05	-.05	.01	-.01	-.06	-.05	.08	.08	.07	-.01	.02

RESULTS OF THE ESTIMATION OF THE SEASONAL DUMMY MODEL ON FIRST DIFFERENCES.

DUMMY VARIABLE	1	2	3	4	5	6	7	8	9	10	11	12
COEFFICIENT	-3.13	-2.02	.34	-1.78	-.44	-1.59	-1.77	0.09	3.36	3.81	.60	-1.13
T-STATISTICS	-3.0	-2.1	.35	-1.82	-.45	-1.63	-1.81	.09	3.4	3.9	.61	-1.16

RSS = 1018.5

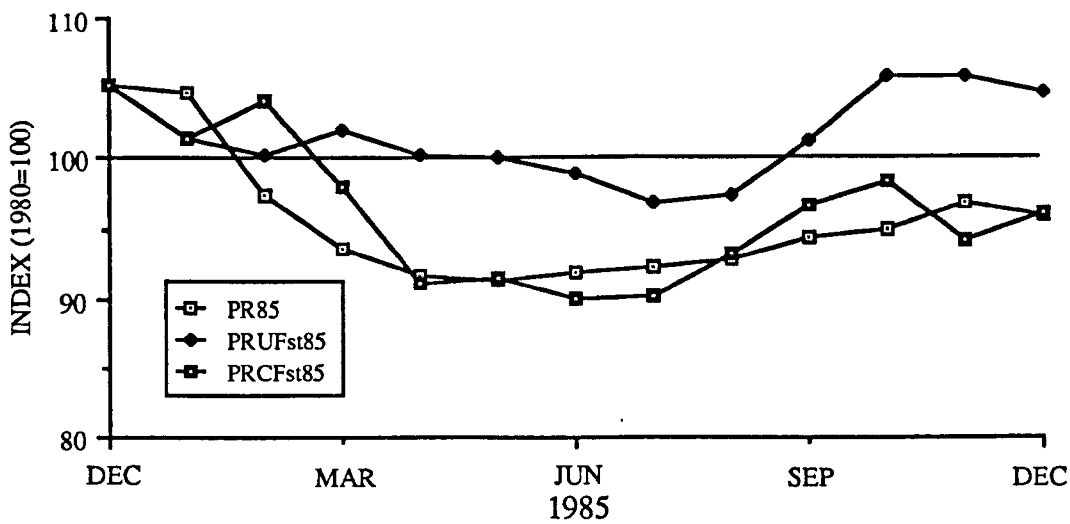
IN-SAMPLE FORECAST RESULTS
CMSFE = 6.32
UMSFE = 28.1

The in-sample conditional and unconditional MSFE statistics from the Harvey model are smaller than the equivalent statistics from the SARIMA model. The

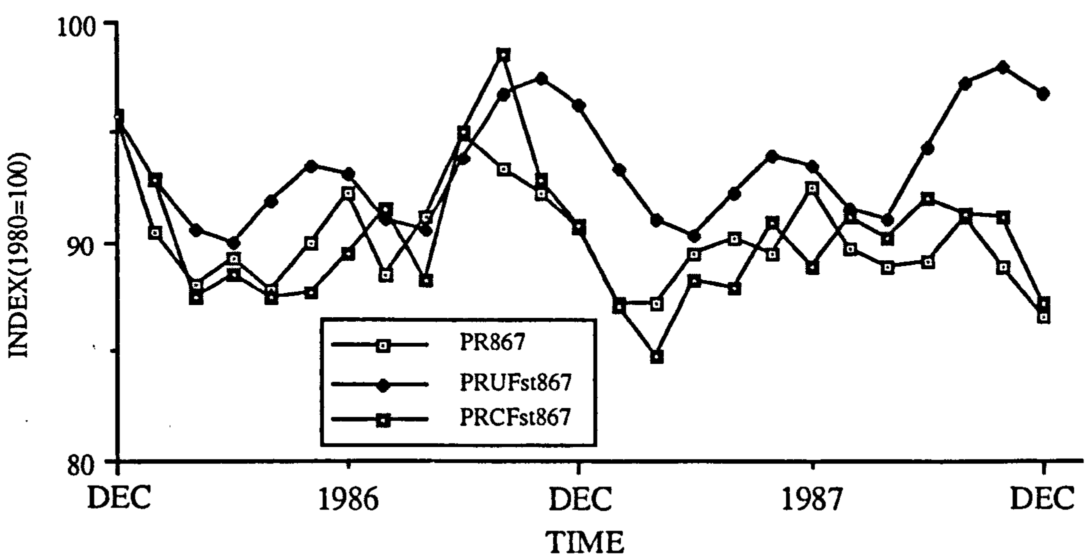
5. See footnote 1

unconditional SARIMA forecasts over-forecast the actual figures for all except the January index. The conditional forecasts are relatively good although the model fails to forecast the sharp fall in profits in February and March.

Figure 6.11
a. The Conditional and Unconditional In-Sample Forecasts Of The Profit Ratio For 1985 From The SARIMA Model Estimated On 1976:1 to 1985:12



b. The Conditional and Unconditional Out-of-Sample Forecasts of The Profit Ratio For 1986-7 From The SARIMA Model Estimated on 1976:1-1985:12.



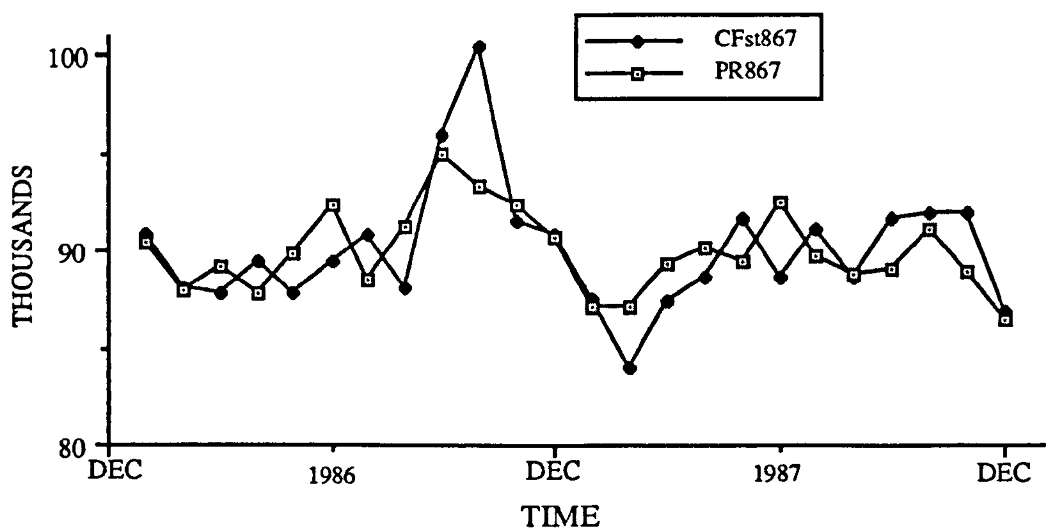
The out-of-sample SARIMA model forecasts for 1986 and 87 are better than the in-sample forecasts when comparing the MSFEs. The unconditional forecasts pick up the general movements in profits quite well though the forecasts are characterised by over-forecasting from around the end of the 1986. The MSFE statistics for the unconditional and the conditional forecasts are 13.26 and 4.46 respectively. The equivalent statistics for the Harvey seasonal dummy model are 16.13 and 5.50, indicating the superiority of the Box-Jenkins model over the seasonal dummy model in the stricter test of out-of sample forecasting performance.

6.7 Derived Versus Actual Profit Forecasts.

To round off this chapter, the forecasts of the model built in the previous section are compared with those derived by taking the forecasts of the real AAPP and the real compound feed price, produced by the models developed in sections 6.4 and 6.5 respectively, and deriving profit ratio forecasts. The analysis is done by comparing the short term forecasting abilities of the models by analysing the conditional one-step forecasts in the out-of-sample period. The forecasts, along with the actual profit ratio index are presented in figure 6.12 below.

Figure 6.12

The Conditional One-Step Forecasts of the Derived Profit Ratio for the Out-Of-Sample Period 1986-7



The forecasts in figure 6.12 can be compared directly with those of figure 6.11. What is immediately obvious from a visual comparison of the two sets of plots is that they are very similar. The only general observation which can be made is that is does appear that the derived forecast errors look slightly larger than they are for the forecasts made from having modelled the actual profit ratio. In order to quantify the difference in the forecasting abilities of the actual and derived forecasts, the MSFE of the conditional forecasts was calculated so that it could be compared with the value of 4.46. The latter statistic is the CMSFE from having forecast using the actual profit ratio model discussed in the previous section. The calculated value of 5.87 suggests that the direct forecasts are slightly superior to the derived forecasts for the short term forecasts. This result, coupled with the fact that it is obviously easier to forecast profits directly using model 6.6.1, suggest that the direct route to forecast the profit ratio should be preferred to the derived route.

6.8 Conclusions

In this chapter univariate statistical models have been built using the Box-Jenkins methodology outlined in chapter two. The slaughter and price data which have been modelled were in monthly form, which made identification of the models considerably easier than was experience when building the quarterly breeding herd models of chapter three. The models were estimated over a comparable period to the biological and econometric trimestic models of chapters three and four, although the frequency of the data means a much larger number of observations are available. The price data for 1975 is not included in the identification and estimation procedures due to its turbulent behaviour, caused by the sharp increase in world commodity prices at that time. Also, the AAPP was not available at the start of the sample period although provision was made to deal with this situation.

All five models estimated were both first and seasonally differenced in order to achieve stationarity, and features common to most of the identified models were non-seasonal autoregressive and seasonal moving average components. In contrast to the quarterly breeding herd models, none of the monthly models had cyclical structures. Most of the series exhibited seasonality, picking up features such as the increased slaughter in the pre-Christmas period. The presence of seasonality is almost certainly a reason for the negative Harvey $\hat{R}_s^2$ statistics in four out of the five models, the exception being the compound feed model in which seasonality is least prevalent. Comparing the in-sample and out-of-sample forecasting performances of the five SARIMA models with those of the Harvey seasonal dummy models, the latter were on the whole the better in-sample but in the stricter test of out-of-sample forecasting the SARIMA models came out on top.

At the end of the chapter a brief analysis of the relative forecasting performances of the profit ratio model and the forecasts derived by forecasting the two components of the ratio, favoured the use of the ratio model itself both in terms of forecasting ability and convenience. This univariate model for the profit ratio is, therefore, used as the sole model for forecasting profits in the thesis, and will be used every time models including profit as an explanatory variable are used for unconditional forecasting. In the following chapter, the models developed for the two slaughter categories and the profit ratio model will be combined to produce bivariate models using the Box-Jenkins methodology.

The short and longer term forecasting performance of the univariate SARIMA models compared with equivalent biological and bivariate Box-Jenkins models for culling and fat pig slaughter will be analysed in chapter eight, the forecasting chapter.

CHAPTER SEVEN

BIVARIATE BOX-JENKINS MODELS FOR SOW AND BOAR CULLING AND FAT PIG SLAUGHTER

7.1 Introduction.

A natural extension of the Box-Jenkins univariate model building methodology outlined in chapter two is the multivariate Box-Jenkins analysis which utilises the univariate models built using the same methodology in order to derive models relating two or more variables. More specifically, a variable $X_{1,t}$ is related to past values of itself through AR terms, past and present values of a second variable $X_{2,t}$ and past and present values of a moving-average error term ε_{1t}, such as that illustrated in the transfer function 7.1.1.

$$\Omega_1 (B) X_{1,t} = \Omega_2 (B) X_{2,t} + \Omega_3 (B) \zeta_{1,t}. \qquad (7.1.1)$$

where B is the usual backshift operator and $\xi_{1,t}$ is a white noise error term. It was deemed appropriate and of interest to consider building such models for relevant key variables modelled in the thesis, introducing profits as an explanatory variable, so that the bivariate Box-Jenkins models can be regarded as alternatives to more traditional econometric approaches. The analysis presented in this chapter, is confined to bivariate models only, for which the theory underlying the building of such models is outlined in section 7.2 and discussed more fully in Granger and Newbold (1977).

Like the univariate analysis, multivariate model building methodology requires suitably long and consistent time series data in order that a serious attempt at modelling can be undertaken. Given that the profit ratio and the monthly culling and fat pig slaughter series have been modelled from 1975 onwards, the bivariate analysis is confined to this period. Given these two constraints it was not considered sensible to attempt model identification for the pseudo-quarterly census data, especially given the possibility of potentially long lags on the profit variables. The bivariate Box-Jenkins analysis is thus confined to the series for which there are monthly data, namely, sow and boar cullings and fat pig slaughter.

The models to be built will be of interest in that their long and short term forecasting abilities can be compared with the other monthly models for the two slaughter categories built in previous chapters. Ex ante, one would expect the models to perform as well as, if not better than, the univariate models in the short term because of the inclusion of the additional profit variable. The medium/long term forecasting should also be superior to the univariate models for the same reason, however, forecasts greater than one month ahead require forecasts of the explanatory profit variable to be made first and as such, the longer term performance of the models will depend upon the longer term forecasting ability of the univariate model for the profit ratio developed in chapter six.

In addition to an interest in the forecasting models themselves, the Box-Jenkins bivariate model building process has a secondary use in that it helps to identify the nature of the lags in the relationships between the two variables modelled. Thus, if the relationships between culling and profits and fat pig slaughter and profits are unknown a priori, the analysis would be useful in indicating the possible length of the lags involved in the action of culling/ slaughtering the pigs following a given change in the level of profits. In the following section, the theory of bivariate Box-Jenkins modelling is outlined.

7.2 Bivariate Box-Jenkins Modelling:- The Theory.

Much of the theory underpinning bivariate Box-Jenkins methodology is obviously related to that used in the univariate modelling procedure and as such, the discussion of the theory of bivariate modelling requires less detailed discussion than that which was devoted to the univariate methodology of Chapter two. The terminology and notation used in this chapter follows that used in chapter two.

As is the case in univariate modelling, bivariate/multivariate modelling involves the three stages of model identification, estimation and diagnostic checking. At the identification stage of the model building process, univariate models for the included variables are built and the error terms cross correlated in order to infer relationships between the variables concerned. Having identified a relationship, the bivariate model can be estimated using a relevant non-linear estimation package. The appropriateness of the estimated model can then be checked in a similar fashion to that for the univariate models by consideration of factors such as the significance of the estimated parameters and the randomness or otherwise of the estimated residuals.

A common way of examining the possibility of relationships between two variables is to examine the degree of linear correlation between them, and so this would seem an appropriate way to help identify a bivariate Box-Jenkins model. Unfortunately, a common feature of economic time series is that they are subject to the influences of time trends, and as such a high correlation coefficient between two variables is very possibly spurious in that it is largely a result of the two variables trending together over time, whether or not they are related to each other. The consequence of this is that observation of the correlation between $X_{1,t}$ and $X_{2,t}$ is likely to lead to mistaken inference, especially where economic data are concerned. A suggested way to get around this problem is to conduct a transformation of the data referred to as 'pre-whitening'. Having built a univariate ARMA model of the form,

$$\phi(B)\, X_{1,t} = \theta(B)\, \varepsilon_{1,t}. \qquad (7.2.1)$$

the estimated residuals from estimating the model should, assuming the model has been identified correctly, exhibit white-noise properties. The significance of this discussion is that the estimated residual is random and is not, therefore, subject to influences such as trending, so that the possibility of spurious correlation being inferred is removed. Given this property, the residual term makes a useful proxy for the variable $X_{1,t}$. This process of transforming a variable into its residual error term is the aforementioned pre-whitening process, and in the case of the general ARMA model represented by equation 7.2.1 is achieved by dividing both sides of the equation by the moving-average polynomial term term $\theta(B)$ as indicated below.

$$\varepsilon_{1t} = \frac{\phi(B)}{\theta(B)} X_{1t} \qquad (7.2.2)$$

The estimated residuals from any univariate model can thus be used to identify relationships between two variables that one wishes to model. Given a second univariate model,

$$\Phi(B)\, X_{2,t} = \Theta(B)\, \varepsilon_{2,t}. \qquad (7.2.3)$$

the two pre-whitened series $\varepsilon_{1,t}$ and $\varepsilon_{2,t}$ can be cross-correlated for an appropriate number of lags, where the latter is determined by the frequency of the time series data. More specifically, a monthly model, for example, requires examination of cross correlations for a larger number of lags than would a model concerned with annual time series data. In general the values of the cross correlations $CORR(\varepsilon_{1,t}, \varepsilon_{2,t-i})$ are calculated for both positive and negative values of i in order to observe the possibility of feedback in the model. As with the univariate models, a general rule of thumb for measuring the significance of the cross correlations is to count the Quenouille statistic, that is, $(2 \div \sqrt{n})$ as the accepted critical value for the level of significance. If, for example the cross-correlation between $\varepsilon_{1,t}$ and $\varepsilon_{2,t-j}$ were larger than the Quenouille statistic, this would suggest that $\varepsilon_{2,t}$ has a causal effect on $\varepsilon_{1,t}$ with a lag of j time periods, and furthermore, this suggests that $X_{2,t}$ has a causal effect on $X_{1,t}$ with the same time lag. Thus the central importance of the cross correlation statistics in the identification of a bivariate Box-Jenkins model has been illustrated. There is nothing to stop more than one cross correlation being significant and, moreover, it is possible for correlations to be significant for positive values of i. Were the latter true, this implies that the causal relationship runs in the opposite direction to that described in the example above. Where there is significant cross correlation for both positive and negative values of i, a situation of feedback is said to exist where the causal relationships run in both directions. If the cross correlation at lag zero is significant simultaneity is present in the relationship and the picture is much less clear as far as the implications for the direction of causality are concerned and, as yet, no theory exists which enables interpretation in such an event.

When feedback is present in a bivariate relationship, the whole process of identifying and estimating the model, which is likely to contain a large number of parameters, is made much more complicated. Where feedback does not exist, or where it is assumed away, the whole process of model building is greatly simplified, one of the transfer functions reducing to a univariate ARIMA model. Because I am not concerned with modelling the profit variable itself, the possibility of feedback will be assumed away, and as such, the remainder of this section dealing with the theory behind the Box-Jenkins methodology will only consider uni-directional causality models.

Having observed the cross correlations between $\varepsilon_{1,t}$ and $\varepsilon_{2,t-i}$ for positive values of i, therefore, a general uni-directional causality model between the two pre-whitened variables can be identified taking the form given in 7.2.4,

$$\varepsilon_{1,t} = \frac{\omega_2(B)}{\omega_1(B)}\,\varepsilon_{2,t} + \frac{\omega_3(B)}{\omega_1(B)}\,\zeta_{1,t} \tag{7.2.4}$$

where $\xi_{1,t}$ is the white noise error term and $\omega_1(B)$ is a finite polynomial. The order of the polynomials $\omega_2(B)$ and $\omega_3(B)$ is inferred by the cross-correlation identification stage of modelling described above. Having identified the above transfer function for the pre-whitened variables, the model can be translated into a model in terms of $X_{1,t}$ and $X_{2,t}$ in the following way. Substituting for $\varepsilon_{1,t}$ and $\varepsilon_{2,t}$ in 7.2.4 using transformations of 7.2.2 and 7.2.3, the following transfer function results.

$$\frac{\phi(B)}{\theta(B)}X_{1,t} = \frac{\omega_2(B)}{\omega_1(B)}\frac{\Phi(B)}{\Theta(B)}X_{2,t} + \frac{\omega_3(B)}{\omega_1(B)}\zeta_{1,t} \tag{7.2.5}$$

Multiplying through both sides of the equation by $\theta(B) / \phi(B)$ gives:-

$$X_{1,t} = \frac{\theta(B)\,\omega_2(B)\,\Phi(B)}{\phi(B)\,\omega_1(B)\,\Theta(B)}\,X_{2,t} + \frac{\theta(B)\,\omega_3(B)}{\phi(B)\,\omega_1(B)}\,\zeta_{1,t} \tag{7.2.6}$$

which is the final transformation required, the transfer function in 7.2.6 expressing the dependent variable $X_{1,t}$ in terms of past and present values of $X_{2,t}$ and the white noise error term $\zeta_{1,t}$. Although the modelling process was simplified by excluding the possibility of feedback in the model, the transfer function in 7.2.6 still looks rather complicated, consisting of 7 polynomials, the order of which could take any value. In practice the order of the polynomials is unlikely to be greater than two and it is possible that the function could be simplified by the cancellation of common roots. A common way of simplifying the whole process further is to bypass much of the transforming of transfer functions and to use the cross-correlation function identification of the pre-whitened series to infer a relationship between the two X variables, estimate the latter relationship and allow the model to be adapted at the diagnostic checking stage of the model building process.

Box and Jenkins note that this method of pre-whitening both the input and the output series is somewhat indirect and, where a large number of coefficients exist, the above method can cause over-parameterisation. The latter problem initiates more work at the diagnostic checking stage of the model building procedure. A more direct method not pursued in this thesis but advocated by Box and Jenkins to reduce the diagnostic work, for modelling when feedback does not exist in the relationship is outlined in Granger and Newbold.[1] Briefly, the method involves pre-whitening only the input variable $X_{2,t}$, and using the estimated coefficients from the univariate model for this series to transform the output variable $X_{1,t}$ in order to define a new variable Z_t. Rewriting 7.2.6. as:-

$$X_{1,t} = V_1^*(B)\,X_{2,t} + \Psi_1^*(B)\,\zeta_{1,t} \tag{7.2.7}$$

then Z_t is defined as:-

$$Z_t = \frac{\Phi(B)}{\Theta(B)}\,X_{1,t} \tag{7.2.8}$$

where $\Phi(B)$ and $\Theta(B)$ are as defined in 7.2.3. It can be shown, using simple

1. Granger and Newbold (1977). pp. 243f

algebra, that:-

$$V^*_{1,i} = \mathrm{CORR}\,(Z_t, \varepsilon_{2,t-i})\left[\frac{\mathrm{VAR}\,(Z_t)}{\mathrm{VAR}\,(\varepsilon_{2,t})}\right]^{0.5} \qquad (7.2.9)$$

Thus, each $V^*_{1,i}$ is simply a constant multiple of the cross correlation between Z_t and $\varepsilon_{2,t-i}$. The cross correlations between Z_t and $e_{2,t-i}$ where the latter are the estimated residuals from having estimated 7.2.3, are then calculated so that the estimates of $V^*_{1,i}$ can be derived. The latter are then used to identify the form of $V^*_1(B)$ in 7.2.7, which is then estimated assuming $\varepsilon_{1,t}$ to be white noise. The residual autocorrelation function is then allowed to suggest the appropriate form of the error structure. The coefficients of 7.2.7 can then be estimated in the usual fashion.

Having estimated the identified transfer function for $X_{1,t}$ reached by whatever route is thought most appropriate, the model is estimated using a non-linear regression package. The model is then checked for inadequacies in much the same way as are the univariate models, and the appropriate modifications made. Any estimated parameters with insignificant t-statistics can be removed from the initial identification if it is thought appropriate. The residual autocorrelations should be observed for large correlations at an appropriate number of lags determined by the frequency of the time series data, again using the Quenouille statistic as a yardstick. A portmanteau statistic such as the Box-Ljung statistic should also be calculated to examine the white noise properties of the residuals as a whole. Given significantly large residual autocorrelation statistics the initial model should be augmented in the usual manner by the addition of AR or MA terms as deemed appropriate by the modeller, the usual rules of augmenting one side of the equation only applying as they did in the univariate modelling procedure.

As is true of univariate Box-Jenkins modelling, the model building procedure is unlikely to be as clear cut as the theory suggests and the results of model identification and diagnostic checking are often a matter for the modeller to decide what he feels to be the most appropriate action to take. Consequently, different modellers will often produce different models given identical sets of data, which may be thought to be an advantage or a disadvantage of the Box-Jenkins approach to other time series methodologies. Having said this, it is also true that the results of forecasting with the differently structured models is more than likely to produce similar forecasts. The theory outlined above is now applied to build bivariate models for the culling and fat pig slaughter time series using the profit ratio as the causal variable.

7.3 A Bivariate Box-Jenkins Model For The Culling Of Sows and Boars.

The first task in the Identification process is to pre-whiten the culling and the profit series. This is achieved by obtaining the residuals from the univariate models for the two series concerned. From chapter six it is known that the identified and estimated univariate models are given as follows:-

$$(1 + 0.22\, B^4)(1 + 0.36\, B^{12})(1 - B)(1 - B^{12})\, M_t = (1 - 0.85\, B^{12})\, Me_t \qquad (7.3.1)$$

where Me_t is the white noise error term for the culling model, and

$$(1 - 0.1854\, B)(1 - B)(1 - B^{12})\, PR_t = (1 - 0.8600\, B^{12})\, PRe_t. \qquad (7.3.2)$$

where PRe_t is the white noise error term for the profit model.

The Observed errors Me_t and the PRe_t were obtained from the estimated univariate models by simple transformations of 7.3.1 and 7.3.2. Thus,

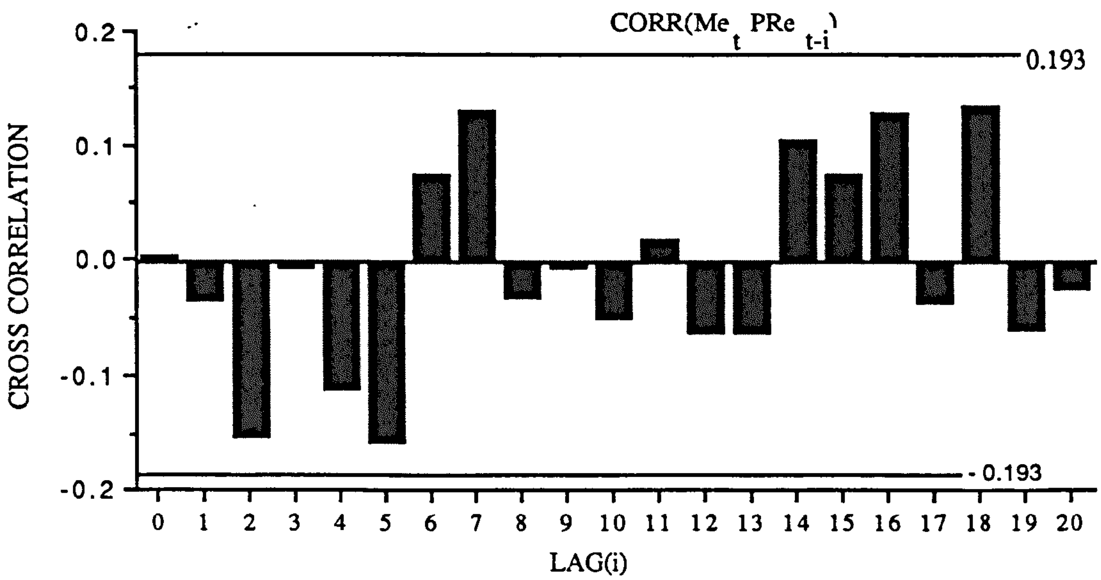
$$Me_t = (1 + 0.22\, B^4)(1 + 0.36\, B^{12})(1 - B)(1 - B^{12})\, M_t + 0.85\ Me_{t-12} \qquad (7.3.3)$$

and

$$PRe_t = (1 - 0.1854\, B)(1 - B)(1 - B^{12})\, PR_t + 0.8600\, PRe_{t-12}. \qquad (7.3.4)$$

Having obtained the observed residuals, Me_t and PRe_t, the two were cross correlated producing the following cross correlation correlogram in which only the lags for positive values of i are observed, our interest being in the relationship in which profit is the causal variable.

