FORECASTING THE MONTHLY ELECTRICITY CONSUMPTION OF MUNICIPALITIES IN KWA-ZULU NATAL

ALISON NORMA WALTON

B.Sc. (Hons)

Submitted in partial fulfilment of the academic

requirements for the degree of

MASTER OF SCIENCE

in the Department of Statistics and Biometry

University of Natal

Pietermaritzburg 1997

ABSTRACT

Eskom is the major electricity supplier in South Africa and medium term forecasting within the company is a critical activity to ensure that enough electricity is generated to support the country's growth, that the networks can supply the electricity and that the revenue derived from electricity consumption is managed efficiently. This study investigates the most suitable forecasting technique for predicting monthly electricity consumption, one year ahead for four major municipalities within Kwa-Zulu Natal.

PREFACE

The experimental work described in this dissertation was carried out in the Department of Statistics and Biometry, University of Natal, Pietermaritzburg, from January 1994 to March 1997 under the supervision of Professor Linda Haines.

These studies represent original work by the author and have not otherwise been submitted in any form for any degree or diploma to any University. Where use has been made of the work of others it is duly acknowledged in the text.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS

I would like to express my gratitude to my supervisor Professor Linda Haines for her time and tireless guidance, assistance and encouragement in the preparation and presentation of this thesis.

I would also like to thank Eskom for the time and resources made available for this study as well as their financial assistance.

1. GENERAL INTRODUCTION

The aim of this study is to find the most suitable forecasting technique for predicting monthly electricity consumption, one year ahead for the major municipalities within Kwa-Zulu Natal.. The group of customers used in the present study tend to display fairly stable, repetitive electricity consumption patterns which lend themselves to statistical modelling methods. The higher electricity consumption during the winter months is caused by an increase in heating and irrigation applications and the colder the area, the more exaggerated this increase. The three forecasting methods which have been studied in depth in the present study are exponential smoothing, ARIMA and state space modelling. The exponential smoothing method is a simple, well established method, ARIMA modelling requires more skill to apply than exponential smoothing and the application of Kalman filtering techniques to state space models is straight forward, delivering pleasing results. Various options within each method are explored and using the time series of monthly electricity consumption for major municipalities, the results of these methods are analysed and compared.

Chapter 2 introduces the theory and modelling techniques for the exponential smoothing method, ARIMA and state space models and briefly explores the relationships between these three methods. Chapter 3 introduces the time series used in this study and then looks at the application of the above mentioned methods to these series and compares their forecasting accuracy. The conclusions drawn from the study are presented in Chapter 4.

2. THEORY

2.1 INTRODUCTION

This thesis is concerned with time series involving monthly data which exhibit a trend and multiplicative seasonality, i.e. seasonality that is proportional to the level of the series. The theory discussed in the present chapter is therefore related primarily to such series.

A complete time series is denoted by $Y_1,\ldots,Y_t,\ldots,Y_T$ where T represents the length of the series. The forecast of an observation Y_{t+k} at k lags ahead of a time t, given the series $Y_1,\ldots,Y_t$, is denoted by $\hat{Y}_{t-k|t}$, and the one-step ahead forecast error at time t is expressed as $e_t = Y_t - \hat{Y}_{t|t-1}$.

2.2 EXPONENTIAL SMOOTHING

2.2.1 INTRODUCTION

The exponential smoothing method involves the calculation of forecasts based on a weighted average of past observations, with more weight being placed on the recent than on the early observations in the series. The method was introduced by Brown and Holt in the 1950's in the context of constant series and extended to time series with trend and seasonality by Holt and Winters (see Chatfield, 1978; Gardner, 1985; Chatfield and Yar, 1988)

The method of exponential smoothing is well established and widely used (Granger and Newbold, 1977; Chatfield, 1989; Janacek and Swift, 1993). Its main advantages are that it is easy to implement, that the amount of data storage and computation required is minimal and that no complicated procedures involving model identification are necessary. Its chief disadvantage is its very simplicity in that there is no obvious model implied by the method and thus that confidence limits to predictions and forecasts cannot be clearly formulated. Ad hoc procedures for finding such confidence limits have been reported by Chatfield and Yar (1991), but are not well established.

2.2.2 SIMPLE EXPONENTIAL SMOOTHING

Consider a time series $Y_1,\ldots,Y_t$ that does not exhibit trend or seasonality. A sensible one-step-ahead forecast at time t is then given by the weighted average

$$\hat{Y}_{t+1|t} = \alpha Y_t + \alpha(1-\alpha)Y_{t-1} + \alpha(1-\alpha)^2 Y_{t-2} + \ldots\ldots \alpha(1-\alpha)^j Y_{t-j} + \ldots\ldots$$

$$= \alpha Y_t + (1-\alpha)\hat{Y}_{t|t-1}$$

where α is termed the smoothing parameter and lies between 0 and 1, i.e. $0 < \alpha < 1$.

The weights $\alpha(1-\alpha)^j$, j = 0, 1, 2,......, are exponentially decreasing as j increases, hence the term exponential smoothing, and sum to 1, i.e.

$$\alpha + \alpha(1-\alpha) + \alpha(1-\alpha)^2 \ldots\ldots = \sum_{j=0}^{\infty} \alpha(1-\alpha)^j = 1 .$$

For values of α close to 1 most weight is placed on recent observations and for values of α close to 0, more weight on past observations.

In practice, for a given value of α , the one-step-ahead forecast at time t is computed as

$$\hat{Y}_{t+1|t} = \alpha Y_t + (1-\alpha)\hat{Y}_{t|t-1}$$

where the initial value $\hat{Y}_{1|0}$ is unknown and is usually taken to be the first observation, Y_1, or the average of the first few observations. However, the value of α is generally unknown and must therefore be estimated. A sensible, albeit ad hoc approach to its estimation is to choose that value of α to minimise a suitable criterion involving the forecast error, such as the mean sum of squared one-step-ahead errors, written

$$\text{M.S.E.} = \frac{1}{T-m+1}\sum_{t=m}^{T}(Y_t - \hat{Y}_{t|t-1})^2 \qquad \textbf{(2.1)}$$

or the mean absolute percentage error, which does not penalise extreme values as severely as the M.S.E., expressed as

$$\text{M.A.P.E.} = \frac{1}{T-m+1}\sum_{t=m}^{T}\left|\frac{Y_t-\hat{Y}_{t|t-1}}{Y_t}\right|. \qquad \textbf{(2.2)}$$

Note that the first m-1 points are excluded from the calculation of these criteria in order to reduce the effect of the initial value, $\hat{Y}_{1|0}$.

2.2.3 HOLT-WINTERS METHOD

The Holt-Winters method of forecasting takes into account the level, trend and seasonality of a time series and is a generalisation of simple exponential smoothing. There are two such methods, one for additive seasonality and the other for multiplicative seasonality and only the latter is considered here. The level, trend and seasonality of the smoothed series are updated as new observations become available in a manner similar to that of simple exponential smoothing. Specifically for a time t and monthly seasonality, the level is updated according to the equation

$$L_t = \alpha(Y_t/S_{t-12}) + (1-\alpha)(L_{t-1}+T_{t-1}),$$

the trend as

$$T_t = \gamma(L_t - L_{t-1}) + (1-\gamma)T_{t-1},$$

and the seasonal term as

$$S_t = \delta(Y_t/L_t) + (1-\delta)S_{t-12},$$

where α , γ and δ are smoothing parameters for updating the level, trend and seasonal indices respectively, and are restricted to lie between 0 and 1. The closer a parameter is to 1, the more weight that is given to recent data when updating the corresponding level, trend or seasonal terms. These three updating equations are invoked successively to provide, at time t, the one-step-ahead prediction

$$\hat{Y}_{t+1|t} = (L_t + T_t)S_{t-12+1}$$

and the k-steps-ahead prediction

$$\hat{Y}_{t+k|t} = (L_t + kT_t)S_{t-12+k}\ .$$

As with simple exponential smoothing, appropriate initial values L_0, T_0 and S_0 are required and there are a number of options available for calculating these (Chatfield, 1988). For example, data from the first year can be used to provide the estimates

$$L_0 = \frac{\sum_{t=1}^{12} Y_t}{12}, \quad T_0 = 0, \quad \text{and} \quad S_j = \frac{12Y_j}{\sum_{t=1}^{12} Y_t} \qquad j = 1, \ldots\ldots, 12, \qquad \textbf{(2.3)}$$

data for the first two years to provide the values

$$L_0 = \frac{\sum_{t=1}^{24} Y_t}{24}, \quad T_0 = \frac{\sum_{t=13}^{24} Y_t/12 - \sum_{t=1}^{12} Y_t/12}{12}, \quad S_j = \frac{12(Y_j + Y_{j+12})}{\sum_{t=1}^{24} Y_t} \qquad j = 1, \ldots, 12 \qquad \textbf{(2.4)}$$

or all the data can be used to calculate the starting values,

$$L_0 = \frac{\sum_{t=1}^{T} Y_t}{T}, \quad T_0 = \frac{\left(\sum_{t=T-s-1}^{T} Y_t - \sum_{t=1}^{12} Y_t\right)}{12(p-1)}, \quad \text{and} \quad S_j = \frac{12\sum_{t=0}^{p-1} Y_{12t-j}}{\sum_{t=1}^{T} Y_t} \quad j = 1, \ldots, 12 \qquad \textbf{(2.5)}$$

where p is the number of years in the series. The latter approach is used by a number of statistical packages including Statistica, but is clearly not suited to series in which the initial trend is steeply upwards or downwards compared to the average trend for the complete series. For large values of α , γ and δ, or if a series is extremely long, the effect of the starting parameters on the forecast is very small. If, on the other hand, the parameters are small, the starting values will influence the forecast significantly.

The parameters α , γ and δ are also unknown and must be estimated. As for simple exponential smoothing, an empirical approach to selecting parameters , based on minimising the forecast error criteria M.S.E. or M.A.P.E. as given in expressions (2.1) and (2.2), is invoked. For seasonal data, a forecast is often required for the ensuing twelve months and thus it would seem sensible to minimise the error of forecasting over that period (Chatfield and Yar, 1988) using for example the mean sum of squared twelve-steps-ahead error defined by

$$\text{M.S.E. (12)} = \left(\frac{1}{T-12-m+1}\right)\left(\frac{1}{12}\right)\sum_{t=m}^{T}\sum_{j=1}^{12}(Y_{t+j}-\hat{Y}_{t+j|t})^2 \qquad \textbf{(2.6)}$$

2.3 ARIMA MODELS

2.3.1 INTRODUCTION

Autoregressive integrated moving averages (ARIMA) models were developed in 1970 by Box and Jenkins as powerful and flexible tools for modelling time series. The methodology underpinning these models is well established (see for example Vandaele, 1983; Cryer, 1986), and is outlined briefly below.

2.3.2 MODEL OVERVIEW

Consider a time series Y_t, t = 1,...,T, which is weakly stationary, i.e. for which the mean and variance are constant through time. Then an ARMA model comprising p autoregressive and q moving average terms can be represented by

$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \ldots + \phi_p Y_{t-p} + Z_t - \theta_1 Z_{t-1} - \theta_2 Z_{t-2} - \ldots - \theta_q Z_{t-q},$$

where the series Z_t, t = 1,...,T, is a sequence of independent, identically distributed random variables i.e. white noise, and the terms ϕ_i, i = 1,...,p and θ_j, j = 1,...q are autoregressive and moving average parameters respectively. The model can be expressed more succinctly as $\phi(B)Y_t = \theta(B)Z_t$ where B is the backward shift operator defined by $BY_t = Y_{t-1}$ and the roots of the polynomials $\phi(B)$ and $\theta(B)$ are restricted to lie outside the unit circle in order to ensure stationarity and invertibility respectively. Such a model is denoted ARMA (p,q).

A non-stationary time series exhibiting a trend can be transformed into a stationary series by differencing, i.e. by introducing $W_t = \nabla^d Y_t$ where $\nabla = 1 - B$, and the series W_t can then be modelled as an ARMA(p,q) model. Such a model is termed an autoregressive integrated moving average model and is denoted ARIMA(p,d,q). If the variance of a time series is non-stationary, then it is common to transform the series into a stationary one by taking logarithms of the observations.

ARIMA models can be extended quite naturally to incorporate seasonality. In particular, the general multiplicative seasonal ARIMA model is given by

$$\phi_p(B)\Phi_P(B^{12})W_t = \theta_q(B)\Theta_Q(B^{12})Z_t,$$

where $W_t = \nabla^d \nabla_{12}^D Y_t$, D represents the order of the seasonal difference operator and $\nabla_{12} = (1 - B^{12})$. The terms $\Phi_P(B^{12})$ and $\Theta_Q(B^{12})$ are polynomials in the seasonal lags of order P and Q respectively and the roots of these polynomials are again restricted to lie outside the unit circle in order to satisfy stationarity and invertibility requirements respectively. Such a model is termed ARIMA(p,d,q)x(P,D,Q)$_{12}$.

In addition to the autoregressive and moving average parameters, ARIMA models can also include a constant corresponding to the mean of the series when there are no autoregressive parameters in the model and to the intercept otherwise. The constant can be included in the ARIMA model by replacing W_t with $W_t - \delta$.

2.3.3 MODELLING

The Box-Jenkins methodology for ARIMA modelling of a time series consists of three stages,

1. Model identification.
2. Parameter estimation.
3. Diagnostic checking and model validation.

If the model is found to be unacceptable after checking the diagnostics, the procedure is repeated from stage 1.

Identification

The model identification step relies on the autocorrelation and partial autocorrelation functions. The autocorrelation ρ_k is the correlation between observations a given time k apart and is defined by

$$\rho_k = Corr(Y_t, Y_{t+k}) = \frac{Cov(Y_t, Y_{t+k})}{\left[Var(Y_t)Var(Y_{t+k})\right]^{1/2}} \qquad \text{for } k = 0, \pm 1, \pm 2, \ldots\ldots$$

and a graph of the autocorrelations ρ_k against the lag k is termed the autocorrelation function (ACF). In practice, the sample autocorrelation is calculated as

$$r_k = \frac{\sum_{t=1}^{T-k} (Y_t - \bar{Y})(Y_{t+k} - \bar{Y})}{\sum_{t=1}^{T} (Y_t - \bar{Y})^2} \qquad \text{for } k = 0, \pm 1, \pm 2, \ldots\ldots$$

where T is the length of the series. For a white noise series the autocorrelations ρ_k are all zero and in practice, for large T, r_k is approximately normally distributed as N(0, $\frac{1}{T}$), and an approximate 95% confidence interval for an individual r_k is thus given by $\left(-2/\sqrt{T}, 2/\sqrt{T}\right)$. Alternatively, the approximation for the standard error of r_k can be further refined by to $\sqrt{\frac{1}{T}\left(\frac{T-k}{T-2}\right)}$ which is the method used in this study. The partial autocorrelation is the correlation between Y_t and Y_{t+k} after the effect of the intervening variables $Y_{t+1}, \ldots Y_{t+k-1}$ has been removed and a graph of the partial autocorrelation against the lag is known as the partial autocorrelation function (PACF). For a white noise series, approximate 95% confidence intervals for the sample partial autocorrelations are given by $\left(-2/\sqrt{T}, 2/\sqrt{T}\right)$.

For a stationary series the ACF decays rapidly, but in contrast for a series exhibiting trend and therefore requiring differencing, the ACF decays slowly with increasing lag. For a series exhibiting a seasonal trend, and therefore requiring seasonal differencing, the autocorrelations at lags which are multiples of the seasonal periodicity, decay slowly. It is clearly possible to use these observations to difference a given series until the resultant series is stationary. It should be noted, however, that not all series can be transformed to stationarity using differencing and that this is a major shortcoming of the ARIMA models.

The values of p, q, P and Q can be determined from the pattern of the ACF and PACF of the differenced series. Characteristic features of an MA(q) model are an ACF that cuts off at lag q, and a slowly decaying PACF. An AR(p) model has a slowly decaying ACF and a PACF which cuts off after lag p. Seasonal models are more difficult to identify and examples of the ACF and PACF for a range of such models are given in Box and Jenkins (1970, pp 329-333). In particular, it should be emphasised that the sample ACF and PACF are frequently difficult to interpret because they are only estimates of the population ACF and PACF.

Estimation

Once a suitable model has been identified, estimates of the parameters need to be obtained. For this purpose, the assumption that the error terms, Z_t, t = 1,...T, are independently and normally distributed as $N(0,\sigma_z^2)$, is introduced and the parameters are estimated by maximising the likelihood function or equivalently its logarithm

$$-\frac{T}{2}\log 2\pi - \frac{T}{2}\ln \sigma_z^2 - \frac{1}{2}\sum_{t=1}^{T} z_t^2 / \sigma_z^2 .$$

It should be noted that this maximisation is not straight forward (see Box and Jenkins, 1970 pp 269-284). Another efficient option of deriving parameter estimates is to place the ARIMA model in state space form and this will be discussed later. Other methods of obtaining estimates of the parameters, which require less computation, include minimising the conditional or the unconditional least squares functions, but these are rarely used today (Cryer, 1986).

Diagnostics

Various diagnostics are available for checking that the model provides a good fit to the data. In particular, the residuals

$$e_t = Y_t - \hat{Y}_{t|t-1} , \quad t = 1, \dots T$$

should be random and a graph of the residuals against time will highlight any trends or outliers which are not accounted for in the model. In addition, the ACF is a useful tool for examining residuals. In particular, if the residual series is white noise, 95% confidence intervals for the individual sample autocorrelations r_k are given by $\left(-2/\sqrt{T}, 2/\sqrt{T}\right)$. However, it should be noted that when considering k autocorrelations for a white noise series, the probability of concluding that at least one autocorrelation is significantly different from zero at the 5% level, is $1\text{-}0.95^k$. Thus a more satisfactory test for white noise is the portmanteau test of Lung, Box and Pierce which tests the hypothesis that the first k autocorrelations are zero using the test statistic

$$Q^* = T(T+2)\sum_{t=1}^{k} e_t^2 / (T-t)$$

For large T under the null hypothesis of white noise, the statistic Q^* is approximately chi-squared with k-p-q-P-Q degrees of freedom (Cryer, 1986).

Parameters of the model that are not significantly different from zero are identified using tests based on the appropriate t-ratio. By successively excluding parameters for which the absolute t ratio is smallest from the model, an appropriate model can be derived. It should be noted however that a hierarchy is retained in that in an ARIMA(p,d,q) model all AR parameters of order less than or equal to p and all MA parameters of order less than or equal to q are necessarily present in the model.

Very often a number of models may be deemed appropriate and it then becomes necessary to compare these models. Two criteria in particular have been developed for this purpose, namely Akaike's Information Criterion (AIC) and Schwartz's Bayesian Criterion (SBC) These criteria penalise the likelihood function by the number of parameters in the model, thus favouring parsimonious models, and are defined as

AIC = -2 (log likelihood) + 2 (number of parameters)

SBC = -2 (log likelihood) + (number of parameters) x log (number of observations).

In both cases models which minimise these criteria are sought.

One possible systematic approach to model selection is to fit an over parameterised model, for example of the form ARIMA(2,d,2)x(2,D,2)$_{12}$, to the series and to drop parameter estimates not significantly different from zero from the model. This process is repeated for all possible models and the associated AIC and SBC statistics compared. In addition, the ACF and PACF of the stationary series must be examined to ensure that the final model chosen is appropriate.

2.3.4 FORECASTING

The forecast k steps ahead of time T for an ARMA(p, q) model is given, quite simply, by

$$\hat{Y}_{T+k|T}=\phi_1\hat{Y}_{T+k-1|T}+\phi_2\hat{Y}_{T+k-2|T}+\dots\dots\dots+\phi_p\hat{Y}_{T+k-p|T}+Z_{T+k}-\theta_1 Z_{T+k-1}-\theta_2 Z_{T+k-2}-\dots-\theta_q Z_{T+k-q}.$$

In practise the values of $\phi_1,\dots\phi_p$ and $\theta_1,..\theta_q$ are unknown and thus estimates from the modelling process are substituted into the above equation. For t less than T, $\hat{Y}_{t|T}$ is replaced with the actual value at time t and the terms $Z_t, Z_{t-1},\dots$ are replaced with the corresponding residuals $Y_t-\hat{Y}_{t|t-k}$, $Y_{t-1}-\hat{Y}_{t-1|t-k-1},\dots$ respectively. For t greater than T, Z_t is taken to be zero since Z_t, t = 1,...,T, is a white noise process.