Figure 7.1
The Results Of Cross Correlating The Observed Culling Residuals
With The Observed Profit Residuals: 1976:2-1985:12.



The figure illustrates the fact that the errors from the univariate culling and profit models are uncorrelated at lag zero, the value of the calculated coefficient of correlation being 0.004, which compared with a Quenouille statistic of 0.193 is nowhere near significant. This result is good in that it expels any possibility of contemporaneous correlation in the model which does not invalidate our assumption of uni-directional causality. Comparing the cross-correlations with the two horizontal lines - representing the Quenouille levels of significance for positive and negative cross correlations - none of the individual correlations is significant. If, however, one observes the cross correlations as a group rather than individually a couple of interesting points arise. The first observation is that the first five cross correlations are negative and, those at lags 2,4 and 5 are all greater than one standard error below zero. This phenomenon is compatible with the lag used in the econometric model of chapter 5 in that culling was related to a weighted average of profits in the previous five months. The correlation is negative because an increase in profits, for example, reduces the subsequent number of cullings as producers build up the breeding herd. The second outstanding feature of the right hand side of the diagram is that the cross correlations for lag 14 to 18, with the exception of that at lag 17, are all positive and relatively large. For the given example of an increase in profits, although producers curb culling initially, the resultant increase in the breeding herd and the number of fat pigs eventually produces an increase in cullings. It is conceivable that this positive relationship occurs 14 to 18 months after the initial change in profit although there is no reason, a priori, for expecting the relationship to occur at this time, nor would one expect it to occur over such a concentrated time span. Having said this, the Box-Jenkins approach is an empirical one and so the first attempt at estimation took the following form:-

$$Me_t = \alpha_1 \,(0.2\, PRe_{t-1,t-5}) + \alpha_2 \,(0.2\, PRe_{t-14,t-18}) + \varepsilon_t \qquad (7.3.5)$$

where ε_t is a white noise error term,
and where $PRe_{t-1,t-5} = PRe_{t-1} + PRe_{t-2} + PRe_{t-3} + PRe_{t-4} + PRe_{t-5}$
and correspondingly for $PRe_{t-14,t-18}$.
In other words, the simple assumption of equal weights on each of the profit variables lagged 1 to 5 months and 14 to 18 months is made. The results of estimation of the above are presented in the expression below.

$$\underset{(-1.70)}{Me_t = -0.292} \,(0.2\, PRe_{t-1,t-5}) + \underset{(0.98)}{0.158} \,(0.2\, PRe_{t-14,t-18}) + e_t \qquad (7.3.6)$$

The t-statistics in the parentheses indicate that the estimated parameter on the longer lagged profit residual variable is not even as large as one standard error above zero. Consequently, the longer lagged profit residual variable was dropped from the estimation procedure. The results of regressing the culling residuals on profit residuals lagged only 1 to 5 months is given below.

$$Me_t = \underset{(-2.44)}{-0.295} \,(0.2\, PRe_{t-1,t-5}) + e_t \qquad (7.3.7)$$

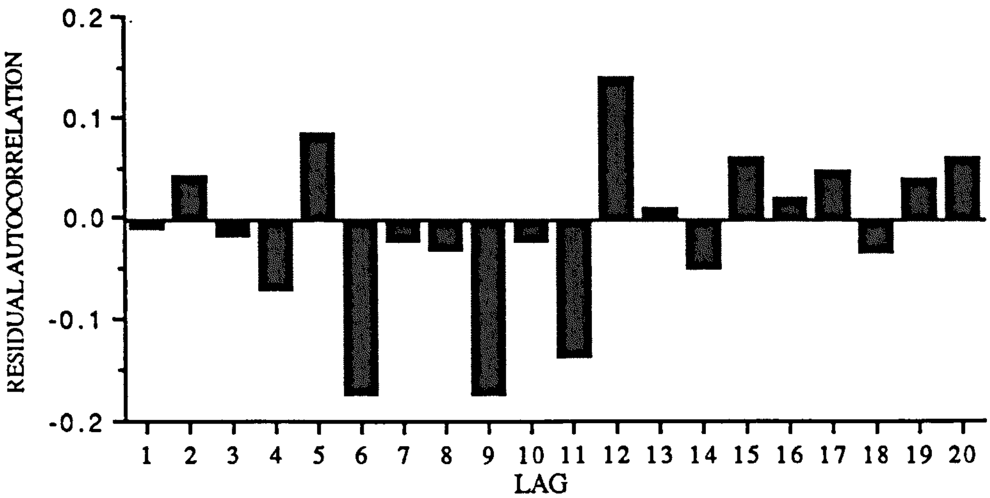
The t-statistic indicates the significance at the 2% level of the estimated parameter which is, of course, of the correct sign. None of the residual auto correlations - the first 20 of which are presented in the figure below - is significant as measured by the Quenouille statistic and the Box-Ljung Q-statistic[2] of 14.64 indicates that the null hypothesis of white noise residuals cannot be rejected even at the 50% level of significance when comparing Q with the Chi-Squares tables for 19 degrees of freedom.

---

2. See footnote on page 2.12 for explanation of Box-Ljung statistic.

Figure 7.2

Plot Of Residual Autocorrelations From The Estimated Equation 7.3.7



Having obtained a satisfactory relationship between the error series for culling and profit, the next task is to identify and estimate a model for 'actual' culling and profit. In order to do this, use is made of the two univariate models 7.3.1 and 7.3.2 and the bivariate model for the errors of these two models, 7.3.7. Replacing the estimated coefficients for these models by variables a_1 to a_6, and defining (1-B)(1-B^{12})M_t in 7.3.1 and (1-B)(1-B^{12})PRt in 7.3.2 as M*t and PR*t respectively, 7.3.1 ,7.3.2 and 7.3.7 can be re-written as 7.3.8, 7.3.9 and 7.3.10 respectively.

$$(1 - a_1 B^4)(1 - a_2 B^{12})M^*_t = (1 - a_3 B^{12}) Me_t. \tag{7.3.8}$$

$$(1 - a_4 B)\, PR^*_t = (1 - a_5 B^{12})\, PRe_t. \tag{7.3.9}$$

$$Me_t = a_6 (0.2\, PRe_{t-1,t-5}) + \varepsilon_t \tag{7.3.10}$$

Substituting 7.3.10 into 7.3.8 gives:-

$$(1 - a_1 B^4)(1 - a_2 B^{12})M^*_t = (1 - a_3 B^{12}) [a_6 (0.2\, PRe_{t-1,t-5}) + \varepsilon_t] \tag{7.3.11}$$

The appropriate transformation of 7.3.9 can then be substituted into 7.3.11 giving:-

$$(1 - a_1 B^4)(1 - a_2 B^{12}) M^*_t = (1-a_3 B^{12})\{ a_6\, 0.2 \frac{(1-a_4 B)}{(1-a_5 B^{12})} PR^*_{t-1,t-5} + \varepsilon_t \} \tag{7.3.12}$$

At this stage, the relationship looks quite complicated on the right hand side of the equality, however, referring back to the results of parameter estimation, a_3 and a_5 -

the parameters for the moving-average terms in the univariate culling and profit models respectively - are different only by .01. By making the not unreasonable assumption that these two parameters are equal 7.3.12 can be rewritten as 7.3.13.

$$(1 - a_1 B^4)(1 - a_2 B^{12})M^*_t = a_6 (1 - a_4 B)0.2PR^*_{t-1,t-5} + (1 - a_5 B^{12})\varepsilon_t \qquad (7.3.13)$$

It is possible to develop the model further by expanding the differenced variables M^*_t and $PR^*_{t-1,t-5}$, and dividing through by $(1-B)(1-B^{12})$ so that the latter equation expresses actual cullings in terms of an average of lagged actual profits as given in the following expression.

$$(1 - a_1 B^4)(1 - a_2 B^{12}) M_t = 0.2\, a_6 (1 - a_4 B)\, PR_{t-1,t-5} + \frac{(1 - a_5 B^{12})}{(1 - B)(1 - B^{12})}\varepsilon_t \qquad (7.3.14)$$

The major drawback with the above equation is the complex polynomial expressing the structure of the moving-average error term. Although many varied attempts were made at trying to estimate (7.3.14), none were successful in that no model was found in which all the estimated parameters were significant or took feasible values. In many cases the signs on the parameters were different from what one would expect from the initial estimates in equations 7.3.8, 7.3.9 and 7.3.10. Also, no model could be found in which the error structure could confidently be described as white noise.

In the end, therefore, it was necessary to return to equation 7.3.13 in order to attempt an estimation. Representing 7.3.13 without the backshift operator is given in equation 7.3.15 below.

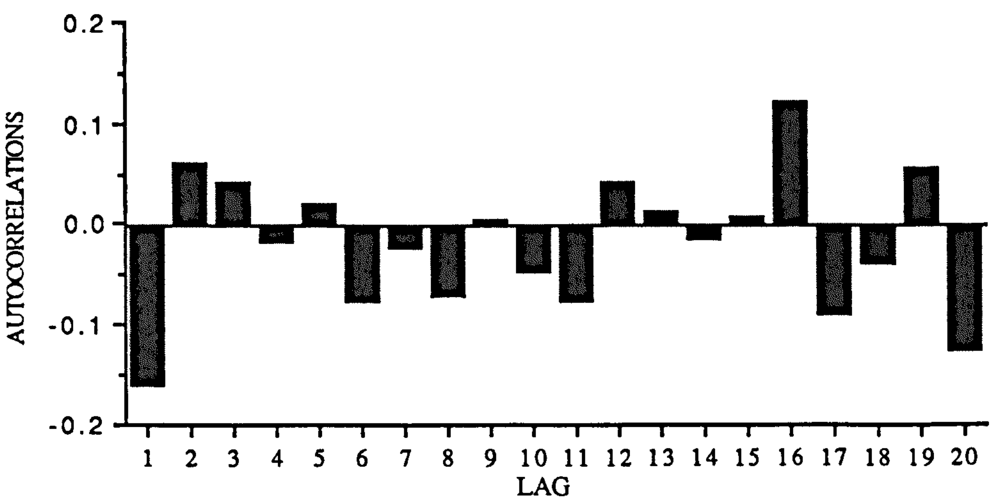
$$\begin{aligned} &M^*_t - a_1 M^*_{t-4} - a_2 M^*_{t-12} + a_1 a_2 M^*_{t-16} = \\ &0.2a_6 (PR^*_{t-1,t-5} - a_4 PR^*_{t-2,t-6}) + \varepsilon_t - a_5\varepsilon_{t-12} \end{aligned} \qquad (7.3.15)$$

Estimating the restricted expression (7.3.13) in order to reduce the number of parameters to be estimated produces the following regression.

$$\begin{aligned} &(1 + \underset{(-2.65)}{0.268}\, B^4)(1 + \underset{(-3.62)}{0.361}\, B^{12})\, M^*_t = \\ &- \underset{(-2.04)}{0.414}\, (1 - \underset{(0.70)}{0.269}\, B)\; 0.2PR^*_{t-1,t-5} + (1 - \underset{(5.38)}{0.581}\, B^{12})\, e_t \end{aligned} \qquad (7.3.16)$$

where e_t is the observed residual at time t. As the t-statistics in parentheses indicate, all but the a_4 parameter is significant at the 5% level. The residual autocorrelations - the plot of which is given below - are all below significance levels as determined by the Quenouille statistics. As the plot clearly indicates, the only potential problem is for the residual autocorrelation at lag 1, which although not significant, is considerably higher than the those at higher lags. The regression has a M.S.E. of 2.832.

Figure 7.3
Plot of The First Twenty Residual Autocorrelations Resulting From Regression 7.3.16.



In order to allow for the autocorrelation at lag 1, a first order moving-average term was added to the latter regression, the results of which are presented below.

$$(1+\underset{(2.76)}{0.270}\,B^4)(1+\underset{(3.56)}{0.351}B^{12})M^*_t=$$

$$-\underset{(-1.67)}{0.330}(1-\underset{(-0.06)}{0.035}B)0.2PR^*_{t-1,t-5}+(1-\underset{(-1.70)}{0.141}\,B-\underset{(-6.08)}{0.629}\,B^{12})e_t \qquad (7.3.17)$$

Although the additional parameter is only significant at the 10% level, it is considered high enough to remain in the model, for which the M.S.E. is now 2.796. The addition of the first order moving-average term in the model has reduced the significance of the lower lagged profit variable, but, the outstanding problem with the estimated regression is the insignificant parameter on the longer lagged profit variable. Consequently the parameter was dropped from the model, a change which proved to be the final adjustment required. The results of this final estimation are presented in equation 7.3.18 below.

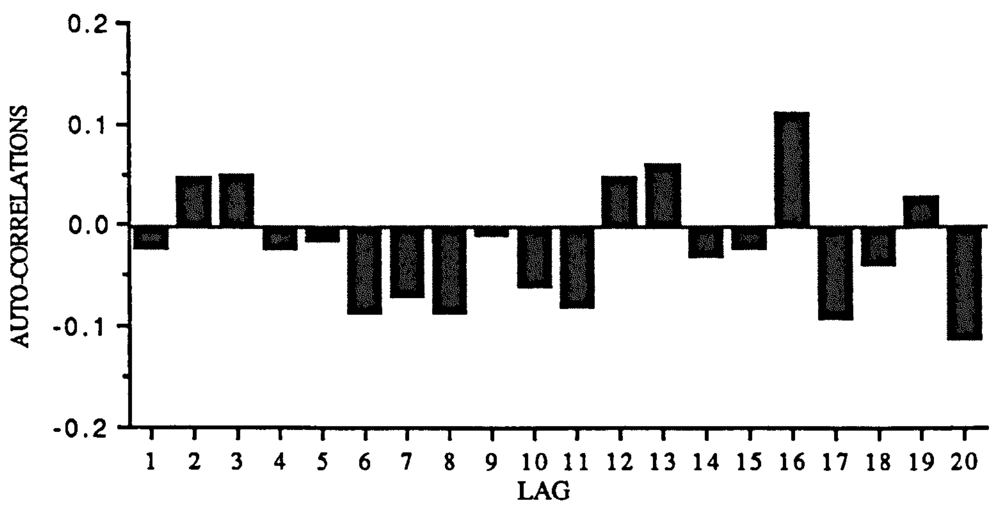
$$(1 + 0.271\,B^4)(1 + 0.349\,B^{12})\,M^*_t = \underset{(2.85)\quad\quad(3.60)}{}$$

$$- \underset{(-4.70)}{0.320}\,(0.2PR^*_{t-1,t-5}) + (1 - \underset{(-1.76)}{0.142}\,B - \underset{(-6.34)}{0.632}\,B^{12})\,e_t \qquad (7.3.18)$$

The M.S.E. for this final model is the lowest of all three estimated models taking a value of 2.768 and all of the t-statistics of the individual parameters has increased from that of the preceding estimation. The profit variable is now significant at the 1% level. There is no apparent problem with the residuals as indicated by the residual autocorrelation plot below and the fact that the first 20 residual autocorrelations have a Box-Ljung Q-statistic of 8.28 which is nowhere near significance at the 10% level when measures against the Chi-Square distribution for 15 degrees of freedom.

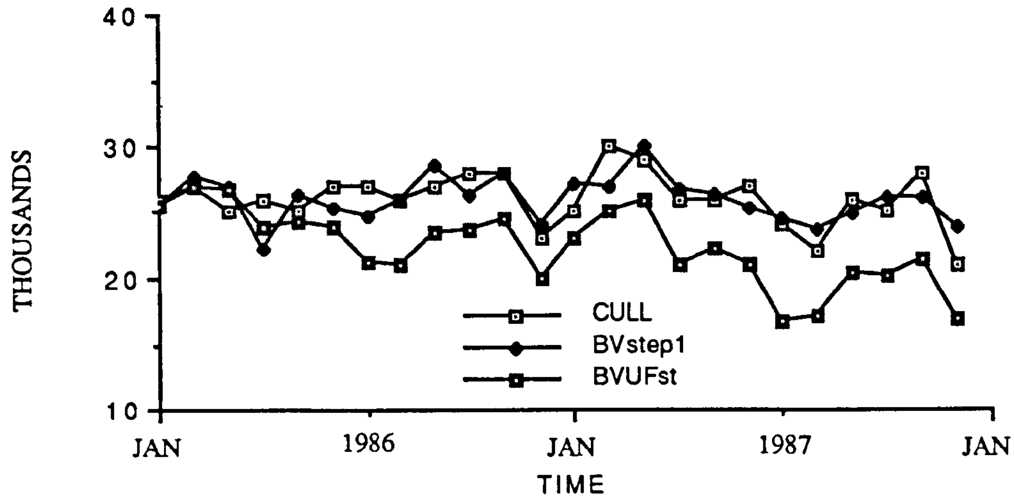
Figure 7.4

Plot of The First Twenty Residual Autocorrelations Resulting From Regression 7.3.18



Having built the model it can then be used for forecasting. Having said this, the long-run equilibrium coefficient on the profit ratio is -0.037, which indicates that the long run effect of profits on culling is small but negative. This result is not consistent with a finite life of breeding sow and, therefore, the model is essentially a short run model as far as forecasting is concerned. Because the forecasting ability of the model in the in-sample period was not of prime concern as a criterion for choosing the best model, only the 24 one-step conditional forecasts of the out-of-sample period and a 24-step unconditional forecast for the December of 1987 are illustrated in figure 7.5 below. The forecasts were derived from a computer program developed especially for the task. Firstly, univariate forecasts of profit are derived from the univariate Box-Jenkins model built in chapter six; these forecasts are then fed in to the bivariate model in order to obtain the culling forecasts.

Figure 7.5
Conditional and Unconditional Forecasts for the out-of-sample period 1986-87



The figure illustrates that the one-step conditional forecasts appear to forecast the out-of-sample period rather well, picking up the general fluctuations in cullings, though exhibiting a somewhat smoother appearance than the actual data. The MSFE for the conditional forecasts is 2.958 which implies an approximate 7% mean absolute error. Turning to the unconditional forecast for the out-of-sample period, these forecasts also pick up the general trend of cullings for the given period very well although the expected under-forecasting is realised. A look at the equivalent univariate model forecasts of cullings presented in chapter six, reveals that the univariate model forecasts are also prone to under-forecasting in this period though to a smaller extent. The exaggerated effect in the bivariate model is almost certainly, therefore, a result of the inclusion in the bivariate model of the additional profit variable. The unconditional out-of-sample univariate forecasts for the profit ratio, presented in figure 6.12 of chapter six illustrate the fact that they have a tendency to over-forecast profit from towards the end of 1986 onwards. Because of the negative coefficient on the profit ratio in the bivariate model, the exaggerated downward effect on the cull forecasts is explained by the univariate model's over-forecasting of profit for the out-of-sample period. The latter phenomenon illustrates the potential drawback of the bivariate model whose forecasting ability is dependent not only on the forecasting ability of itself but of the univariate model for the explanatory variable as well. This problem is discussed further in chapter eight.

7.4 A Bivariate Box-Jenkins Model For Fat Pig Slaughter.

Having built a model for the culling of sows and boars, the same methodology is employed to construct an equivalent model for the adjusted monthly fat pig slaughter figures. As the methodology employed in this model has been fully explained in the previous section, the discussion of the identification and estimation stages will be much briefer.

Once again, the second variable will be the profit ratio, as used in the culling model, which means that the univariate model 7.3.2 is again applicable. The equivalent univariate model for fat pigs is that identified and estimated in chapter 6, that is, equation 6.3.1, reproduced below as equation 7.4.1.

$$(1-B)(1-B^{12})FP_t = \underset{(-3.33)}{(1-0.295\,B)}\,\underset{(-28.0)}{(1-0.87\,B^{12})}\,FPe_t. \qquad (7.4.1)$$

where FPe_t is the white noise error term from the univariate fat pig model. Having used the univariate models to obtain the two pre-whitened series, the first step of identifying a relationship between the observed error terms FPe_t and PRe_t could commence by cross correlating them. The correlation at lag zero was nowhere near accepted levels of significance, indicating the desirable result of no simultaneity relationship between the two variables. The resultant cross-correlations for the observed fat pig and profit error variables, $Corr(FPe_t, PRe_{t-i})$, suggested the possibility of both a short term and a long term effect of profits on fat pig slaughter. As was the case in the bivariate culling model, the first 5 cross correlations were negative and relatively large, although none was individually significant compared with the Quenouille statistic for 107 observations. These correlations imply that an increase in profits, for example, will, in the short term, decrease the number of fat pigs slaughtered presumably because of transfers into the breeding herd in order to build it up as quickly as possible. To model this relationship, a variable which is an average of profits lagged 1 to 5 months has to be included on the right hand side of the model. The second pattern emerging from the cross correlations was a string of positive cross correlations from lags 11 to 18 inclusive, the exception being a small negative correlation at lag 13. Once again, none of these correlations was individually significant. Given that a sow's gestation period is approximately four months and given that fat pigs can be slaughtered from about the age of five months onwards, it is quite conceivable that a sharp increase in profits in month t could produce an increase in fat pig numbers some 11 months plus later. As with the short run effect, a simple average of profits from lag 11 to 18 inclusive will be included as a variable on the right hand side of the model. The identified model which was subsequently estimated, therefore, took the following form.

$$FPe_t = \gamma_1 \, (0.2 \, PRe_{t-1,t-5}) + \gamma_2 \, (0.125 \, PRe_{t-11,t-18}) + \varepsilon_t \qquad (7.4.2)$$

where ε_t is a white noise error term, and where $PRe_{t-1,t-5}$ and $PRe_{t-11,t-18}$, are the sums of the profit ratio index lagged 1 to 5 months and 11 to 18 months respectively.

The results of estimation are presented in equation 7.4.3 below.

$$FPe_t = \underset{(-1.80)}{-4.142} \, (0.2 \, PRe_{t-1,t-5}) + \underset{(1.62)}{4.629} \, (0.125 \, PRe_{t-11,t-18}) + e_t \qquad (7.4.3.)$$

Although neither parameter is significant at the 5% level, both were considered high enough to justify their inclusion in the model. The residual autocorrelations at lags 10 and 19 are significantly below zero, but, since there are no reasons, a priori, for including such variables in the regression, nothing was done about the potential problem. In addition, the Box-Ljung Q-statistic of 23.48 is not significant at the 10% level of significance, implying that the residuals as a whole would appear to be a white noise process.

Having obtained a bivariate relationship between the error structures of the fat pig and profit variables, the task remained to convert the relationship into one between actual fat pig slaughter and profits. Following the problems experienced at the equivalent stage when modelling the bivariate culling model, the decision was taken to achieve this by modelling the first and seasonally differenced series, defined as FP* and PR*. Re-defining the two relevant univariate models in terms of the differenced variables, and replacing the estimated coefficients with symbols results in equations 7.4.4 and 7.4.5 presented below.

$$FP^*_t = (1 - \beta_1 B)(1 - \beta_2 B^{12}) \, FPe_t. \qquad (7.4.4)$$

$$(1 - \beta_3 B) \, PR^*_t = (1 - \beta_4 B^{12}) \, PRe_t. \qquad (7.4.5)$$

Following a similar procedure to that carried out when deriving a model for M^*_t, equation 7.4.5 is substituted into 7.4.2, the resulting expression being substituted into equation 7.4.4. The end result of the two substitutions is equation 7.4.6.

$$FP^*_t = (1-\beta_1 B)(1-\beta_2 B^{12})\left[\frac{(1-\beta_3 B)}{(1-\beta_4 B^{12})}\right](0.2\,\gamma_1 PR^*_{t-1,t-5} + 0.125\,\gamma_2 PR^*_{t-11,t-18}) +$$
$$(1-\beta_1 B)(1-\beta_2 B^{12})\,\varepsilon_t \qquad (7.4.6)$$

As was the case with the culling model the fortunate result occurs that the estimated parameters of β_2 (0.874) and of β_4 (0.860) are close enough to be assumed equal, thus allowing the cancellation of the terms $(1-\beta_2 B^{12})$ and $(1-\beta_4 B^{12})$ on the right hand side of equation 7.4.6. The expression resulting from this cancellation of terms is reproduced in equation 7.4.7.

$$FP^*_t = (1-\beta_1 B)\{(1-\beta_3 B)(\gamma_1\, 0.2PR^*_{t-1,t-5} + \gamma_2\, 0.125PR^*_{t-11,t-18})$$
$$+ (1-\beta_2 B^{12})\, e_t\} \qquad (7.4.7)$$

Having obtained the latter expression, constrained estimation of the five included parameters could take place. The results of estimation are given in equation 7.4.8.

$$FP^*_t = (\underset{(-4.13)}{1-0.40\, B})\{(\underset{(-1.46)}{1-0.38\, B})(\underset{(-1.75)}{-7.10}*0.2PR^*_{t-1,t-5} + \underset{(2.29)}{7.84}*0.125PR^*_{t-11,t-18}) +$$

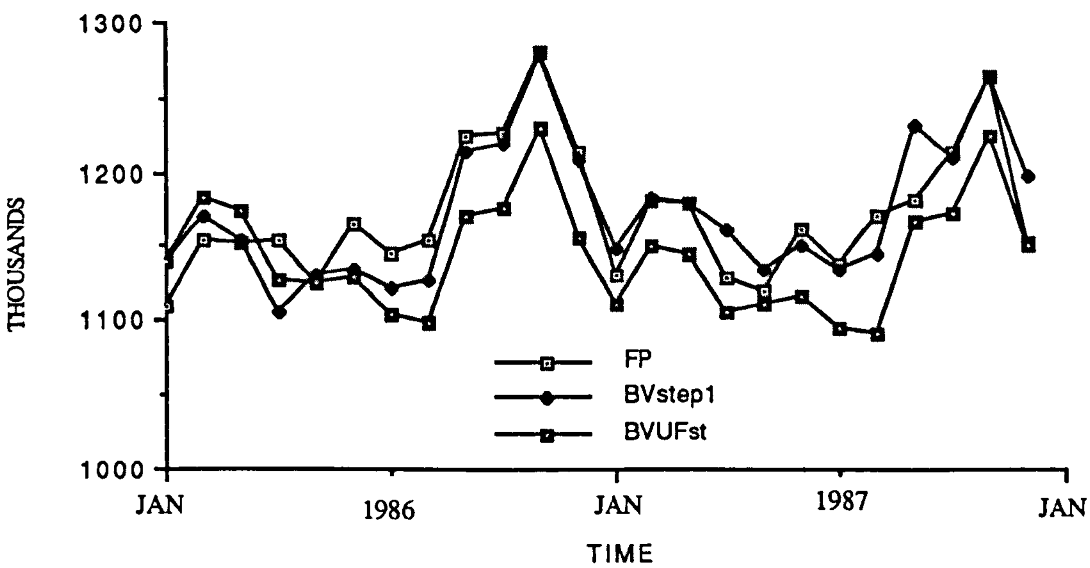
$$(\underset{(2.69)}{1-0.29\, B^{12}})\, e_t\} \qquad (7.4.8)$$

There is no reason to believe that the residual autocorrelation structure of 7.4.8 is other than white noise, and the regression has a MSE of 877.22. The estimated parameters are all of the expected sign and all but β_3 and γ_1 are significant at least at the 5% level. Because the t-statistic for the estimated coefficient of β_3 is below accepted levels of significance, the decision was taken to remove the coefficient form the regression. The results of re-estimation minus the said coefficient are presented in equation 7.4.9.

$$FP^*_t = (\underset{(-5.53)}{1-0.498\, B})\{(\underset{(-2.06)}{-3.85}*0.2PR^*_{t-1,t-5} + \underset{(2.04)}{5.587}*0.125PR^*_{t-11,t-18}) +$$
$$(\underset{(-5.18)}{1-0.466\, B^{12}})\, e_t\} \qquad (7.4.9)$$

The significance of each of the included parameters has increased from the previous model so that all except the estimated coefficient of γ_1 and γ_2 - which are significant at 5% - is significant at the 1% level. The MSE of the estimated regression is also lower than the previous regression at 834.62. Only one of the residual autocorrelations is significantly above zero compared with the Quenouille statistic for 100 observations and this is at lag 9. Because there is no apparent reason for including this lag in the model, the decision was taken not to adjust the model in any way. The Box-Ljung statistic of 14.40 gives no reason to suspect that the residuals are not a white-noise process when checked against the chi-square tables for 16 degrees of freedom. The estimated long run coefficient measures -0.036 indicating a small negative long run effect of profits on fat pig slaughter which is not consistent with our a priori expectations. Consequently, and as was the case with the bivariate culling model, the model is essentially a short run model, and the longer term forecasts are liable to under-forecast. The 24 one-step forecasts and the 24 step forecast for December 1987 are presented in figure 7.6 below.

Figure 7.6
Conditional and Unconditional Forecasts for the out-of-sample period 1986-87



The MSFE for the conditional one-step forecasts measures 552.5 which implies a mean absolute forecasting error of 23.5 which is approximately 2% of the average four week fat pig slaughter figures. The 24 one-step forecasts appear to forecast the two year period very well, picking up nearly all of the seasonal fluctuations in slaughterings. The 24 -step forecast for December 1987 illustrates that the unconditional forecasts also follow the seasonal slaughter fluctuations very closely, although it is apparent that in this case the two year forecast generally under-forecasts. Compared with the equivalent forecasts from the univariate fat pig model of chapter six, which are not prone to such under-forecasting it would appear that the problem is again a result of the influence in the model of the univariate profit ratio forecasts and the long run negative profit coefficient. The 24-step over-forecast of the profit ratio already discussed in the culling model has had the effect of reducing the slaughter forecast through the negative coefficient on the shorter lagged profit variable which has not been compensated by the positive coefficient of the profit variable lagged 11 to 18 months. Had the univariate profit model consistently over-forecast for 1986 as well as 1987, the net effect would have been to reduce the under-forecasting of fat pigs by the bivariate model. This discussion further highlights the importance for the forecasting abilities of the bi-variate models of the reliability of the longer term forecasting abilities of the univariate profit model.

7.5 Conclusion

In this chapter, the theory underlying the methodology for bivariate Box-Jenkins model building has been outlined, concentrating on the case of uni-directional causality. The theory was then applied to build monthly models for the culling and fat pig slaughter time series in order to derive alternative forecasting models to the univariate and biological models built in previous chapters. The exercise was of interest in that the methodology provides a framework for modelling two variables in which the direction of causality, and the lags involved in the causal relationships, need not be known prior to modelling, in order for a workable model to be built. In this chapter, however, the direction of causality was assumed known and the process of identification was undertaken bearing in mind prior knowledge of the biological lags existing in the breeding herd system. The derived models were used to generate one-step conditional forecasts and an unconditional forecast for the out-of-sample period. All the forecasts are quite satisfactory, except that there is indication that the models might struggle to forecast the correct level of slaughterings in the longer term due to the nature of the univariate forecasts of profits from the model built for that purpose in chapter six and the long run negative profit coefficient. The one-step, 12-step and 24-step forecasting ability of the bivariate Box-Jenkins culling fat pig models will be compared with the equivalent forecasts of the other monthly models in chapter eight.

CHAPTER EIGHT

FORECASTING ANALYSIS: TRIMESTIC AND MONTHLY MODELS

8.1 Introduction

In this chapter of the thesis, the relative forecasting abilities of the various models are compared in respect of their relative abilities to forecast the variable concerned, both in terms of level and, to a lesser extent, the predicted seasonal movements. The forecasting performance with respect to levels will be measured using the mean square forecasting error, MSFE, and the percentage root mean square forecasting error, RMSFE, of the average level of the variable concerned in the out-of-sample forecasting period. Models are derived for the trimestic breeding herd and the monthly sow and boar culling and fat pig slaughtering using the models built in the preceding five chapters.

The analysis is concerned with the forecasting performance of the models in the short to medium term and so it is proposed to look at three types of forecast. The short term forecasting abilities are assessed by conditional one trimester ahead forecasts for the breeding herd and one month ahead forecasts for the culling and fat pig series. For the more medium to long term forecasts, one year ahead and two year ahead forecasts will be analysed. The period which is to be forecast is the out-of-sample period of 1986-87 inclusive, which was left out of the estimation process specifically for this purpose. The breeding herd, culling and fat pig models are discussed separately, starting with the breeding herd forecasts.

8.2 Forecasting The Breeding Herd

Three types of breeding herd forecasting model have been built, namely, the univariate Box-Jenkins models of chapter three, the biological model of chapter four and the econometric model of chapter five. Because the latter two approaches have been estimated using the equally spaced trimestic time intervals between the April, August and December sample censuses, the forecasting analysis will take place on that basis, despite the fact that the univariate models are pseudo-quarterly, having been estimated including the June census. The consequence of this for the univariate model forecasts is that the one-step forecast for August is a forecast from June and not April and the one year and two years ahead unconditional forecasts are four-step and eight-step forecasts respectively.

The breeding herd forecasts from the biological and the econometric models are derived using the following identity as discussed in chapters four and five,

$$HB_t = HB_{t-1} + 1.0863\ PG_t - M_{t-1,t} \qquad (8.2.1)$$

The breeding herd forecasts, therefore require a forecasting model for the inflow and outflow variables, that is, pregnant gilts and culling respectively. The pregnant gilt biological forecasting model is that discussed in section 4.5g of chapter four and the equivalent econometric model is that presented in section 5.4 of chapter five. The biological and econometric culling models are those discussed in sections 4.5d and 5.5 respectively. The only unresolved question for forecasting the breeding herd concerns the univariate Box-Jenkins models. The question is to decide whether to use the aggregate univariate model for breeding sows or whether to aggregate the forecasts from the three component models, the chosen forecasts being added to the forecasts of the boar herd in order to derive the required breeding herd forecasts. This question is resolved in the following section, after which the comparative performances of the three approaches are analysed.

8.2a The Box-Jenkins Forecasts and the April 1987 Census Data

In this section analysis is conducted to find out whether the breeding sow herd is best forecast using the univariate Box-Jenkins model built for this series in chapter three, or whether it is better to aggregate the forecasts from the univariate models built for the three components of the breeding sow herd, also presented in chapter three. Discussion is also given to the question of whether or not the suspect figure for April 1987 should be replaced by a more acceptable figure and, if so, what that figure should be.

The data are pseudo-quarterly in that the June census is included as a data point as well as the three sample censuses utilised in the biological and econometric models. With two years of out-of-sample data to forecast, eight sample points are available for comparison. Because univariate models are primarily useful for their short term forecasting abilities it is the one- step forecasting performances of the two approaches for forecasting the breeding sows which will be used to determine the 'best' method.[1] All forecasts from the univariate Box-Jenkins models are produced as required by the TSP package.

1. Although the aggregate versus the disaggregate comparisons could have been analysed for the pregnant pig herd also, the fact that the pregnant pig herd was found not to be of direct interest, after having built the biological models, the decision was taken to confine the analysis to the key variable, the breeding sow herd.

The mean square forecasting error, MSFE, of the one-step forecasts from the breeding sow model presented in section 3.2a produced a value of 714.92, which compares with an equivalent statistic of 413.26 from the aggregate forecasts produced by the component models. The error statistics clearly favour aggregation from the component forecasts: both sets of forecasts, however, are dominated by large positive errors in April and June of 1987[2]. The square of the residual error for the June 1987 forecast contributes 58.3% of the overall sums of squares of forecasting errors from the pregnant sow model and 42% in the case of the aggregate forecasts. Both errors arise as a result of the highly suspect sample census figures produced by the April sample census of that year, as discussed in section 3.5 of chapter three. Inspection of the three component models indicates that the problem appears to lie with the data as recorded for the pregnant sow and the barren sow series, which together account for over 85% of the breeding sow herd total. Because of the analysis presented in section 3.5, the decision was taken to analyse the forecasting performance of the two approaches, replacing the April 1987 sample census figures by the relevant one-step forecasts produced by the breeding sow, pregnant sow and barren sow models. The pregnant gilt figure does not appear to be affected and, therefore, no adjustment to April 1987 is made to the pregnant gilt data.

The MSFE's of the one-step forecasts are compared after having made the April adjustment and ignoring the error for the April 1987 which would otherwise bias the analysis in favour of the aggregation approach. Having done this, the results still favour the aggregation approach since the MSFE statistics for the breeding sow model forecasts and the aggregate forecasts respectively are 252.45 and 196.46, illustrating the relative superiority of the aggregate forecasts. Both sets of 1987 forecasts are considerably better than those produced not having made the adjustment for April 1987, as indicated by comparing the appropriate MSFE statistics presented above. Although the results of the analysis are not presented, comparing the forecasts from the pregnant pig herd model and the aggregate of the forecasts from the pregnant sow and pregnant gilt herd models tells an almost identical story to that presented for the breeding sow forecast analysis. It should be noted that the four-step and eight-step forecasts for the chosen out-of-sample period 1986-7, will not be affected by the suspect April figure for 1987, this observation appearing only three periods before the end of the out-of-sample forecasting period. On the basis of the above analysis it would appear that it is the aggregation of the forecasts from the component models which forecast the better in the short term. These forecasts, when added to the forecasts from the univariate boar model will therefore be used to represent the Box-

[2]. Error = actual minus forecast

Jenkins univariate model forecasts of the UK breeding herd. The recorded boar figure for April 1987 is also substituted by the one-step forecast for that period from the univariate boar model. No consideration of the abilities of the two approaches to model the directional movements of the breeding herd is necessary because the movements forecast were identical.

In the light of the analysis presented above and in section 5.3, for any sensible discussion of the relative forecasting abilities of the various models for the breeding herd to take place, it seems advisable to replace the April 1987 sample census figure for the breeding herd by a forecasted figure. The need for adjustment is also justified when the implications of non-adjustment for comparing the relative forecasting performances of the various modelling approaches are analysed. Considering non-adjustment, the high April census figure produces a large over-forecast one-step ahead. For the univariate models this means an over-forecast for June 1987, and because the trimestic models are not concerned with June, it is the forecast for August which is affected in the biological and econometric models. It should be said that, though the June error from the univariate model will feed through via the moving average error term to lower the one-step univariate forecast for August, the forecast will be based principally upon the more reliable June census figure. Because the Biological and Econometric models are estimated on trimestic time periods, the analysis of forecasting comparisons has to be made on a trimestic basis. Although the pseudo-quarterly data will be used for estimating and forecasting the breeding herd in the univariate model, the June forecast will not enter into the forecasting comparison analysis. The consequence of this is that the analysis would be biased in favour of the Box-Jenkins methodology if no adjustment were made for the suspect April figures.

On the grounds that Box-Jenkins univariate models are expected to perform best in the short run, it is the one-step aggregate forecast of 828.25 from the univariate breeding sow component models and the one-step forecast of 45.29 for the boar herd which will be aggregated and substituted for the suspect sample census observations for these two series in April 1987. Thus, all subsequent analysis which involves forecasting using the breeding herd figure for April 1987 as an independent variable will have used the two forecast figures given above, and not the figure as recorded from the sample census. The implications of this, if any, for the comparison of the Box-Jenkins and the biological and econometric model forecasts are discussed at the end of section 8.3.

8.3a The One-Step Conditional Forecasting Results

Having determined that the April census figure for the breeding herd is to be replaced by the one-step conditional forecasts for that period produced by the Box-Jenkins models for the four components of the breeding herd, the one trimester step conditional forecasts for the breeding herd from the univariate, trimestic biological and the trimestic econometric model can be produced and compared. To produce the biological and econometric forecasts, software for a micro-computer had to be developed. Although the forecasts for one trimester ahead have been labeled as conditional forecasts this is not strictly true for the econometric model. Making the assumption that the census results are known immediately after the census is taken, a one-trimester ahead forecast of the breeding herd using the econometric model requires a four month ahead unconditional forecast of the profit ratio using the univariate Box-Jenkins model derived for the profit ratio in chapter Six. Having forecast the profit ratio, the forecast figures are incorporated into the relevant profit ratio variables as required by the trimestic culling and pregnant gilt econometric models.

Figure 8.1
The One-Step 'Conditional' Forecasts For The Breeding Herd By the Univariate, Biological and Econometric Models

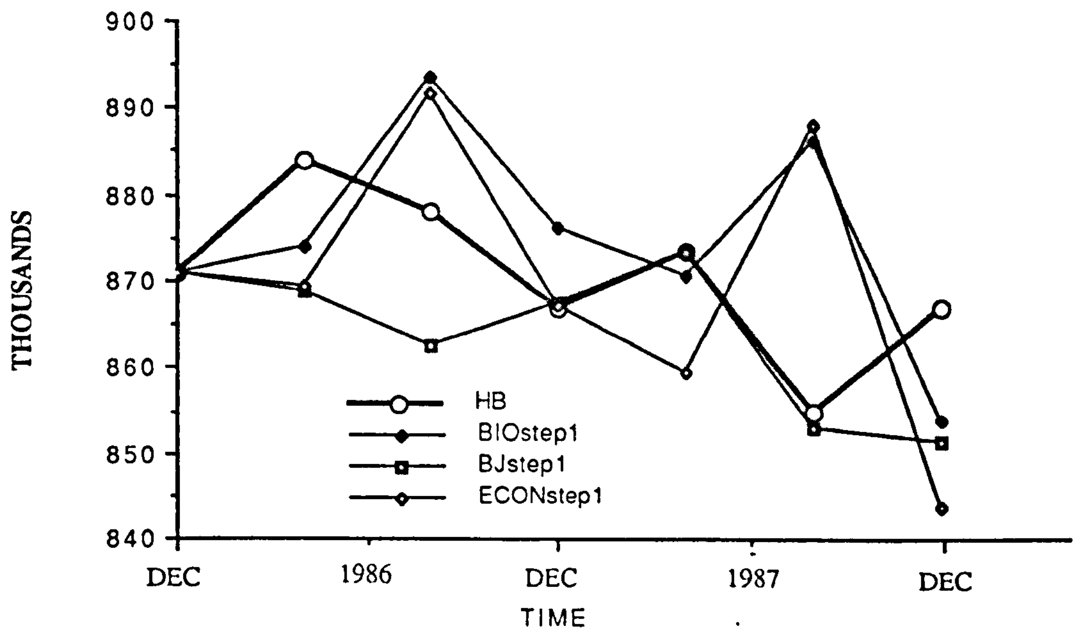


Figure 8.1 - which includes the univariate forecasts for April 1987 rather than the actual recorded figure from that sample census - illustrates that the breeding herd over the given out-of-sample period experiences a general downward trend and has a mean value of 870.76 thousand pigs. There are three downward movements and three upward movements and there are four census to census changes of direction. The 'pseudo one-step conditional' forecasts using the Box-Jenkins univariate models

correctly predicts three of the directional movements and two of the four turning points. The MSFE measures 143.3 thousand which converts to a RMSFE percentage of 1.4% of the breeding herd in the out-of sample period. The latter figures have been calculated excluding the result for April 1987 for which the error has been fixed at zero for the univariate models and is therefore, biased in favour of the Box-Jenkins models if included.

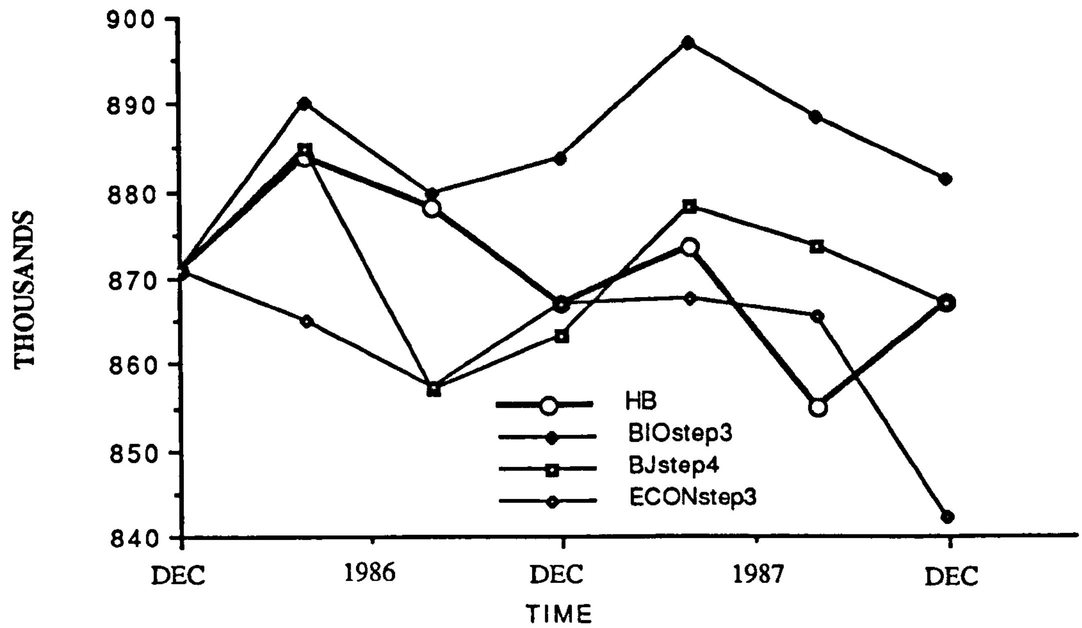
The one-step forecasts from the biological and the econometric models are very similar which is not surprising given that the econometric model includes biological terms as well as the profit variables. Also as expected, the forecasting performance of the biological and econometric models is not as good in the short term as that of the univariate Box-Jenkins models. The biological and econometric models correctly predict the direction of movement three times and once respectively, the biological model correctly predicting one of the turning points whereas the econometric model fails to forecast any. The MSFE statistics for the two sets of forecasts are 311.3 and 406.2 respectively, and these imply 2.1% and 2.3% absolute mean forecast errors using the root value of the MSFE as the relevant statistic. Of the two non-univariate models, therefore, it is the less sophisticated biological model which is marginally the better of the two, although they are both out-performed by the univariate model.

8.3b The One Year Ahead Forecasts of the Breeding Herd

The required one year ahead unconditional forecasts were generated using the TSP package for the univariate models and the software developed for the econometric and the biological models, the econometric forecasts requiring a 12-month ahead forecast of the profit ratio using the univariate model for this series. The results of the various forecasting procedures are illustrated in figure 8.2 below.

In terms of the models' abilities to forecast seasonal movements, the univariate model correctly forecasts four of the six seasonal movements and two of the four turning points. The MSFE of 134.8 the root of which implies a 1.3% average absolute forecast error, marginally better than the one-step forecast result although the MSFE statistic has been calculated having included all six of the out-of-sample observations which was not so in the one-step case. The obvious comment to make about the biological model's ability to forecast one year ahead is that all the forecasts over-forecast the actual figure to varying degrees. Having said this, the MSFE of 361.5 is not much higher than the 311.3 from the one-step forecasts, and implies an average 2.2% absolute forecasting error. The biological model correctly forecasts four of the seasonal movements and two of the changes of direction as did the one year ahead univariate forecasts.

Figure 8.2
The One-Year Ahead Unconditional Forecasts For The Breeding Herd By the Univariate,Biological and Econometric Models



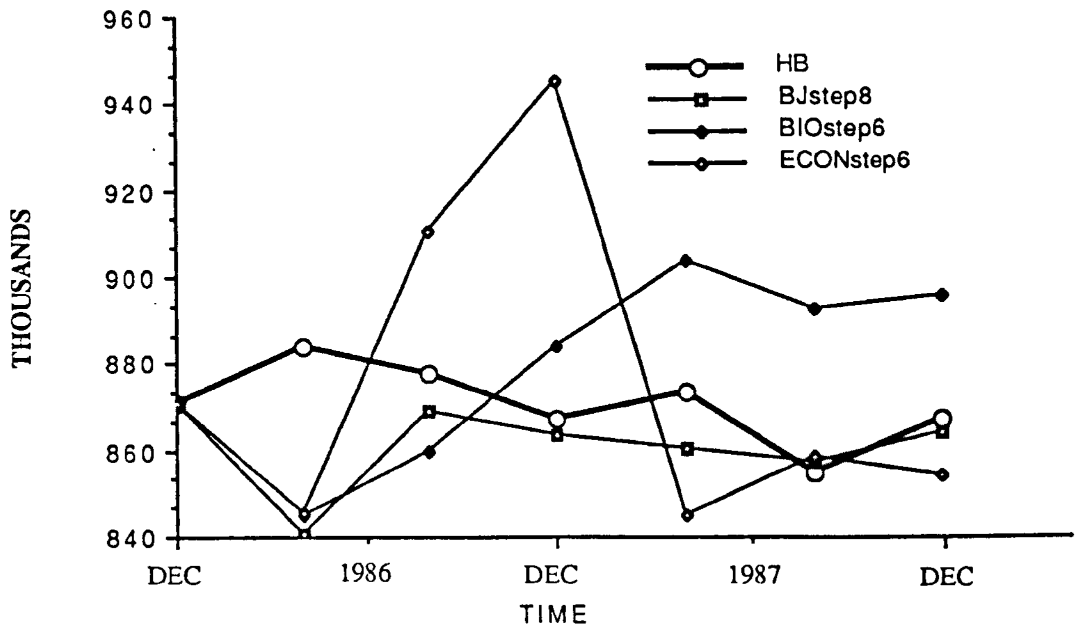
Turning to the econometric model forecasts, they are now more dependent on the forecasting ability of the univariate profit ratio model. The MSFE is lower than for the biological model, but not as good as the univariate model, taking a value of 257.5 which converts to a root mean square error equivalent to 1.8% of the breeding herd. The forecasts are not as good as those of the univariate and biological models in term of ability to forecast the seasonal movements.

The conclusion for the one year ahead forecasts then, is that on grounds of both the MSFE statistic and the directional analysis, the univariate model is again the best of the three models for the given period. This is somewhat surprising in view of the fact that we are dealing with a one year period for which we might expect the biological and the econometric model to perform the better.

8.3c The Two Years Ahead Forecasts of the Breeding Herd

The two-year ahead unconditional forecasts of the breeding herd from the three types of models were generated in a similar fashion to the one year ahead forecasts, the univariate forecasts being generated using an eight-step unconditional forecast and the econometric forecasts having made the relevant 24-month ahead univariate model forecasts of the profit ratio. The resulting forecasts from the three sets of models are presented in figure 8.3.

Figure 8.3
The Two-Years Ahead Unconditional Forecasts For The Breeding Herd
From the Univariate, Biological and Econometric Models



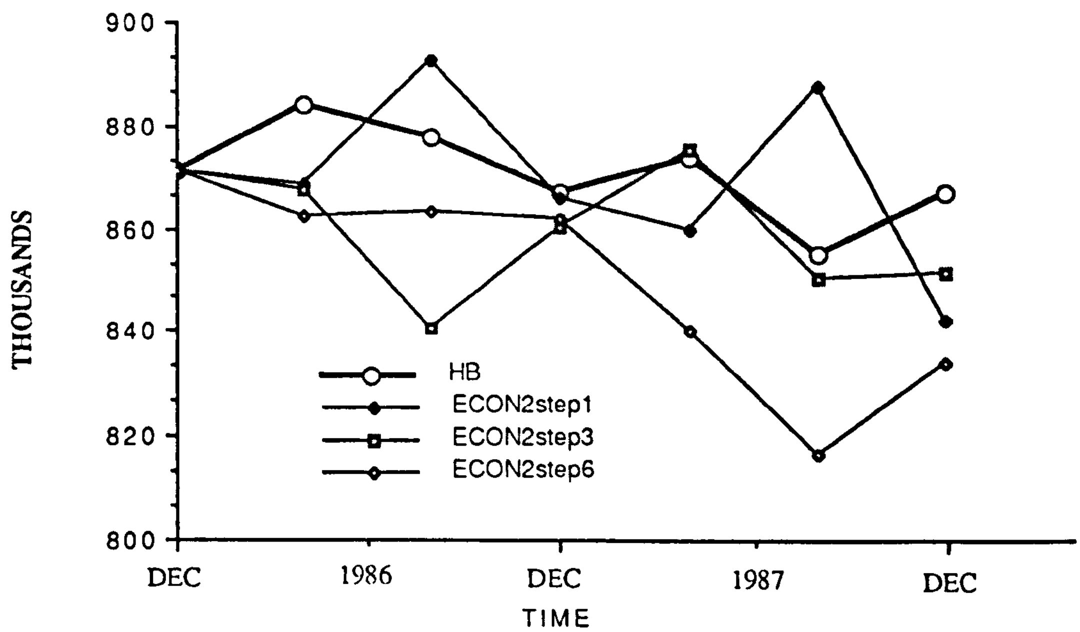
The univariate model has an MSFE statistic of 360.7 which is more than twice the equivalent values for the univariate model forecasts, one trimester and one year ahead, so that the univariate model is significantly worse at forecasting two years ahead than one year ahead. This is supported to some extent by the fact that the univariate forecasts correctly forecast only one of the four actual changes in direction and gets the actual seasonal movement correct on three occasions. The six-step forecasts from the biological model produce a MSFE of 881.6 which is also greater than twice its value when forecasting one year ahead and suggests an absolute forecasts error of 3.4%. This again is larger than the equivalent statistic from the univariate model forecasts, mainly due to over-forecasting by the biological model in 1987. Having said this, the biological forecasts are as good as the univariate forecasts in predicting the seasonal movements of the breeding herd and better in terms of directional changes in that the biological forecasts correctly predict two of the four direction changes. The two years ahead forecasts using the econometric model have a very large relative MSFE of 1609.2, the root of which accounts for 4.6% of the breeding herd for the out-of-sample period. This large error variance is due mainly to the large over-forecasts of the herd in August and especially in the December of 1986, over-forecasts which are the result of the 24-step over-forecasts produced by the univariate profit ratio model. In addition to the large error variance, the forecasts from the econometric model fail to pick up any of the seasonal movements in the breeding herd in the given two year period.

A notable feature of the two-year forecasts from all three models is the relative closeness of all three forecasts of the April 1986 breeding herd, all of which under-forecast the actual recorded figure and wrongly predict the fact that the recorded figure is an increase on the previous December. Overall, the two years ahead forecasts are worse in respect of the size of errors than is the case for one-step and one year ahead forecasts. The result is to be expected but the fact that the error variance of the univariate Box-Jenkins model forecasts is lower than that of the biological model is rather surprising. Of course, these results are only applicable to the specific time period available for out-of-sample testing, although there is little doubt about the relative superiority of the univariate model over both the biological and the econometric forecasts for both the short and the medium/long term.