Similar considerations apply for an ARIMA(p,d,q)x(P,D,Q)$_{12}$ process. For example, the model ARIMA(1,1,1)x(1,1,1)$_{12}$ written as

$$W_t=\phi_1 W_{t-1}+\Phi_1 W_{t-12}+Z_t-\theta_1 Z_{t-1}-\Theta_1 Z_{t-12}-\phi_1\Phi_1 W_{t-13}+\theta_1\Theta_1 Z_{t-13}$$

or equivalently as

$$\left[(Y_t-Y_{t-1})-(Y_{t-12}-Y_{t-13})\right]=\phi_1\left[(Y_{t-1}-Y_{t-2})-(Y_{t-13}-Y_{t-14})\right]+\Phi_1\left[(Y_{t-12}-Y_{t-13})-(Y_{t-24}-Y_{t-25})\right]$$

$$+Z_t-\theta_1 Z_{t-1}-\Theta_1 Z_{t-12}-\phi_1\Phi_1\left[(Y_{t-13}-Y_{t-14})-(Y_{t-25}-Y_{t-26})\right]+\theta_1\Theta_1 Z_{t-13}$$

can be expressed as

$$Y_t = (1+\phi_1)Y_{t-1} - \phi_1 Y_{t-2} + (1+\Phi_1)(Y_{t-12} - Y_{t-13}) - \phi_1\Phi_1(Y_{t-13} - Y_{t-14} - Y_{t-25} + Y_{t-26})$$
$$- \phi_1(Y_{t-13} - Y_{t-14}) - \Phi_1(Y_{t-24} - Y_{t-25}) + Z_t - \theta_1 Z_{t-1} - \Theta_1 Z_{t-12} + \theta_1\Theta_1 Z_{t-13}.$$

Then the forecast k steps ahead of time T is calculated using the equation

$$\hat{Y}_{T+k|T} = (1+\phi_1)\hat{Y}_{T-1+k|T} - \phi_1 \hat{Y}_{T-2+k|T} + (1+\Phi_1)(\hat{Y}_{T-12+k|T} - \hat{Y}_{T-13+k|T})$$
$$- \phi_1\Phi_1(\hat{Y}_{T-13+k|T} - \hat{Y}_{T-14+k|T} - \hat{Y}_{T-25+k|T} + \hat{Y}_{T-26+k|T}) - \phi_1(\hat{Y}_{T-13+k|T} - \hat{Y}_{T-14+k|T})$$
$$- \Phi_1(\hat{Y}_{T-24+k|T} - \hat{Y}_{T-25+k|T}) + Z_{T+k} - \theta_1 Z_{T-1+k} - \Theta_1 Z_{T-12+k} + \theta_1\Theta_1 Z_{T-13+k}$$

where, for t less than T, $\hat{Y}_{t|T}$ is replaced with the actual value at time t, the $Z_t, Z_{t-1}, \ldots$ are replaced with the residuals $Y_t - \hat{Y}_{t|t-k}$, $Y_{t-1} - \hat{Y}_{t-1|t-k-1}, \ldots$ and for t greater than T, Z_t is taken to be zero.

Prediction limits for the forecast $\hat{Y}_{T+k|T}$ are approximated by

$$\hat{Y}_{T+k|T} \pm z_{(1-\frac{\alpha}{2})}\sqrt{Var(e_{T+k})}$$

where 1 - α is the required confidence level and $z_{(1-\frac{\alpha}{2})}$ is the critical value (Vandaele, 1983).

2.3.5 INTERVENTION ANALYSIS

There are often factors which cause a sudden change in the structure of a time series and intervention analysis allows these changes to be incorporated into a forecasting model.

There are various types of intervention that can occur in a time series, but only two are considered in this study.

(i) A single event intervention at time t_I is modelled by a pulse indicator as

$$I_t = \begin{cases} 0 & for \quad t \neq t_I \\ 1 & for \quad t = t_I \end{cases}$$

(ii) An intervention at time t_I which results in a permanent change in the level of the time series is modelled by a step indicator of the form

$$I_t = \begin{cases} 0 & for \quad t < t_I \\ 1 & for \quad t \geq t_I \end{cases}.$$

The intervention events frequently alter the ACF and PACF, making it difficult to identify the underlying ARIMA model. Thus for a stationary, non-seasonal time series, the model $\phi(B)Y_t = \theta(B)Z_t$ which can be expressed as $Y_t = \dfrac{\theta(B)}{\phi(B)}Z_t$, is initially identified using the time series prior to the intervention. Thereafter the model $Y_t = \lambda I_t + \dfrac{\theta(B)}{\phi(B)}Z_t$, where λ is a constant and the indicator I_t represents the intervention event, is fitted to the complete series (Deadman and Pyle, 1989).

2.4 STATE SPACE MODELS

2.4.1 INTRODUCTION

State space models were originally introduced by Kalman in 1960, and used by control engineers in aerospace related applications. They were adapted with great success in 1976 by Harrison and Stevens (1976) to model time series. An excellent introduction to the topic is given by Janaceck and Swift (1993), while Harvey (1989) provides a more in-depth analysis.

Once a problem is formulated in state space form, the Kalman filter can be invoked to derive optimal estimates of the current state of the system and to calculate forecasts. A further refinement of this approach is the calculation of maximum likelihood estimates of the unknown parameters either by direct maximisation or by using the EM algorithm. With a minor adjustment, intervention events can be incorporated into the Kalman filtering process.

2.4.2 THE STATE SPACE FORM

The state space model is defined using two equations known as the observation and the state equations. The observation equations relate the observed univariate time series Y_t to an unknown d-dimensional vector α_t, termed the state vector, as

$$Y_t = h_t^T \alpha_t + \varepsilon_t \qquad t = 1,\dots,T$$

where h_t is a given d-dimensional vector and the error terms ε_t are independent and satisfy $\varepsilon_t \sim N(0, \sigma_\varepsilon^2)$. The state equations in turn relate the unknown state vector α_t to its previous values according to

$$\alpha_t = \Phi_t \alpha_{t-1} + \eta_t \qquad t = 1,\dots,T$$

where Φ_t is a d x d transition matrix and the d-dimensional error vectors, η_t, are independent and satisfy $\eta_t \sim N(0, \Sigma)$. The two error terms ε_t and η_t are assumed to be independent and in order to initiate the model, it is usual to take $\alpha_0 \sim N(\mu, C_0)$, for specific values of μ and C_0. In the present study, the terms h_t and Φ_t in the observation and state equations are assumed to be time invariant and are thus referred to as h and Φ respectively.

2.4.3 THE KALMAN FILTER

Once a time series model has been formulated in state space form, the Kalman filter provides a method for calculating the minimum mean square estimate of α_t, and hence an estimate for $\hat{Y}_{t|t-1}$, where the parameters σ_ε^2, Σ, μ and C_0 are taken as known . This can be done either by filtering, where the parameters are estimated using only the observations available up to the time point t, or by smoothing recursions using the complete set of observations in the estimation process.

Filtering

An outline of the derivation of the Kalman filter is presented here following Meinhold and Singpurwalla (1983). Let

$$\hat{\alpha}_{t|s} = E(\alpha_t|Y_1,\ldots.Y_s)$$

$$\hat{\alpha}_t = E(\alpha_t|Y_1,\ldots.Y_t)$$

$$\hat{Y}_{t|t-1} = E(Y_t|Y_1,\ldots Y_{t-1})$$

$$C_{t|t-1} = E\{(\alpha_t-\hat{\alpha}_{t|t-1})(\alpha_t-\hat{\alpha}_{t|t-1})^T|Y_1,\ldots\ldots Y_{t-1}\}$$

$$C_t = E\{(\alpha_t-\hat{\alpha}_t)(\alpha_t-\hat{\alpha}_t)^T|Y_1,\ldots\ldots Y_t\}$$

and $\quad e_t = Y_t - \hat{Y}_{t|t-1}$.

The Kalman filter prediction equations prior to observing Y_t, are given by

$$E(\alpha_t|Y_1,\ldots Y_{t-1}) = \alpha_{t|t-1} = \Phi\hat{\alpha}_{t-1}$$

and $\quad Var(\alpha_t|Y_1,..Y_{t-1}) = C_{t|t-1} = \Phi \bar{C}_{t-1}\Phi^T + \Sigma$.

Once the observation Y_t becomes available the Kalman filter updating equations can be applied. To derive these, the following well known result from multivariate statistics is used (see Anderson, 1958, pp. 28-29).

Result : Let X_1 and X_2 have a bivariate normal distribution such that

$$\begin{bmatrix} X_1 \\ X_2 \end{bmatrix} \sim N\left(\begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}, \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}\right) \tag{2.7.1}$$

Then $X_1|X_2=x_2 \sim N(\mu_1+\Sigma_{12}\Sigma_{22}^{-1}(x_2-\mu_2), \Sigma_{11}-\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21})$ (2.7.2)

and $X_2|X_1=x_1 \sim N(\mu_2+\Sigma_{21}\Sigma_{11}^{-1}(x_1-\mu_1), \Sigma_{22}-\Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12})$ (2.7.3)

Thus for $\quad e_t = Y_t - \acute{Y}_{t|t-1} = Y_t - h^T\Phi\hat{\alpha}_{t-1}$

$$= h^T\alpha_t + \varepsilon_t - h^T\Phi\hat{\alpha}_{t-1}$$

it follows that $\quad e_t|\alpha_t, Y_1,..Y_{t-1} \sim N(h^T(\alpha_t - \Phi\hat{\alpha}_{t-1}), \sigma_\varepsilon^2)$,

and hence from (2.7.1) and (2.7.3) that

$$\begin{bmatrix} \alpha_t \\ e_t \end{bmatrix} |Y_1,\ldots\ldots Y_{t-1} \sim N\left(\begin{bmatrix} \hat{\alpha}_{t|t-1} \\ 0 \end{bmatrix}, \begin{bmatrix} C_{t|t-1} & C_{t|t-1}h \\ h^T C_{t|t-1} & \sigma_\varepsilon^2 + h^T C_{t|t-1}h \end{bmatrix}\right)$$

It further follows from (2.7.1) and (2.7.2) that

$$\alpha_t|e_t, Y_1,..Y_t = \alpha_t|Y_1,..Y_t \sim N(\hat{\alpha}_{t|t-1} + C_{t|t-1}h\frac{e_t}{f_t}, C_{t|t-1} - \frac{C_{t|t-1}hh^T C_{t|t-1}}{f_t})$$

where $f_t = \sigma_\varepsilon^2 + h^T C_{t|t-1}h$ is the error variance.

Thus the Kalman filter method can be summarised as follows.

The prediction equations :	$\hat{\alpha}_{t\|t-1} = \Phi\hat{\alpha}_{t-1}$
	$C_{t\|t-1} = \Phi C_{t-1}\Phi^T + \Sigma$
The updating equations :	$f_t = h^T C_{t\|t-1} h + \sigma_\varepsilon^2$
	$C_t = C_{t\|t-1} - C_{t\|t-1} hh^T C_{t\|t-1} / f_t$
	$\hat{\alpha}_t = \hat{\alpha}_{t\|t-1} + C_{t\|t-1} h(Y_t - h^T\hat{\alpha}_{t\|t-1}) / f_t$
The one step ahead error :	$e_t = Y_t - h^T\hat{\alpha}_{t\|t-1}$

BOX 2.4.1 : Kalman filter equations

Kalman smoothing

The Kalman smoothing or backward recursions extend the Kalman filtering procedure by making use of all the data available at time T to estimate the state vector α_t. After the forward recursions given in Box 2.4.1 are calculated, the backward recursion equations given in Box 2.4.2 below are applied. (Shumway and Stoffer, 1982)

Starting with $C_{T,T-1|T} = (I - C_{T|T-1}\dfrac{hh^T}{f_t})\Phi C_{T-1}$

where $C_{t,t-1|T} = E[(\alpha_t - \hat{\alpha}_{t|t-1})(\alpha_{t-1} - \hat{\alpha}_{t-1|t-2})^T | Y_1, \ldots. Y_T]$

$$= C_t C^{*T}{}_{t-1} + C_t^*(C_{t+1,t|T} - \Phi C_t)C_{t-1}^{*T}$$

with $C_t^* = C_t \Phi^T C_{t+1|t}^{-1}$,

calculate for t = T-1,...1,

$$\hat{\alpha}_{t|T} = \hat{\alpha}_t + C_t^*(\hat{\alpha}_{t+1|T} - \Phi\hat{\alpha}_t)$$

$$C_{t|T} = C_t + C_t^*(C_{t+1|T} - C_{t+1|t})C_t^{*T}.$$

BOX 2.4.2 : Kalman backward recursions

The Kalman filtering and smoothing recursions clearly require starting estimates μ and C_0. Janacek and Swift (1993) recommend taking μ to be 0 and assume little is known about the initial variance by taking $C_0 = MI$, for some large number M and the identity matrix I. The model parameters h and Φ are assumed to be known and thus do not need to be estimated.

2.4.4 MAXIMUM LIKELIHOOD

In implementing the Kalman filter process, suitable values for the unknown parameters Σ and σ_ε^2 need to be set, but this is rather subjective. A more satisfactory approach is to estimate the parameters by the method of maximum likelihood. There are two methods of obtaining such estimates, the one involving direct maximisation of the likelihood and the other the EM algorithm. The parameters μ and C_0 are usually fixed as discussed previously, although it is possible to obtain a maximum likelihood estimate of μ by incorporating μ into the likelihood function as an additional parameter and maximising the likelihood directly.

Direct maximisation

The likelihood can be expressed as the product of the conditional probability density functions of Y_t given $Y_1, \ldots Y_{t-1}$ as

$$L(\sigma_\varepsilon^2, \Sigma | Y_1, Y_2, \ldots Y_{t-1}) = \prod_{t=1}^{T} f(Y_t | Y_1, Y_2, \ldots Y_{t-1}).$$

Thus,
$$\ln L(\sigma_\varepsilon^2, \Sigma | Y_1, Y_2, \ldots Y_{t-1}) = \sum_{t=1}^{T} \ln f(Y_t | Y_1, Y_2, \ldots Y_{t-1})$$

and since $Y_t | Y_1, \ldots\ldots Y_{t-1} \sim N(\hat{Y}_{t|t-1}, f_t)$

$$\ln f(Y_t | Y_1, \ldots\ldots Y_{t-1}) = -\frac{1}{2}\ln 2\pi - \frac{1}{2}\ln f_t - \frac{1}{2f_t}(Y_t - \hat{Y}_{t|t-1})^2,$$

and
$$\ln L(\sigma_\varepsilon^2, \Sigma | Y_1, \ldots Y_{t-1}) = -\frac{T}{2}\ln 2\pi - \frac{1}{2}\sum_{t=1}^{T}\ln f_t - \frac{1}{2}\sum_{t=1}^{T}(Y_t - \hat{Y}_{t|t-1})^2 / f_t,$$

where the values of f_t and $\hat{Y}_{t|t-1}$ for t = 1, ...,T are calculated using the Kalman filter. The effect of the starting parameters μ and C_0 can be reduced by ignoring the first few iterations of the Kalman filter in the calculation of the log likelihood function. Thus the function

$$\ln L(\sigma_\varepsilon^2, \Sigma | Y_1, \ldots Y_{t-1}) = -\frac{T-d}{2}\ln 2\pi - \frac{1}{2}\sum_{t=d+1}^{T}\ln f_t - \frac{1}{2}\sum_{t=d+1}^{T}(Y_t - \hat{Y}_{t|t-1})^2 / f_t$$

where d is the number of initial iterations ignored, can be maximised with respect to σ_ε^2 and the elements of Σ using a non-linear optimisation routine.

The covariance matrices, namely C_t, $C_{t|t-1}$ and the error variance f_t, often converge quickly to fixed, steady state values. In such cases, the speed of the Kalman filtering routine can be improved by using the steady state values of these covariance matrices. The efficiency of the routine, when maximising the likelihood function directly, can be improved further by concentrating out a parameter. This only applies to structural models which are introduced later in this chapter and the approach will be discussed there.

The EM algorithm

Shumway and Stoffer (1982) developed an alternative method of maximising the likelihood function by invoking the EM algorithm. The algorithm applies forward and backward Kalman filter recursions on the data successively until the change in the likelihood function is small. EM is an acronym for Expectation-Maximisation and describes the procedure of first calculating the expected values of a complete data likelihood function conditional on the observed data and then maximising that function.

The complete likelihood function of $\alpha_0, \alpha_1, \ldots \alpha_T$, $Y_1, \ldots Y_T$ is given by

$$L(\sigma_\varepsilon^2, \Sigma | Y_1, \ldots Y_T) = \prod_{t=1}^{T} f(Y_t | \alpha_t) f(\alpha_t | \alpha_{t-1}) f(\alpha_0)$$

$$= \prod_{t=1}^{T} f(\alpha_t|\alpha_{t-1}) \prod_{t=1}^{T} f(Y_t|\alpha_t) f(\alpha_0).$$

where $\alpha_0, \alpha_1, \ldots \alpha_T$ are regarded as unobserved or missing values and $Y_1, \ldots Y_T$ are observed values. It follows from the observation equation that $Y_t|\alpha_t \sim N(h^T \alpha_t, \sigma_\varepsilon^2)$ and from the state equation that $\alpha_t|\alpha_{t-1} \sim N(\Phi\alpha_{t-1}, \Sigma)$, and $\alpha_0 \sim N(\mu, C_0)$ where μ and C_0 are held constant. Thus the probability distribution functions embedded in the likelihood function can be written as follows :

$$f(\alpha_0) = \frac{1}{(2\pi)^{d/2} |C_0|^{1/2}} \exp[-\frac{1}{2}(\alpha_0 - \mu)^T C_0^{-1} (\alpha_0 - \mu)]$$

$$f(\alpha_t|\alpha_{t-1}) = \frac{1}{(2\pi)^{d/2} |\Sigma|^{1/2}} \exp[-\frac{1}{2}(\alpha_t - \Phi\alpha_{t-1})^T \Sigma^{-1} (\alpha_t - \Phi\alpha_{t-1})]$$

$$f(Y_t|\alpha_t) = \frac{1}{(2\pi\sigma_\varepsilon^2)^{\frac{1}{2}}} \exp[-\frac{1}{2\sigma_\varepsilon^2}(Y_t - h^T\alpha_t)^2],$$

and the log likelihood function, with constants omitted from the equation, is given by

$$\ln L_C(\sigma_\varepsilon^2, \Sigma | Y_1, \ldots Y_T) = -\frac{1}{2}\ln|C_0| - \frac{1}{2}(\alpha_0 - \mu)^T C_0^{-1} (\alpha_0 - \mu)$$

$$-\frac{T}{2}\ln|\Sigma| - \frac{1}{2}\sum_{t=1}^{T}(\alpha_t - \Phi\alpha_{t-1})^T \Sigma^{-1} (\alpha_t - \Phi\alpha_{t-1})$$

$$-\frac{T}{2}\ln\sigma_\varepsilon^2 - \frac{1}{2\sigma_\varepsilon^2}\sum_{t=1}^{T}(Y_t - h^T\alpha_t)^2 .$$

The terms $\alpha_0, \alpha_1, \ldots \alpha_T$ are unobserved and thus taking expectations of the above expression with respect to the $\alpha_0, \alpha_1, \ldots \alpha_T$ conditional on the values $Y_1, \ldots Y_T$ and using the results of Appendix A.1, gives

$$E[\ln L_C(\sigma_\varepsilon^2, \Sigma | Y_1, \ldots Y_T)] = -\frac{1}{2}\log|C_0| - \frac{1}{2} tr\{C_0^{-1}(C_{0|T} + (\hat{\alpha}_{0|T} - \mu)(\hat{\alpha}_{0|T} - \mu)^T\}$$

$$-\frac{T}{2}\ln|\Sigma| - \frac{1}{2} tr\{\Sigma^{-1}[\sum_{t=1}^{T}(C_{t|T} + \hat{\alpha}_{t|T}\hat{\alpha}_{t|T}^T) - \sum_{t=1}^{T}(C_{t,t-1|T} + \hat{\alpha}_{t|T}\hat{\alpha}_{t-1|T}^T)\Phi^T]\}$$

$$-\frac{1}{2} tr\{\Sigma^{-1}[\Phi\sum_{t=1}^{T}(C_{t,t-1|T} + \hat{\alpha}_{t|T}\hat{\alpha}_{t-1|T}^T)^T + \Phi\sum_{t=1}^{T}(C_{t-1|T} + \hat{\alpha}_{t-1|T}\hat{\alpha}_{t-1|T}^T)\Phi^T]\}$$

$$-\frac{T}{2}\ln\sigma_\varepsilon^2 - \frac{1}{2\sigma_\varepsilon^2}\sum_{t=1}^{T}[(Y_t - h^T \hat{\alpha}_{t|T})^2 + h^T C_{t|T} h].$$

The function $E[\ln L_C(\sigma_\varepsilon^2, \Sigma | Y_1, \ldots Y_T)]$ is maximised by setting the derivatives with respect to σ_ε^2 and Σ to zero, letting $\mu = \hat{\alpha}_{0|T}$ and solving for σ_ε^2 and Σ. The resultant estimates are given below and more details are provided in Appendix A.2.

$$\hat{\Sigma} = T^{-1}[\sum_{t=1}^{T}(C_{t|T} + \hat{\alpha}_{t|T}\hat{\alpha}_{t|T}^T) - \sum_{t=1}^{T}(C_{t,t-1|T} + \hat{\alpha}_{t|T}\hat{\alpha}_{t-1|T}^T)\Phi^T - \Phi\sum_{t=1}^{T}(C_{t,t-1|T} + \hat{\alpha}_{t|T}\hat{\alpha}_{t-1|T}^T)^T$$

$$+ \Phi\sum_{t=1}^{T}(C_{t-1,T} + \hat{\alpha}_{t-1|T}\hat{\alpha}_{t-1|T}^T)\Phi^T]$$

and $\hat{\sigma}_\varepsilon^2 = T^{-1}\sum_{t=1}^{T}[(Y_t - h^T \hat{\alpha}_{t|T})^2 + h^T C_{t|T} h]$.