Because it was felt that the forecasting ability of the econometric model was handicapped to some extent by the ability of the univariate profit model to forecast the profit ratio, and particularly in the longer term, it was thought to be of interest to look at the econometric forecasts using the actual profit ratio, rather than the unconditional univariate forecasts. Figure 8.4 illustrates the results of these forecasts.

Figure 8.4

The Forecasts For 1986-7 From the Econometric Model Using Actual Profits



The one trimester ahead forecasts of the breeding herd using the actual profit ratio rather than the four month ahead unconditional profit ratio forecasts produces very similar forecasts, although the MSFE of 427.5 is surprisingly larger than that having used the forecasts of the profit ratio.The one year ahead forecasts have an MSFE of 328.7, the root of which accounts for 2.1% of the breeding herd. This is superior to the one-step equivalent figure and is lower than the three-step forecast errors from the biological model. Having said this, the somewhat surprising result is that the MSFE is actually better than that of the econometric three-step forecasts, when the profit forecasts and not the actual profit were used. On the other hand the three-step forecasts using actual profit are superior to the 'true' three-step econometric forecasts in terms of forecasting seasonality in the breeding herd, four of the six seasonal movements and two of the four changes in direction being forecast correctly.

Looking at the six-step econometric forecasts using actual rather than predicted profit, the effect of the profit forecasts on the forecasting ability of the econometric model becomes more apparent. The forecasts for 1986 are considerably better than the large over-forecasts produced when forecasting profits although there is a tendency for the model to under-forecast 1987. The MSFE of 735.0 is less than half of what it was when profits were forecast and the seasonal movements are forecast correctly on three occasions, and one of the four changes in direction is picked up as opposed to none being correct when profits were forecast. The MSFE which converts to a root mean absolute error of 3.1% is now lower than the equivalent statistic for the biological model, although it is still slightly more than double the MSFE of the univariate two-year forecast errors. The MSFE and the RMSFE percentage from each of the various short and longer term forecasts are summarised in table 8.1 below.

Table 8.1[3]

The Error Statistics from the Trimestic Breeding Herd Forecasts

	MSFE's AND RMSFE PERCENTAGES			
STEP	UNIVARIATE	BIOLOGICAL	ECONOMETRIC 1	ECONOMETRIC 2
1-STEP	**143.3 (1.4%)**	**311.3 (2.1%)**	406.2 (2.3%)	**427.5 (2.4%)**
1-YEAR	134.8 (1.3%)	361.5 (2.2%)	257.5 (1.8%)	328.7 (2.1%)
2-YEAR	360.7 (2.2%)	881.6 (3.4%)	1609.2 (4.6%)	735.0 (3.1%)

In conclusion, therefore, the univariate model appears the best for all types of forecasts studied up to two-years ahead when comparing the MSFE statistics. This result was not expected for the one-year ahead and especially for the two-year ahead

3. All figures in thousands of pigs
Figures in bold indicate the fact that the April 1987 forecasts error has been removed from the analysis to remove the bias in favour the univariate forecasts.
RMSFE = The Root Mean Square Forecast Error.
ECONOMETRIC 1 = profits are forecast.
ECONOMETRIC 2 = actual profits used.

unconditional forecasts. The unconditional forecasting ability of the econometric model is clearly affected by the ability of the univariate profit ratio model to forecast profits two years ahead although the picture is less clear one year ahead where the results of comparing the MSFEs and the ability of the models to predict seasonal movements in the breeding herd conflict. Removing the effects of forecasting by using the recorded profit ratio rather than the forecast figure the econometric model performs better than the biological model three-step and six-step ahead. Having said this, the forecaster will not know the future profit ratio figure and so as a working model, he may prefer the biological to the econometric model, unless a better model for forecasting profits in the longer term can be found, or alternatively, rather than modelling profits at all, the modeller may prefer to use expertise in the field to give a prediction of profits based on his knowledge of the market. All these conclusions refer to the results of the analysis on the given out-of-sample period of 1986-7 and further analysis as more data become available would enhance the robustness of these conclusions.

To end this discussion of the trimestic breeding herd forecasting performance it should be stressed that any bias towards the Box-Jenkins 'one-step' forecasts from having chosen the forecasts from the univariate component models to represent April 1987 is relative only. Of the three one-step forecasts for April 1987, the univariate and the biological model forecasts are very similar although the univariate is marginally the larger of the two. Had there been no adjustment to the recorded figure for April 1987, therefore, the univariate forecast would have been the closest anyhow and moreover the large over-forecasts by the biological and the econometric model for the following August would have been even larger. The ex-post knowledge that the univariate model is the best forecaster for all lead periods also helps to justify the univariate forecast for April 1987 as the correct choice.

8.4 Forecasting Monthly Sow and Boar Culling

Attention is now turned to analysing the relative forecasting performance of the three types of model which have been built for the monthly cull series. Three sets of forecasts will be compared: the one-step, that is, one month ahead conditional forecasts to analyse the short term forecasting ability, and for the medium to long term, 12-month and 24-month unconditional forecasts are made. The three types of model include the univariate Box-Jenkins model discussed in Section 6.3 of chapter Six, the forecasts from which are produced automatically by the TSP package. The extension of the univariate Box-Jenkins model for culling is the bivariate Box-Jenkins model using profit as the independent variable built in section 7.3 of the previous chapter. Because both the one-step conditional and the unconditional forecasts using the bivariate model require forecasts of the profit ratio using the univariate profit ratio model, software for a microcomputer was again built in order to feed into the model the required profit forecasts. The reason for the need for a one-step profit forecast when forecasting culling one month ahead is that the profit regressor includes the value of the profit ratio at lag zero, therefore the forecast is only pseudo-conditional. The remaining cull forecasting model is the biological model discussed in section 4.6a of chapter six, the actual model being presented in equation 4.6a.1d of Appendix 4.6. In order to make unconditional forecasts using the monthly biological model for culling, unconditional forecasts of the UK breeding herd - the independent regressor - are required. The decision was taken to use the unconditional forecasts of the breeding herd as produced by the trimestic biological model, thereby keeping the forecasts wholly biological in nature.

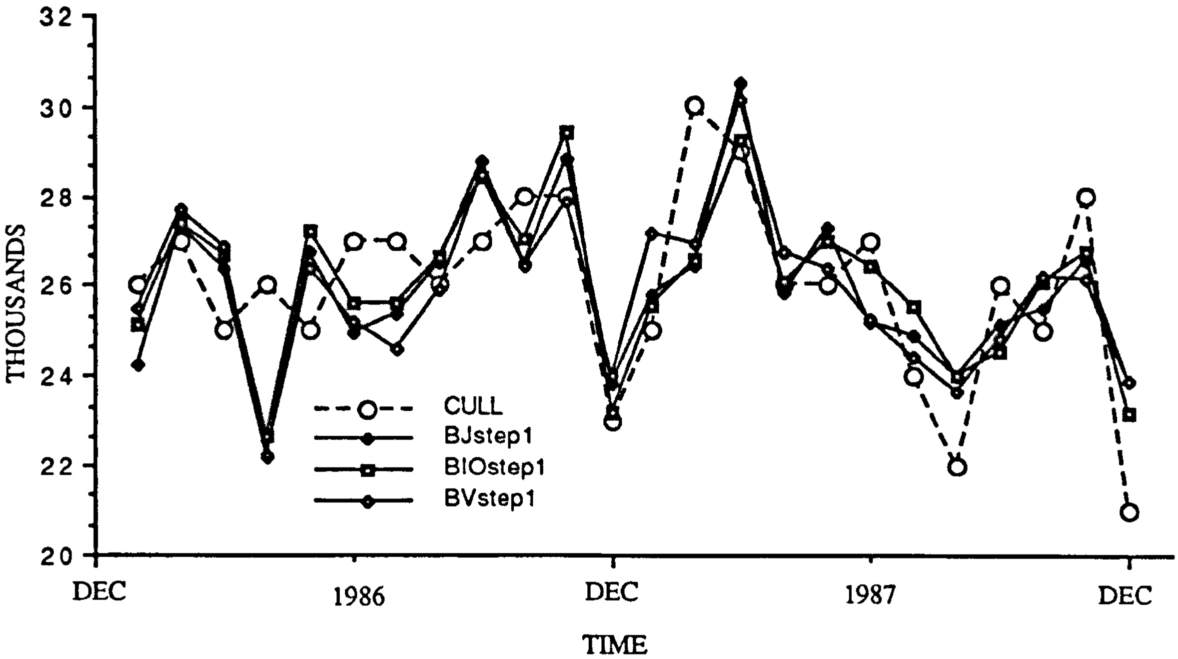
Where the April 1987 breeding herd figure is required to generate forecasts, the forecasts from the univariate models is used instead of the suspect recorded figure from that sample census as is the case throughout this chapter. Software was again developed so that the required forecasts could be fed into the monthly cull forecast generating function. Having built the relevant software for the bivariate and the biological models, the conditional and the unconditional forecasts were made, the results of which are discussed below.

8.4a The One-Month Ahead Conditional Forecasts For Culling

The 24 one month ahead forecasts were generated as described above for each of the three types of approaches to modelling, the results being presented in graphical form in figure 8.5. The actual recorded culling figures, all of which represent four week culling periods and are rounded to the nearest thousand, show a series which is clearly affected by seasonality. The series experiences 11 month to month increases, 10 decreases and 3 none movements in direction, and there are 15 recorded month to month changes of direction.

Figure 8.5 illustrates one-step forecasts from the three types of model which appear to be quite similar to one another. A couple of observations stand out, namely April 1986 and February 1987, when the conditional forecasts have relatively large and very similar errors. Because these errors are not repeated for the same months in the other out-of-sample period, it may suggest that the recorded figures are possibly suspect or that something unusual is going on in those particular months. Because the culling data are collated by the slaughter houses, sampling errors are much less likely than in the farm sample censuses although there is always a possibility of a recording error. Because 24 observations are available for the analysis it was thought unnecessary to correct for these possible outliers.

Figure 8.5
The One-Month Ahead Conditional Forecasts For Culling From the Univariate, Biological and Bivariate Box-Jenkins Models

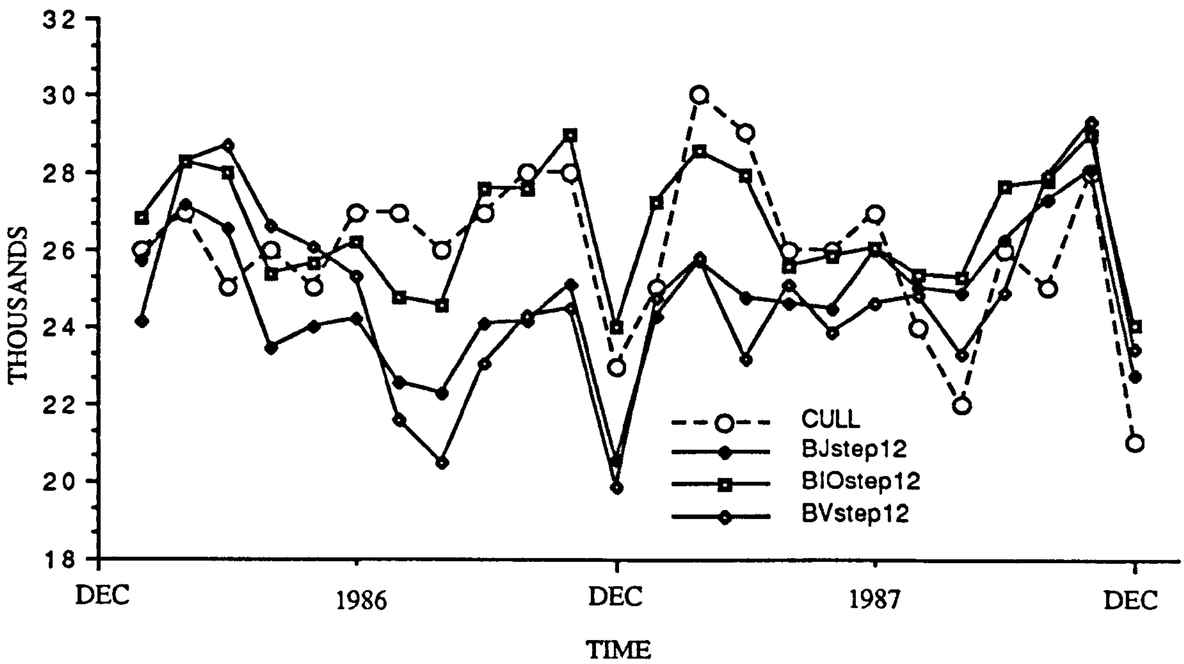


The univariate Box-Jenkins one-step forecasts have a MSFE of 3.05, the root of which converts to 6.7 percent of the average monthly cull figures for the out-of-sample period. The forecasts correctly predict 16 of the 21 seasonal movements. The pseudo-conditional forecasts of the bivariate Box-Jenkins model give an MSFE of 2.96, which implies an absolute error of 6.6%, an improvement on the performance of the univariate equivalent and they correctly predict the same number of seasonal movements. Turning to the alternative biological model, it is somewhat surprising that this model performs better than the two Box-Jenkins models at forecasting one-step ahead, though having said this, the lags in the biological model are short. The MSFE measures 2.45, the root of which implies a 6.0% average absolute forecast error and is lower than the equivalent statistics for the two Box-Jenkins model forecasts. The biological model one-step forecasts are also better at predicting the month to month movements in the recorded series correctly predicting 18 of the 21 movements. In conclusion, therefore, although the three sets of conditional forecasts are very similar in terms of actual forecasts and their ability to forecast the month to month movements, the biological model would appear to be the better of the three over the given out-of-sample period.

8.4b The One-Year Ahead Unconditional Forecasts For Culling

The one year ahead unconditional forecasts involve twenty-four 12-step forecasts using each of the three types of model, the results of which are given in figure 8.6 below.

Figure 8.6
The One-Year Ahead Unconditional Forecasts For Culling From the Univariate, Biological and Bivariate Box-Jenkins Models



The univariate Box-Jenkins forecasts have a MSFE of 6.23 which implies an average absolute error of 9.6%. Although this is worse than the one-step conditional forecasts' equivalent, the forecasts are good at predicting the seasonal changes of direction, getting 10 of the 15 correct. As figure 8.6 illustrates these forecasting results are explained by the fact that the univariate model is under-forecasting actual cullings for much of the latter half of 1986 and the fist half of 1987, but picking up the seasonal movements in the series rather well.

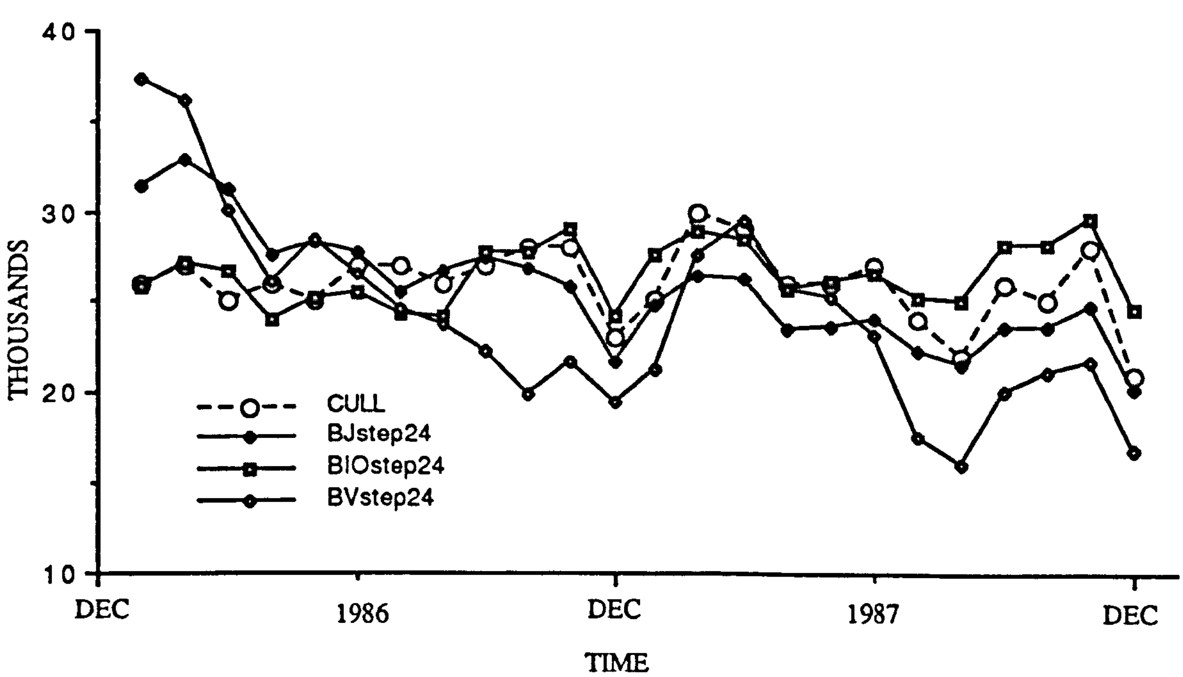
Unlike the results of the one-step forecasting procedure, the 12-step unconditional forecasts of the bivariate Box-Jenkins model are not as good as the univariate model both in terms of size of error and the ability of the model to pick up the month to month fluctuations.The MSFE of 9.03 converts to a RMSFE of 11.5% and only six of the 15 seasonal changes of direction are forecast correctly. Apart from the first five months of 1986, when the bivariate forecasts are somewhat higher than those of the univariate model, the forecast levels from the bivariate model are not that different from the levels forecast by the univariate model. On these grounds it would appear that the univariate profit forecasts are not having that great an influence on the forecast of culling levels, but the effect on directional change forecasting is somewhat greater and an adverse influence rather than good.

Analysing the biological 12-step forecasts, figure 8.6 clearly illustrates the expected superiority of the biological model over the Box-Jenkins models in forecasting the level of cullings. This is supported statistically by the MSFE of 2.72 which is less than half that of the next best univariate model at 6.23. The RMSFE is 6.3% of the average culling figure over the out-of-sample period and the forecasts correctly predict 9 of the 15 directional changes, one less than that of the univariate model. In conclusion, there is little doubt that the biological model is the best at forecasting one year ahead when the MSFE and the ability of the models to forecasts directional changes are considered together. For all three models' forecasts, the errors are larger than they had been when conditionally forecasting one month ahead as one would expect.

8.4c The Two-Year Ahead Unconditional Forecasts For Culling

The two years ahead unconditional forecasts, consist of twenty-four 24-step forecasts from the univariate and bivariate Box-Jenkins models and the biological model, the resulting forecasts appearing in figure 8.7.

Figure 8.7
The Two-Year Ahead Unconditional Forecasts For Culling From the Univariate, Biological and Bivariate Box-Jenkins Models



Apart from some relatively large over-forecasts for the first three months of the out-of-sample period, the univariate forecasts are not that much worse than the 12-step unconditional forecasts discussed above. The MSFE of 8.08 implies an average absolute error of 10.9% which is only 1.3% larger than in the 12-step case. The forecasts appear to pick up the seasonal movements in the series quite well, and 7 of the 15 changes of direction are forecast correctly. The 24-step forecasts from the bivariate model are the worst set of forecasts in terms of ability to forecast the correct level with notable over-forecasting at the start of 1986 and under-forecasting at the end of both 1986 and 1987. The MSFE measures 25.98, the RMSFE being the equivalent of 19.6% of the average monthly culling period. Having said this, and bearing in mind that the bivariate forecasts also require a 24-step univariate forecast of the profit ratio to be fed into the forecasting procedure, the forecasts still predict many of the month to month movements quite well. Once again it is the biological model which turns out to be the best forecasting model for the culling series. The MSFE of 3.02 is less than half that of the univariate model 24-step forecasts and implies an average absolute forecast error of 6.7%. The biological model is also the best at forecasting the seasonal movements, correctly forecasting 10 of the 15 seasonal changes of direction, which is as good as any other set of forecasts, and better than the biological model's own performance when forecasting one-step and 12-steps ahead.

Table 8.2[4]
The Error Statistics from the Monthly Culling Forecasts

	MSFE's AND RMSFE PERCENTAGES					
STEP	UNIVARIATE		BIVARIATE		BIOLOGICAL	
1-STEP	3.05	(6.7%)	2.96	(6.6%)	2.45	(6.0%)
1-YEAR	6.23	(9.6%)	9.03	(11.5%)	2.72	(6.3%)
2-YEAR	8.08	(10.9%)	26.0	(19.6%)	3.02	(6.7%)

In conclusion, in terms of both the MSFE statistics, which are summarised in table 8.2 above, and in terms of the models' abilities to forecast month to month changes in direction, the biological model clearly comes out on top when forecasting both short and medium/long term. This result is expected in the longer and medium term, but was not the result expected for the one month ahead forecasts, where the Box-Jenkins models were expected to perform the best. Comparing the relative performance of the two Box-Jenkins models, the bivariate model, has a lower MSFE for the one-step ahead forecasts indicating that the inclusion of the profit variable has added some explanatory power to the univariate model. For the two unconditional forecasts,

4. All figures in thousands of pigs

however, the story is reversed, and indicates that the combination of having to forecast profits and the culling series also, has a detrimental effect on the relative forecasting ability of the bivariate model. Compared with the trimestic breeding herd models, the monthly culling models show an improved ability to forecast the step by step changes of direction, although the absolute error percentages are, on the whole more than double the equivalent percentage in the breeding herd models. This is almost certainly a result of the fact that the monthly cull figures are smaller in magnitude than the breeding herd figures and both are rounded to the nearest whole figure, so that the proportion of the forecast error accountable to any rounding error will be greater in the culling forecasts.

8.5 Forecasting Monthly Fat Pig Slaughter

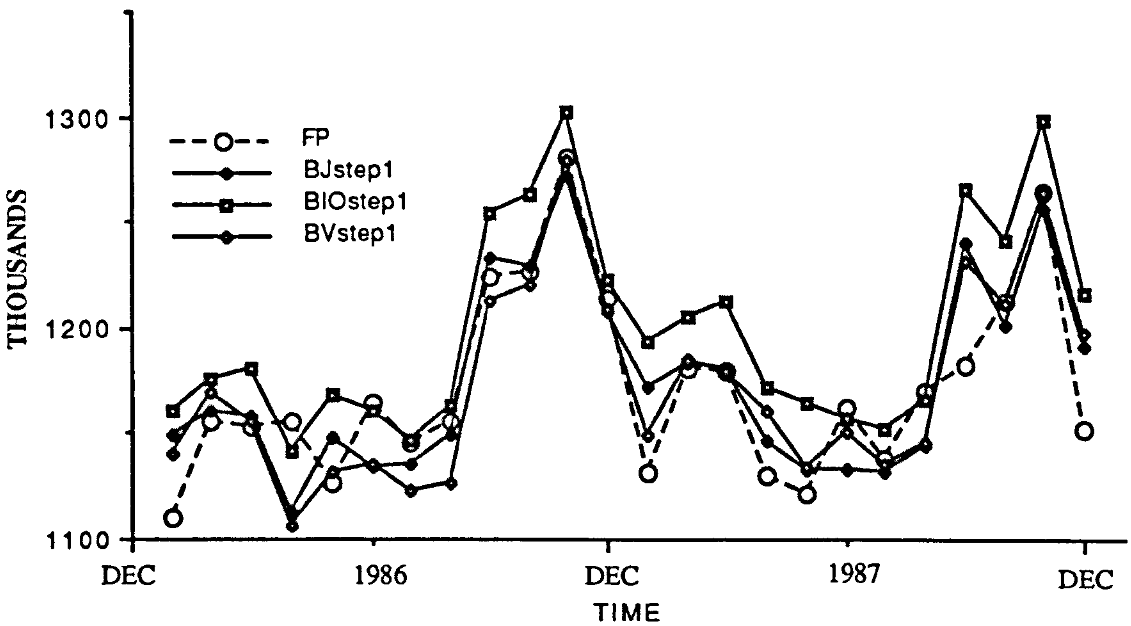
The conditional and unconditional forecasts for the fat pig slaughter series are obtained using the same three approaches used to produce forecasts of monthly cullings. The univariate Box-Jenkins model is that built in section 6.3 of chapter six and the corresponding bivariate model is that discussed in section 7.4 of chapter 7. Again, all the forecasts using the bivariate model require forecasts of the profit ratio using the univariate model developed for that series in chapter six. The biological fat pig model is that developed in section 4.6c of chapter four, the model itself being presented in equation 4.6c.4d of appendix A4d. The reader may remember that the biological model contains within it a positive time trend which models the fact that the breeding herd has become more productive in terms of fat pigs reared per litter per annum over the estimation period. The forecasts of the independent breeding herd variable required to unconditionally forecast the fat pigs with the biological model are once again the unconditional forecasts from the trimestic biological breeding herd model. As was the case when forecasting monthly culling, forecasting fat pigs with the biological and the bivariate Box-Jenkins models required micro computer programs to be developed. Having achieved this, the forecasts for one month, 12 months and 24 months ahead were made, the results of which are discussed in turn below. The fat pig slaughter numbers represent four week accounting periods and are rounded to the nearest thousand, and as the forecast diagrams will show, the slaughter figures for the out-of-sample period are clearly subject to seasonal fluctuations, the number of slaughterings clearly rising in September, October and November. The month to month movements in the slaughter series can be summarised by saying there are 11 decreases, 13 increases and 14 changes of direction.

5.5a The One-Month Ahead Conditional forecasts for Fat Pig Slaughter.

The term conditional is again used to denote the fact that the forecasts are produced for one-month ahead of the latest observed value, although the inclusion of a zero lagged profit variable in the bivariate Box-Jenkins model means that the bivariate forecast does depend on a one-step forecast of the profit ratio by the previously mentioned univariate model for this series. The one-step forecasts are reproduced in graphical form in figure 8.8 below.

Figure 8.8

The One-Month Ahead Conditional Forecasts For Fat Pigs From the Univariate, Biological and Bivariate Box-Jenkins Models



The figure indicates that all three sets of forecasts appear to pick up the general and seasonal trends in the fat pig series although there is evidence that the biological forecasts are, on the whole, over-forecasting the actual level of slaughterings. The univariate forecasts have a MSFE of 573.5 the root of which accounts for 2.0% of the average slaughter figure for 1986 and 1987 combined, and the model correctly predicts 9 of the 14 turning points. Comparing these results with those of the bivariate Box-Jenkins model, the bivariate forecasts are very similar to univariate forecasts as one would expect one-step ahead: however, the evidence, as was so with the culling model equivalents, favours the more sophisticated bivariate model. The latter model's MSFE of 522.5 is an approximate 9% improvement on that of the univariate model, and MSFE implies a 1.9% average absolute error. The bivariate forecasts are also slightly better in terms of their ability to forecast the seasonal movements of the actual series, 10 of the 14 direction changes being correctly predicted.

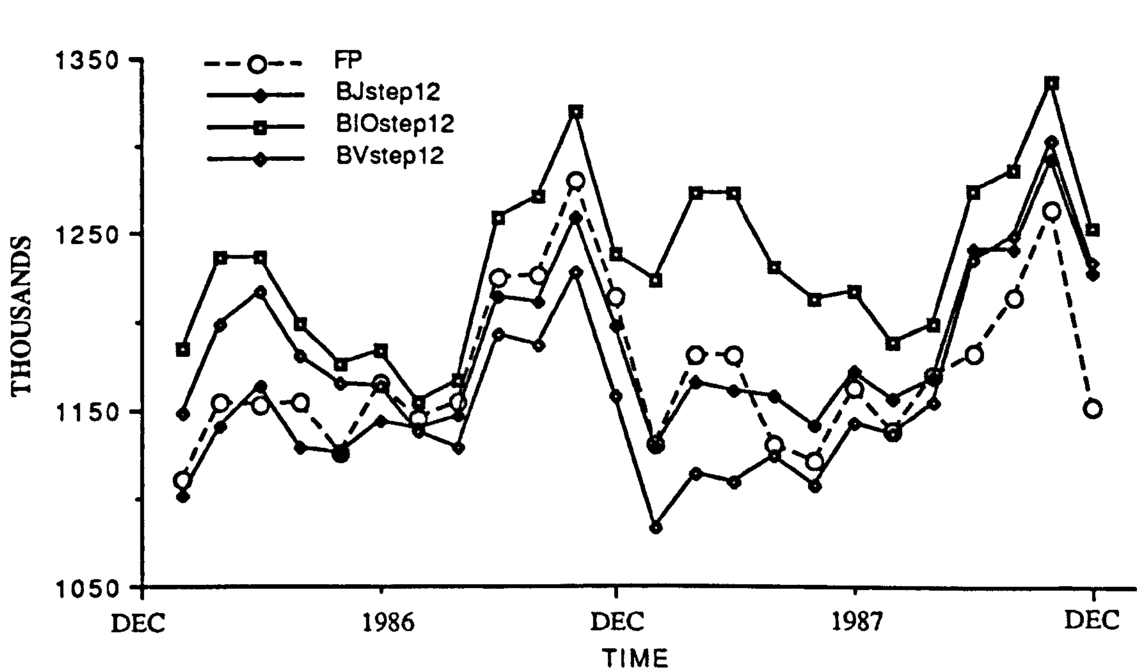
As one would expect ex-ante, the biological forecasts are the least good at forecasting one-step ahead, the error and turning point criterion confirming the picture presented in figure 8.8. The MSFE measures 1298.7, which is over twice as large as the equivalent statistic from the Box-Jenkins model forecasts. The MSFE converts to a RMSFE equivalent to 3.1% of fat pig slaughterings for that period and only 7 of the turning points are picked up by the one-step biological forecasts. In conclusion, the results of conditionally forecasting fat pig slaughter in the out-of-sample period specified are very much as one would expect, the two statistical models coming out on top and the more sophisticated bivariate model being somewhat superior to the univariate equivalent. The biological forecasts show a tendency to over-forecast which may well imply that the positive time trend variable is having too great an effect, and indicating that the rate of breeding herd productivity experienced during the estimation period may have slowed.

5.5b The One-Year Ahead Unconditional forecasts for Fat Pig Slaughter.

The one year, 12-step fat pig forecasts are obtained using the relevant models in a similar fashion to the 12-step culling forecasts , the resulting three sets of forecasts being presented in figure 8.9.

Figure 8.9

The One-Year Ahead Unconditional Forecasts For Fat Pigs From the Univariate, Biological and Bivariate Box-Jenkins Models



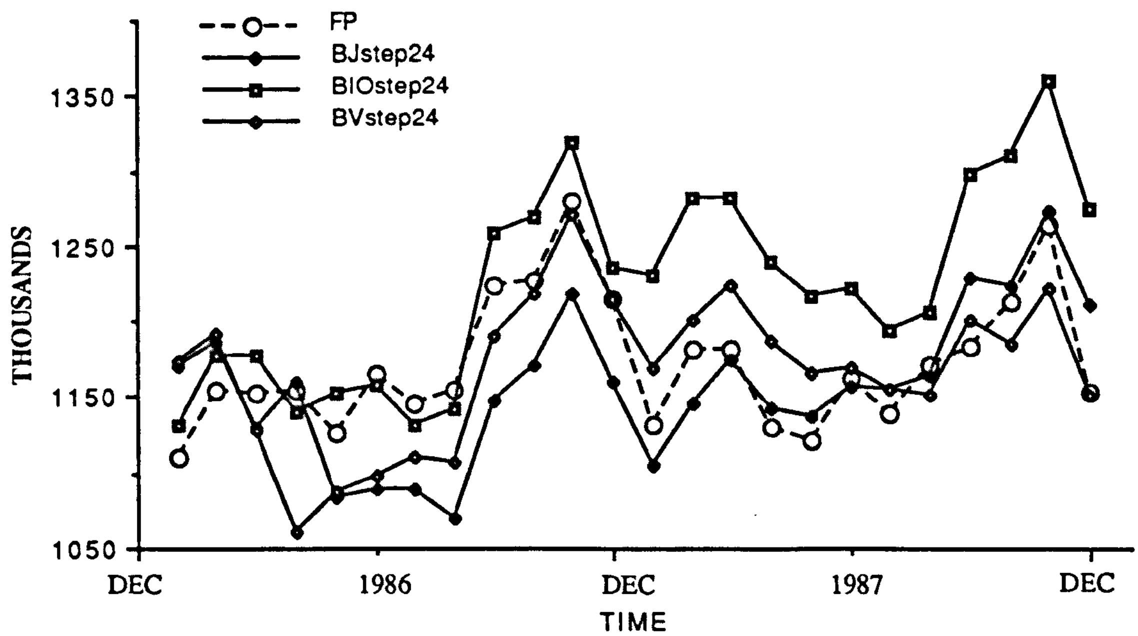
The figure above gives an even clearer indication of the fact that the biological model is over-forecasting the out-of-sample period, every one of the forecasts errors being negative, although the model still forecasts the seasonal movements very well. It would appear that the univariate model is now superior to the bivariate model in forecasting the level of cullings, a phenomenon born out by the statistics. The MSFE of the univariate model has increased by a relatively low 12.5% to 645.4 which implies an average absolute error of 2.2% and the forecasts correctly predict 11 of the 14 turning points, 21 of the 24 directional movements being forecast correctly. Whereas the abilities of the univariate model appear to have improved relative to its performance one-step ahead, the opposite is true of the bivariate Box-Jenkins forecasts. The MSFE of the bivariate forecasts more than triples to a value of 1820.4, and converts to a RMSFE equivalent to 3.6% of fat pig slaughter. The ability of the bivariate model to forecast month to month directional changes decreases from 10 in the one-step case to 8 in the 12-step case. The deterioration in the ability of the bivariate model to forecast the medium term must be largely the result of the forecasts of profit by the univariate profit ratio model. As figure 8.8 illustrates, the biological model forecasts are in fact the worst of the three, for forecasting one year ahead. The MSFE of 4548.7 is over 7 times larger than the univariate model's equivalent when forecasting 12-step ahead, and suggests an average absolute error of 5.7%. The latter statistic is more than double the univariate forecast equivalent and masks the fact that all the biological forecast errors are negative. The biological model forecasts still forecast the directional movements of the series well, getting 10 of the 14 turning points correct.

In conclusion, the univariate model is clearly the best at forecasting 12-step ahead, the deterioration of the bivariate forecasts being largely a result of the effect of the implicit univariate forecasts of profit. The forecasting results of the biological model confirm the idea gained from the conditional forecasts that the positive time trend is clearly causing the biological model to over-forecast the out-of-sample period, the effect being exaggerated by the tendency of the trimestic biological model to over-forecast the breeding herd also. The directional forecasting of the univariate and the biological models are very good.

5.5c The Two-Year Ahead Unconditional forecasts for Fat Pig Slaughter.

And so to the final set of forecasts to be analysed; the two year ahead, that is, the 24-step unconditional forecasts of fat pig slaughter, the results of which might be expected to re-iterate those obtained when unconditionally forecasting 12-steps ahead. The forecasts, which are presented in figure 8.10 illustrate that the results for the biological model are as expected, in that they appear to further over-forecast the recorded figures, though it should be noted that four of the earlier forecasts in 1986 are under rather than over-forecasts. The picture concerning the univariate and the bivariate model is less clear, however; the bivariate model appearing to forecast better than the univariate model in 1986 and the reverse occurring in 1987. Computing and comparing the relevant statistics helps to clear the picture somewhat.

Figure 8.10
The Two-Years Ahead Unconditional Forecasts For Fat Pigs From the Univariate, Biological and Bivariate Box-Jenkins Models.



The MSFE of 4760.9 for the biological forecasts is worse than was the case when forecasting 12-steps ahead, although the deterioration is not that great relative to the deterioration from one-step to 12-step. The biological model correctly forecasts 9 of the 14 seasonal changes of direction. Turning to the univariate model forecasts, the MSFE of 1975.3 is a considerable deterioration for the equivalent statistic of 645.4 when forecasting 12-steps ahead, although it is still less than half the 24-step MSFE statistic for the biological forecasts. The univariate forecasts correctly predict 11 of the 14 turning point. The most surprising results are those for the bivariate model, which while only getting half of the turning points correct it has a MSFE statistic of 1600.7, the lowest value for the three sets of 24-step forecasts. This then is a complete contrast to the longer term forecast results obtained for the bivariate Box-Jenkins longer term forecasts for monthly cullings and the 12-step forecasts for fat pigs, all of which implied a deterioration in performance relative to the univariate model. Furthermore, the 24-step MSFE, which converts to an average absolute error of 3.4% is an improvement on the bivariate model's forecasting performance 12-steps ahead. Although these results are in contrast to the longer term performance of previous bivariate forecasts, the longer lagged profit variable does include the monthly profit ratio lagged 11 to 18 months so that the forecasting performance of the bivariate model might be expected to improve relative to the univariate and biological forecasts, nevertheless, this does not explain the improved performance of the bivariate model 24-steps ahead relative to its performance 12-steps ahead. Summarising the result of the 24-step forecasts, in terms of forecasting the correct level of fat pig slaughter, the bivariate model is clearly the best, although figure 8.10 indicates that the univariate forecasts are the better of the two in 1987. In terms of forecasting turning points, the bivariate Box-Jenkins model is not as good as the other two, the univariate model performing very well, and despite the continued over-forecasting, the biological model forecasts seasonal movements quite well also.

Looking at the fat pig models' forecasting performance as a whole, the out-of-sample period chosen has thrown up some interesting results. The forecast error statistics, which are summarised in table 8.3 below, are good in terms of percentage error compared with the equivalent results of the monthly culling models, though the larger percentage error in the culling case may largely be attributed to rounding errors as mentioned in section 8.4c above.

Table 8.3[5]
The Error Statistics from the Monthly Fat Pig Forecasts

	MSFE's AND RMSFE PERCENTAGES		
STEP	UNIVARIATE	BIVARIATE	BIOLOGICAL
1-STEP	573.5 (2.0%)	522.5 (1.9%)	1298.7 (3.1%)
1-YEAR	645.4 (2.2%)	1820.4 (3.6%)	4548.7 (5.7%)
2-YEAR	1975.3 (3.8%)	1600.7 (3.4%)	4760.9 (5.9%)

In terms of the one-step forecasts, the results were very much as one would expect ex-ante. One of the main interests from the two-year analysis lies in the results of the unconditional forecasting performance of the biological and the bivariate models and what these might suggest. The biological model, clearly over-forecasts persistently and yet it performs relatively well when the ability to forecast directional change is observed. On the other hand, the bivariate model when forecasting 24-steps ahead performs better with respect to the level of fat pigs than the univariate and biological models and yet it is the worst at forecasting seasonal movements. Given the above arguments, and given the evidence from the performances of the trimestic

5. All figures in thousands of pigs

econometric and the bivariate monthly culling models, together with the fact that the 24-step MSFE is lower than the 12-step MSFE for the bivariate fat pig forecasts, all of which suggest that the longer term forecasting ability of the profit ratio model appears to be suspect, it might well be inferred that the relative forecasting performance of the three types of fat pig models 24-steps ahead might not be the norm. Having said this, time alone will tell whether or not the number of fat pigs produced by a given size of breeding herd will be subject to the positive time trend which is apparent over the estimation period or whether the apparent halt in increased sow productivity is a temporary phenomenon. If the former is true then the biological forecasting model, were it to be used as a working model, may well have to have the time trend re-estimated or even removed from it.

8.6 Conclusion

The analysis presented in this chapter has been concerned with assessing the relative forecasting performances of the models built in this thesis, from the point of view of their ability to forecast the correct levels of the variables concerned, although note of their ability to forecast seasonal movements has been taken as this is often of interest to forecasters.

The forecasting analysis of the models for the breeding herd census data was rather interesting in its outcome. The analysis of the Box-Jenkins univariate models for the breeding sow herd supported the use of the component models as opposed to the the model developed for the breeding sow herd itself, the relevant forecasts from the component models being used as a replacement for the suspect sample data of April 1987. The results of the one trimester step forecasts were as expected in that is was the Box-Jenkins model forecasts which were clearly the best. One year and two-years ahead, it was rather surprising that it was again the univariate forecasts which were better at forecasting the correct level of the breeding herd, the biological model showing a tendency to over-forecast and the two years ahead forecasts of the econometric model clearly being affected by the unconditional univariate profit forecasts. An attempt to overcome this handicap by replacing the profit forecast by the actual profit data surprisingly lead to the econometric forecasts being worse in terms of MSFE for the one trimester and three trimesters ahead forecasts. The advantage of using the actual profits showed through clearly in the comparison of the two years ahead forecasts in terms of level and seasonal movements: though not as good as the univariate model, the econometric model using actual profits improved on the performance of the biological model for the two sets of unconditional forecasts.

For the forecasts of the sow and boar culling series over the given out-of-sample period it appears to be the biological model which is the best forecasting model. The error statistics, which are larger in percentage terms than those of the trimestic breeding herd forecasts, due largely to rounding errors, favour the biological model for all three sets of forecasts undertaken. The bivariate model out-performs the univariate Box-Jenkins equivalent when forecasting one-month ahead, but the effect of having to unconditionally forecast profit using the specified univariate model clearly hampers the forecasting performance of the bivariate model over the two longer periods. All three models forecast seasonal movements in the culling data reasonably well.

In terms of the monthly fat pig model forecasts, things were very much as expected for the one month ahead forecasts: however, the longer term forecasts presented a few problems. The biological model shows a clear tendency to over-forecast, probably due to the effect of the positive time trend included in the biological model, although the trimestic breeding herd forecasts which feed into the biological model also tend to over-forecast. Having said this, the same breeding herd forecasts are used in the biological model for culling, the best model for this series. Suggestions for future work which might improve the biological fat pig model are discussed in the final chapter of the thesis. The performance of the bivariate Box-Jenkins model forecasts are curious in that they have a lower MSFE for the two years ahead forecasts than do the forecasts one year ahead. This, together with the fact that it is the worst model in terms of predicting seasonal movements, which are generally very good for the other two types of model, suggests that this result may be untypical. Further analysis as more data become available is required to provide an answer to this question.

The comments above all refer to the forecasting results for the out-of-sample period from 1986 to 1987 inclusive, and although the number of observations available is reasonable as far as the monthly models are concerned, a larger number of years would have been preferred for the trimestic breeding herd analysis. The analysis is also subject to the peculiarities of the out-of-sample period chosen, and again it is the trimestic analysis which suffered due to the apparently suspect sample data for April 1987. All the conclusions draw are done so in the light of these latter comments. In the final chapter, suggestions for further work will be accompanied by a discussion of how forecasts from different models might be combined in the hope of incorporating useful information contained by each method.

CHAPTER NINE

SUMMARY AND SUGGESTED FURTHER WORK

The aim of this research has been to build and compare the relative short term and longer term forecasting abilities of statistical time series models - univariate and bivariate - with more traditional modelling approaches in the context of the UK pig meat sector. The analysis has concentrated to a large extent on building forecasting models for the UK breeding herd, which is universally accepted as *the* key variable in the sector, in that it determines potential future supplies of pigmeat and also indicates the likely size of the future breeding herd.

In chapters three, four and five, semestic models for the breeding herd were built using univariate Box-Jenkins time series, biologically based and econometric methodologies respectively, the latter two both treating the breeding herd as an inflow-outflow system. In terms of knowledge of the sector and the number of variables required to build the models, the three approaches became progressively more sophisticated. The univariate models which were built purely on statistical grounds used the least amount of information, followed by the biological model, built on assumptions concerning the nature of the biological lags in the breeding herd, and finally the econometric model which introduced the economic phenomenon of profit and implicitly and explicitly incorporated biological information. Seasonality was a common feature of all of the approaches, modelled in the time series approach both by seasonal differencing and the appropriate seasonal autoregressive and moving-average components, and modelled in the biological and econometric approaches using seasonal dummies.

The period on which most of the model building analysis was conducted was the post 1973 era, a period in which the pig sector and the key variables modelled were clearly affected by internal and external influences. Externally, the UK joined the EEC which caused the shift in timing of the March and September farm censuses to April and August respectively, directly affecting the frequency of observations with which to work. Time series plots of the breeding herd and its components indicated structural changes in the form of a stabilisation of the mean and variance of the various series post EEC entry, later confirmed by Chow test analysis on the time series models. Another external influence on the sector, following shortly after the UK's EEC entry, was the sharp increase in feed prices partly exacerbated by an increase in world commodity prices in 1973, which appears to have affected some of the key variables in the industry well into 1974 and beyond. The latter influence meant that the start of the estimation period was pushed back to 1975 and even then it was thought advisable to apply intervention dummies in the biological and econometric models to some of the earliest observations in the estimation period. Internally, the estimation period is a period of great change in the industry in terms of its increasingly concentrated structure and changes in key technical coefficients. The major consequence of the latter has been to increase the productivity of the breeding sows, an effect which required the introduction of time trend variables into the relevant biological models. Other disruptive influences in the estimation period were the Aujezky eradication campaign of 1983 and the introduction of a temporary government subsidy on pig meat in the first half of 1977, both of which were modelled in the biological and econometric models by intervention dummies.

The frequency and quality of the data used in the model building analyses had a considerable influence on the types of analyses conducted and much time was devoted to discussing these data problems and how they might be resolved. Perhaps the most fundamental data influence was the 1974 sample census timing change. The change meant that the biological and econometric models had to be built using the four-monthly - trimestic - data from the April, August and December sample censuses. Because of software availability, the time series analysis had to be conducted on a pseudo-quarterly basis post 1973 including the recorded data from the June census, thereby breaking the rules of time series analysis by not having equally spaced intervals between observations. Analysis was done which implied that Box-Jenkins models identified and estimated using data going back to 1957 were better at forecasting an out-of-sample period than equivalent models built solely on the post 1973 data. This result was very interesting in that it implied that the statistical information used to build the models gained from the longer period of data was more useful for forecasting the out-of-sample period chosen, despite the apparent structural change on the series post 1973. It would appear likely that the influences of the eradication campaign and the 1977 subsidy had a detrimental effect on the forecasting performance of the forecasting models estimated on the shorter period. The relative forecasting performances of the longer and shorter period univariate models for future periods, when more data become available on which to estimate the latter period models, will make interesting future research. Another interesting result of out-of-sample forecasting analysis using the univariate Box-Jenkins models, was the apparent superiority of the breeding sow component models over that of the Box-Jenkins model built for the breeding sow total itself. This is possibly a result of the fact that the diametrically opposed technical and seasonal features of the pregnant sow and barren sow herds become masked when aggregated and begs the question as to whether or not future modelling of the pig sector, by whatever method, should not switch attention away from the breeding herd total towards its component parts.

The biological and econometric models for the breeding herd both required autocorrelation corrections of various kinds for the key inflow and outflow variables, pregnant gilts and culling and imply that decisions to increase or decrease these variables take longer than one trimestic period to implement. A possible reason for this, is the presence of adjustment costs as discussed briefly in chapter five. As well as producing a recursive forecasting model, the other biological relationships examined in chapter four illustrated the nature of the biological system in and between the breeding and feeding herds. Significant time trend variable parameters illustrated changing technical features such as the shortening of the weaning period and the resultant increase in sow productivity. The econometric models produced structures with lags and estimated parameters very much as expected. In view of the methodological imperfections of Savin's work in 1977 and the apparent misspecification of the MLC econometric models, I am satisfied that my models, which explicitly include a biological element, provide an improvement on existing forecasting models. Although a logit approach to modelling limited dependent variables was researched, the method was found to provide little or no improvement on the models eventually used.

Before out-of-sample forecast comparisons could be made, it was deemed advisable to replace the April 1987 sample census data, which appeared to give consistently high values for nearly all the breeding herd data, by a one-step forecast value from the univariate model. Given the three one-step forecasts for 1987 from the different types of model, the choice of the univariate forecast to replace the figure for April 87, did not appear to bias unduly the analysis in favour of the Box-Jenkins models. For the one step conditional forecasts and the one year and two years ahead unconditional forecasts, the univariate models were the best as measured by the Mean Square Forecast Error statistic. This was the result expected for the short term forecasts but was very much unexpected in the longer term. Whether this is a result of the fact that the univariate models have been estimated and forecast produced using information from the June full census is a matter of conjecture. The biological model showed a tendency to over-forecast one year and two years ahead and the long term forecasting ability of the econometric model was clearly hampered by its reliance on univariate unconditional forecasts of the profit ratio. Having said this, the replacement of the profit forecasts by actual profits, although a clear help when forecasting two years ahead, was an apparent handicap when forecasting one year ahead. When using actual rather than profit forecasts, the econometric model outperformed the biological model when forecasting one and two years ahead.

The other series for which forecasting models were built comprised the two monthly slaughter series, sow and boar cullings and fat pig slaughter. Models were built for these two categories using the univariate Box-Jenkins and biological methodologies and, as an alternative to, an econometric model, bivariate Box-Jenkins models were also constructed. Univariate time series models were also built in chapter six for the AAPP and the compound feed indices, and the profit ratio created by the ratio of the two said price indices. The one-step forecasting analysis implied the superiority of using the ratio model itself for forecasting profits rather than forecasting the individual price indices and making the appropriate transformation of the resulting forecasts. This result contrasts somewhat with the aforementioned aggregate versus dissaggregate forecasting analysis of the univariate models for the pseudo-quarterly breeding sow herd series. This univariate profit ratio model was the only model used in the thesis to forecast profits and was used to provide forecasts for the semestic econometric model and the bivariate monthly models. The five univariate monthly models built for the two slaughter categories and the price and profit variables were considerably easier to identify than the univariate semestic models had been, due to the nature of the seasonality involved. All five models were first and seasonally differenced in order to obtain stationary series, and none of the models were cyclical, in contrast to the pseudo-quarterly breeding herd models.

The biological models built in chapter four along with the semestic biological models are basically the same in structure as the equivalent semestic models, but applied to a monthly data set. Both models were estimated on the assumption that the breeding herd at any particular census also represented the size of the breeding herd in the previous three months. Future work on a monthly biological model could investigate the validity of this assumption by testing the forecasting ability of models based on different assumptions, for example, interpolating the size of the breeding herd so that the breeding herd is deemed to change smoothly over time between the census dates.

Rather sophisticated, though still simplified bivariate Box-Jenkins methodology was employed, using the previously estimated univariate models for culling, fat pigs and the profit ratio, to build forecasting models for the two slaughter series, using profit as the explanatory variable. These models made an interesting alternative approach modelling the monthly culling series on the grounds that they help to identify, using empirical statistical techniques, the nature of the lags involved between culling, slaughtering and profit. Having gained such knowledge from the bivariate analysis, future work on econometric modelling of these monthly series could take on board the lag structures identified in the bivariate analysis. Indeed it would have been desirable to apply the bivariate time series analysis to the trimestic breeding herd models, but the lack of sufficient data rendered identification infeasible.

Each of the three different types of model was used to forecast one month, one year, and two years ahead, again using the data for 1986-7 as the out-of-sample period, to investigate their respective short and longer term forecasting abilities. For all sets of forecasts of culling it was the biological model which was the best in terms of forecast error statistics and the ability to forecast seasonality. Although the superiority of the biological model over the univariate model was expected in the longer term, the one month conditional superiority was rather unexpected, though it should be said that the lags in the biological model were short, possibly helping to explain the relatively good short term forecasting results. The bivariate forecasts were similar, though slightly superior to those of the univariate model, suggesting that the addition of the profit variable was a useful one in term of explanatory and forecasting ability. For the longer term forecasts, however, the performance of the bivariate model is clearly affected for the worse by the univariate profit forecasts, and especially so when forecasting two years ahead. Although the bivariate model forecasts the month to month seasonal changes reasonably well, the usefulness of the bivariate model as a longer term forecasting model is clearly dependent on a better profit forecasting model being found.

In terms of the fat pig forecasting analysis, the results were very much as expected. In the short term, the bivariate Box-Jenkins forecasts were the best, closely followed by similar univariate forecasts: the biological model, with its longer lags, trailed in third place. The apparent reason for the poor MSFE statistic of the biological forecasts was the fact that the model appeared to be over-forecasting. This over-forecasting is even more evident in the longer term unconditional biological forecasts, and can be explained in part by the fact that the forecasts from the trimestic breeding herd biological model, which feed into the unconditional monthly biological forecasting function are, on the whole, over-forecasts themselves. However, the latter cannot be cited as the cause of the one-step ahead conditional over-forecasting of the monthly biological model which must, therefore, be explained by the presence of the positive time trend in the biological model which takes account of the increase in sow productivity over the estimation period. In an attempt to rectify this problem, the one-step conditional forecasts from the biological model were re-made stopping the time trend at the end of the estimation period, December 1985. Although the MSFE of the adjusted biological forecasts was reduced considerably to 865.8

compared with the equivalent unadjusted forecast figure of 1298.7, it was evident that the tendency to over-forecast still persisted. The conclusion from this simple analysis was that the problem caused by the size of the coefficient on the positive time trend originated further back into the estimation period, even though this was not apparent at the end of the plot of residuals produced having estimated the model. A clue to the likely period of change in the productivity time trend in the biological model can be found in chapter one where it was evident that the reduction in the improvement in the weaning period has slowed down the rate of increase in sow productivity. The consequence of the forecasting analysis for the fat pig biological model is that it would appear to be wise for future work on such a model to allow for a break in the increase in sow productivity around the start of the 1980's by allowing the time trend parameter to be lowered, possibly by the inclusion of a dummy variable to represent the weaning period. In the meantime the biological model as presented in this thesis could be used, adjusting the biological forecasts downward by an appropriate percentage calculated by the mean percentage error of the appropriate forecasts in the out-of-sample period. Another curious result from the fat pig forecasting analysis was the superiority of the bivariate Box-Jenkins model in terms of the MSFE statistic when forecasting two years ahead. The two year ahead forecasts were better than those of the univariate model, which could be explained by the relatively long lags on the profit variables, although the latter argument falls somewhat when one takes into consideration the fact that the two-year ahead forecasts for the profit ratio adversely affected the two year ahead forecasting performance of the trimestic econometric and the bivariate monthly culling model. Also, the MSFE for the two year ahead bivariate fat pig forecasts is actually lower than that of the equivalent one year ahead forecasts. These results imply that the relative superiority of the bivariate fat pig model when forecasting two years ahead may be atypical, though this hypothesis can only be tested as more data become available in the future. Indeed all the results of the forecasting analysis of chapter eight must be seen in the context of the out-of-sample period chosen, and given that the census information from the April 1987 period has been treated with considerable suspicion, some of the conclusions drawn from the analysis are made tentatively. Further analysis as more data become available will be informative and should help to make the conclusions more robust.