Box 2.4.3 : Optimal estimates for Σ and σ_ε^2

Note that Kalman smoothing results are used in the estimation of the above parameters and that the standard error estimates for σ_ε^2 and Σ can be calculated using various methods such as the Louis Method (Tanner, 1993). Overall therefore the EM algorithm can be summarised in the following steps:

1. Adopt sensible initial values for σ_ε^2 and Σ.
2. Use the Kalman filter recursions given in Box 2.4.1, for t = 1, ...T, and then use the backward recursions given in Box 2.4.2 for t = T, T-1, ...1 to calculate the log likelihood as

$$\ln L(\sigma_\varepsilon^2, \Sigma | Y_1, \ldots Y_T) = -\frac{T-d}{2}\ln 2\pi - \frac{1}{2}\sum_{t=d+1}^{T}\ln f_t - \frac{1}{2}\sum_{t=d+1}^{T}(Y_t - \hat{Y}_{t|t-1})^2 / f_t.$$

3. Calculate estimates for σ_ε^2 and Σ as in Box 2.4.3.
4. Repeat steps 2 and 3 until satisfactory convergence of the algorithm is attained.

BOX 2.4.4 : EM algorithm

The main advantages of using the EM algorithm as opposed to an optimising routine are that derivatives need not be calculated and the likelihood function is guaranteed to increase with every iteration of the algorithm. However, the EM algorithm is notoriously slow to converge (Shumway and Stoffer, 1982). One possible approach is to use the EM algorithm to estimate starting values for the unknown parameters and then to refine these estimates using a discrete optimisation routine.

2.4.5 FORECASTING

The one-step-ahead forecast is calculated by direct substitution in the observation equation as

$$\hat{Y}_{T+1|T} = h^T \hat{\alpha}_{T+1} = h^T \Phi \alpha_T ,$$

and by repeated substitution, the forecast k steps ahead of time T is given by

$$\hat{Y}_{T+k|T} = h^T \Phi^k \hat{\alpha}_T .$$

Confidence limits for these forecasts are derived using the one-step-ahead prediction error variance,

$$\text{Var}(f_t \mid Y_1, \ldots Y_{t-1}) = E[(Y_t - \hat{Y}_{t|t-1})^2]$$

$$= E[(h^T(\alpha_t - \hat{\alpha}_{t|t-1}) + \varepsilon_t)^2]$$

$$= h^T C_{t|t-1} h + \sigma_\varepsilon^2 \text{ since } \varepsilon_t \text{ is independent of } h^T(\alpha_t - \hat{\alpha}_{t|t-1}) .$$

Thus $\text{Var}(f_{T+1} \mid Y_1, \ldots Y_T) = h^T C_{T+1|T} h + \sigma_\varepsilon^2$

$$= h^T(\Phi C_T \Phi^T + \Sigma) h + \sigma_\varepsilon^2$$

and more generally

$$\text{Var}\ (f_{T+k} \mid Y_1, \ldots Y_T) = h^T(\Phi^k C_T (\Phi^k)^T + \sum_{i=1}^{k} \Phi^{k-i} \Sigma (\Phi^{k-i})^T) h + \sigma_\varepsilon^2 .$$

The $100(1-\alpha)\%$ confidence limits for the forecast $\hat{Y}_{T+k|T}$ are thus approximated by

$$\hat{Y}_{T+k|T} \pm z_{(1-\frac{\alpha}{2})} \sqrt{h^T[\Phi^k C_T(\Phi^k)^T + \sum_{i=1}^{k} \Phi^{k-i}\Sigma(\Phi^{k-i})^T]h + \sigma_\varepsilon^2}$$

where $z_{(1-\frac{\alpha}{2})}$ is the critical value for the N(0,1) distribution.

2.4.6 INTERVENTION ANALYSIS

The state space form can easily be adapted to model intervention events by including appropriate indicator terms in the model. In particular, as for ARIMA models, a single event intervention j at time t_I for j = 1,...,J where J is the number of intervention events, is modelled by a pulse indicator as

$$I_{t,j} = \begin{cases} 0 & for \quad t \neq t_I \\ 1 & for \quad t = t_I \end{cases}$$

and an intervention at time t_I which results in a permanent change in the level of the time series is modelled by a step indicator of the form

$$I_{t,j} = \begin{cases} 0 & for \quad t < t_I \\ 1 & for \quad t \geq t_I \end{cases}.$$

The observation equation is now written as

$$Y_t = h^T \alpha_t + \varepsilon_t + \sum_{j=1}^{J} I_{t,j}\lambda_j$$

where λ_j is a constant associated with the indicator variable.

Letting $\alpha_t^* = \begin{pmatrix} \alpha_t \\ \lambda_1 \\ \cdot \\ \cdot \\ \lambda_J \end{pmatrix}$ and $h^* = \begin{pmatrix} h \\ I_{t,1} \\ \cdot \\ \cdot \\ I_{t,J} \end{pmatrix}$ for t = 1,...,T, the observation equation becomes

$$Y_t = h^{*T}\alpha_t^* + \varepsilon_t \tag{2.8}$$

and the state equation can be written as

$$\alpha_t^* = \begin{pmatrix}\alpha_t \\ \lambda\end{pmatrix} = \begin{pmatrix}\Phi & 0 \\ 0 & \mathrm{I}\end{pmatrix}\begin{pmatrix}\alpha_{t-1} \\ \lambda\end{pmatrix} + \begin{pmatrix}\eta_t \\ 0\end{pmatrix} \tag{2.9}$$

where $\lambda = \begin{pmatrix}\lambda_1 \\ \cdot \\ \cdot \\ \lambda_J\end{pmatrix}$, I is the identity matrix and $\mathrm{var}\begin{pmatrix}\eta_t \\ 0\end{pmatrix} = \begin{pmatrix}\Sigma & 0 \\ 0 & 0\end{pmatrix}$. The equations (2.8) and (2.9) describe a state space model which can be fitted to the data as described in the previous section using Kalman filtering and maximum likelihood estimates for the parameters σ_ε^2, Σ and μ^*.

2.4.7 STRUCTURAL MODELS

Structural models constitute a specific class of state space models in which the observations are modelled as the sum of separate components such as trend and seasonality. Some examples of structural models relevant to the present study are given below.

Random walk plus noise

This model, also known as the steady state model, is one of the simplest state space models. The observation equation is given by

$$Y_t = \alpha_t + \varepsilon_t$$

where α_t follows a random walk and $\varepsilon_t \sim N(0, \sigma_\varepsilon^2)$. Thus the state equation is given by

$$\alpha_t = \alpha_{t-1} + \eta_t ,$$

where $\eta_t \sim N(0, \sigma_\eta^2)$. Note that in this case the terms h, Φ and Σ in the observation and state equations are 1, 1 and σ_η^2 respectively.

Local linear trend model

This model is described by the observation equation

$$Y_t = \mu_t + \varepsilon_t$$

together with the state equations

$$\mu_t = \mu_{t-1} + \beta_{t-1} + \eta_t$$

$$\beta_t = \beta_{t-1} + \varsigma_t$$

where β_t represents the slope at time t, $\varepsilon_t \sim N(0,\sigma_\varepsilon^2)$, $\eta_t \sim N(0,\sigma_\eta^2)$ and $\varsigma_t \sim N(0,\sigma_\varsigma^2)$.

These equations can be expressed more succinctly in state space form as

$$Y_t = (1 \quad 0)\begin{pmatrix} \mu_t \\ \beta_t \end{pmatrix} + \varepsilon_t$$

$$\alpha_t = \begin{pmatrix} \mu_t \\ \beta_t \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}\begin{pmatrix} \mu_{t-1} \\ \beta_{t-1} \end{pmatrix} + \begin{pmatrix} \eta_t \\ \varsigma_t \end{pmatrix}.$$

Basic structural models

These models are examples of structural models which contain trend, seasonal and irregular components and are thus appropriate for the monthly time series used in the present study. The basic structural model (BSM) can be represented by the set of equations

$$Y_t = \mu_t + \gamma_t + \varepsilon_t$$

$$\mu_t = \mu_{t-1} + \beta_{t-1} + \eta_t$$

$$\beta_t = \beta_{t-1} + \varsigma_t$$

$$\gamma_t = -\sum_{j=1}^{s-1} \gamma_{t-j} + \omega_t$$

where μ_t is the local linear trend, β_t is the slope, γ_t is the seasonal component and the terms ε_t, η_t, ς_t and ω_t are mutually uncorrelated, irregular components such that $\varepsilon_t \sim N(0,\sigma_\varepsilon^2)$, $\eta_t \sim N(0,\sigma_\eta^2)$, $\varsigma_t \sim N(0,\sigma_\varsigma^2)$ and $\omega_t \sim N(0,\sigma_\omega^2)$. The random terms η_t, ς_t and ω_t allow μ_t, β_t and γ_t respectively to evolve over time. Note that for the seasonal components, $E(\sum_{j=0}^{s-1}\gamma_{t-j})=0$, where s is the number of seasons. Thus a monthly time series model can be expressed in state space form as follows :

$$Y_t = (1\ \ 0\ \ 1\ \ 0\ \ 0\ \ 0\ \ 0\ \ 0\ \ 0\ \ 0\ \ 0\ \ 0\ \ 0)\begin{pmatrix}\mu_t\\ \beta_t\\ \gamma_t\\ \gamma_{t-1}\\ \gamma_{t-2}\\ \gamma_{t-3}\\ \gamma_{t-4}\\ \gamma_{t-5}\\ \gamma_{t-6}\\ \gamma_{t-7}\\ \gamma_{t-8}\\ \gamma_{t-9}\\ \gamma_{t-10}\end{pmatrix}+\varepsilon_t$$

$$\alpha_t=\begin{pmatrix}\mu_t\\ \beta_t\\ \gamma_t\\ \gamma_{t-1}\\ \gamma_{t-2}\\ \gamma_{t-3}\\ \gamma_{t-4}\\ \gamma_{t-5}\\ \gamma_{t-6}\\ \gamma_{t-7}\\ \gamma_{t-8}\\ \gamma_{t-9}\\ \gamma_{t-10}\end{pmatrix}=\begin{pmatrix}1&1&0&0&0&0&0&0&0&0&0&0&0\\ 0&1&0&0&0&0&0&0&0&0&0&0&0\\ 0&0&-1&-1&-1&-1&-1&-1&-1&-1&-1&-1&-1\\ 0&0&1&0&0&0&0&0&0&0&0&0&0\\ 0&0&0&1&0&0&0&0&0&0&0&0&0\\ 0&0&0&0&1&0&0&0&0&0&0&0&0\\ 0&0&0&0&0&1&0&0&0&0&0&0&0\\ 0&0&0&0&0&0&1&0&0&0&0&0&0\\ 0&0&0&0&0&0&0&1&0&0&0&0&0\\ 0&0&0&0&0&0&0&0&1&0&0&0&0\\ 0&0&0&0&0&0&0&0&0&1&0&0&0\\ 0&0&0&0&0&0&0&0&0&0&1&0&0\\ 0&0&0&0&0&0&0&0&0&0&0&1&0\end{pmatrix}\begin{pmatrix}\mu_{t-1}\\ \beta_{t-1}\\ \gamma_{t-1}\\ \gamma_{t-2}\\ \gamma_{t-3}\\ \gamma_{t-4}\\ \gamma_{t-5}\\ \gamma_{t-6}\\ \gamma_{t-7}\\ \gamma_{t-8}\\ \gamma_{t-9}\\ \gamma_{t-10}\\ \gamma_{t-11}\end{pmatrix}+\begin{pmatrix}\eta_t\\ \varsigma_t\\ \omega_t\\ 0\\ 0\\ 0\\ 0\\ 0\\ 0\\ 0\\ 0\\ 0\\ 0\end{pmatrix}$$

The seasonal component of the BSM can also be modelled using trigonometric terms in the model. The seasonal effect at time t is given by $\gamma_t=\sum_{j=1}^{\lfloor \frac{s}{2}\rfloor}\gamma_{jt}$ where

$$\begin{pmatrix} \gamma_{jt} \\ \gamma^*_{jt} \end{pmatrix} = \begin{pmatrix} \cos\frac{2\pi j}{s} & \sin\frac{2\pi j}{s} \\ -\sin\frac{2\pi j}{s} & \cos\frac{2\pi j}{s} \end{pmatrix} \begin{pmatrix} \gamma_{jt-1} \\ \gamma^*_{jt-1} \end{pmatrix} + \begin{pmatrix} \omega_{jt} \\ \omega^*_{jt} \end{pmatrix}$$

for j = 1, ..., $\left\lfloor \frac{s}{2} \right\rfloor$ where γ^*_{jt} is introduced as an artefact to generate γ_{jt} and $\left\lfloor \frac{s}{2} \right\rfloor$ denotes defined as the integer part of $\frac{s}{2}$. The white noise disturbances ω_{jt} and ω^*_{jt} allow the seasonality to evolve over time and are assumed to be uncorrelated and to follow a normal distribution. If s is even, then the sine term with j = $\frac{s}{2}$ is zero, and thus the number of trigonometric parameters is s - 1.

Because the BSM with trigonometric terms for monthly data, i.e. for s=12, is very cumbersome to write out in full, the model for quarterly data represented in state space form is given below. Thus

$$Y_t = (1 \quad 0 \quad 1 \quad 0 \quad 1) \begin{pmatrix} \mu_t \\ \beta_t \\ \gamma_{1t} \\ \gamma^*_{1t} \\ \gamma_{2t} \end{pmatrix} + \varepsilon_t$$

and $$\alpha_t = \begin{pmatrix} \mu_t \\ \beta_t \\ \gamma_{1t} \\ \gamma^*_{1t} \\ \gamma_{2t} \end{pmatrix} = \begin{pmatrix} 1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & \cos\left(\frac{\pi}{2}\right) & \sin\left(\frac{\pi}{2}\right) & 0 \\ 0 & 0 & -\sin\left(\frac{\pi}{2}\right) & \cos\left(\frac{\pi}{2}\right) & 0 \\ 0 & 0 & 0 & 0 & -1 \end{pmatrix} \begin{pmatrix} \mu_{t-1} \\ \beta_{t-1} \\ \gamma_{1t-1} \\ \gamma^*_{1t-1} \\ \gamma_{2t-1} \end{pmatrix} + \begin{pmatrix} \eta_t \\ \varsigma_t \\ \omega_{1t} \\ \omega^*_{1t} \\ \omega_{2t} \end{pmatrix}.$$

Concentrating out a parameter

The computation of the parameter estimates by maximising the likelihood directly can be made more computationally efficient when applied to the structural model, and the BSM in particular, by "concentrating" out a parameter, resulting in one less parameter being estimated. This is done by selecting one of the noise variances as a scaling variance, for example take $\sigma_{\varepsilon}^{2} = \sigma^{*2}$. The optimal estimate of σ^{*2} is derived by differentiating the likelihood function with respect to σ^{*2} and setting the result equal to zero to give

$$\sigma^{*2} = \frac{1}{T-d} \sum_{t=d+1}^{T} e_t^{2} / f_t .$$

Substituting this result back into the likelihood function results in

$$\ln L_C^{*}(\sigma_{\varepsilon}^{2}, \Sigma | Y_1, \ldots, Y_T) = -\frac{1}{2} \sum_{t=d+1}^{T} \ln f_t - \frac{T-d}{2} \ln \sigma^{*2},$$

which is known as the concentrated likelihood function. This function is then maximised with respect to the unknown parameters σ_{η}^{2}, σ_{ς}^{2} and σ_{ω}^{2}, using the Kalman filtering equations as before, but scaling σ_{η}^{2}, σ_{ς}^{2}, σ_{ω}^{2}, C_t, $C_{t|t-1}$ and f_t by σ^{*2} and fixing the scaling variance to 1 (Janaceck and Swift, 1993; Jones, 1993).

2.5 RELATIONSHIPS BETWEEN METHODS

There are various cases in which exponential smoothing and ARIMA models and ARIMA and state space models are found to be equivalent. Examples of such cases are discussed below.

2.5.1 EXPONENTIAL SMOOTHING AND ARIMA MODELS

The simple exponential smoothing method has the same updating equations and forecasting functions as ARIMA(0,1,1) models. Similarly exponential smoothing with a trend can be shown to be equivalent to an ARIMA(0,2,2) model. Further details of this are given in Appendix A.3. For monthly seasonality, the ARIMA model equivalent to the additive Holt-Winters exponential smoothing method is given by $(1-B)(1-B^{12})Y_t = \theta_{13}(B)Z_t$, where θ_{13} is a moving average parameter, but this is so complex that it would never be identified in practice. Details on this relationship are proved in Box and Jenkins (1976). There is no ARIMA model that is equivalent to the multiplicative Holt-Winters method. However it can be shown that for certain cases, by imposing non-linear restrictions on the coefficients of the ARIMA model, the same forecast functions but not the same updating equations as the Holt-Winters method are obtained (Abraham and Ledolter, 1986).

2.5.2 ARIMA MODELS IN GENERAL STATE SPACE FORM

It can be shown that all ARMA models can be placed in the state space form and thus maximum likelihood estimates of the parameters are easily calculated. Letting d = max(p, q+1), the model ARMA(p,q) can be expressed in state space as

$$Y_t = \begin{pmatrix} 1 & 0 & \cdots\cdots & 0 \end{pmatrix} \alpha_t$$

$$\alpha_t = \begin{pmatrix} \phi_1 & 1 & 0 & 0 & 0 \\ \phi_2 & 0 & 1 & 0 & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & 1 \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \phi_d & 0 & 0 & 0 & 0 \end{pmatrix} \alpha_{t-1} + \begin{pmatrix} 1 \\ \theta_1 \\ \theta_2 \\ \cdot \\ \cdot \\ \theta_{d-1} \end{pmatrix} \eta_t$$

where $\phi_i = 0$ for all i > p and $\theta_j = 0$ for all j > q and $\{\eta_t\}$ is a scalar white noise sequence which satisfies $\eta_t = N(0,\sigma^2)$ for t=1,...T (Abraham and Ledolter, 1986).

BSM and MA(q) models

The random walk plus noise model is equivalent to an ARIMA(0,1,1) model where the moving average parameter θ is constrained as $-1 \leq \theta \leq 0$ and the linear trend model is equivalent to an ARIMA(0,2,2) model, with various restrictions placed on the moving average parameters θ_1 and θ_2 (Abraham and Ledolter, 1986; Janacek and Swift, 1993). From this it can thus be deduced that the simple exponential smoothing method has the same updating functions and forecasting equations as the structural random walk plus noise model and that the exponential smoothing method with a trend is equivalent to the linear trend model. Furthermore the BSM with dummy seasonal components is equivalent to the ARIMA(0,1,1)x(0,1,1)$_{12}$ model when the seasonal moving average parameter is taken as $\Theta_1 = -1$ and the noise variances σ^2_ω and σ^2_ς are exactly zero (Janacek and Swift, 1993).

3.APPLICATIONS

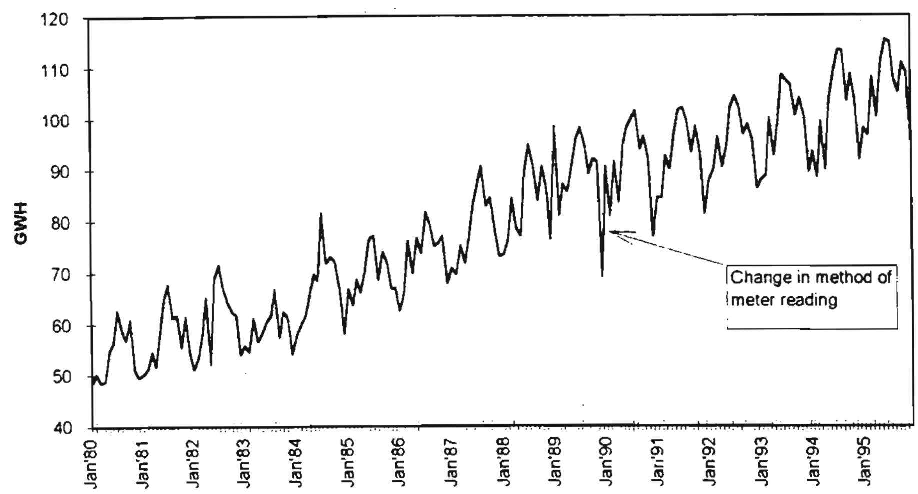
3.1. TIME SERIES

The time series introduced in the present study involve the monthly electricity consumption, measured in Giga Watt hours (GWH), for selected municipalities in Kwa-Zulu Natal, between the years 1980 and 1995. The complete data sets are given in Appendix B. To maintain client confidentiality, the municipalities are not identified but are simply referred to as Municipalities A, B, C and D. All individual series studied exhibited a trend and multiplicative seasonality and specific features of the data are discussed below. It should be noted that the last twelve months of each series was withheld from the modelling process, and used as a test set for assessing the forecasting results.

3.1.1. MUNICIPALITY A

The monthly electricity consumption between 1980 and 1995 of Municipality A is exhibited as a time series plot in Figure 3. 1. 1. Prior to January 1990, monthly readings were taken manually on a working day close to the 24th day of the month. From January 1990 onwards, the meter was read electronically at midnight on the last day of each month. The manual meter reading method resulted in a variable number of hours of electricity consumption recorded within each month. A trading day adjustment was considered, but, since the dates and times at which the meters were read prior to 1990 were unknown, this was not implemented. Thus the raw data was used in all subsequent analyses and cognisance was taken of the fact that the nature of the series might have changed after the electronic metering system was installed.

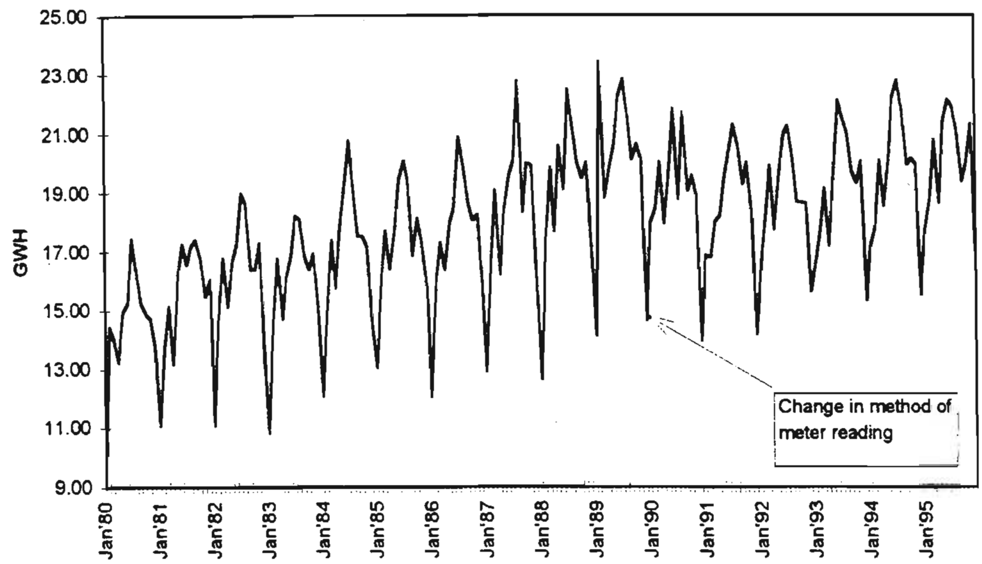
Figure 3. 1. 1. : Time series of monthly electricity consumption for Municipality A



3.1.2. MUNICIPALITY B

The monthly electricity consumption of this municipality between January 1980 and December 1995 is exhibited as a time series plot in Figure 3. 1. 2. It should be noted that an electronic meter reading system was installed in January 1990, and that no trading day adjustments were introduced to accommodate the irregular number of days within the billing months prior to this when analysing the data.

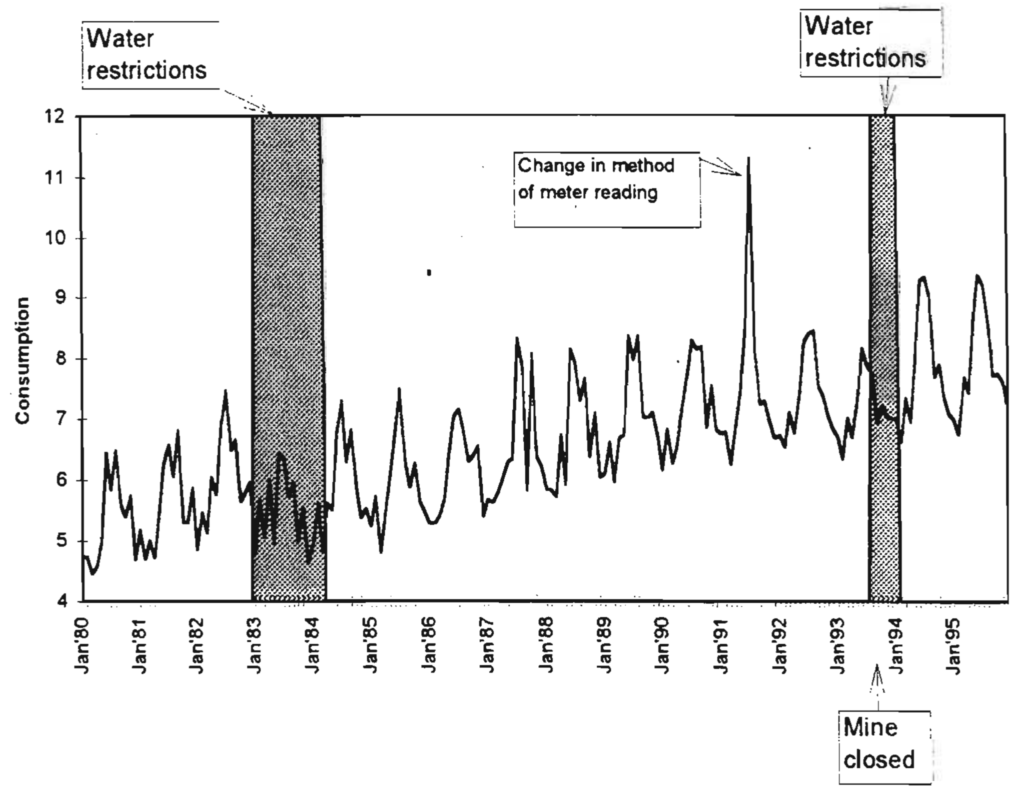
Figure 3. 1. 2.: Time series of monthly electricity consumption for Municipality B



3.1.3. MUNICIPALITY C

A time series plot of the monthly electricity consumption of Municipality C is shown in Figure 3.1.3. The municipality imposed water restrictions on their customers between January 1983 and March 1984 and again between August 1993 and January 1994 and in addition there was a long billing month of 40 days in July 1991 when the meter reading system changed from manual to electronic. These features are shown in Figure 3. 1. 3. Furthermore, a large mine just outside the municipality closed down permanently in August 1993 and it was thought that its satellite industries within the municipality would consequently consume less electricity.

Figure 3. 1. 3. : Time series of monthly electricity consumption for Municipality C



No trading day adjustments were invoked in subsequent analyses. The effect of the water restrictions, the long billing month and the mine closure were investigated using intervention techniques.

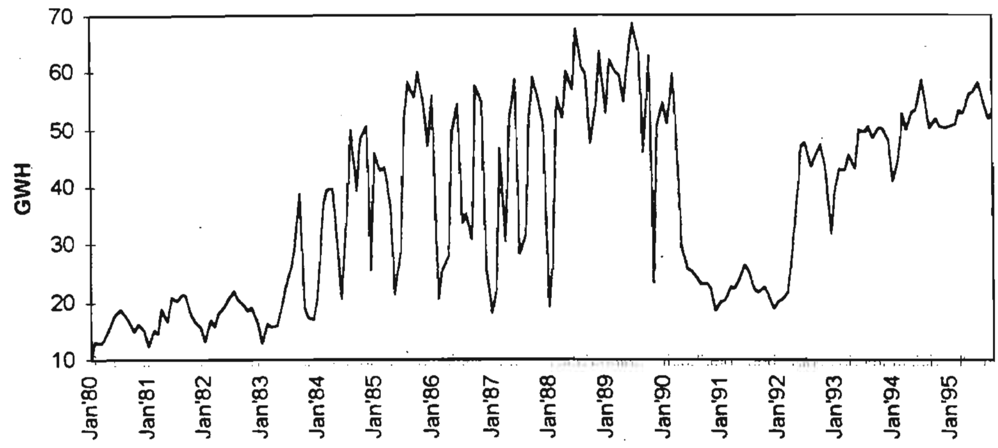
3.1.4. MUNICIPALITY D

A time series plot of the monthly electricity consumption of Municipality D is given in Figure 3.1.4. A large factory has operated in the municipality since 1983 and at present accounts for approximately half of the electricity consumed. A time series plot of the electricity consumption for this factory is included in Figure 3.1.4 and the actual data is given in Appendix B.

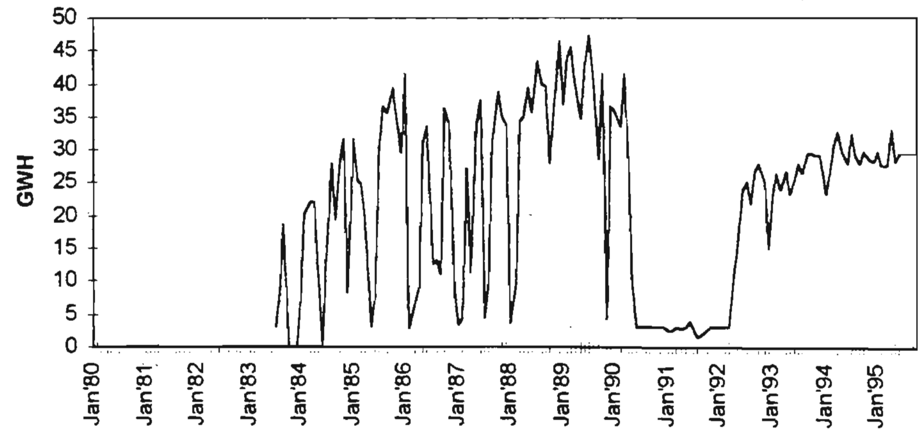
It is clear from Figure 3.1.4 that the electricity consumption of the factory is very erratic. In particular the factory started production in July 1983, but only produced on demand. This resulted in wild fluctuations in electricity consumption and as a consequence Eskom introduced a tariff incentive scheme in March 1988 to encourage a more consistent consumption pattern. The scheme was effective but in May 1990 the market for the factory's products collapsed and it closed. The plant was sold, adapted to a different manufacturing process and production from the new plant started in June 1992 and has been reasonably stable since then. The monthly electricity consumption for the factory exhibits no trend or seasonality.

Figure 3. 1. 4. : Time series of monthly electricity consumption for Municipality D

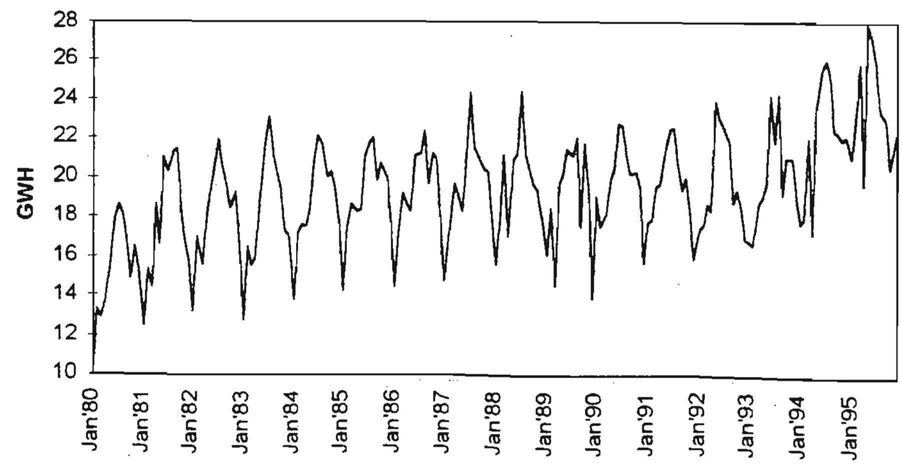
a) Municipality including the factory



b) Factory alone



c) Municipality excluding the factory



3.2. EXPONENTIAL SMOOTHING

All the time series studied here exhibit trend and multiplicative seasonality, and the Holt-Winters method of smoothing is therefore appropriate. The results of applying this method for Municipality A are presented in detail below and those for the other municipalities, which are similar, are summarised thereafter.

The Holt-Winters procedure was implemented using the programming language Gauss in order to introduce a flexibility into the analyses which is not available in packages such as Statistica, SAS and Forecast Pro.

3.2.1. MUNICIPALITY A

The time series of monthly electricity consumption for Municipality A between 1980 and 1994 was regarded as a complete series and the twelve observations for 1995 were used as a test set for evaluating forecasts.

Three different sets of initial values for L_0, T_0 and S_j j = 1,....12, based on the first years data, the first two years data and all the data and calculated using equations (2.3), (2.4) and (2.5) respectively, were used in the smoothing procedure. In each case estimates of the smoothing parameters α , γ and δ were obtained by minimising three different criteria. These are the mean squared error criterion given in equation (2.1) and specified here by

$$\text{M.S.E.} = \frac{1}{T-36}\sum_{t=37}^{T}(Y_t - \hat{Y}_{t:t-1})^2 ,$$

the mean absolute percentage error defined in equation (2.2) and given by

$$\text{M.A.P.E.} = \frac{1}{(T-36)}\sum_{t=37}^{T}\left|\frac{Y_t - \hat{Y}_{t:t-1}}{Y_t}\right|$$

and the mean squared error criterion for twelve months ahead specified in equation (2.6) and calculated here as

$$\text{M.S.E.(12)} = \left(\frac{1}{T-36}\right)\left(\frac{1}{12}\right)\sum_{t=37}^{T}\sum_{j=1}^{12}(Y_{t-j} - \hat{Y}_{t-j:t})^2 .$$

The adequacy of the various starting value options and estimation criteria was evaluated by forecasting the observations of the test set, and using the criteria

$$\text{M.S.E.(F)} = \left(\frac{1}{12}\right)\sum_{t=T+1}^{T+12}\left(Y_t - \hat{Y}_{t|T}\right)^2$$

and

$$\text{M.A.P.E.(F)} = \left(\frac{1}{12}\right)\sum_{t=T+1}^{T+12}\frac{\left|Y_t - \hat{Y}_{t|T}\right|}{Y_t}$$

to measure the accuracy of these forecasts.

The complete set of results are summarised in Table 3. 2. 1. It is interesting to observe that in all cases the best forecasts, as gauged by the particular criterion minimised, were obtained by using initial values based on all the data, but that this is not true when forecasts are evaluated using the criteria M.S.E.(F) and M.A.P.E.(F) based on the test set. Comparisons between the results for the different minimisation criteria can be made on the basis of M.S.E.(F) and M.A.P.E.(F) and in particular it is clear that the results obtained by minimising M.S.E. provide the best forecasts for the test set. Since both the M.S.E. and the M.A.P.E. criteria measure the one-step-ahead forecast errors, minimising the M.S.E. is easier to implement and the results are better than for minimising M.A.P.E., only the minimisation criteria M.S.E. and M.S.E.(12) will be used in further comparisons.

Table 3. 2. 1. Summary of Results for Municipality A

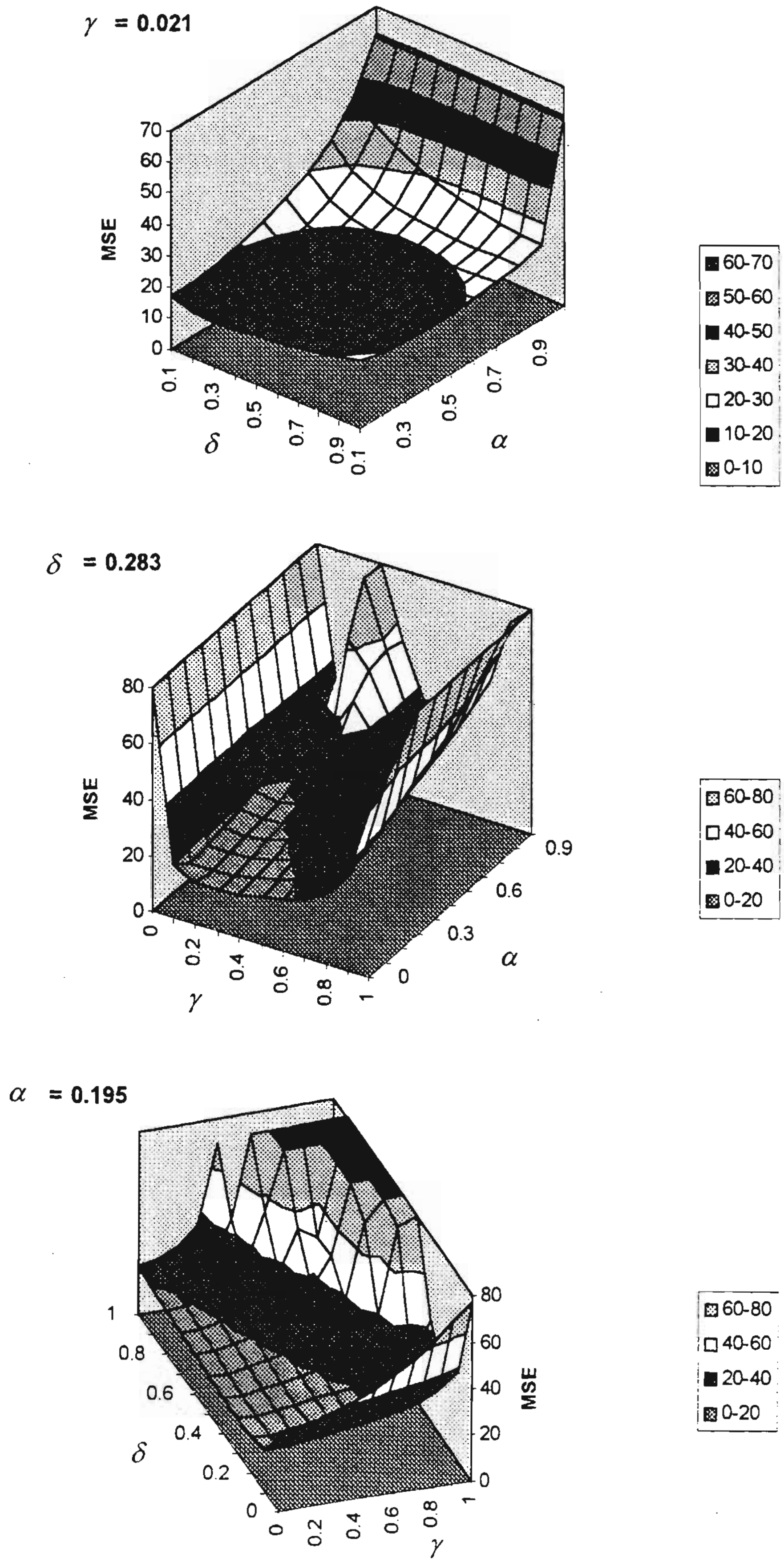
ERROR FUNCTION	START-UP VALUES	ESTIMATES α	γ	δ	MINIMISED ERROR FUNCTION	M.S.E.(F)	M.A.P.E.(F)
M.S.E.	1 year	0.195	0.021	0.283	13.836	**7.567**	1.94%
	2 years	0.197	0.023	0.285	14.338	7.684	**1.19%**
	All data	0.186	0.001	0.075	**12.812**	11.733	2.69%
M.A.P.E(*)	1 year	0.144	0.024	0.295	3.32%	7.958	2.08%
	2 years	0.126	0.035	0.245	3.37%	8.314	2.15%
	All data	0.179	0.003	0.097	**3.19%**	10.959	2.61%
M.S.E.(12)	1 year	0.045	0.091	0.245	17.373	11.304	2.70%
	2 years	0.121	0.026	0.261	17.243	8.220	2.26%
	All data	0.126	0.000	0.080	**15.350**	11.283	2.29%

(*) Some problems in convergence, due to the nature of the function, were encountered.