As well as the comments already made concerning the direction of useful future research, the work carried out in the course of researching this thesis points to further areas of development. Because of the data problems encountered the trimestic models have not been compared on equal terms, in the sense that the univariate models were built and forecast using pseudo-quarterly rather than trimestic data. Despite this, the univariate model has been seen to perform relatively well, although it has had the advantage of including information from the full June census, which should be more reliable than the sample census data in that the data are not subject to sampling errors. A true comparison of the univariate and biological and econometric models requires a semestic univariate model to be built: however, given that the Box-Jenkins methodology requires a long time series before any meaningful analysis can be undertaken, a few more years of good quality data are still required.

Both the biological and the econometric semestic breeding herd models were introduced using steady state equilibrium as a theoretical framework in which to explain the models. Given that the UK breeding herd is unlikely to err from such an equilibrium relationship in the long run, there is good reason for estimating the biological and econometric model parameters within such an equilibrium framework. Yet again, however, it was the quantity and quality of semestic data available which deterred such an exercise but provide yet more fodder for future analysis.

The research has been concerned with the relative forecasting abilities of the various models built: were a forecaster interested in obtaining a 'best' forecast for the variable with which he is concerned, work by Bates and Granger (1969) on the combining of forecasts, suggests that the forecaster should not necessarily concern himself with finding and using a single 'best' model, thereby ignoring information provided by other models produced. Instead, by combining the forecasts from some or all the models using appropriate weights, which may or may not change over time, it is possible to produce forecasts with an error variance lower than that of the best individual set of forecasts. Although it was one of the original aims of the thesis to analyse such combining methods in depth, the relative shortness of the out-of-sample forecasting period, chosen largely as a result of the constraints placed on the analysis by the data available for estimation, and the suspect quality of the data towards the end of the sample period, namely the April 1987 census data and the presence of the effects of the 1983 Aujezky eradication campaign, rendered the usefulness of such analysis questionable. Even so, the methodology and an application to the one-step forecasts for fat pigs is presented briefly in appendix 9 in order to give the reader an appreciation of what could be done in the future as more, hopefully problem free, data become available.

Following on from the combination idea, and in an attempt to find 'the best' forecasts for a given variable, it would be possible to build a recursive model so that the forecasts deemed to be the best for a given variable, whether produced by a single model or a combined weighting system, could be fed into a given model in which the given variable is an independent variable, in order to produce forecasts of a second variable. For example, consider forecasting the breeding herd using a recursive model such as the biological or the econometric model having derived a relevant weighting system. The one-step forecast for the breeding herd requires a one-step forecast of culling. The relevant forecasts of culling can be produced by the biological and econometric semestic models and indeed by the various monthly models for culling, the latter being aggregated up to the relevant scale and the weightings then applied in order to derive a best one semester step cull forecast. This forecast could then be fed into the semestic biological and the econometric models, and assuming the one semester step forecast for gilts is given, the models could produce their respective forecasts of the breeding herd for the next semester, and so on. The latter process would, of course, require complex software to be developed.

In conclusion then, this empirical exercise has succeeded to a certain extent in building and comparing the short and medium-long term forecasting abilities of models built using methodologies requiring various levels of prior information and sophistication. A major flaw for the models requiring forecasts of profit was the lack of a good model for long term profit forecasting, casting serious doubt on the wisdom of using models including profit as an explanatory variable to forecast in the longer term unless, of course, a superior long term profit forecast model can be developed. The conclusions of the analyses were varied in terms of actual and expected results, and some analyses were inconclusive. Problems were caused by the fact that the models were estimated in a period when the sector was volatile and undergoing much technical and structural change. Together with the problem of suspect sample data in the out-of-sample forecasting period, the conclusions from the analysis can only be tentative and repeated analysis in a few years time would make interesting reading.

APPENDIX 2

EXAMPLES OF MODEL BUILDING USING BOX-JENKINS METHODOLOGY

This appendix aims to illustrate how three non-seasonal time series models were identified and estimated using Box-Jenkins methodology. A colleague randomly generated the data for all three examples on a computer from time series models which were, of course, known to him. I then attempted to identify the correct models using the procedures outlined in Chapter two. All models are eventually identified correctly.

	EXAMPLE ONE		EXAMPLE TWO		EXAMPLE THREE	
K	r_k	$\hat{a}_{kk}$	r_k	$\hat{a}_{kk}$	r_k	$\hat{a}_{kk}$
0	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
1	0.4969	0.4969	-0.2211	-0.2211	0.6731	0.6731
2	0.0937	-0.2034	0.1391	0.0948	0.3680	-0.1554
3	-0.0425	0.0034	0.0031	0.0551	0.2732	0.1717
4	-0.1669	-0.1716	0.0250	0.0246	0.1984	-0.0591
5	-0.2296	-0.0921	0.1355	0.1463	0.0816	-0.0724
6	-0.1857	-0.0432	-0.0547	-0.0043	-0.0635	-0.1503
7	-0.0053	0.1227	-0.2072	-0.2763	-0.1774	-0.1056
8	0.0749	-0.0224	0.1112	0.0141	-0.2348	-0.0765
9	-0.0214	-0.1354	-0.0983	-0.0188	-0.2362	-0.0090
10	-0.1258	-0.1308	0.0411	-0.0053	-0.1984	0.0275
11	-0.1487	-0.0593	-0.0706	-0.0176	-0.1542	0.0204
12	-0.2616	-0.2276	-0.0615	0.0203	-0.1767	-0.1181
13	-0.2848	-0.0908	-0.0415	-0.1055	-0.1285	0.0805
14	-0.2188	-0.1602	0.0216	-0.0268	-0.0251	0.0232
15	-0.1381	-0.1467	0.0529	0.1210	-0.0426	-0.1544
16	-0.0263	-0.0860	0.0366	0.0754	-0.1190	-0.1102
17	0.1135	0.0249	0.0800	0.1270	-0.1624	-0.1033
18	0.1561	-0.0916	0.0246	0.0467	-0.1353	0.0005
19	0.1419	-0.0324	0.1300	0.0862	-0.0723	0.0462
20	0.1197	-0.0272	0.0471	0.0215	-0.0424	0.0173

Having generated the data the identification package was employed to generate the sample auto and partial autocorrelation statistics, up to and including the twentieth lag, as shown above. With a sample size of 80 observations, two standard deviations from zero, as measured by the Quenouille statistic is 0.2236. Observing the sample autocorrelations for the first ten lags reveals that all three series appear to be stationary so that they do not need to be differenced. An initial identification can now be made of the three series, starting with example one.

EXAMPLE ONE

The sample autocorrelations for example one reveal that all the autocorrelations are small with the exception of that at lag one which has a value of 0.4969, after which there is an immediate and sustained reduction in the value of the sample autocorrelations of higher order. Given this information and given that the partial autocorrelations tail-off more slowly, it would appear that an MA(1) is the best choice for an initial identification. An initial estimate of b_1 is obtained by solving $r_1 = b_1 / (1+b_1^2)$, which is then input to a parameter estimation program producing the following results.

		b_1
PARAMETER ESTIMATE	=	0.6067
STANDARD ERROR	=	0.0937
MEAN SQUARE ERROR	=	0.8253
MEAN OF RESIDUALS	=	-0.1928
VARIANCE OF RESIDUALS	=	0.7877

AUTOCORRELATIONS OF RESIDUALS

τ	$r_\tau(e)$
0	1.0000
1	0.0287
2	0.0968
3	-0.0549
4	-0.1176
5	-0.1055
6	-0.1464
7	0.0340
8	0.0689
9	0.0180
10	-0.1444

The standard error of the parameter estimate indicates that the parameter b_1 is highly significant with a t-statistic of 6.475. The mean value of residuals is not significantly different from zero given the low residual variance of 0.7877. Finally, analysis of the autocorrelations of the residuals indicates that the model identification is acceptable as there is no evidence of serial correlation, none of the residual autocorrelations being significantly different from zero. Hence, the estimated model is;

$$x_t = (1 + 0.6067B)\, e_t.$$

EXAMPLE TWO

At first sight this example appears to be white-noise, because none of the sample statistics lies further than two standard deviations away from zero. However, because r_1

and $\hat{a}_{11}$ are very close to being significant the model was overfitted by an AR(1) and by an MA(1) to see what results would occur. Although the results are not shown here, the conclusion from the diagnostic checking stage was to reject both overfitted models because neither of the additional parameters was significantly different from zero. In order to give an actual example of the problem of overfitting on both sides of the equality an ARMA(1,1) was estimated and diagnostically checked, producing the following results.

		a_1	b_1
PARAMETER ESTIMATES	=	-0.8379	0.8091
STANDARD ERRORS	=	0.0894	0.1219
MEAN SQUARE ERROR	=	0.7935	
MEAN OF RESIDUALS	=	-0.0290	
VARIANCE OF RESIDUALS	=	0.7926	

AUTOCORRELATIONS OF RESIDUALS

τ	$r_\tau(e)$
0	1.0000
1	-0.0928
2	0.0538
3	0.1117
4	- 0.0718
5	0.2046
6	-0.1099
7	-0.2058
8	0.0900
9	-0.0965
10	0.0259

The estimation program shows results which might well suggest that the model is indeed an ARMA(1,1) model with an AR parameter of -0.8379 and an MA parameter of 0.8091. Both these parameters are significantly different from zero, the mean value of the errors is low and the autocorrelations of the residuals appear to be white noise. This example illustrates the point that two-sided overfitting may well lead to the inducement of unnecessary parameters, which is not good for reasons of parsimony. Having said this, writing the estimated ARMA(1,1) with the parameters expressed to one decimal place gives the following equation:-

$$(1 + 0.8B)\, x_t = (1 + 0.8B)\, e_t.$$

Dividing through by (1 + 0.8B) results in the true white noise generating series indicating that the two-sided overfitting of the model is not always as harmful as it might initially appear.

EXAMPLE THREE.

At first sight an AR(1) appears to be the best identification for example three, due mainly to the fact that the sample autocorrelations fall at a steady rate rather than experiencing a sudden drop. The partials on the other hand, all lie below two standard deviations from zero except for that at lag one ($\hat{a}_{11}$), although it is also true that the partials at lags two and three are not too low. Fitting an AR(1) with a starting parameter of 0.6731 gives the following results.

		a_1
PARAMETER ESTIMATE	=	0.6878
STANDARD ERRORS	=	0.0739
MEAN SQUARE ERROR	=	0.9135
MEAN OF RESIDUALS	=	-0.0429
VARIANCE OF RESIDUALS	=	0.9117

AUTOCORRELATIONS OF RESIDUALS

τ	$r_\tau(e)$
0	1.0000
1	0.0999
2	-0.2429
3	0.0235
4	0.0946
5	0.0091
6	-0.0267
7	-0.2424
8	-0.1380
9	0.0299
10	-0.0301

The results of diagnostic checking appear to be favourable in that the AR(1) parameter is significantly different from zero and the error variance and residual mean are low. The only indication of any problem is that the residual autocorrelations at lags two and seven are significantly different from zero. The high autocorrelation at lag two may well suggest that an MA parameter should be added to the model. The results of fitting an ARMA(1,1) are given below.

		a_1	b_1
PARAMETER ESTIMATES	=	0.5583	0.3341
STANDARD ERRORS	=	0.1096	0.1351
MEAN SQUARE ERROR	=	0.8903	
MEAN OF RESIDUALS	=	-0.0569	
VARIANCE OF RESIDUALS	=	0.8870	

AUTOCORRELATIONS OF RESIDUALS

τ	$r_\tau(e)$
0	1.0000
1	-0.0496
2	-0.1279
3	0.0821
4	0.0867
5	-0.0345
6	0.0350
7	-0.2285
8	-0.1176
9	0.0229
10	-0.0523

Analysis of the results shows both the MA and the AR parameters to be significant and the error variance and the variance of the residuals to be lower than what they were under the AR(1) identification. The mean of the residuals has increased slightly although the increase and the actual value are so small that there is no need for concern. In order to check whether or not the ARMA(1,1) specification can be improved at all, the model was overfitted with both an extra MA and then an extra AR parameter. Although the results are not shown, both the overfitted models were rejected because the additional parameters did not prove to be significantly different from zero, and the mean square errors and variance of residuals increased in value. Therefore, the correct model appears to be an ARMA(1,1) with an MA parameter of 0.3341 and an AR parameter of 0.5583, that is;

$$(1 - 0.5583B)\, x_t = (1 + 0.3341B)\, e_t$$

APPENDIX 3a.

The Breeding Herd Data Used in the Building of the Box-Jenkins SARIMA Models.

U.K. Pig Breeding Sow Herd Census Data (1957:1-1987:4)*

TIME	PS	PG	BS	PP	H	B	UG
1957:1	362	142	208	504	712		
1957:2	383	153	208	536	744		
1957:3	361	151	246	512	758		
1957:4	424	152	237	576	813		
1958:1	408	159	256	567	823		
1958:2	414	132	230	546	776		
1958:3	420	104	264	524	788		
1958:4	420	98	230	518	748		
1959:1	382	105	227	487	714		
1959:2	388	112	205	500	705		
1959:3	360	100	234	460	694		
1959:4	369	113	206	482	688		
1960:1	344	126	223	470	693	40	
1960:2	375	142	208	517	725	40	
1960:3	363	133	244	496	740	42	
1960:4	385	126	228	511	739	41	
1961:1	376	133	242	509	751	44	
1961:2	401	148	224	549	773	43	
1961:3	396	140	271	536	807	45	
1961:4	430	137	251	567	818	45	
1962:1	423	145	272	568	840	45	
1962:2	456	147	255	603	858	46	
1962:3	454	134	297	588	885	48	
1962:4	478	125	274	603	877	47	
1963:1	445	125	297	570	867	48	
1963:2	481	142	253	623	876	47	
1963:3	453	130	301	583	884	48	
1963:4	480	122	265	602	867	47	
1964:1	441	150	281	591	872	49	
1964:2	466	174	263	640	903	47	
1964:3	468	168	306	636	942	50	
1964:4	512	152	280	664	944	49	
1965:1	493	156	310	649	959	51	
1965:2	517	147	281	664	945	49	
1965:3	510	119	328	629	957	50	
1965:4	513	100	287	613	900	49	
1966:1	460	100	288	560	848	44	
1966:2	463	110	249	573	822	41	
1966:3	436	109	266	545	811	44	
1966:4	450	111	247	561	808	44	
1967:1	429	125	256	554	810	44	
1967:2	445	136	243	581	824	44	

Cont.

TIME	PS	PG	BS	PP	H	B	UG
1967:3	437	135	268	572	840	45	
1967:4	466	122	263	588	851	44	
1968:1	454	130	277	584	861	45	
1968:2	482	151	254	633	887	44	
1968:3	477	142	295	619	914	45	
1968:4	517	129	271	646	917	44	
1969:1	489	133	293	622	915	44	
1969:2	507	141	267	648	915	45	
1969:3	501	132	302	633	935	47	
1969:4	518	130	274	648	922	46	
1970:1	490	138	282	628	910	47	
1970:2	524	159	270	683	953	46	
1970:3	521	155	313	676	989	47	
1970:4	548	142	296	690	986	44	
1971:1	554	128	314	682	996	45	
1971:2	570	121	292	691	983	45	
1971:3	556	111	305	667	972	48	
1971:4	566	106	285	672	957	46	
1972:1	548	118	298	666	964	48	
1972:2	557	128	276	685	961	48	
1972:3	545	138	304	683	987	51	
1972:4	571	138	277	709	986	46	
1973:1	552	148	300	700	1000	43	
1973:2	577	156	282	733	1015	44	
1973:3	565	157	310	722	1032	42	
1973:4	579	136	287	715	1002	40	
1974:1	503	109	287	612	899	37	96
1974:2	521	107	262	628	890	40	80
1974:3	504	92	269	596	865	42	81
1974:4	498	84	234	582	816	41	73
1975:1	453	97	259	550	809	41	75
1975:2	485	104	225	589	814	43	87
1975:3	475	113	228	588	816	43	95
1975:4	496	122	226	618	844	43	102
1976:1	458	133	250	591	841	42	112
1976:2	512	137	235	649	884	41	101
1976:3	511	138	236	649	885	40	106
1976:4	537	111	238	648	886	41	90
1977:1	483	105	239	588	827	42	89
1977:2	503	103	222	606	828	41	76
1977:3	479	89	222	568	790	44	81
1977:4	502	102	218	604	822	42	91
1978:1	486	110	235	596	831	44	105
1978:2	510	118	214	628	842	43	90
1978:3	497	120	225	617	842	42	100
1978:4	534	109	222	643	865	42	90
1979:1	(498)	(111)	(241)	(609)	(850)	(43)	(88)
1979:2	528	109	215	637	852	42	82
1979:3	517	102	218	619	837	43	77
1979:4	520	92	208	612	820	42	91
1980:1	497	99	213	596	809	43	88

Cont.

TIME	PS	PG	BS	PP	H	B	UG
1980:2	517	109	204	626	830	43	84
1980:3	510	110	204	620	824	43	94
1980:4	514	101	203	615	818	43	89
1981:1	517	102	202	619	821	44	91
1981:2	522	112	203	634	837	45	87
1981:3	520	107	205	627	832	45	87
1981:4	532	108	197	640	837	46	90
1982:1	533	109	204	642	846	47	97
1982:2	543	122	200	665	865	45	89
1982:3	536	118	203	654	857	45	96
1982:4	558	114	204	672	876	43	92
1983:1	547	119	210	666	876	42	89
1983:2	542	110	204	652	856	42	82
1983:3	532	103	197	635	832	43	82
1983:4	510	96	184	606	790	43	75
1984:1	487	99	184	586	770	44	80
1984:2	518	105	178	623	801	44	77
1984:3	504	105	186	609	795	44	89
1984:4	526	107	181	633	814	45	84
1985:1	523	111	192	634	826	46	87
1985:2	530	112	187	642	829	44	80
1985:3	521	105	195	626	821	45	86
1985:4	537	102	187	639	826	45	85
1986:1	531	108	199	639	838	46	92
1986:2	534	108	182	642	824	44	79
1986:3	540	105	188	645	833	45	87
1986:4	533	106	183	639	822	45	86
1987:1	544	110	197	654	851	47	87
1987:2	528	105	180	633	813	44	80
1987:3	522	102	186	624	810	45	85
1987:4	536	104	182	640	822	45	79

* All Figures in thousands of pigs.

Numbers in parentheses indicate forecast value.

Source: M.A.F.F., obtained from M.L.C. Economics Department.

APPENDIX 3b
Methodology of Data Collection

This appendix outlines the methodology of data collection employed by each member country of the U.K. as defined by Ministerial boundaries. Prior to 1974 the data were collected on a quarterly basis, the censuses being carried out at the beginning of each of the months of March, June, September and December. Since 1974 however, the one third sample censuses previously performed in March and September have been carried out at the beginning of April and August respectively, in order to fall in line with the rest of the E.E.C. M.A.F.F. continued to perform the March and September censuses in addition to the April and August ones, up to and including 1977. Because of industrial action carried out by the Civil Service, no census was performed in April 1979. The details of the methodologies of data collection for each of the three ministerial bodies are presented on the following page.

Sources:-

Agricultural Statistics U.K.. (various years)
Agricultural Statistics England and Wales. (various years)
Statistical Review of Northern Ireland Agriculture. (various years)
Economic Report on Scottish Agriculture. (various years)

Notes.

S.M.D.= Standard Man Day. (Eight hours work performed buy one unit of labour)
E.S.U. = European Size Unit.
(The equivalent of 1000 E.U.A. of standard gross margins at average 1972-74 values)

TIME	ENGLAND AND WALES	NORTHERN IRELAND	SCOTLAND
Pre-1969	The June census records all livestock on holdings larger than one acre in size unless the holding is deemed to be statistically insignificant as far as economic activity is concerned. Estimates of livestock numbers on these minor holdings are made and added to the census records. Estimated numbers for March, September and December are raised from one third sample surveys and are subject to sampling errors.	Livestock returns collected from all holdings whatever size and complete censuses are performed in both June and December.	As for England and Wales, except that complete censuses are performed in both June and December, and no estimate of livestock numbers on minor holdings are added to the the census returns.
1969	The definition of small holdings changed to those holdings with:- a. < 10 acres of crops/grass. b. a labour requirement of <26 S.M.D.'s and c. No full time worker. The changes increased the number of minor holdings by 34,000. 13,000 holdings no longer operated as farms.		
1970			The definition of a minor holding was given as a holding requiring less than 26 SMD's per annum. The result was a loss of 16,000 holdings from the main censuses.
1973	Minor holding SMD labour requirements increased to <40 SMD's per annum.	Minor holdings now defined as in England and Wales. Resulted in a net loss of 5,500 holdings from census.	Minor holding definition is changed to include all holdings with a labour requirement of <40 SMD's. Negligible effect on figures.
1980	New higher threshold for definition of minor holdings. a. Total area of < 6H b. No full time worker. c. <100 SMD's d. <100 sqm. glasshouse area e. Occupier farm no other holding.	As for England and Wales. Loss of a further 6,000 holdings to minor sector.	
1981		December census dropped in favour of sample survey. Minor holdings now include:- a. Total area of < 6H. b. No full time worker. c. Farm business size of < 1 ESU	
1983			Sample surveys replace December census

Appendix 3c

Chow Tests On The SARIMA-Models For The Breeding Sow Herd Series Estimated On 1957:1- 1985:4

Having estimated the five breeding sow herd SARIMA models for the 1957:1-1985:4 period, the decision was taken to investigate the significance of the influence, if any, of the U.K.'s membership of the EEC which resulted, amongst other things, in a change in the timings of the Spring and Autumn sample censuses. A look at the 1957-85 plots for the breeding sow herd series' reveals that all the series show changes in behaviour post 1973/4. Sows in-pig numbers fall dramatically during 1973/4 due to falls in profitability and adjacent reductions in in-pig gilt numbers. The series beyond 1974 appears to be much more stable than it had been before 1974. The number of in-pig gilts is also a series which oscillates less about a lower mean level post 1974, whereas the barren sow series experiences a dramatic and sustained fall. The effect of all these changes upon the aggregated series, "total breeding sow herd', is that from 1974 onwards, the series look much more stable and stationary than they had been previously.

The models, as estimated on the sample 1957:1-1985:4 were re-estimated on the data from 1957:1-1973:4 and from 1974:1-1985:4, 1974 being the first year of the new census timings. The appropriate statistics were calculated using the RSS values from each of the three estimations and the Chow statistics were then measured against critical values from the appropriate F-distribution. The results of the tests are presented in the table below, showing the Chow test to be highly significant for all four models.

<u>Results of Chow Tests Performed on the Breeding Sow Herd SARIMA-Models (1957:1-1985:4)</u>

Series	Chow Value	$F^{.995}_{k,w}$ *
PS	5.20	3.08
PG	6.31	3.54
BS	4.59	3.54
PP	6.63	3.23
H	4.59	3.23

* k = No. of estimated parameters.
W = Degrees of freedom for the aggregate model derived from the addition of the two component models.

APPENDIX 3d

An analysis of the In-Sample and Out-Of-Sample MSFE Statistics From The Univariate Box-Jenkins Models For The Breeding Herd and its Components

	In-Sample (1983:1-85:4) Forecasts		Out-of-Sample (1986:1-87:4) Forecasts			
	1 Step Ahead		8 Step Ahead		1 Step Ahead	
Sample	57:1-85:4	75:1-85:4	57:1-85:4	75:1-85:4	57:1-85:4	75:1-85:4
PS	248.75	187.34	150.38	312.66	185.90	203.51
PG	25.22	28.02	17.49	21.15	14.96	20.67
PP	287.42	231.91	188.74	236.27	358.69	245.45
BS	26.55	23.22	21.77	51.26	35.83	37.47
H	323.73	252.26	338.66	281.93	741.19	592.55

The MSFE's are generally smaller in the in-sample period for the 1975-85 model, although this was not so in the case of the pregnant gilt model. Furthermore, for the four 1975-85 models in which the in-sample MSFE was the smaller, the removal of the worst forecast from both the shorter and longer period models produced MSFE statistics for the 57-85 model forecasts which were equally as good, if not better, than those of the shorter period model. These latter results indicate that the longer period model is therefore equally as good, if not better, at forecasting the in-sample period if it were not for one relatively bad forecast. Analysing the forecast errors more closely it is apparent that this bad forecast occurs in the third quarter of 1983, the quarter in which the consequences of the Aujezky disease eradication campaign are felt most heavily in the breeding herd. All five models estimated on the longer sample over-forecast the said quarter - though this is only just the case in the pregnant gilt model - and by an amount which adversely affects the MSFE significantly more than is the case for the shorter sample models. It would appear, therefore, that despite the fact that the Aujezky period has a greater influence on the identification and the parameter estimates of the shorter sample model, apart from the aforementioned quarter, the longer sample models are equally as good or better than the shorter sample models at forecasting the period 1983:1-1985:4. Almost certainly, the reason for the lesser effect of the eradication campaign on the pregnant gilt herd, is that they can be replaced by, for example, transfers from the feeding herd with very little time lag. This is not true of pregnant sows and barren sows and consequently these two herds were reduced by the eradication campaign and could not recover anywhere nearly as fast as was the case for the gilt herd. This latter phenomenon plays some part in explaining the superiority of the longer period model over the shorter period model in forecasting the pregnant gilt herd over the in-sample period.

Turning to the forecast results of the out-of-sample period, the 8 step ahead unconditional forecast results show heavily in favour of the longer period model. All the MSFE's, except for that of the total breeding herd model, are lower for the 1957-85 model. Furthermore, the higher MSFE of the breeding herd model is the result of one bad under forecast of the figure for the first quarter of 1987, a figure which both the shorter and the longer sample models under forecast. Removing the forecast for this period for both models produces an MSFE statistic for the longer period model which is lower than that of the 1975-85 model. The majority of in-sample MSFE statistics, therefore, favour the longer period models and even in the case where the opposite is true, the reason would appear to be the result of one large under-forecast.

The story for the one step ahead conditional out-of-sample forecasts is similar to that of the eight step unconditional forecasts except that it is only the three component models estimated on the longer period which out-forecast the shorter period models. The MSFE statistics for the two aggregate series models are lower for the 75-85 model. However, once again the reason for the better statistic of the shorter period model is the appearance of one - or in the case of the total breeding herd model two - bad forecast at the start of 1987. Removing these forecasts produces a superior MSFE for the longer period model.