Overall, the estimates of the smoothing parameters α, γ and δ varied slightly with choice in initial values and in the criterion to be minimised. However, the seasonal parameter δ is much smaller when the initial values are calculated using all the data as opposed to the first one or two years data. This low value is a result of initial seasonal estimates being good approximations and, apart from the initial few years data, there being little change in the seasonal pattern of the series. It is interesting to note that in all cases the estimate for γ was close to zero, suggesting that changes in the trend are very slow.

In addition, for the case in which the criterion M.S.E. is minimised, with initial values calculated from the first years data, a check on the nature of the optimum was made by plotting M.S.E. against values of each pair of parameters, with the third parameter fixed at its optimum. The plots for the data of Municipality A are shown in Figure 3. 1. 5 and clearly indicate a single global minimum for the criterion.

Figure 3. 1. 5. : Global minimum for M.S.E. criterion found by applying exponential smoothing for Municipality A



For given values of the smoothing parameters, the time series Y_t can be decomposed into the four component series of level, trend, seasonality and error, calculated as L_t, T_t, S_t and e_t, for t = 1,...T, respectively. The decomposition of the time series of monthly electricity consumption for Municipality A is illustrated in Figure 3. 1. 6. for the optimal parameter values α = 0.195, γ =0.021 and δ = 0.283 obtained by minimising the M.S.E. criterion and using initial values based on the first years data. The residual series is shown in Figure 3. 1. 7. The high residual value in January 1989 is due to an unusually long billing month of 34 days and the low value associated with January 1990 coincides with the installation of an electronic metering system which resulted in a short billing month. Otherwise this error series appears to be random indicating that the Holt-Winters method has captured the systematic variation of the original time series.

Figure 3. 1. 6. Municipality A : Decomposition of the time series

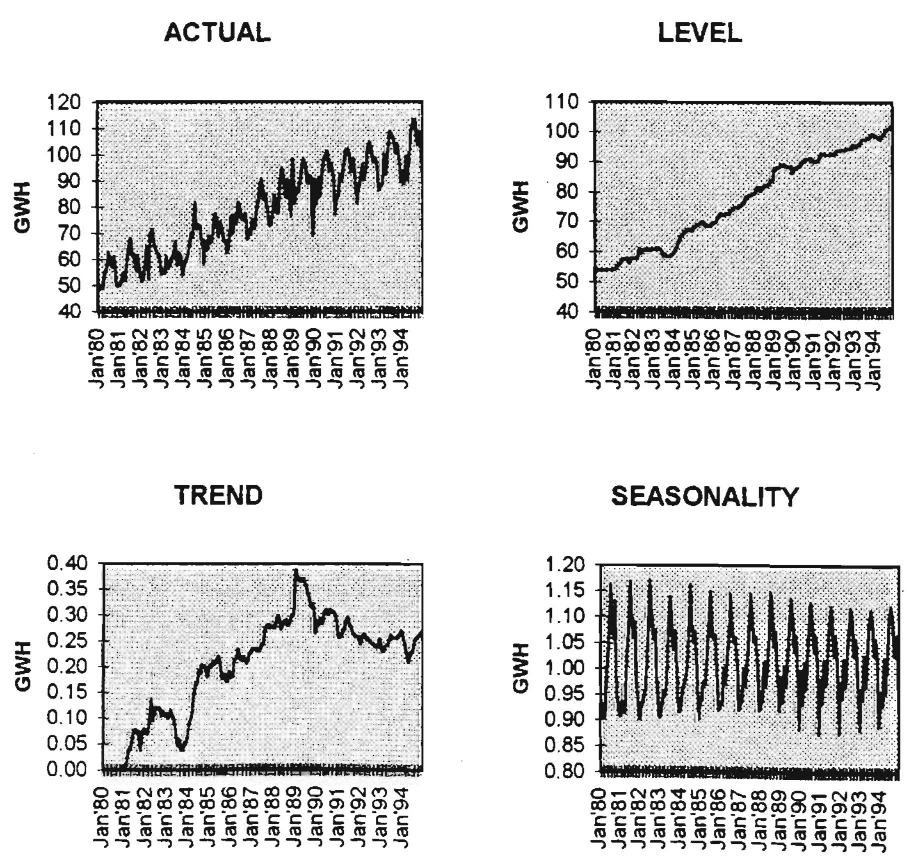
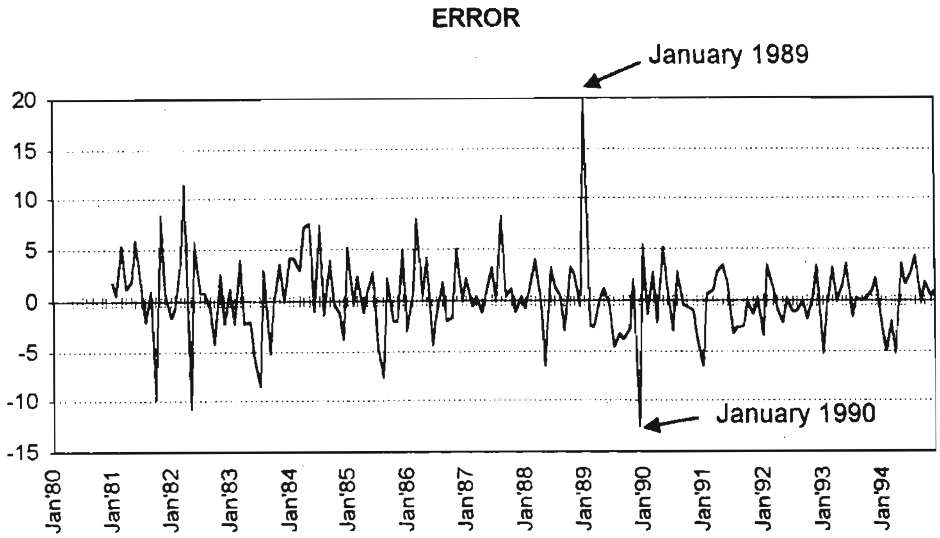


Figure 3. 1. 7 Municipality A : Residual error for exponential smoothing



In addition to analysing the full series, the sub-series between January 1990 and December 1994 was considered in isolation, in order to investigate whether or not the electronically metered sub-series would result in better forecasts. However the sub-series was too short to perform any meaningful analysis.

3.2.2. MUNICIPALITIES B AND C

The Holt-Winters exponential smoothing procedure was implemented for the time series of monthly electricity consumption for Municipalities B and C in a manner similar to that of Municipality A and the results are summarised in Tables 3. 2. 2. and 3. 2. 3. respectively. Again a low parameter value δ was derived when calculating the initial values using the whole series, indicating a stable seasonal pattern.

Table 3. 2. 2. : Summary of results for Municipality B

ERROR FUNCTION	START-UP VALUES	ESTIMATES α	γ	δ	MINIMISED ERROR FUNCTION	M.S.E.(F)	M.A.P.E.(F)
M.S.E.	1 year	0.145	0.003	0.444	1.165	0.429	2.88%
	2 years	0.161	0.040	0.462	1.207	0.409	2.87%
	All data	0.161	0.003	0.000	**1.154**	1.118	4.43%
M.S.E.(12)	1 year	1.82	0.005	0.472	1.421	0.407	2.82%
	2 years	0.189	0.033	0.491	1.507	**0.401**	2.44%
	All data	0.219	0.000	0.000	**1.372**	1.133	**2.34%**

Table 3. 2. 3. : Summary of results for Municipality C

ERROR FUNCTION	START-UP VALUES	ESTIMATES α	γ	δ	MINIMISED ERROR FUNCTION	M.S.E.(F)	M.A.P.E.(F)
M.S.E.	1 year	0.133	0.008	0.226	0.264	0.140	3.34%
	2 years	0.175	0.055	0.227	0.279	**0.138**	3.87%
	All data	0.45	0.044	0.010	**0.230**	0.220	4.85%
M.S.E.(12)	1 year	0.095	0.013	0.206	0.296	0.146	3.39%
	2 years	0.176	0.034	0.216	0.326	0.142	**2.48%**
	All data	0.144	0.001	0.010	**0.255**	0.218	2.53%

3.2.3. MUNICIPALITY D

As mentioned earlier, the large factory within the boundaries of Municipality D has a dominating effect on the monthly electricity consumption in that municipality. As a consequence, the full time series for Municipality D was split into two series, electricity consumption excluding the factory and the electricity consumption of the factory itself and each series was analysed separately. The series which excludes the factory consumption exhibits trend and seasonality, forecasts for it were obtained in the same way as those for Municipalities A, B and C and the results are summarised in Table 3.2.4. The time series of monthly electricity consumption for the factory exhibited no systematic trend or seasonality and forecasts were therefore obtained by simple exponential smoothing. In addition, two time series were analysed, the complete time series as well as only the new factory's electricity consumption from July 1992 to December 1994. The results are summarised in Table 3.2.5 and clearly using the complete time series results in more accurate forecasts. Note that as a result of the large fluctuations in the time series prior to July 1992, the minimisation criterion M.S.E. is much larger when using the complete time series as opposed to using the time series only between July 1992 and December 1994.

Table 3. 2. 4 : Summary of results for Municipality D excluding the factory

ERROR FUNCTION	START-UP VALUES	ESTIMATES α	γ	δ	MINIMISED ERROR FUNCTION	M.S.E.(F)	M.A.P.E.(F)
M.S.E.	1 year	0.084	0.008	0.529	1.778	**5.427**	**7.85%**
	2 years	0.076	0.203	0.542	1.778	7.816	10.85%
	All data	0.120	0.023	0.000	**1.822**	8.761	11.22%
M.S.E.(12)	1 year	0.040	0.009	0.492	1.774	6.752	8.84%
	2 years	0.077	0.115	0.524	1.905	6.126	11.31%
	All data	0.001	1.00	0.306	**1.748**	11.033	11.43%

Table 3. 2. 5. : Summary of results for the factory

TIME SERIES	α	MINIMISED M.S.E.	M.S.E.(F)	M.A.P.E.(F)
Whole series	0.425	108.878	**17.048**	**11.33%**
Between July 1992 and December 1994	0.413	6.139	17.990	11.86%

3.2.4.COMMENTS

The optimal method of calculating the initial values is not clear, although using the first years data appears to give good results generally and is therefore the preferred option. The optimisation criterion M.A.P.E. was awkward to calculate and the results were poor compared to the M.S.E. criterion. In addition, the optimisation criterion M.S.E. was simpler to calculate than the criterion M.S.E.(12) and the results are better as measured by the forecasting criteria M.S.E.(F) and M.A.P.E.(F). Thus the optimisation criterion M.S.E. is taken as the most suitable option.

3.4 ARIMA MODELS

ARIMA models were fitted to each of the time series in this study using the Box-Jenkins approach and the resultant models were used to provide forecasts. The package SAS was used for all the modelling processes.

3.4.1 MUNICIPALITY A

Plots of the ACF's for Y_t, the time series of monthly electricity consumption for Municipality A, and the differenced time series ∇Y_t and $\nabla\nabla_{12}Y_t$ are given in Figure 3. 3. 1. It is clear from these that first order and seasonal differencing are appropriate and thus that the model will be of type ARIMA(p,1,q)x(P,1,Q)$_{12}$. The initial model fitted after studying the pattern of the ACF and the PACF of the differenced series $\nabla\nabla_{12}Y_t$, given in Figures 3. 3. 1 and 3. 3. 2 respectively, was an ARIMA(2,1,1)x(0,1,1)$_{12}$ model. However, the t ratios for testing whether the parameters of this model are zero, given in Table 3. 3. 1 below, suggested that the parameter estimate for ϕ_2 was unnecessary and thus that the model ARIMA(1,1,1)x(0,1,1)$_{12}$ should be examined.

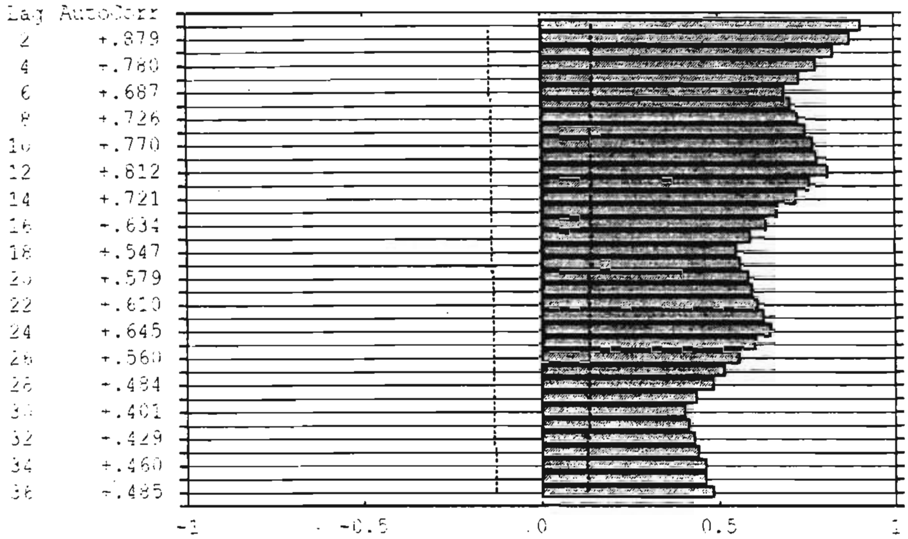
Table 3. 3. 1 Municipality A : Parameter estimates for the ARIMA (2,1,1)x(0,1,1)$_{12}$ model

Parameter	Estimate	t ratio
θ_1	0.74794	8.41
Θ_1	0.78942	11.37
ϕ_1	-0.16049	-1.37
ϕ_2	0.05081	0.48

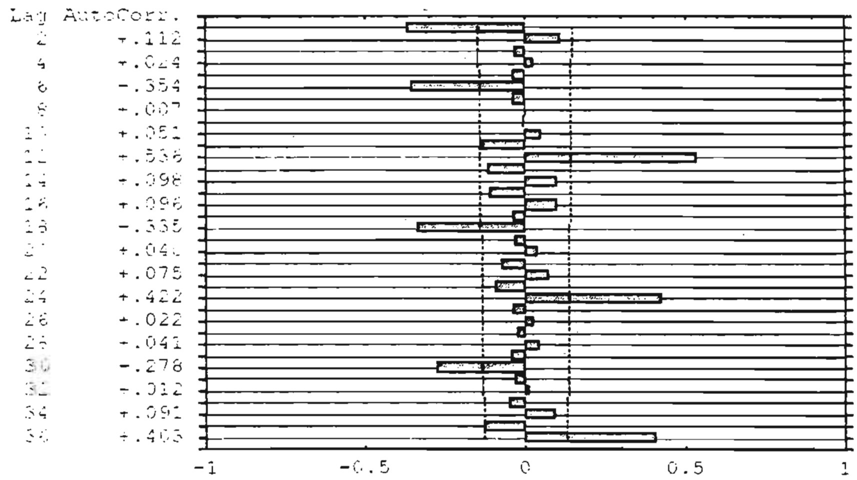
The associated results and diagnostics for the model ARIMA(1,1,1)x(0,1,1)$_{12}$ are summarised in Table 3. 3. 2. The t-ratios for the parameters are all greater than 1.96 and

Figure 3. 3. 1 Municipality A : ACF's of Y_t, ∇Y_t and $\nabla\nabla_{12}Y_t$

Autocorrelations for Y_t



Autocorrelations for ∇Y_t



Autocorrelations for $\nabla\nabla_{12}Y_t$

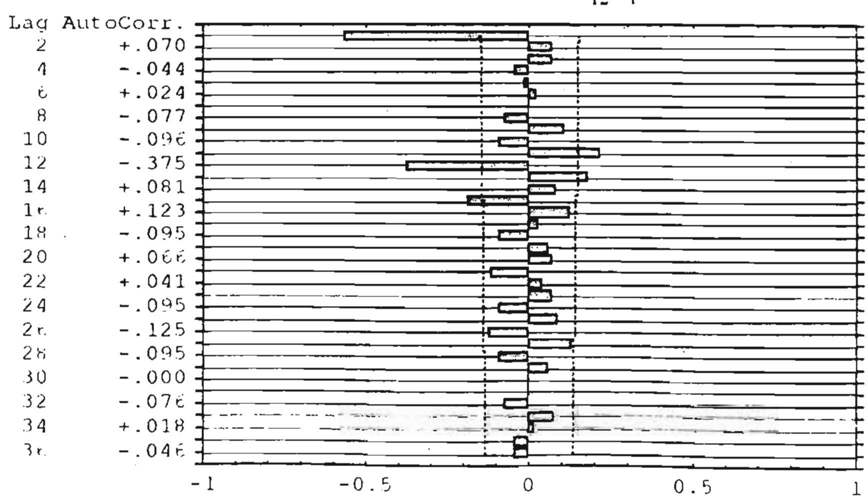
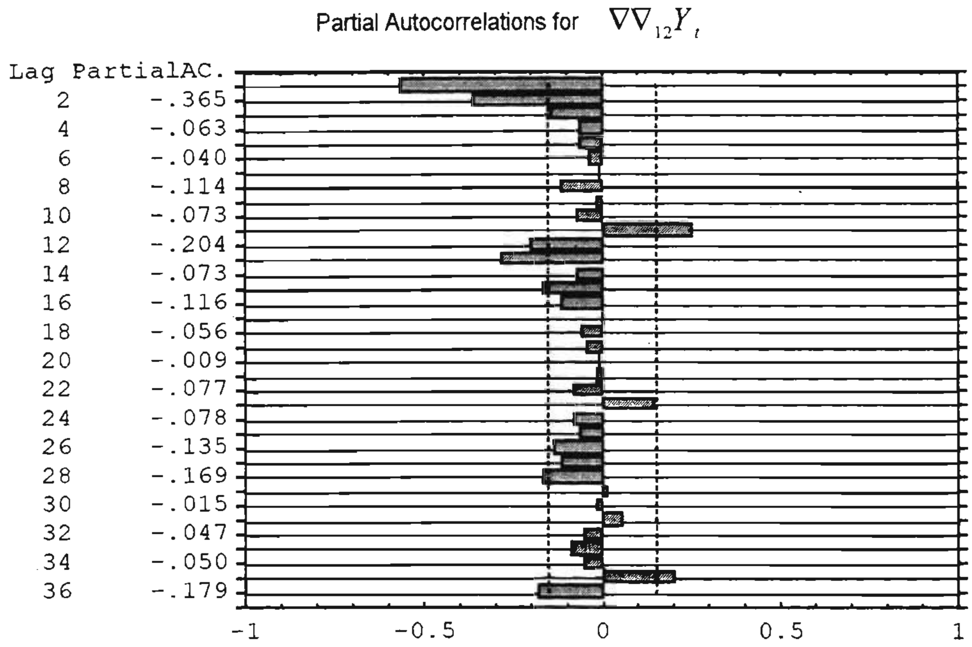


Figure 3. 3. 2 Municipality A : PACF of $\nabla\nabla_{12}Y_t$



thus the parameters are significantly different from zero at the 5% level of significance. It should be noted that a high correlation between the parameter estimates for θ_1 and ϕ_1 is an indication that the model could be over-parameterised, but the AIC statistic did not improve by fitting models with fewer parameters. The ACF of the residuals given in Figure 3. 3. 3 together with the portmanteau test results suggest that the residuals are random and thus that the model ARIMA(1,1,1)x(0,1,1)$_{12}$ is acceptable. The model adopted can thus be summarised as

$$W_t = -0.21833W_{t-1} + Z_t - 0.69143Z_{t-1} - 0.7869Z_{t-12} + 0.54409Z_{t-13}$$

where $W_t = \nabla_{12}\nabla_1 Y_t$.

Table 3. 3. 2 Municipality A : Results for fitting an ARIMA $(1,1,1)x(0,1,1)_{12}$ model

Parameter estimates using MLE :

Parameter	Estimate	t ratio
θ_1	0.69143	9.97
Θ_1	0.78690	11.37
ϕ_1	-0.21833	-2.35

The Portmanteau test for white noise :

Lags	Chi Square	DF	P-value
1-6	3.38	3	0.337
1-12	8.08	9	0.526
1-18	13.87	15	0.536
1-24	16.95	21	0.714
1-30	18.54	27	0.886

Correlations of the Estimates :

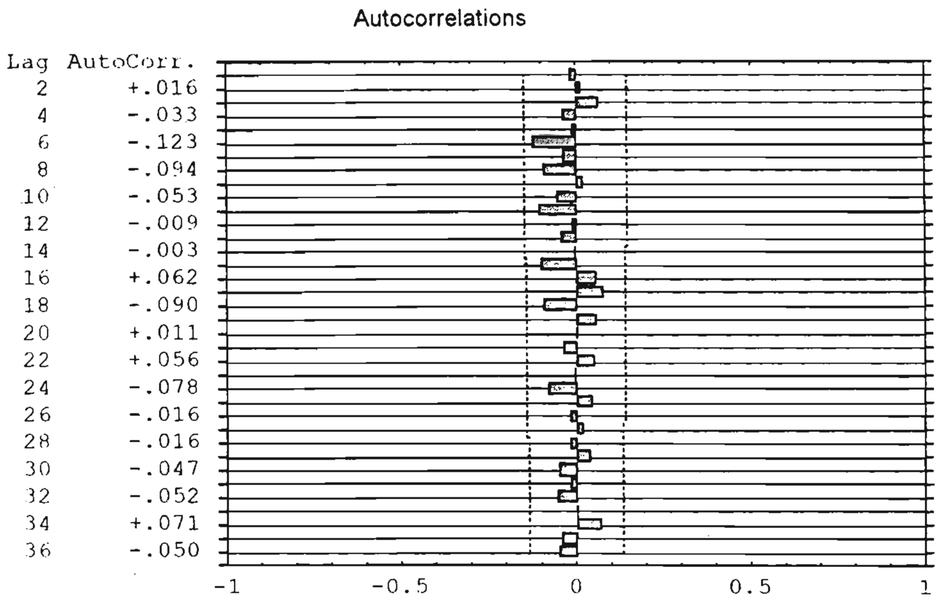
Parameter	θ_1	Θ_1	ϕ_1
θ_1	1.000	0.049	0.607
Θ_1	0.049	1.000	-0.068
ϕ_1	0.607	-0.068	1.000

Model comparison statistics : AIC = 928.836 SBC = 938.190

Test set forecasting results : M.S.E. (F) = 6.795

M.A.P.E.(F) = 1.96%

Figure 3. 3. 3. Municipality A : Residual error resulting from fitting an ARIMA(1,1,1)x(0,1,1)$_{12}$ model



An alternative approach to identifying the most appropriate model to that described above is to fit an over-parameterised model to the series and then to reduce it by successively dropping parameters, until all the parameters are significantly different from zero. Because the values of p,q, P and Q rarely exceed 2, the model ARIMA(2,1,2)x(2,1,2)$_{12}$ was initially fitted to the time series. Reducing the model until all the t-ratios in the model were significant resulted in the model ARIMA(1,1,1)x(0,1,1)$_{12}$ which is consistent with the model selected above.

The test set of the final twelve months electricity consumption was forecast using the model ARIMA(1,1,1)x(0,1,1)$_{12}$ and the forecasting error was measured as before using

$$\text{M.S.E.(F)} = \left(\frac{1}{12}\right)\sum_{t=T+1}^{T+12}(Y_t-\hat{Y}_{t|T})^2$$

and

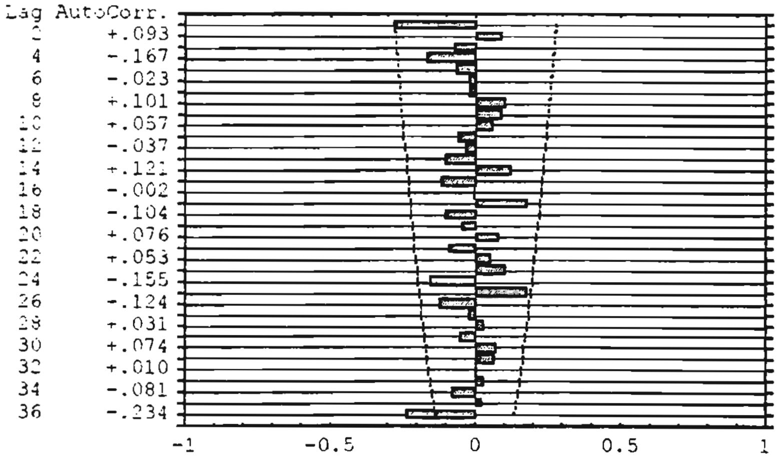
$$\text{M.A.P.E.(F)} = \left(\frac{1}{12}\right)\sum_{t=T+1}^{T+12}\frac{\left|Y_t-\hat{Y}_{t|T}\right|}{Y_t}$$

where T=180, the length of the times series used in the modelling process. The results are included in Table 3. 3. 2.

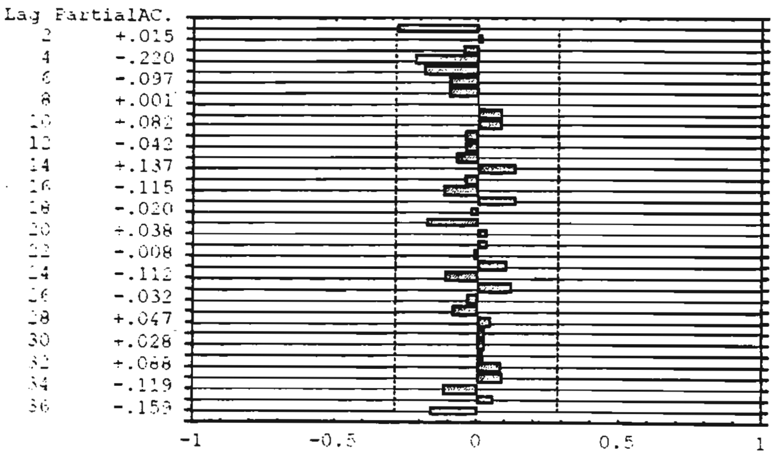
The sub-series of monthly electricity consumption of Municipality A, between January 1990 and December 1994, when the meters were read electronically, was considered separately to ascertain whether or not this time series would result in more accurate forecasts. First order and seasonal differencing were again appropriate and the ACF and PACF of the resultant differenced series are presented in Figure 3. 3. 4 .

Figure 3. 3. 4 Municipality A : ACF and PACF of $\nabla\nabla_{12}Y_t$ for the sub-series corresponding to electronic metering

Autocorrelations of $\nabla\nabla_{12}Y_t$



Partial Autocorrelation $\nabla\nabla_{12}Y_t$



Clearly, there are no significant autocorrelations or partial autocorrelations indicating that either the differenced series is white noise or that the time series is too short to derive any meaningful results. Overall it was therefore not deemed sensible to pursue modelling this time series any further.

3.4.2 MUNICIPALITY B

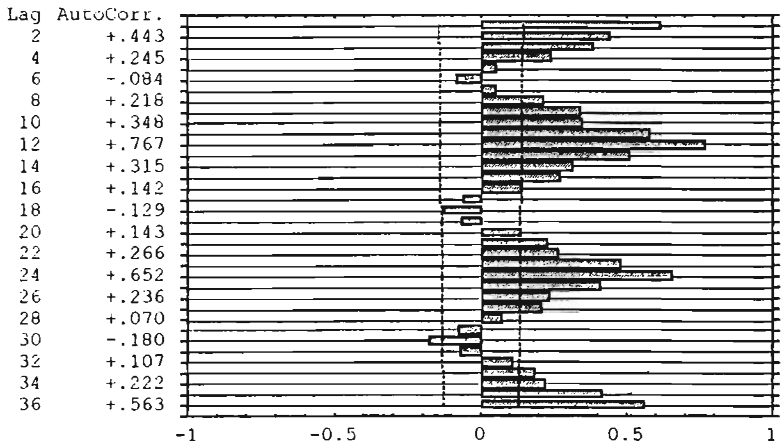
The ACF's of Y_t, the time series for monthly electricity consumption of Municipality B, and of the differenced series $\nabla_{12}Y_t$, and $\nabla\nabla_{12}Y_t$ given in Figure 3. 3. 5 clearly suggest a model of the form ARIMA(p,0,q)x(P,1,Q)$_{12}$. The PACF of the seasonally differenced series is given in Figure 3. 3. 6. Various suitable models suggested by the ACF and PACF patterns were investigated, but a model that satisfied all the diagnostic checks could not be found. After considerable investigation, the most suitable model was deemed to be ARIMA(1,0,2)x(0,1,1)$_{12}$. The ACF of the residual errors for this model given in Figure 3. 3. 7, are acceptable but the results which are summarised in Table 3. 3. 3 clearly show that the portmanteau test for white noise is not satisfactory. In addition, the high correlation between the MA parameter estimates for θ_1 and θ_2 suggests that the model could well be over-parameterised. The model ARIMA(0,0,0)x(0,1,1)$_{12}$ was also fitted to the time series but the ACF of the associated residuals given in Figure 3. 3. 8 was clearly unsatisfactory. Another alternative model considered was ARIMA(2,0,1)x(0,1,1)$_{12}$ but a correlation of -0.891 between the parameter estimates for ϕ_1 and ϕ_2 was deemed to be unacceptably high. Fitting an over-parameterised model and systematically eliminating the parameters according to the t-ratios resulted in the model ARIMA(1,0,2)x(0,1,1)$_{12}$ which is consistent with the model deduced from the patterns of the ACF and PACF. Thus the model

$$W_t = 0.93347W_{t-1} + Z_t - 1.05609Z_{t-1} + 0.34432Z_{t-2} - 0.64602Z_{t-12} + 0.68226Z_{t-13} - 0.22243Z_{t-14} + 0.35555$$

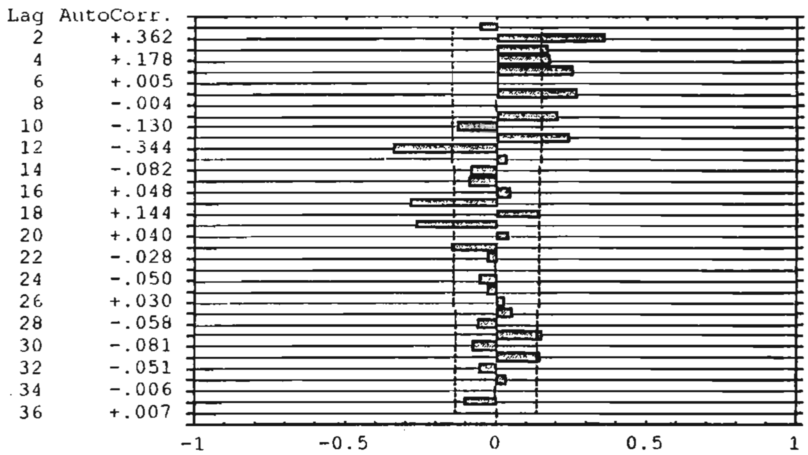
where $W_t = \nabla_{12}Y_t$, is taken to be the most appropriate model for the time series of monthly electricity consumption for Municipality B.

Figure 3. 3. 5 Municipality B : ACF's of Y_t, $\nabla_{12}Y_t$ and $\nabla\nabla_{12}Y_t$

Autocorrelations for Y_t



Autocorrelations for $\nabla_{12}Y_t$



Autocorrelations for $\nabla\nabla_{12}Y_t$

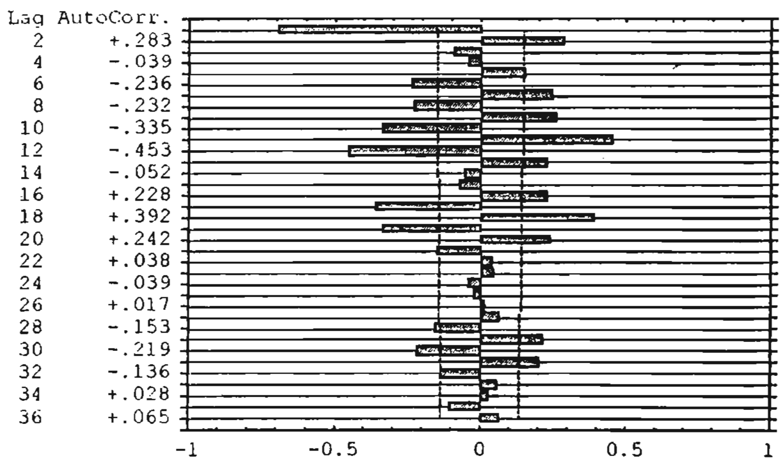


Figure 3. 3. 6 Municipality B : PACF of $\nabla_{12}Y_t$

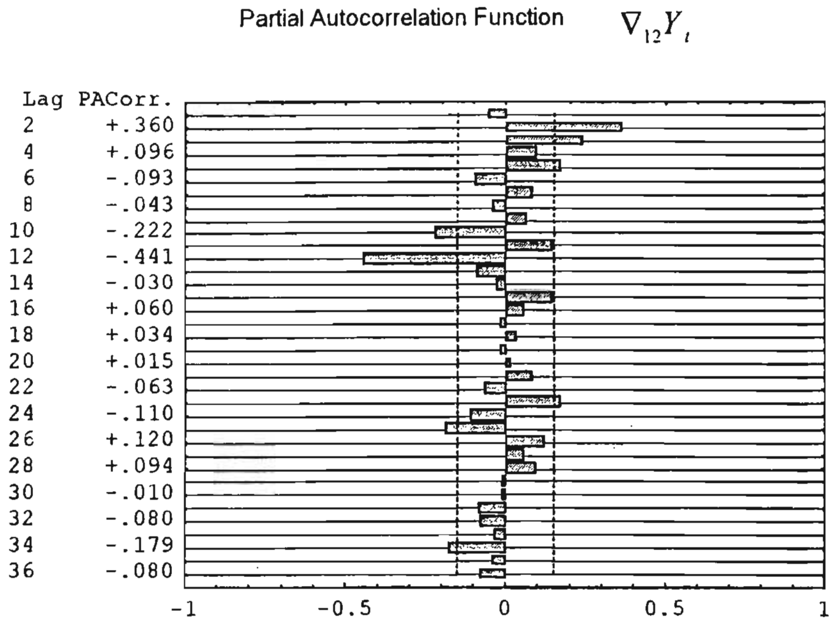


Figure 3. 3. 7 Municipality B : ACF of the residual errors when fitting an ARIMA(1,0,2)x(0,1,1)$_{12}$ model.

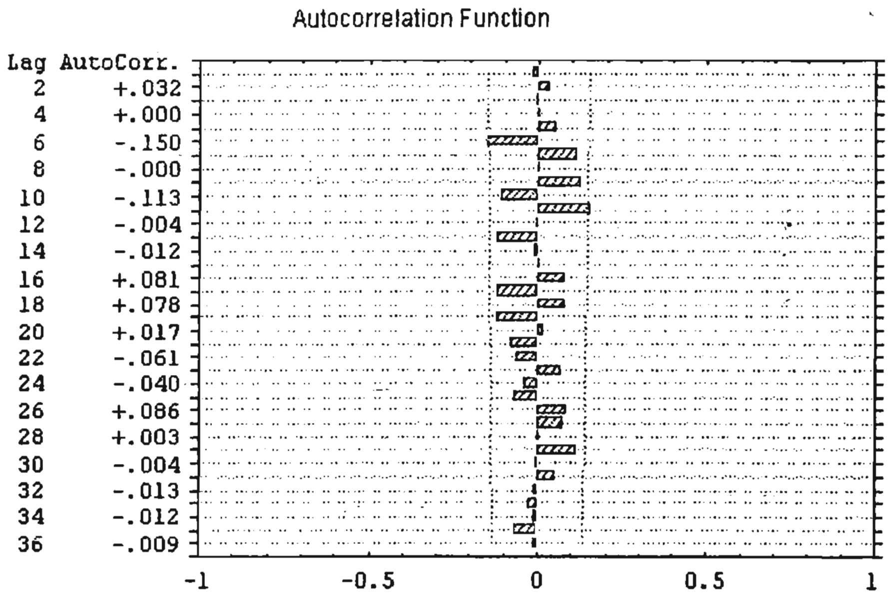


Figure 3. 3. 8 Municipality B : ACF of the residual errors when fitting an ARIMA(0,0,0)x(0,1,1)$_{12}$ model

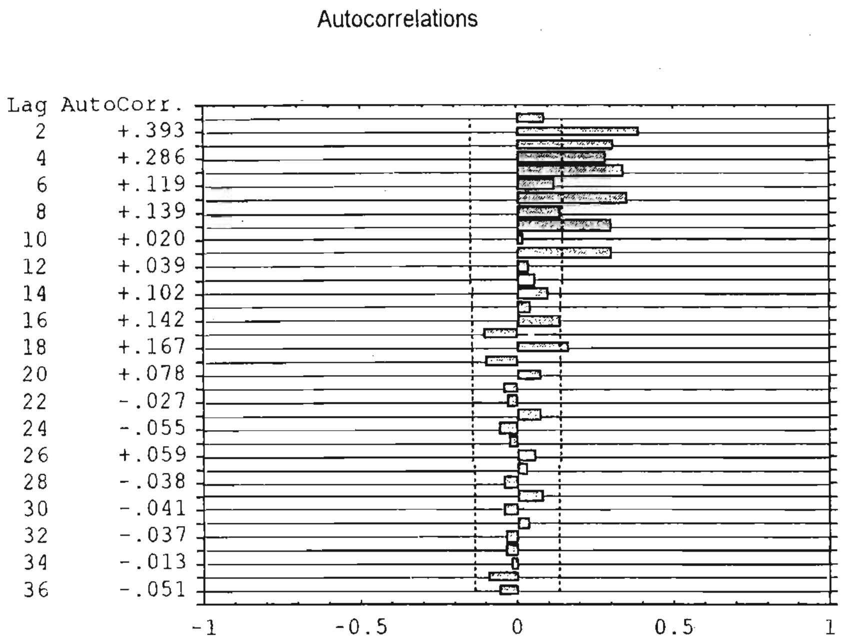


Table 3. 3. 3 Municipality B : Results when fitting an ARIMA(1,0,2)x(0,1,1)$_{12}$ model

Parameter estimates using MLE :

Parameter	Estimate	t ratio
δ	0.35555	2.88
θ_1	1.05609	13.57
θ_2	-0.34432	-4.54
Θ_1	0.64602	9.26
ϕ_1	0.93347	23.34

The Portmanteau test for white noise :

Lags	Chi Square	DF	P-value
1-6	4.86	2	0.088
1-12	17.32	8	0.027
1-18	25.99	14	0.026
1-24	32.64	20	0.037
1-30	39.11	26	0.048

Correlations of the Estimates :

Parameter	δ	θ_1	θ_2	Θ_1	ϕ_1
δ	1.000	-0.011	0.008	0.022	-0.026
θ_1	-0.011	1.000	-0.633	-0.014	0.342
θ_2	0.008	-0.633	1.000	0.193	0.256
Θ_1	0.022	-0.014	0.193	1.000	0.294
ϕ_1	-0.026	0.342	0.256	0.294	1.000

Model comparison statistics : AIC = 489.661 SBC = 505.281

Test Set Forecasting Results : M.S.E.(F) = 0.432 M.A.P.E.(F) = 2.72%

The sub-series of electricity consumption for Municipality B which was measured electronically from January 1990 onwards was modelled to ascertain if a more satisfactory model could be obtained. The ACF in Figure 3. 3. 9 indicates only seasonal differencing of the series is required. After examining the ACF and PACF of the differenced series, various models including ARIMA(0,0,1)x$(0,1,1)_{12}$, ARIMA(1,0,0)x$(0,1,1)_{12}$ and ARIMA(0,0,0)x$(0,1,1)_{12}$ were fitted, and the model ARIMA(1,0,1)x$(0,1,1)_{12}$ was found to be the most appropriate. The associated results for this model, which are given in Table 3. 3. 5, are more acceptable than for those for the best model derived when modelling the complete time series. This is probably as a result of the time series being more regular once the meters were read electronically.