APPENDIX 4a

The Data Used to Estimate the Trimestic Biological Models*

TIME**	PS	PG	PP	BS	H	UG	B	HB	I	M	FP
1974:1	503	109	612	287	899	96	43	942		178.43	5034.8
1974:2	504	92	596	269	865	81	42	907	123.26	156.96	4982.4
1974:3	498	84	582	234	816	73	40	856	88.70	140.71	4868.6
1975:1	453	97	550	259	809	75	37	846	119.29	129.49	4463.9
1975:2	475	113	588	228	816	95	42	858	114.59	103.04	4003.0
1975:3	496	122	618	226	844	102	41	885	130.51	102.98	4075.4
1976:1	458	133	591	250	841	112	41	882	105.02	108.08	4109.0
1976:2	511	130	641	236	877	106	43	920	136.74	100.15	4064.7
1976:3	537	111	648	238	886	90	43	929	136.63	127.45	4633.1
1977:1	483	105	588	239	827	89	42	869	84.74	142.52	4674.2
1977:2	479	89	568	222	790	81	40	830	90.50	128.07	4463.3
1977:3	502	102	604	218	822	91	41	863	159.06	125.41	4653.5
1978:1	486	110	596	235	831	105	42	873	126.43	116.24	4460.8
1978:2	497	120	617	225	842	100	44	886	123.26	110.74	4346.8
1978:3	534	109	643	222	865	90	42	907	138.67	117.26	4665.7
1979:1	(498)	(111)	(609)	(241)	(850)	(88)	(44)	(894)	99.92	113.18	4626.0
1979:2	517	102	619	218	837	77	42	879	106.89	121.33	4760.9
1979:3	520	92	612	208	820	91	42	862	111.14	128.47	4961.4
1980:1	497	99	596	213	809	88	43	852	108.08	118.28	4765.7
1980:2	510	110	620	204	824	94	43	867	117.48	103.04	4611.6
1980:3	514	101	615	203	818	89	42	860	99.92	107.06	4890.1
1981:1	517	102	619	202	821	91	43	864	118.27	114.20	4808.5
1981:2	520	107	627	205	832	87	43	875	111.70	101.11	4613.5
1981:3	532	108	640	197	837	90	43	880	117.26	112.16	4875.8
1982:1	533	109	642	204	846	97	44	890	121.33	111.70	4805.2
1982:2	536	118	654	203	857	96	45	902	120.37	108.81	4683.8
1982:3	558	114	672	204	876	92	46	922	146.82	126.43	5159.2
1983:1	547	119	666	210	876	89	47	923	133.57	132.55	5113.3
1983:2	532	103	635	197	832	82	44	876	115.56	160.81	5114.3
1983:3	510	96	606	184	790	75	43	833	99.92	143.77	5314.2
1984:1	487	99	586	184	770	80	42	812	101.96	123.37	4977.7
1984:2	504	105	609	186	795	89	43	838	124.22	99.19	4694.4
1984:3	526	107	633	181	814	84	43	857	123.37	104.00	4989.0
1985:1	523	111	634	192	826	87	44	870	118.27	105.02	4903.3
1985:2	521	105	626	195	821	86	44	865	104.00	108.81	4874.5
1985:3	537	102	639	187	826	85	45	871	116.24	110.12	5165.3
1986:1	531	108	639	199	838	92	46	884	122.35	109.10	4977.7
1986:2	540	105	645	188	833	87	45	878	108.81	114.60	4974.7
1986:3	533	106	639	183	822	86	45	867	107.06	118.30	5293.8
1987:1	544	110	654	197	851	87	47	898	140.71	109.10	5086.8
1987:2	522	102	624	186	810	85	45	855	70.30	111.70	4929.4
1987:3	536	104	640	182	822	79	45	867	121.33	109.10	5232.6

* All Figures in thousands of pigs.Numbers in parentheses indicate a forecast value.

** 1,2,3 refer to April, August, December or periods 1,2 and 3

Source: M.A.F.F., obtained from M.L.C. Economics Department.

APPENDIX 4b:- The Data Used to Estimate the Monthly Biological Models*

TIME	M	FP	TIME	M	FP
1975:1	29	1021	1978:12	24	1048
1975:2	33	1034	1979:1	26	1044
1975:3	30	1037	1979:2	27	1077
1975:4	26	975	1979:3	28	1107
1975:5	24	925	1979:4	26	1062
1975:6	23	899	1979:5	28	1077
1975:7	22	892	1979:6	30	1131
1975:8	21	858	1979:7	28	1126
1975:9	26	972	1979:8	27	1096
1975:10	23	961	1979:9	28	1158
1975:11	25	974	1979:10	30	1146
1975:12	27	952	1979:11	33	1180
1976:1	23	927	1979:12	25	1091
1976:2	25	947	1980:1	28	1082
1976:3	25	972	1980:2	29	1115
1976:4	23	948	1980:3	27	1115
1976:5	25	967	1980:4	24	1061
1976:6	23	924	1980:5	24	1043
1976:7	22	916	1980:6	24	1088
1976:8	25	998	1980:7	23	1066
1976:9	28	1043	1980:8	22	1054
1976:10	30	1097	1980:9	25	1143
1976:11	34	1132	1980:10	25	1136
1976:12	30	1057	1980:11	27	1179
1977:1	33	1040	1980:12	25	1130
1977:2	35	1118	1981:1	26	1068
1977:3	34	1115	1981:2	27	1114
1977:4	29	1052	1981:3	27	1137
1977:5	31	1073	1981:4	22	1067
1977:6	30	1000	1981:5	25	1075
1977:7	29	998	1981:6	25	1093
1977:8	27	1022	1981:7	22	1031
1977:9	30	1110	1981:8	24	1059
1977:10	29	1070	1981:9	25	1095
1977:11	30	1094	1981:10	26	1147
1977:12	26	1071	1981:11	28	1194
1978:1	27	1019	1981:12	22	1103
1978:2	28	1032	1982:1	28	1091
1978:3	26	999	1982:2	27	1121
1978:4	26	1022	1982:3	27	1126
1978:5	24	984	1982:4	24	1050
1978:6	27	997	1982:5	25	1100
1978:7	25	1005	1982:6	26	1098
1978:8	25	1007	1982:7	26	1083
1978:9	28	1088	1982:8	26	1104
1978:10	27	1090	1982:9	30	1196
1978:11	28	1119	1982:10	30	1196

Cont.

TIME	M	FP
1982:11	31	1265
1982:12	25	1167
1983:1	29	1139
1983:2	33	1218
1983:3	36	1206
1983:4	38	1214
1983:5	39	1210
1983:6	39	1195
1983:7	33	1110
1983:8	31	1142
1983:9	35	1249
1983:10	34	1226
1983:11	33	1289
1983:12	27	1179
1984:1	29	1136
1984:2	30	1166
1984:3	28	1117
1984:4	23	1068
1984:5	24	1095
1984:6	23	1094
1984:7	22	1081
1984:8	24	1065
1984:9	24	1152
1984:10	24	1168
1984:11	24	1216
1984:12	20	1149
1985:1	24	1093
1985:2	26	1136
1985:3	27	1158
1985:4	25	1119
1985:5	25	1115
1985:6	26	1135
1985:7	24	1130
1985:8	23	1130
1985:9	26	1197
1985:10	26	1197
1985:11	27	1248
1985:12	22	1186
1986:1	26	1110
1986:2	27	1155
1986:3	25	1153
1986:4	26	1155
1986:5	25	1126
1986:6	27	1165
1986:7	27	1145
1986:8	26	1155

TIME	M	FP
1986:9	27	1224
1986:10	28	1226
1986:11	28	1280
1986:12	23	1214
1987:1	25	1131
1987:2	30	1181
1987:3	29	1180
1987:4	26	1130
1987:5	26	1121
1987:6	27	1162
1987:7	24	1138
1987:8	22	1170
1987:9	26	1182
1987:10	25	1213
1987:11	28	1264
1987:12	21	1152

All figures in thousands of pigs.
All months represent 4 week periods.

Source:- M.L.C. Economics Dept.

Appendix 4c

The Results of Regressing The Chosen Trimestic Biological Models Including The Relevant Subsidy and Outlier Dummies

The regressions presented in this appendix are the results of the trimestic regression models presented in the main text of chapter 4 including all relevant subsidy and outlier dummies. The equation numbers are the same as those of the equivalent expressions presented in the main text followed by the additional letter c.

$$PS_t = \underset{(151.3)}{(0.5720\,H_t} + \underset{(1.74)}{0.0056\,AugH_t} + \underset{(4.7)}{0.0164\,DecH_t)}(1+ \underset{(12.2)}{0.0037\,T}) - \underset{(-3.9)}{30.11\,D76{:}1} \qquad (4.5c.4c)$$

Obs. = 33 RSS = 1410.5 $\hat{R}^2$ = 0.92 DW = 1.46

$$PG_t = \underset{(37.6)}{0.1331\,H_t} + \underset{(0.68)}{0.0016\,AUGH_t} - \underset{(-2.71)}{0.0061\,DECH_t} - \underset{(-2.14)}{12.69\,D75{:}1} - \underset{(-1.63)}{8.16\,A83{:}2} -$$

$$\underset{(-0.70)}{4.03\,D77{:}1} - \underset{(-2.60)}{16.62\,D77{:}2} - \underset{(-0.31)}{1.77\,D77{:}3} + U_t$$

$$U_t = \underset{(4.5)}{0.6472\,U_{t-1}} \qquad (4.5c.6c)$$

Obs. = 33 RSS = 797.0 $\hat{R}^2$ = 0.98

$$M_{t-1,t} = \underset{(45.9)}{0.1308\,HB_{t-1}} - \underset{(-2.55)}{0.0071\,AugHB_{t-1}} + \underset{(0.38)}{0.0011\,DecHB_{t-1}} +$$

$$\underset{(6.38)}{41.41\,A83{:}2} + \underset{(3.78)}{26.18\,A83{:}3} + \underset{(2.19)}{14.11\,A84{:}1} + \underset{(2.77)}{18.03\,D77{:}1} + \underset{(2.70)}{18.74\,D77{:}2} + \underset{(2.19)}{14.02\,D77{:}3} + U_t$$

$$U_t = \underset{(2.46)}{0.4458\,U_{t-1}} \qquad (4.5d.1c)$$

Obs. = 32 RSS = 876.8 $\hat{R}^2$ = 0.992

$$M_{t-1,t} = \underset{(48.4)}{0.2211\ PS_{t-1}} + \underset{(5.00)}{36.45\ A83{:}2} + \underset{(3.30)}{26.01\ A83{:}3} + \underset{(1.88)}{13.68\ A84{:}1} +$$

$$\underset{(2.35)}{17.13\ D77{:}1} + \underset{(2.25)}{17.64\ D77{:}2} + \underset{(2.24)}{16.24\ D77{:}3} + U_t$$

$$U_t = \underset{(2.49)}{0.4376\ U_{t-1}} \qquad (4.5e.1c)$$

Obs. = 32 RSS = 1295.6 $\hat{R}^2$ = 0.989

$$M_{t-1,t} = \underset{(8.1)}{1.2974\ M_{t-2,t-1}} - \underset{(-4.9)}{0.5755\ M_{t-3,t-2}} + REG_t(\underset{(3.2)}{0.5386}, \underset{(3.8)}{0.5709}, \underset{(-2.9)}{-0.0544}, \underset{(-1.03)}{-0.0188}) -$$

$$1.2974\ REG_{t-1}(0.5386, 0.5709, -0.0544, -0.0188) -$$
$$0.5755\ REG_{t-2}(0.5386, 0.5709, -0.0544, -0.0188) +$$
$$\underset{(6.2)}{41.781\ A83{:}2} - \underset{(-3.0)}{31.58\ A83{:}3} + \underset{(3.0)}{20.01\ D79{:}2} + \underset{(0.04)}{0.35\ D79{:}3} - \underset{(-1.29)}{12.26\ D80{:}1} - \underset{(-1.39)}{11.29\ D80{:}2} \qquad (4.5f.1c)$$

Where $REG_t(a_1,a_2,b_1,b_2) = (a_1\ PG_{t-7} + a_2\ PG_{t-8})(1 + b_1\ AUG_t + b_2\ DEC_t)$,
and $REG_{t-1}(a_1,a_2,b_1,b_2) = (a_1\ PG_{t-8} + a_2\ PG_{t-9})(1 + b_1\ AUG_{t-1} + b_2\ DEC_{t-1})$,
and $REG_{t-2}(a_1,a_2,b_1,b_2) = (a_1\ PG_{t-9} +\ _2\ PG_{t-10})(1 + b_1\ AUG_{t-2} + b_2\ DEC_{t-2})$,

Obs. = 23 RSS = 361.0 $\hat{R}^2$ = 0.84 H = -3.16

$$PG_t = \underset{(42.3)}{0.1243\ HB_{t-3}} - \underset{(-1.80)}{12.70\ A83{:}2} - \underset{(-2.53)}{19.54\ A83{:}3} - \underset{(-2.06)}{14.59\ A84{:}1} + \underset{(3.47)}{27.63\ D76{:}1} + \underset{(3.22)}{22.89\ D76{:}2} -$$

$$\underset{(-0.74)}{5.25\ D77{:}1} - \underset{(-3.36)}{25.97\ D77{:}2} - \underset{(-2.04)}{14.47\ D77{:}3} + U_t$$

$$U_t = \underset{(2.53)}{0.4708\ U_{t-1}} \qquad (4.5g.1c)$$

Obs. = 30 RSS = 1028.4 $\hat{R}^2$ = 0.99

$$PG_t = \underset{(43.4)}{0.1752\ PP_{t-3}} - \underset{(-1.81)}{13.86\ A83{:}2} - \underset{(-2.64)}{21.78\ A83{:}3} - \underset{(-2.02)}{15.53\ A84{:}1} + \underset{(4.35)}{36.20\ D76{:}1} + \underset{(3.38)}{25.88\ D76{:}2} -$$
$$\underset{(-0.003)}{0.0220\ D77{:}1} - \underset{(-3.02)}{24.84\ D77{:}2} - \underset{(-1.89)}{14.48\ D77{:}3} + U_t$$

$$U_t = \underset{(2.04)}{0.4034\ U_{t-1}} \qquad (4.5h.1c)$$

Obs. = 30 RSS = 1191.4 $\hat{R}^2$ = 0.99

$$PG_t = 1.1260\ UG_{t-1} - 0.0217\ AugUG_{t-1} - 0.0696\ DecUG_{t-1})(1 + 0.0045\ T) + 29.80\ D75{:}2 +$$

(35.8) (-0.89) (-2.82) (3.50) (5.6)

$$20.73\ D75{:}3 + 16.58\ D76{:}1 + 0.90\ D77{:}1 - 12.40\ D77{:}2 + 13.33\ D77{:}3 \qquad (4.5i.1c)$$

(3.74) (2.91) (0.17) (-2.35) (2.55)

Obs. = 32 RSS = 529.4 $\hat{R}^2$ = 0.75 DW = 2.60

$$UG_t = 0.2614\ UG_{t-1} + REG_t(0.1069, -0.0014) - 0.2614\ REG_{t-1}(0.1069, -0.0014) -$$

(2.12) (26.4) (-0.77)

$$12.12\ A83{:}2 - 16.75\ A83{:}3 - 4.94\ A84{:}1 + 17.48\ D76{:}1 -$$

(-2.76) (-3.62) (-1.00) (3.74)

$$7.68\ D77{:}1 - 15.21\ D77{:}2 + 3.58\ D77{:}3 + 17.53\ D78{:}1 - 16.95\ D79{:}2 \qquad (4.5j.1c)$$

(-1.59) (-3.05) (0.67) (3.9) (-3.7)

Where $REG_t(\alpha,\gamma) = \alpha\ HB_{t-2}(1 + \gamma\ T)$,

and $REG_{t-1}(\alpha,\gamma) = \alpha\ HB_{t-3}(1 + \gamma\ \{T-1\})$

Obs. = 30 RSS = 316.91 $\hat{R}^2$ = 0.73 H = -0.94

$$UG_t = 0.1916\ UG_{t-1} + REG_t(0.1609, -0.0043) - 0.1916\ REG_{t-1}(0.1609, -0.0043) -$$

(1.73) (31.8) (-3.32)

$$13.14\ A83{:}2 - 17.08\ A83{:}3 - 6.14\ A84{:}1 + 16.22\ D76{:}1 -$$

(-3.29) (-4.04) (-1.37) (3.78)

$$10.52\ D77{:}1 - 17.56\ D77{:}2 + 3.82\ D77{:}3 + 17.52\ D78{:}1 - 19.15\ D79{:}2 \qquad (4.5k.1c)$$

(-2.44) (-3.8) (0.78) (4.34) (-4.7)

Where $REG_t(\alpha,\gamma) = \alpha\ PP_{t-2}(1 + \gamma\ T)$,

and $REG_{t-1}(\alpha,\gamma) = \alpha\ PP_{t-3}(1 + \gamma\ \{T-1\})$

Obs. = 30 RSS = 260.51 $\hat{R}^2$ = 0.78 H = -0.53

$$FP_{t-1,t} = (3.4644\ HB_{t-2} + 1.5622\ HB_{t-3})\ (1 - 0.0308\ AUG + 0.0271\ DEC)\ (1 + 0.0058\ T) -$$

(4.84) (2.13) (-4.18) (3.62) (12.9)

$$210.15\ D76{:}1 - 256.01\ D76{:}2 + 53.59\ A83{:}2 - 116.64\ A83{:}3 - 152.30\ A84{:}1 -$$

(-2.88) (-3.42) (0.74) (-1.65) (-1.77)

$$180.87\ D77{:}2 - 89.61\ D77{:}3 + 54.83\ D78{:}1 \qquad (4.5l.1c)$$

(-2.49) (-1.04) (0.67)

Obs. = 30 RSS = 70376.8 $\hat{R}^2$ = 0.95 DW = 1.52

$$FP_{t-1,t} = (4.4969\ PP_{t-2} + 2.9695\ PP_{t-3})\ (1 - 0.0381\ AUG + 0.0260\ DEC)\ (1 + 0.0027\ T) -$$
$$\qquad (5.06) \qquad (3.21) \qquad (-4.70) \qquad (2.56) \qquad (5.91)$$

$$179.79\ D76{:}1 - 310.99\ D76{:}2 + 46.99\ A83{:}2 - 133.79\ A83{:}3 - 177.61\ A84{:}1 -$$
$$(-2.21) \qquad (-3.75) \qquad (0.60) \qquad (-1.71) \qquad (-1.91)$$

$$232.09\ D77{:}2 - 108.86\ D77{:}3 + 80.26\ D78{:}1 \qquad (4.5m.1c)$$
$$(-2.88) \qquad (-1.18) \qquad (0.93)$$

Obs. = 30 RSS = 86061.9 $\hat{R}^2$ = 0.94 DW = 1.07

Appendix 4d

The Results of Regressing The Monthly Biological Models Including The Relevant Subsidy and Outlier Dummies

The regressions presented in this appendix are the results of the monthly regression models presented in the main text of chapter 4 including all relevant subsidy and outlier dummies. The equation numbers are the same as those of the equivalent expressions presented in the main text with the addition of the letter d.

Equation 4.6a.1d

$$M_{t-1,t} = a\ HB_{t-4} + b2\ febHB_{t-4} + b3\ marHB_{t-4} + b4\ aprHB_{t-4} + b5\ mayHB_{t-4} + b6\ junHB_{t-4} +$$

$$b7\ julHB_{t-4} + b8\ augHB_{t-4} + b9\ sepHB_{t-4} + b10\ octHB_{t-4} + b11\ novHB_{t-4} + b12\ decHB_{t-4} +$$
$$D1\ A83{:}3 + D2\ A83{:}4 + D3\ A83{:}5 + D4\ A83{:}6 + E1\ D77{:}2 + E2\ D77{:}3 + E3\ D77{:}4 +$$
$$E4\ D77{:}5 + E5\ D77{:}6 + F1\ D75{:}12 + U_t$$
$$U_t = R1\ U_{t-1} + e_t \qquad (4.6a.1d)$$

Obs. = 128 RSS = 225.1 $\hat{R}^2$ = 0.91

Table 4.6a.1d.
The Results of Estimating the Monthly Culling of Sows and Boars as a Proportion of the Total Breeding Herd Including Subsidy and Outlier Dummies

VARIABLE	COEFFICIENT	ESTIMATE	t-RATIO
U_{t-1}	R1	0.8496	17.3
HB_{t-4}	a	0.0306	28.4
febHB	b2	0.0015	2.68
marHB	b3	0.0010	1.28
aprHB	b4	-0.0021	-2.40
mayHB	b5	-0.0010	-1.06
junHB	b6	-0.0005	-0.49
julHB	b7	-0.0020	-2.17
augHB	b8	-0.0021	-2.36
sepHB	b9	0.0009	1.07
octHB	b10	0.0009	1.06
novHB	b11	0.0025	3.60
decHB	b12	-0.0031	-5.67
A83:3	D1	3.08	2.08
A83:4	D2	7.42	4.13
A83:5	D3	6.73	3.76
A83:6	D4	5.49	3.73
D77:2	E1	0.94	0.63
D77:3	E2	0.65	0.35
D77:4	E3	-1.35	-0.68
D77:5	E4	1.40	0.75
D77:6	E5	-0.14	-0.10
D75:12	F1	7.13	6.12

Equation 4.6b.1d

$M_{t-1,t} = a\ PS_{t-4} + b2\ febPS_{t-4} + b3\ marPS_{t-4} + b4\ aprPS_{t-4} + b5\ mayPS_{t-4} + b6\ junPS_{t-4} +$

$b7\ julPS_{t-4} + b8\ augPS_{t-4} + b9\ sepPS_{t-4} + b10\ octPS_{t-4} + b11\ novPS_{t-4} + b12\ decPS_{t-4} +$

D1 A83:3 + D2 A83:4 + D3 A83:5 + D4 A83:6 + E1 D77:2 + E2 D77:3 + E3 D77:4 +

E4 D77:5 + E5 D77:6 + F1 D75:12 + U_t

$U_t = R1\ U_{t-1} + e_t$ (4.6b.1d)

Obs. = 128 RSS = 233.0 $\hat{R}^2 = 0.91$

Table 4.6b.1d.
The Results of Estimating the Monthly Culling of Sows and Boars as a Proportion of the Pregnant Sow Herd Including Subsidy and Outlier Dummies.

VARIABLE	COEFFICIENT	ESTIMATE	t-RATIO
U_{t-1}	R1	0.8515	17.4
PS_{t-4}	a	0.0513	27.6
febPS	b2	0.0026	2.69
marPS	b3	0.0018	1.34
aprPS	b4	-0.0034	-2.28
mayPS	b5	0.0001	0.07
junPS	b6	0.0008	0.48
julPS	b7	-0.0018	-1.13
augPS	b8	-0.0020	-1.27
sepPS	b9	0.0027	1.81
octPS	b10	0.0026	1.90
novPS	b11	0.0055	4.53
decPS	b12	-0.0043	-4.54
A83:3	D1	3.06	2.03
A83:4	D2	7.47	4.08
A83:5	D3	6.47	3.54
A83:6	D4	5.40	3.60
D77:2	E1	0.99	0.65
D77:3	E2	0.67	0.35
D77:4	E3	-1.43	-0.71
D77:5	E4	1.45	0.77
D77:6	E5	-0.007	-0.005
D75:12	F1	6.87	5.83

Equation 4.6c.4d

$$FP_{t-1,t} = R1\ FP_{t-2,t-1} + REG_t - R1\ REG_{t-1} + D1\ DUMA + E1\ D77{:}1 + E2\ D77{:}3 + E3\ D77{:}4 + E4\ D77{:}5 + E5\ D77{:}6 + e_t \qquad (4.6c.4d)$$

Where $REG_t = (a1\ HB_{t-8} - a2\ DumXHB_{t-8} + a2\ DumXHB_{t-12})(1 + c\ T)(1 + b2\ Feb + b3\ Mar + b4\ Apr + b5\ May + b6\ Jun + b7\ Jul + b8\ Aug + b9\ Sep + b10\ Oct + b11\ Nov + b12\ Dec)$

and $REG_{t-1} = (a1\ HB_{t-9} - a2\ DumXHB_{t-9} + a2\ DumXHB_{t-13})(1 + c\ (T-1))(1 + b2\ Feb(-1) + b3\ Mar(-1) + b4\ Apr(-1) + b5\ May(-1) + b6\ Jun(-1) + b7\ Jul(-1) + b8\ Aug(-1) + b9\ Sep(-1) + b10\ Oct(-1) + b11\ Nov(-1) + b12\ Dec(-1))$

Obs. = 119 RSS = 55681.4 $\hat{R}^2 = 0.89$ H = -1.42

Table 4.6c.1d.

The Results of Estimating the Monthly Slaughterings of Fatpigs as a Proportion of the Total Breeding Herd Including The Relevant Subsidy Dummies.

VARIABLE	COEFFICIENT	ESTIMATE	t-RATIO
$FP_{t-2,t-1}$	R1	0.5165	6.19
HB_{t-8}	a1	1.0556	57.5
HB_{t-12}	a2	0.6568	4.40
T(ime)	c	0.0018	9.75
Feb	b2	0.0325	3.79
Mar	b3	0.0311	2.96
Apr	b4	-0.0031	-0.27
May	b5	-0.0019	-0.16
Jun	b6	0.0013	0.11
Jul	b7	-0.0236	-2.06
Aug	b8	-0.0151	-1.33
Sep	b9	0.0546	4.69
Oct	b10	0.0605	5.38
Nov	b11	0.0999	9.38
Dec	b12	0.0300	3.54
ADUM	D1	-16.24	-2.10
D77:2	E1	20.37	0.79
D77:3	E2	9.42	0.37
D77:4	E3	-17.39	-0.69
D77:5	E4	11.44	0.45
D77:6	E5	-80.47	-3.18

Equation 4.6d.1d

$$FP_{t-1,t} = R1\ FP_{t-2,t-1} + REG_t - R1\ REG_{t-1} + D1\ DUMA + E1\ D77{:}1 + E2\ D77{:}3 + E3\ D77{:}4 + E4\ D77{:}5 + E5\ D77{:}6 + e_t \qquad (4.6d.1d)$$

Where $REG_t = (a1\ PP_{t-8} - a2\ DumXPP_{t-8} + a2\ DumXPP_{t-12})\ (1 + c\ T)\ (1 + b2\ Feb + b3\ Mar + b4\ Apr + b5\ May + b6\ Jun + b7\ Jul + b8\ Aug + b9\ Sep + b10\ Oct + b11\ Nov + b12\ Dcc)$

and $REG_{t-1} = (a1\ PP_{t-9} - a2\ DumXPP_{t-9} + a2\ DumXPP_{t-13})\ (1 + c\ (T-1))\ (1 + b2\ Feb(-1) + b3\ Mar(-1) + b4\ Apr(-1) + b5\ May(-1) + b6\ Jun(-1) + b7\ Jul(-1) + b8\ Aug(-1) + b9\ Sep(-1) + b10\ Oct(-1) + b11\ Nov(-1) + b12\ Dec(-1))$

Obs. = 119 RSS = 62127.9 $\hat{R}^2 = 0.88$ H = -1.30

Table 4.6d.1d.