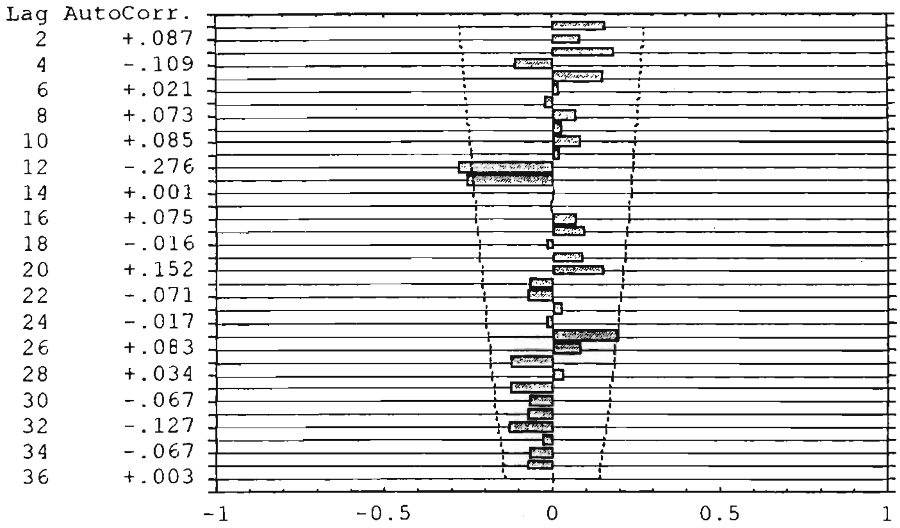
The test set of observations was forecast using the ARIMA(1,0,2)x$(0,1,1)_{12}$ model derived for the whole time series and then using the ARIMA(1,0,1)x$(0,1,1)_{12}$ model derived for the shorter series and the results are compared in Table 3. 3. 4. It is interesting to note that although the model derived using the complete time series was poor, it still produced slightly better forecasting results than when using the model derived using the shorter time series of electronically metered electricity consumption.

Table 3. 3. 4 Municipality B : Comparison of forecast results using the whole time series verses the sub-series corresponding to electronic metering

DATA	MODEL	M.A.P.E.(F)	M.S.E.(F)
1980->1994	ARIMA(1,0,2)x$(0,1,1)_{12}$	2.72%	0.432
1990->1994	ARIMA(1,0,1)x$(0,1,1)_{12}$	3.01%	0.466

Figure 3. 3. 9 Municipality B : The ACF and PACF for the sub-series corresponding to electronic metering

Autocorrelations for $\nabla_{12} Y_t$



Partial Autocorrelations for $\nabla_{12} Y_t$

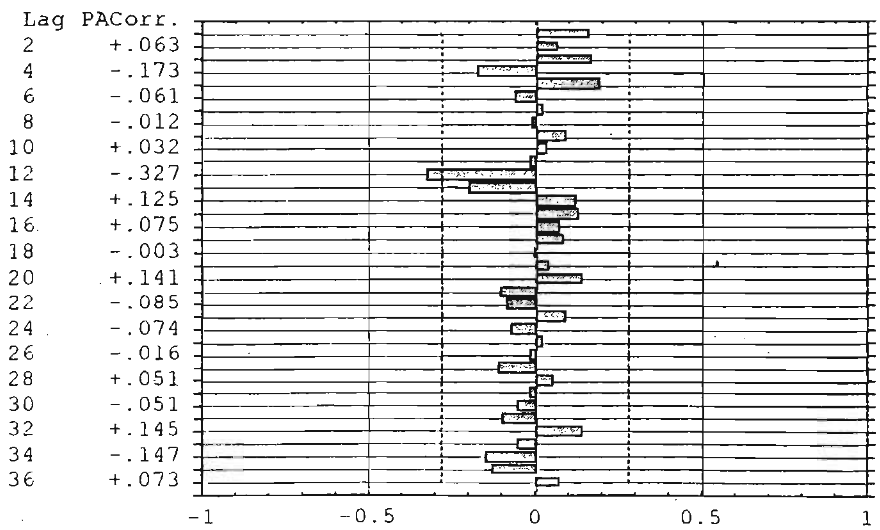


Table 3. 3. 5 Municipality B : Results when fitting an ARIMA(1,0,1)x(0,1,1)$_{12}$ model to sub-series corresponding to electronic metering

Parameter estimates using MLE :

Parameter	Estimate	t ratio
θ_1	0.79055	5.13
Θ_1	0.60940	2.73
ϕ_1	0.95859	9.64

The Portmanteau test for white noise :

Lags	Chi Square	DF	P-value
1- 6	3.60	3	0.309
1-12	4.90	9	0.843
1-18	12.37	15	0.651
1-24	17.33	21	0.691

Correlations of the Estimates :

Parameter	θ_1	Θ_1	ϕ_1
θ_1	1.000	0.174	0.832
Θ_1	0.174	1.000	0.353
ϕ_1	0.832	0.353	1.000

Model comparison statistics : AIC = 124.059 SBC = 129.672

Test set forecasting results : M.S.E.(F) = 0.466 M.A.P.E.(F) = 3.01%

3.4.3 MUNICIPALITY C

A number of events such as water restrictions are thought to have had an impact on the time series of monthly electricity consumption for Municipality C. To assess the improvement in the forecast when including these events in the model, the time series was modelled excluding and then including the intervention events and the results compared.

The ACF's of the time series Y_t, and of the difference time series $\nabla_{12}Y_t$ and $\nabla\nabla_{12}Y_t$, which are given in Figure 3. 3. 10, indicate a model of the form ARIMA(p,0,q)x(P,1,Q)$_{12}$. Identifying the characteristic patterns of the ACF and PACF, which are given in Figure 3. 3. 11, is difficult as they have probably been distorted by intervention events. Thus an over-parameterised model was fitted and parameters not significantly different from zero were successively dropped from the model resulting in the model ARIMA(2,0,1)x(0,1,1)$_{12}$. Details of this models fit are given in Table 3. 3. 6 and the ACF of the residual error is shown in Figure 3. 3. 13. In summary therefore the model represented by

$$W_t = 0.50133W_{t-1} + 0.30844W_{t-2} + Z_t - 0.56911Z_t - 0.89129Z_{t-12} - 0.50724Z_{t-13} + 0.17095$$

where $W_t = \nabla_{12}Y_t$, was adopted.

Table 3. 3. 6 Municipality C : Results for the ARIMA(2,0,1)x(0,1,1)$_{12}$ model

Parameter estimates using MLE :	**Parameter**	**Estimate**	**t ratio**
	δ	0.17095	9.74
	θ_1	0.56911	4.02
	Θ_1	0.89129	9.48
	ϕ_1	0.50133	3.51
	ϕ_2	0.30844	3.94

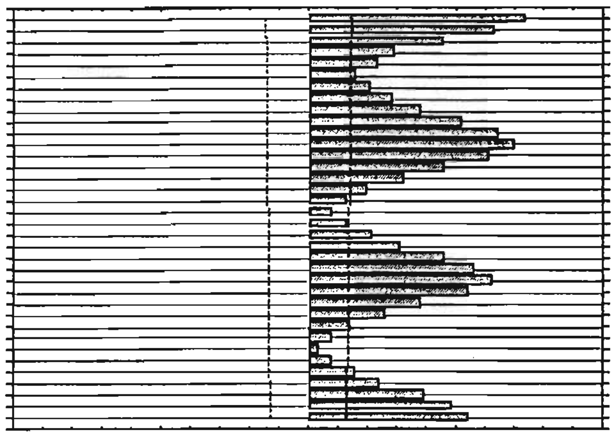
Model comparison statistics : AIC = 246.095 SBC = 261.715

Test set forecasting results : M.S.E.(F) = 0.165 M.A.P.E. = 3.79%

Figure 3. 3. 10 Municipality C : ACF's of Y_t, $\nabla_{12}Y_t$ and $\nabla\nabla_{12}Y_t$

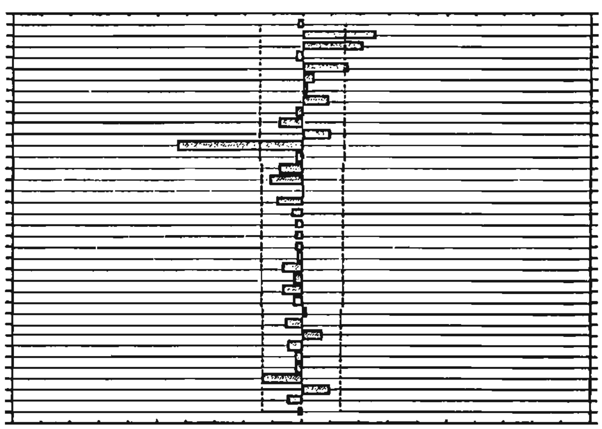
Autocorrelations for Y_t

Lag	AutoCorr.
2	+.628
4	+.289
6	+.157
8	+.284
10	+.518
12	+.702
14	+.460
16	+.195
18	+.081
20	+.216
22	+.461
24	+.623
26	+.379
28	+.143
30	+.037
32	+.161
34	+.391
36	+.541



Autocorrelations for $\nabla_{12}Y_t$

Lag	AutoCorr.
2	+.253
4	-.021
6	+.041
8	+.089
10	-.080
12	-.431
14	-.077
16	+.005
18	-.034
20	-.021
22	-.017
24	-.030
26	-.030
28	-.050
30	-.045
32	-.023
34	+.100
36	-.007



Autocorrelations for $\nabla\nabla_{12}Y_t$

Lag	AutoCorr.
2	+.152
4	-.204
6	-.050
8	+.087
10	-.114
12	-.459
14	-.011
16	+.102
18	+.020
20	+.002
22	+.027
24	+.037
26	-.005
28	-.091
30	-.065
32	+.055
34	+.187
36	-.085

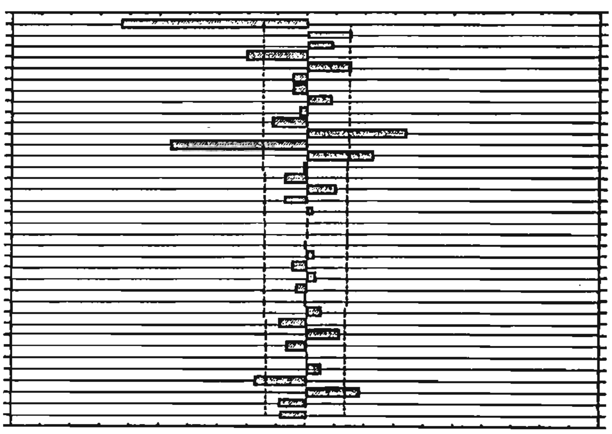
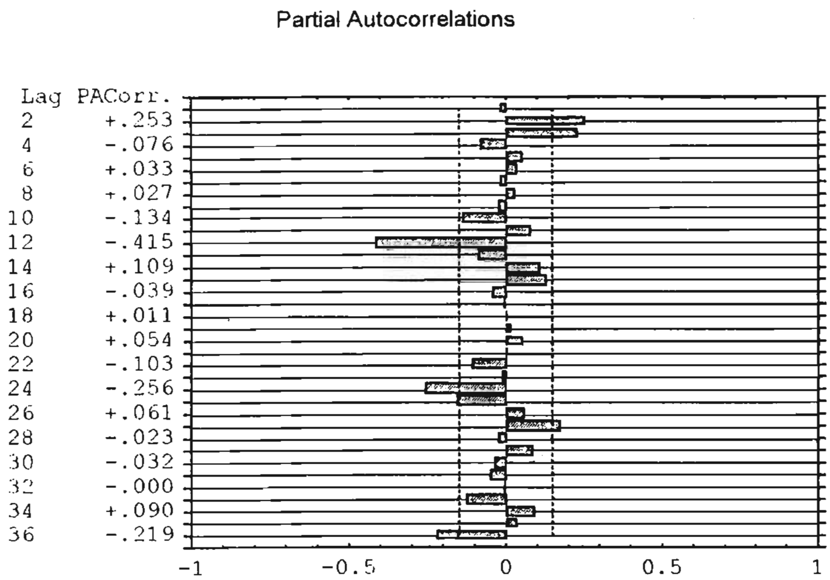


Figure 3. 3. 11 Municipality C : PACF of $\nabla_{12} Y_t$



The intervention events expected to have an impact on the electricity consumption for Municipality C are summarised in Table 3. 3. 7. The two periods of water restrictions were modelled as separate interventions because the severity of the restrictions differed.

Table 3. 3. 7 Municipality C : Summary of intervention events

INTERVENTION SERIES	PARAMETER	DESCRIPTION
$I_{1,t} = \begin{cases} 1 & t = Jan'83 -> Mar'84 \\ 0 & all\ other\ months \end{cases}$	λ_1	Water restrictions between January 198 and March 1984.
$I_{2,t} = \begin{cases} 1 & t = Jul'91 \\ 0 & all\ other\ months \end{cases}$	λ_2	There was a long billing month of 40 days i July 1991 when the meter reading syster changed from manual to electronic.
$I_{3,t} = \begin{cases} 1 & t = Jan'80 - Jul'93 \\ 0 & all\ other\ months \end{cases}$	λ_3	In August 1993 a large mine just outside th municipality's area of supply closed dow permanently.
$I_{4,t} = \begin{cases} 1 & t = Aug'93 - Jan'94 \\ 0 & all\ other\ months \end{cases}$	λ_4	Water restrictions between August 199 and January 1994.

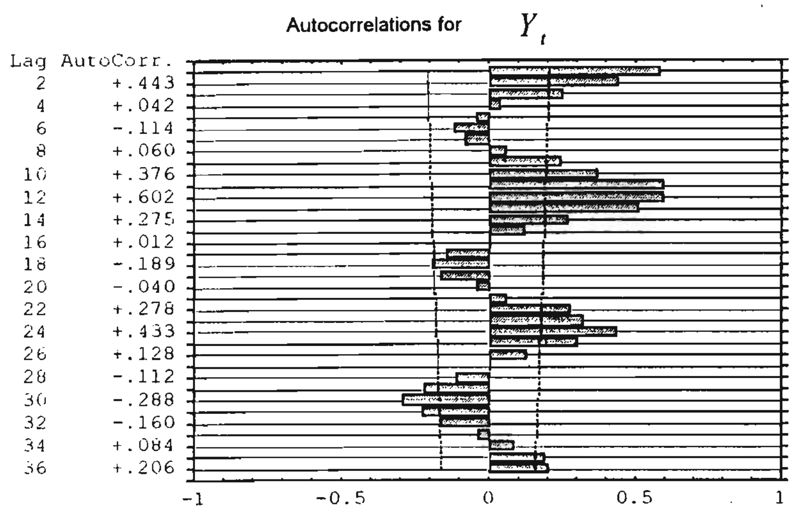
A suitable ARIMA model was developed for the time series unaffected by any interventions, i.e. for the sub-series from April 1984 to June 1991. The model ARIMA(1,0,0)x(1,1,0)$_{12}$ as identified from the ACF and PACF given in Figure 3.3.12 and the model ARIMA(1,0,0)x(2,1,0)$_{12}$ was identified by systematically reducing an over-parameterised model. The results for fitting both models are summarised in Table 3. 3. 8 and clearly there is very little difference, the former performing better according to the SBC statistic and the latter model resulting in a smaller AIC statistic. The ARIMA(1,0,0)x(2,1,0)$_{12}$ model was taken as the most suitable since the AIC statistic is more commonly used than the SBC statistic Thus the model ARIMA(1,0,0)x(2,1,0)$_{12}$ was used in conjunction with the intervention events specified earlier and the results are given in Figure 3. 3. 9. A disturbing feature is that the parameter associated with the mine closure was estimated to be negative, but is expected to be positive. Since this parameter is just significantly different from zero at the 5% level it was therefore decided to remove it from the model. The final results are given in Table 3. 3. 9. A noticeable problem with the residual errors is highlighted by the portmanteau statistic which indicates that the residual errors are not white noise, and this is illustrated in a plot of the ACF of the residual error given in Figure 3. 3. 14. As a point of interest the model ARIMA(1,0,0)x(1,1,0)$_{12}$ including interventions also resulted in similar problems and since no other suitable model could be fitted, the model ARIMA(1,0,0)x(2,1,0)$_{12}$ including interventions was taken as the best fitting model.

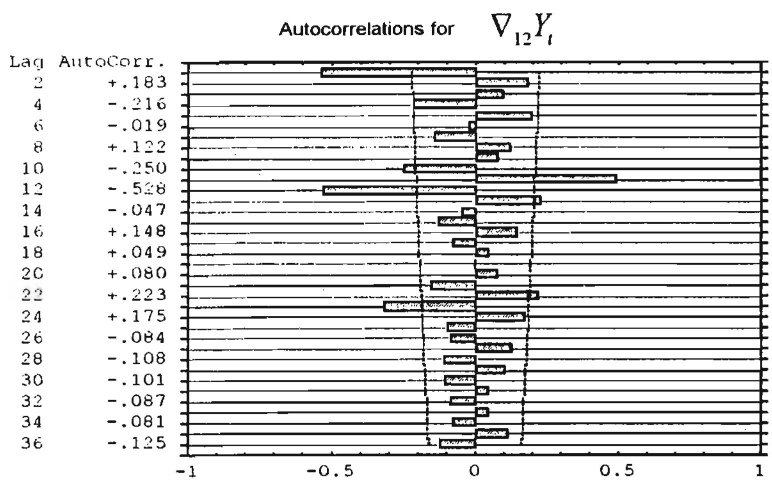
This model can be represented by

$$W_t = Z_t - 0.12390W_{t-1} - 0.54892W_{t-12} - 0.28705W_{t-24} - 0.06801W_{t-13} - 0.03556W_{t-25} + 0.20702 - 0.40136I_{1,t} + 1.03735I_{2,t} - 0.43653I_{4,t}$$

where $W_t = \nabla_{12} Y_t$.

Figure 3. 3. 12 Municipality C : ACF of Y_t and $\nabla_{12}Y_t$ and PACF of $\nabla_{12}Y_t$ resulting from the time series unaffected by intervention events.





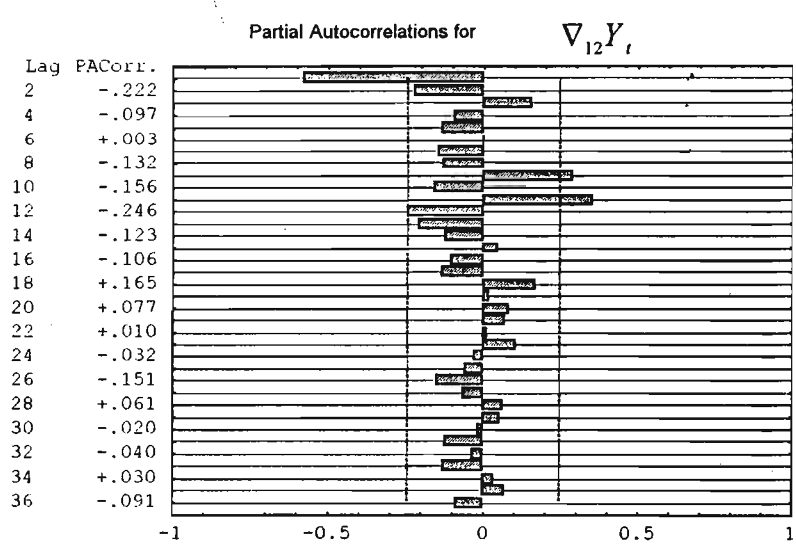


Table 3. 3. 8 Municipality C : Comparison of Model Results fitted to the time series unaffected by intervention events

ARIMA(1,0,0)x(1,1,0) $_{12}$				ARIMA(1,0,0)x(2,1,0) $_{12}$			
Parameter	**Estimate**	**t ratio**		**Parameter**	**Estimate**	**t ratio**	
δ	0.24502	9.52		δ	0.24760	12.32	
ϕ_1	-0.42400	-3.84		ϕ_1	-0.46106	-4.37	
Φ_1	-0.47064	-4.36		Φ_1	-0.59602	-4.94	
				Φ_2	-0.23774	-1.99	
The Portmanteau test for white noise :							
Lags	**Chi Square**	**DF**	**P-value**	**Lags**	**Chi Square**	**DF**	**P-value**
1-6	4.56	4	0.335	1-6	5.76	3	0.124
1-12	16.08	10	0.097	1-12	14.94	9	0.093
1-18	20.68	16	0.191	1-18	18.00	15	0.263
1-24	31.94	22	0.078	1-24	23.81	21	0.303
Model comparison statistics :							
AIC	**SBC**			**AIC**	**SBC**		
96.7852	103.6974			95.52057	104.7905		

Figure 3. 3. 13 Municipality C : Residual errors when fitting an ARIMA(2,0,1)x(0,1,1)$_{12}$ model to the time series unaffected by intervention events