The Results of Estimating the Monthly Slaughterings of Fatpigs as a Proportion of the Pregnant Pig Herd Including The Relevant Subsidy Dummies

VARIABLE	COEFFICIENT	ESTIMATE	t-RATIO
$FP_{t-2,t-1}$	R1	0.5613	6.94
PP_{t-8}	a1	1.6160	52.4
PP_{t-12}	a2	0.9469	4.88
T(ime)	c	0.0009	4.93
Feb	b2	0.0244	2.70
Mar	b3	0.0242	2.16
Apr	b4	-0.0090	-0.75
May	b5	-0.0119	-0.97
Jun	b6	-0.0146	-1.16
Jul	b7	-0.0382	-3.12
Aug	b8	-0.0293	-2.41
Sep	b9	0.0479	3.59
Oct	b10	0.0654	5.35
Nov	b11	0.1057	9.19
Dec	b12	0.0360	3.96
ADUM	D1	-15.74	-1.93
D77:2	E1	4.73	0.17
D77:3	E2	0.33	0.01
D77:4	E3	-25.51	-0.95
D77:5	E4	8.55	0.32
D77:6	E5	-79.62	-2.98

APPENDIX 6

The Monthly Data Used In The Analysis Included In Chapters 5 and 6

TIME	M*	FP*	AAPP**	CF**	PR**
1975:1	29	1021	61.4	60.1	102.2
1975:2	33	1034	60.8	59.3	102.5
1975:3	30	1037	63.8	56.3	113.3
1975:4	26	975	67.9	53.4	127.2
1975:5	24	925	71.0	52.5	135.2
1975:6	23	899	71.4	52.5	136.0
1975:7	22	892	69.2	52.6	131.6
1975:8	21	858	67.9	52.7	128.8
1975:9	26	972	71.6	54.6	131.1
1975:10	23	954	77.2	58.9	131.1
1975:11	25	974	80.4	60.3	133.3
75:12	27	952	80.9	61.4	131.8
1976:1	23	927	79.9	61.8	129.3
1976:2	25	947	79.6	62.1	128.2
1976:3	25	972	79.0	62.5	126.4
1976:4	23	948	78.8	65.0	121.2
1976:5	25	967	77.8	68.6	113.4
1976:6	23	924	75.8	70.0	108.3
1976:7	22	916	71.0	71.9	98.7
1976:8	25	998	72.4	73.5	98.5
1976:9	28	1043	74.5	74.8	99.6
1976:10	30	1097	78.5	78.6	99.9
1976:11	34	1132	80.2	81.5	98.4
1976:12	30	1057	80.7	83.1	97.1
1977:1	33	1040	79.9	83.6	95.6
1977:2	35	1118	82.2	85.1	96.6
1977:3	34	1115	80.4	87.0	92.4
1977:4	29	1052	81.9	87.7	93.4
1977:5	31	1073	84.6	88.0	96.1
1977:6	30	1000	81.6	89.6	91.1
1977:7	29	998	81.6	88.9	91.8
1977:8	27	1022	81.1	86.1	94.2
1977:9	30	1110	84.0	83.7	100.4
1977:10	29	1070	87.0	81.0	107.4
1977:11	30	1094	87.2	79.2	110.1
1977:12	26	1071	87.4	77.9	112.2
1978:1	27	1019	87.0	77.9	111.7
1978:2	28	1032	88.1	77.9	113.1
1978:3	26	999	89.4	78.0	114.6
1978:4	26	1022	90.9	78.5	115.8
1978:5	24	984	92.7	80.7	114.9
1978:6	27	997	91.0	83.2	109.4
1978:7	25	1005	92.2	85.3	108.1
1978:8	25	1007	92.0	85.0	108.2
1978:9	28	1088	93.7	84.1	111.4
1978:10	27	1090	95.2	83.8	113.6

Cont.

TIME	M*	FP*	AAPP**	CF**	PR**
1978:11	28	1119	95.2	83.9	113.5
1978:12	24	1048	94.3	85.2	110.7
1979:1	26	1044	93.6	86.8	107.8
1979:2	27	1077	92.3	89.0	103.7
1979:3	28	1107	91.1	89.9	101.3
1979:4	26	1062	89.8	92.7	96.9
1979:5	28	1077	88.8	93.9	94.6
1979:6	30	1131	87.9	95.2	92.3
1979:7	28	1126	88.6	95.2	93.1
1979:8	27	1096	89.7	94.5	94.9
1979:9	28	1158	93.1	93.1	100.0
1979:10	30	1146	99.7	93.1	107.1
1979:11	33	1180	104.4	94.7	110.2
1979:12	25	1091	104.3	96.3	108.3
1980:1	28	1082	100.9	99.5	101.4
1980:2	29	1115	98.9	100.5	98.4
1980:3	27	1115	98.2	100.8	97.4
1980:4	24	1061	99.3	99.5	99.8
1980:5	24	1043	101.0	98.6	102.4
1980:6	24	1088	100.8	98.5	102.3
1980:7	23	1066	100.4	98.9	101.5
1980:8	22	1054	98.2	99.0	99.2
1980:9	25	1143	97.8	98.9	98.9
1980:10	25	1136	99.9	100.1	99.8
1980:11	27	1179	102.4	102.1	100.3
1980:12	25	1130	102.4	103.9	98.6
1981:1	26	1068	101.0	106.5	94.8
1981:2	27	1114	99.8	108.1	92.3
1981:3	27	1137	102.0	108.1	94.4
1981:4	22	1067	105.3	108.9	96.7
1981:5	25	1075	108.9	109.4	99.5
1981:6	25	1093	111.9	110.0	101.7
1981:7	22	1031	108.8	110.3	98.6
1981:8	24	1059	103.1	110.4	93.4
1981:9	25	1095	104.3	110.1	94.7
1981:10	26	1147	113.2	110.8	102.2
1981:11	28	1194	118.1	111.6	105.8
1981:12	22	1103	119.3	112.8	105.8
1982:1	28	1091	117.8	113.3	104
1982:2	27	1121	116.3	114.2	101.8
1982:3	27	1126	116.1	114.9	101.0
1982:4	24	1050	113.1	115.3	98.1
1982:5	25	1100	111.3	117.1	95.0
1982:6	26	1098	109.4	117.8	92.9
1982:7	26	1083	108.8	118.0	92.2
1982:8	26	1104	106.3	117.0	90.9
1982:9	30	1196	107.8	114.7	94.0
1982:10	30	1196	110.4	114.1	96.8
1982:11	31	1265	110.8	114.4	96.9
1982:12	25	1167	109.9	115.9	94.8
1983:1	29	1139	104.3	117.0	89.1

Cont.

TIME	M*	FP*	AAPP**	CF**	PR**
1983:2	33	1218	99.9	118.9	84.0
1983:3	36	1206	101.1	120.7	83.8
1983:4	38	1214	100.1	123.7	80.9
1983:5	39	1210	103.4	126.0	82.1
1983:6	39	1195	105.3	127.4	82.7
1983:7	33	1110	103.3	127.1	81.3
1983:8	31	1142	102.3	124.9	81.9
1983:9	35	1249	109.7	124.1	88.4
1983:10	34	1226	114.3	126.2	90.6
1983:11	33	1289	114.4	129.7	88.2
1983:12	27	1179	115.0	131.1	87.7
1984:1	29	1136	111.5	133.8	83.3
1984:2	30	1166	115.1	134.2	85.8
1984:3	28	1117	120.7	134.2	89.9
1984:4	23	1068	124.3	134.2	92.6
1984:5	24	1095	125.0	133.9	93.4
1984:6	23	1094	126.1	133.9	94.2
1984:7	22	1081	121.5	132.9	91.4
1984:8	24	1065	122.8	128.1	95.9
1984:9	24	1152	126.7	124.3	101.9
1984:10	24	1168	131.7	120.5	109.3
1984:11	24	1216	131.7	122.8	107.2
1984:12	20	1149	131.0	124.4	105.3
1985:1	24	1093	125.1	119.6	104.6
1985:2	26	1136	118.2	121.3	97.4
1985:3	27	1158	117.0	125.0	93.6
1985:4	25	1119	117.2	127.9	91.6
1985:5	25	1115	117.0	128.3	91.2
1985:6	26	1135	117.1	127.5	91.8
1985:7	24	1130	115.7	125.4	92.3
1985:8	23	1130	114.4	123.3	92.8
1985:9	26	1197	116.6	123.7	94.3
1985:10	26	1197	117.8	124.1	94.9
1985:11	27	1248	120.6	124.4	96.9
1985:12	22	1186	120.1	125.3	95.8
1986:1	26	1110	113.5	125.6	90.4
1986:2	27	1155	110.3	125.4	88.0
1986:3	25	1153	113.6	127.4	89.2
1986:4	26	1155	111.9	127.5	87.8
1986:5	25	1126	113.9	126.7	89.9
1986:6	27	1165	116.2	125.9	92.3
1986:7	27	1145	111.3	125.7	88.5
1986:8	26	1155	112.5	123.4	91.2
1986:9	27	1224	115.4	121.6	94.9
1986:10	28	1226	115.0	123.3	93.3
1986:11	28	1280	114.5	124.1	92.3
1986:12	23	1214	112.9	124.5	90.7
1987:1	25	1131	109.5	125.7	87.1
1987:2	30	1181	110.0	126.3	87.1
1987:3	29	1180	113.9	127.4	89.4
1987:4	26	1130	114.7	127.2	90.2

Cont.

TIME	M*	FP*	AAPP**	CF**	PR**
1987:5	26	1121	113.8	127.2	89.5
1987:6	27	1162	117.4	126.9	92.5
1987:7	24	1138	113.8	126.9	89.7
1987:8	22	1170	110.6	124.5	88.8
1987:9	26	1182	110.2	123.7	89.1
1987:10	25	1213	112.1	123.0	91.1
1987:11	28	1264	110.3	124.1	88.9
1987:12	21	1152	108.0	124.8	86.5

* Data in thousands of pigs and representative of four week months.

** Index with a base year of 1980=100 and deflated by the RPI where necessary

Source:- MAFF through MLC economics Department.

APPENDIX 9

COMBINATION OF FORECASTS

In this appendix, an illustration of forecast combining is presented using the one-step forecasts for fat pigs as an example, based upon the work of Bates and Granger (1969). The method is illustrated on a monthly series because of the lack of observations and the presence of suspect sample data in the trimestic breeding herd data: fat pigs were chosen rather than culling because it was thought that it might be of interest to some readers to see how adjustments could be made for biased forecasts.

When forecasts for a given variable have been derived using various models, as has been done in this thesis, the first reaction of the forecaster might well be to choose the model which appears to be best to use as a working model, discarding the other models. Bates and Granger set out to illustrate that under certain circumstances, using the information from two or more forecasting models by combining their respective forecasts, the resulting combined forecasts could well produce an error variance smaller than that of the forecasts from the best individual model. An important condition which should prevail before combining of forecasts can be considered is that all the individual sets of forecasts are unbiased because the combining of biased with unbiased forecasts is certain to result in biased forecasts. Only the combination of the bivariate and the biological one-step forecasts are considered because of the independence of these two methods, the univariate method being excluded because of its similarity to the bivariate modelling method. The biological model forecasts for fat pig slaughterings are clearly over-forecasting for reasons discussed in chapters eight and nine, and, therefore, a correction for the over-forecasting bias of the biological model is required before combining can be considered. The average percentage error from the biological model forecasts was computed at 2.31%. All the forecasts from the biological model were, therefore, reduced by 2.31% and the resulting adjusted forecasts were computed as having a sum of squares of forecasting error of 13040.2, that is, a MSFE of 543.3, slightly larger than that from the bivariate Box-Jenkins model forecasts of the out-of-sample period 1986-7.

Before a 'best' method for combining the two sets of forecasts is considered, a simple illustration of combining is presented, in which equal weight is given to the biological and the bivariate forecasts. Combining the forecasts in such a way and calculating the MSFE of the 24 forecasts of 505.9, an error variance which is lower than that for either of the two individual sets of forecasts, calculated as 522.5 for the bivariate forecasts and 543.3 for the adjusted biological forecasts. So, by taking no account of the fact that the bivariate forecasts are the better of the two sets in terms of the MSFE statistics, a lower forecasting error variance has been produced, thereby indicating the usefulness of including the information provided by the biological model.

Having shown for illustrative purposes how a the simplest of all combinations can improve the error variance of even the best individual set of forecasts, use of the individual forecasts' error variance can be made in order to give greater weight to the set of forecasts with the lowest MSFE statistic. Denoting the forecast error variance of the two models by $MSFE_{bio}$ and $MSFE_{bv}$, a combined forecast is obtained by a linear combination of the two sets of forecasts multiplying the biological forecasts by the weight W_{bio}, and the bivariate forecasts by the weight $1 - W_{bio}$. The error variance of the combined forecasts can easily be calculated and by differentiating with respect to W_{bio} and setting equal to zero, an error variance minimising expression for W_{bio} can be obtained as given in equation A9.1 below;

$$W_{bio} = \frac{\sigma^2_{bv} - \rho\sigma_{\beta\iota o}\sigma_{bv}}{\sigma^2_{bio} + \sigma^2_{bv} - 2\rho\sigma_{bio}\sigma_{bv}} \qquad (A9.1)$$

where ρ is the correlation coefficient between the two individual sets of forecast errors. It can be shown using equation A9.1 that calculating W_{bio} in this way produces an error variance for the combined forecasts which is no greater than the smallest error variance of the individual sets of forecasts. Assuming ρ to be zero, weights for the two sets of forecasts can be apportioned in the following way; $W_{bio} = MSFE_{bv} / (MSFE_{bio} + MSFE_{bv})$, and likewise for W_{bv}, $(1 - W_{bio})$, the weight for the bivariate forecasts, also obtained by replacing the numerator in the previous expression by $MSFE_{bio}$.

Calculating the weights assuming ρ to be zero, using the latter expressions, 0.509 is the weight given to the slightly superior bivariate forecasts and, therefore, the equivalent weight for the adjusted biological forecasts is given as 0.491. The weights are very similar, reflecting the similarity of the MSFE statistics of the two sets of individual forecasts. Combining the forecasts by applying the weights discussed above produced forecasts with a MSFE statistic of 505.7, very marginally lower than that obtained when giving equal weights to the two sets of forecasts.

Having justified and explained a basis for combining forecasts, Bates and Granger went on to show how different methods could be applied which allowed the weights applied to the various forecasting methods to change over time as information became available and as the relative forecasting abilities of the various approaches changed over time. Thus, the combined forecast for time period t could be written as;

$$C_t = W_{bio} F_{bio,t} + (1 - W_{bio}) F_{bv,t} \ , \qquad \text{(A9.2)}$$

where $F_{i,t}$ is the forecast for time t using forecasting method i.

Granger and Bates illustrated five examples of methods for updating the weights, some of which assumed ρ to be zero and others which did not, and two of the five methods allowed changing relative forecasting abilities of the two sets of forecasts to have a greater influence on the calculated weights by basing the values of the weight only on the most recent of forecasting errors rather than incorporating the errors from all past forecasts. For illustrative purposes I decided to apply three such methods to the fat pig forecasts discussed in this appendix.[1] These were:-

1. and 2.

$$W_{bio,t} = \frac{E_{bv}}{E_{bio} + E_{bv}} \qquad \text{(A9.3)}$$

where E_i is the sum of the square forecast errors from individual forecasts i, summed over the period T-v to T-1, time T being the latest time period to forecasts. Method 1 sums over all past forecast errors so that v = 23, whereas in method 2, v = 6, so that only the 6 most recent forecasting errors are taken into consideration. For both methods 1 and 2 equal weighting is given to the forecasts for time period t=1.

The MSFE for the combined forecast using method 1 was calculated at 500.7, a further reduction on the values obtained using the individual and crude combination methods examined to date. The weight on the bivariate forecasts was 0.36 in period 2, and remained less than 0.5 until about period 18 when it became larger than 0.5, ending up at 0.51 in period 24. These results imply that the bivariate forecasts, while having the lowest overall MSFE of the two sets of individual forecasts, only became superior towards the end of the out-of-sample period. The second method produced a MSFE value of 503.2, slightly greater than method 1, implying that the information gathered over all past forecast errors was more useful than the information gathered from observation of the last 6 errors only.

---

1. The reader who is interested in seeing all the methods employed by Bates and Granger should consult their paper, or alternatively see Granger and Newbold (1977).

Equation A9.4 denotes the third and final method examined,

3.

$$W_{bio,t} = (1-x)\, W_{bio,t-1} + x \frac{e_{bv,t-1}}{e_{bio,t-1} + e_{bv,t-1}} \qquad \text{(A9.4)}$$

where $e_{i,t}$ is the absolute forecast error at time t using forecasting method i, and where x is a constant taking values between 0 and 1. This method was applied, allowing the value of x to vary from 0.25 up to 1, the latter putting all the weight on to the most recent of errors. As x increased in value from 0.25 to 1 the MSFE of the resulting combined forecast fell consistently from 492.5 to 471.1, implying that method 3 was the best method of the three updating methods examined and that the optimal value of x for the forecasting period analysed was a value of one, so that W_t is based solely on the forecast errors from the two individual methods at time t-1, and takes no account of W_{t-1}. This value of one implies that the weights used are very volatile, and do not converge on a single optimal value over time.

These results are somewhat contradictory in that the comparison of the various combining methods indicate that the best method is method three, for all values of x, although the best MSFE statistic is achieved when x=1, that is, the weight for forecasts at time t is determined solely by the forecasts errors at time t-1, the most recent errors. This result conflicts somewhat with a comparison of the MSFEs from methods 1 and 2, which favour the use of all past information, not only the most recent of information.

This appendix has served its purpose of illustrating some simple and more sophisticated methods of combining unbiased forecasts derived from different models, and indicating how such combined forecasts can improve on the forecasting abilities of individual models, by making use of other less successful, but nonetheless useful models.

Bibliography

Ash, J.C.K. and Smyth, D.J., (1973), *Forecasting the UK Economy.* Westmead: Saxon House.

Bates, J.M. and Granger, C.W.J., (1969), *The Combination of Forecasts*, Operations Research Quarterly. Vol 20.

Beech, C.M. and Mackinnon, J.G., (1978), *A Maximum Likelihood Procedure for Regression Containing Autocorrelated Errors.* Econometrica Vol. 46.

Bessler, D.A., (1984), *An Analysis of Dynamic Economic Relationships: An Application to the US Hog Market.* Canadian Journal of Agricultural Economics Vol. 32.

Box, G.E.P. and Jenkins, G.M., (1970), *Time Series Analysis: Forcasting and Control.* San Francisco: Holden-Day.

Box, G.E.P. and Jenkins, G.M., (1970), *Distribution of Autocorrelations in ARIMA Time Series Models.* J. Am. Stat. Assoc. Vol. 65.

Burton, M.P., (1987), *The Pig Sector.* in An Econometric Model of the UK Agricultural Sector. Manchester Working Papers in Agricultural Economics. WP 87/02

Byers, J.D. and Peel, D.A., (1987), *Forecasting Livestock Slaughter:- An Empirical Assessment of MLC Forecasts.* Journal of Agricultural Economics. Vol. xxxviii, No 2.

Calvert, N.W., Jennings, A.N. and Rayner, A.J., (1986), *A Statistical Digest of UK Meat Production, Consumption and Trade, 1973-84.* Departmental Discussion Paper No. 48, University of Nottingham.

Colman, D., (1983), *A Review of the Arts of Supply Response Analysis.* Review of Marketing and Agricultural Economics Vol 51, No. 3.

C.S.O., (various), Monthly Digest of statistics. London: H.M.S.O.

D.A.F.S., (various), *Economic Report on Scottish Agriculture.* Edinburgh: H.M.S.O.

Damsleth, E., (1984), *Forecasting and the Production of Pigs- The Birds and the Bees Revisited.* in Time Series Analysis: Theory and Practice 5. Ed. Anderson, O.D., Amsterdam: North-Holland.

Daniels, L.M. and Savin, D. (1977), *A Model For Forecasting Livestock Producer Prices.*in 'Meat Demand and Price Forecasting'. Procedings of a symposium. Milton Keynes, M.L.C. Information Services.

Davies, N. and Newbold, P., (1979), *Some Power Studies of a Portmanteau Test of Time Series Model Specification.* Biometrica, Vol. 66, No. 1.

Davies, N., Triggs, C.M. and Newbold, P., (1977), *Significance Levels of the Box-Pierce Portmanteau statistic in Finite Samples.* Biometrica, Vol. 64.

Durbin, J., (1960), *The Fitting of Time Series.* Rev. Inst. Int. Stats. Vol. 28.

Durbin, J., (1970), *Testing For Serial Autocorrelation in Least Squares Regression When Some of the Regressors are Lagged Dependent Variables.* Econometrica, Vol 38.

Engle, R.E., and Granger, C.W.J., (1985), *Cointegration and Error Correction: Representation, Estimation and Testing.*University of California, San Diego.

Fair, R.C., (1986), *Evaluating the Predictive Accuracy of Models.* Handbook of Econometrics. Vol 3. Ed. Griliches and Intrilligator. Elsavier Science.

Farebrother, R.W., (1980), *The Durbin-Watson Test for Serial Autocorrelation when there is no intercept in the regression.* Econometrica Vol. 48, No.6.

Godfrey, L.G., (1978), *Testing Against General Autoregressive and Moving Average Error Models When the Regressors Include Lagged Dependent Variables.* Econometrica Vol. 46.

Granger, C.W.J., (1980), *Forcasting In Business and Economics.*London: Academic Press.

Granger, C.W.J. and Hatanaka, M., (1964), *Spectral Analysis of Economic Time Series.*Priceton: Princeton University Press.

Granger, C.W.J. and Newbold, P., (1977), *Forcasting Economics Time Series.* London: Academic Press.

Gujarati. D., (1978), *Basic Econometrics.* New York: McGraw Hill.

Hall, B.H., (1983), *Time Series Processor Version 4.0 Reference Manual.* California: TSP International

Hall, B.H., (1983), *Time Series Processor Version 4.0 Users Manual including an Introductory Guide.* California: TSP International.

Hendry, D.F., Pagan, A.R. and Sargan, J.D., (1984), *Dynamic Specification.* Handbook of Econometrics Vol 2, Ed. Griliches and Intrilligator. Elsavier Science.

Harvey, A.C., (1981), *Time Series Models.* Oxford: Philip Allan.

Harvey, A.C., (1983), *A Unified View of Statistical Forcasting Procedures.* L.S.E. Econometrics Programme Discussion Paper No. A.40

Hill, B.E., (1984), *The Common Agricultural Policy: Past, Present, and Future.* London: Methuen.

Jones, G.T., (1965), *The Influence of Price on Livestock Population Over the Last Decade.* J. Ag. Econ. Vol. xvi. No. 4

Judge, G.G. et al., (1982), *Introduction to the Theory and Practice of Econometrics.* New York: Wiley.

King, M.L.,, (1980), *The Durbin-Watson Bounds Test and Regressions without an intercept.* Monash University Working Paper No. 10/80.

Lon-Mu Liu, (1986), *Identification ofTime Series Models in the Presence of Calendar Variation.* International Journal of Forecasting Vol 2, No. 3.

Ljung, G.M. and Box, G.E.P., (1978), *On a Measure of Lack of Fit in Time Series Models*. Biometrica, Vol. 65 No. 2.

Makridakis, S. and Hibon, M., (1979), *Accuracy of Forecasting: An Empirical Investigation*. J.R.S.S.(A). Vol. 142 part 2.

Mellor, C.J., (1982), *Dynamic Statistical Models Of Agricultural Supply Within An Application To Cereals Production*. PhD. University of Nottingham.

McClements, L.D. (1970), *Note on Harmonic Motion and the Cobweb Theorem*. J. Ag. Econ. Vol. XXI No. 1.

McClements, L.D. (1971), *United Kingdom Pig Industry Statistics 1956-1969*. Thesis, University of Manchester.

McClements, L.D., (1971), *An Econometric Model of The United Kingdom Pig Industry*. Thesis, University of Manchester.

Mühlebach, F., (1988), *Combination of an Animal Breeding Simulation Model with a Sector Model*. Paper given at 16th European Seminar of Agricultural Economists, Bonn, F.R.G., April 1988.

M.A.F.F., (various), *Agricultural statistics. England and Wales- Agricultural Censuses and Production*. London: H.M.S.O. Government statistical service.

M.A.N.I., (various), *Statistical Review of Northern Ireland Agriculture*. Belfast: A Government statistical publication, Economics and Statistics division.

M.L.C., *European Handbook Vol 1: The Common Agricultural Policy*.Milton Keynes: Economic Services and Publications.

M.L.C., *European Handbook Vol 2: EEC and International Statistics*.Milton Keynes: Economic Services and Publications.

M.L.C., (various), *International Market Survey*: Milton Keynes Economic Information Service.

M.L.C., (various), *Pig Improvement Services*. Milton Keynes: Economic Information Service.

M.L.C., (various), *Pig Year Book*. Milton Keynes: Economics, Livestock and Marketing Services.

M.L.C., (various), *U.K. Market Survey*. Milton Keynes: Economic Information Service.

M.L.C., (various), *U.K. Meat and Livestock Statistics*. Milton Keynes: Economic Information Service.

Nelson, C.R., (1973), *Applied Time Series Analysis For Managerial Forcasting*. San Francisco: Holden-Day.

Ness, M. and Colman, D.R., (1976), *Forecasting the Size of the UK Pig Breeding Herd*.Bulletin 157, Dept. of Ag. Econ., Manchester University.

Newbold, P. and Granger, C.W.J., (1974), *Experience with Forecasting Univariate Time Series and the Combination of Forecasts*. J.R.S.S. Series A Vol. 137, part 2.

Norton. D., (1986), *Smuggling Under the C.A.P.: Northern Ireland and the Republic of Ireland.* Journal of Commom Market Studies. Vol.xxiv No. 4.

Pagano, M. and Hartley, M., (1981), *On Fitting Distributed Lag Models subject to Polynomial* Restrictions. J. Econometrics. Vol. 16, North-Holland.

Pindyck, R.S. and Rubinfeld, D.L., (1987), *Econometric Models and Economic Forecasts.* Ed. 2, London: McGraw-Hill.

Quenouille, M.H., (1949), *Approximate Tests of Correlation in Time Series.* J. Roy. Stat. Soc. Vol. B11.

Rayner, A.J. and Young, R.J., (1980), *Information, Hierarchical Model Structures and Forecasting.* European Review of Agricultural Economics. Vol 7.

Revell, B.J. (1977), *Cattle Price Forecasting with a Box-Jenkins Model - Experience in the 1974-76 period.*in 'Meat Demand and Price Forecasting'. Procedings of a symposium. Milton Keynes, M.L.C. Information Services.

Savin, D., (1978), *Forecasting the Pig Breeding Herd. (- an examination of the differential response to changes in profitability.)* in 'Supply Response and the World Meat Situation'. Procedings of a symposium. Milton Keynes, M.L.C. Information Services.

Seber, G.A.F., (1977), *Linear Regression Analysis.* London: John Wiley and Sons.

Stewart, J., (1984), *Undestanding Econometrics.* 2nd Ed. London: Hutchinson.

Stillman, R.P., (1985), *A Quarterly Model of the Livestock Industry.* U.S.D.A. Economics Research Service, Technical Bulletin No. 1711.

University of Cambridge, (Various), *Pig Management Scheme Results.* Agricultural Economics Unit.

University of Exeter, (Various), *Pig Production in South-West England.* Agricultural Economics Unit.

Weisberg, S., (1980), *Applied Linear Regression.* New York: John Wiley and Sons.

Westcott, P.C. and Hull, D., (1985), *A Quarterly Forecasting Model for US Agriculture: Subsector Models for Corn, Wheat, Soya Beans, Cattle Hogs and Poultry.* U.S.D.A. Economic Research Service, Technical Bulletin No. 1700.

Wickens, M. and Breusch, T.S., (1987), *Dynamic Specification, The Long Run and The Estimation of Transformed Regression Models.* C.E.P.R. Discussion Paper No. 154, London.

Young, R.J., (1979), *Quarterly Models Of Milk Supply In England and Wales Including An Application Of Capital Theory.* Ph.D. University of Nottingham.