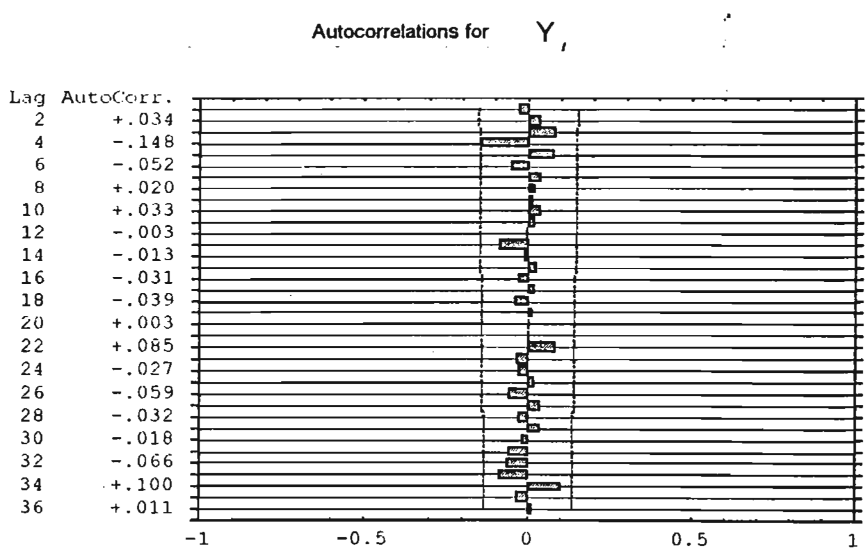


Table 3. 3. 9 Municipality C : Parameter estimates when fitting an ARIMA(1,0,0)x(2,1,0)$_{12}$ model including Interventions

Parameter	Estimate	t ratio	Parameter	Estimate	t ratio
δ	0.50436	3.52	λ_1	-0.75120	-3.12
ϕ_1	-0.16103	-2.04	λ_2	1.17877	2.61
Φ_1	-0.50672	-6.22	λ_3	-0.30749	-2.09
Φ_2	-0.24856	-3.01	λ_4	-0.42672	-4.03

Table 3. 3. 10 Municipality C : Results when fitting an ARIMA(1,0,0)x(2,1,0)$_{12}$ model including intervention events

Parameter estimates using MLE :

Parameter	Estimate	t ratio	Parameter	Estimate	t ratio
δ	0.20702	8.54	λ_1	-0.40136	-2.19
ϕ_1	-0.12390	-1.58	λ_2	1.03735	2.33
Φ_1	-0.54892	-6.87	λ_4	-0.43653	-4.09
Φ_2	-0.28705	-3.57			

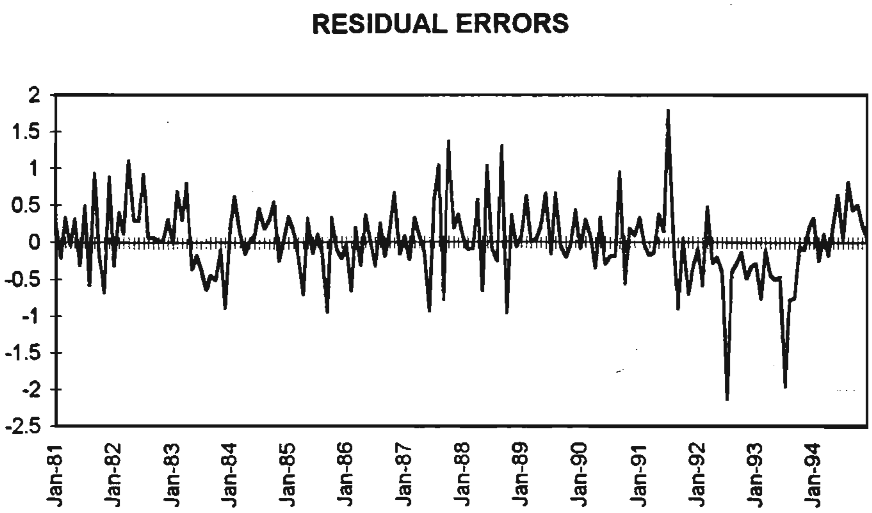
Portmanteau test for white noise :

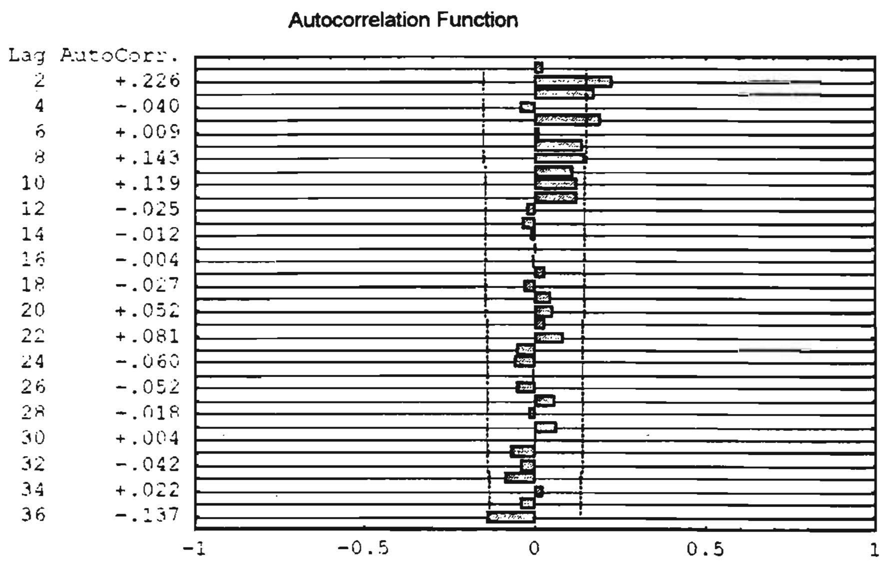
Lags	Chi Square	DF	P-value
1-6	25.65	3	0.000
1-12	34.46	9	0.000
1-18	35.98	15	0.002
1-24	38.92	21	0.010
1-30	41.88	27	0.034

Model comparison statistics : AIC = 272.992 SBC = 294.860

Test set forecasting results : M.S.E.(F) = 0.065 M.A.P.E. = 2.13%

Figure 3. 3. 14 Municipality C : Residual errors resulting from fitting an ARIMA(1,0,0)x(2,1,0)$_{12}$ model including Interventions





The ARIMA(2,0,1)x(0,1,1)$_{12}$ derived when ignoring intervention events and the model ARIMA(1,0,0)x(2,1,0)$_{12}$ including intervention events were both evaluated by forecasting the test set. The results of this are given in Table 3. 3. 11 and it can clearly be seen that the incorporation of interventions improves the model.

Table 3. 3. 11 Municipality C : Comparison of results

MODEL	M.A.P.E.(F)	M.S.E.(F)
$(2,0,1)x(0,1,1)_{12}$	3.79%	0.165
$(1,0,0)x(2,1,0)_{12}$ + interventions	2.13%	0.065

3.4.4 MUNICIPALITY D

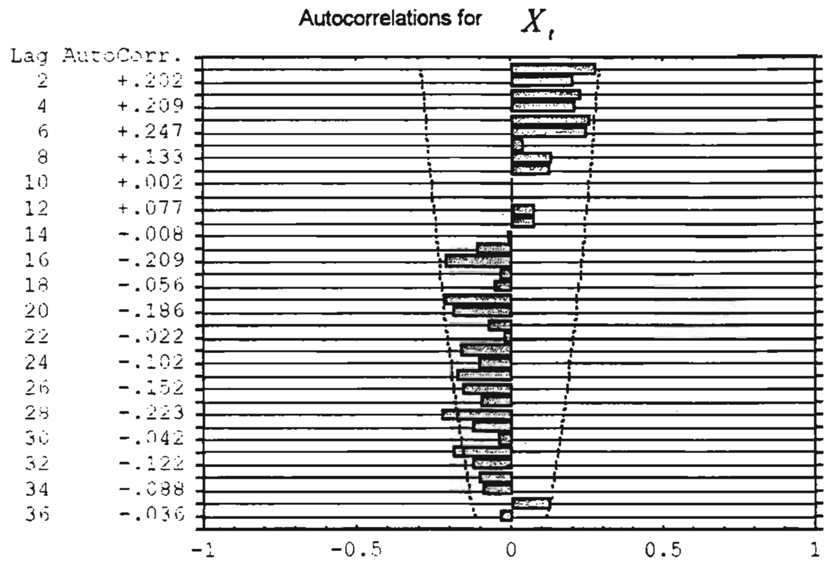
One large factory has a dominating effect on the monthly electricity consumption for Municipality D and thus two time series were modelled separately, one consisting of the factory's electricity consumption and the other the electricity consumption of the municipality excluding the factory. Only the portion of the time series for the factory from June 1992 onwards, when a new production process was introduced, was used in the modelling process. This time series, which consists of only 31 data points, is fairly short. However it is nonseasonal and the modelling results appear to be satisfactory.

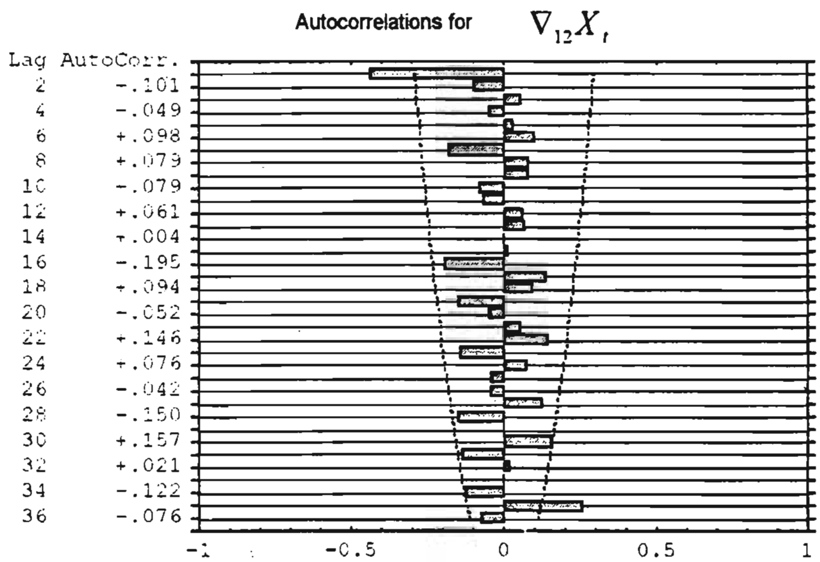
Let X_t represent the non-seasonal time series of monthly electricity consumption for the factory. The ACF's of X_t and ∇X_t as well as the PACF of ∇X_t are shown in Figure 3. 3. 15. Clearly first differencing is enough to ensure that the series is stationary and the model will be of type ARIMA(p, 1, q). The most appropriate ARIMA model was identified as the ARIMA(2,1,0) written as

$$W_t = -0.35718W_{t-1} - 0.43439W_{t-2} + Z_t$$

where $W_t = \nabla_1 Y_t$, and the results of the fitting process are summarised in Table 3. 3. 12.

Figure 3. 3. 15 Factory : ACF's of X_t and ∇X_t and PACF of ∇X_t





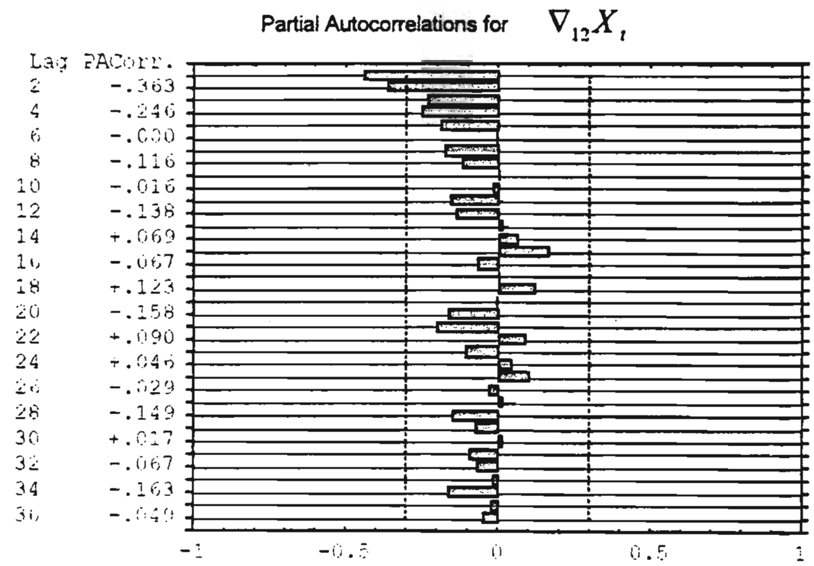


Table 3. 3. 12 Factory : Results when fitting an ARMA(2,1,0) model

Parameter estimates using MLE :

Parameter	Estimate	t ratio
ϕ_1	-0.35718	-2.14
ϕ_2	-0.43439	-2.60

Model comparison statistics : AIC = 157.438 SBC = 160.240

Portmanteau test for white noise :

Lags	Chi Square	DF	P-value
1-6	5.14	4	0.248
1-12	7.48	10	0.680
1-18	19.65	16	0.237
1-24	20.84	22	0.530

A model was also developed for the time series Y_t, the monthly electricity consumption for the Municipality D excluding the factory. The ACF's of Y_t , $\nabla_{12}Y_t$ and $\nabla\nabla_{12}Y_t$, given in Figure 3. 3. 16, indicate that the model is seasonal and of the form ARIMA(p,0,q)x(P,1,Q)$_{12}$. In fact the pattern of the ACF and the PACF of the differenced series given in Figures 3. 3. 16 and 3. 3. 17 respectively, suggest that an appropriate model is ARIMA(2,0,1)x(1,1,1)$_{12}$. The results associated with fitting this model appear in Table 3. 3. 14 and the fitted model can be written as

$$W_t = 0.70904W_{t-1} + 0.29095W_{t-2} + 0.42901W_{t-12} - 0.32448W_{t-13} - 0.12482W_{t-14}$$
$$+ Z_t - 0.75634Z_{t-1} - 0.99338Z_{t-12} + 0.75133Z_{t-13}$$

where $W_t = \nabla_{12}Y_t$.

The test set for the time series of monthly electricity consumption for the factory and Municipality D excluding the factory were forecast using the two models chosen and the results are given in Table 3. 3. 13.

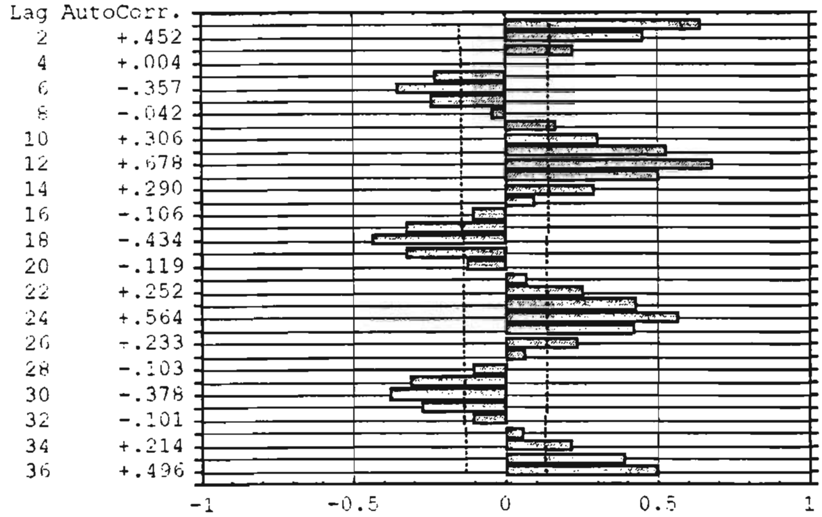
Table 3. 3. 13 Factory and Municipality D excluding Factory : Forecasting errors

DATA	M.A.P.E.(F)	M.S.E.(F)
FACTORY : (2,1,0)	12.25%	17.60
MUNICIPALITY : $(2,0,1)x(1,1,1)_{12}$	8.58%	5.20

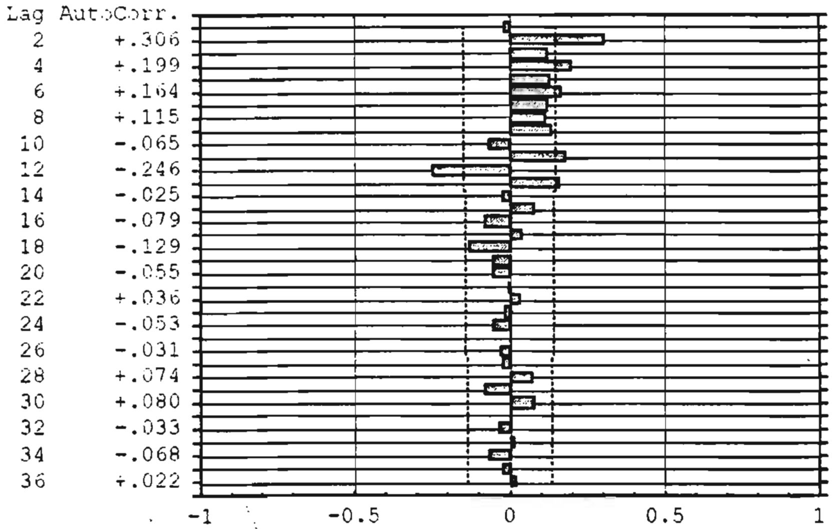
Note that a plant fault at the factory in November 1995 caused a drop in consumption which the forecast could not have predicted. As a consequence the forecasting errors are large.

Figure 3. 3. 16 Municipality D : ACF's of Y_t, $\nabla_{12}Y_t$ and $\nabla\nabla_{12}Y_t$

Autocorrelations for Y_t



Autocorrelations for $\nabla_{12}Y_t$



Autocorrelations for $\nabla\nabla_{12}Y_t$

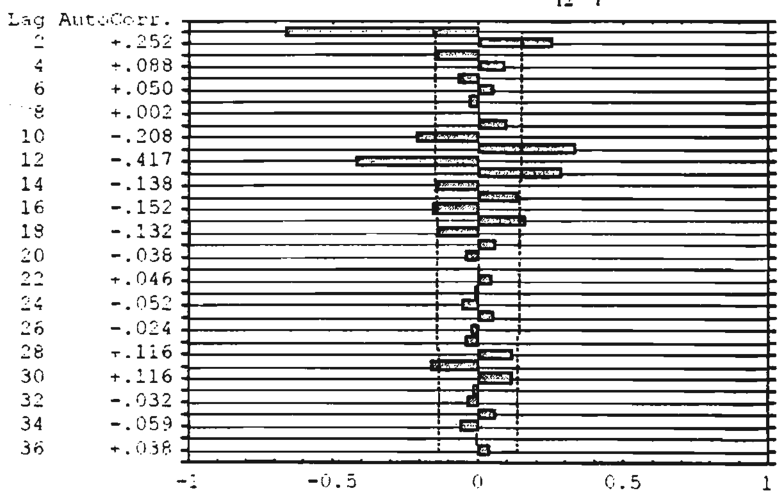


Figure 3. 3. 17 Municipality D : PACF of $\nabla_{12}Y_t$

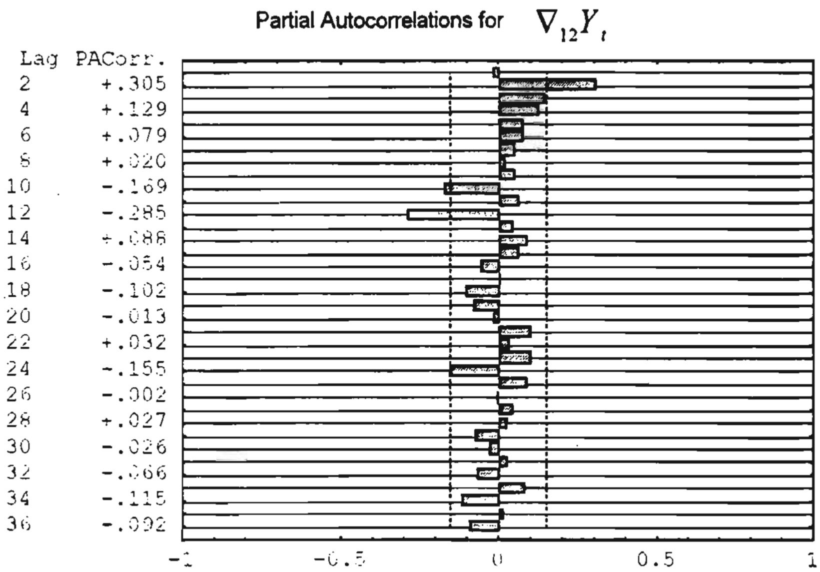


Table 3. 3. 14 Municipality D excluding the factory : Results when fitting an ARIMA(2,0,1)x(1,1,1)$_{12}$ model

Parameter Estimates using MLE :

Parameter	Estimate	t ratio
θ_1	0.75634	12.61
Θ_1	0.99338	53.28
ϕ_1	0.70904	8.59
ϕ_2	0.29095	3.53
Φ_1	0.42901	5.40

Portmanteau test for white noise :

Lags	Chi Square	DF	P-value
1-6	0.53	1	0.465
1-12	5.44	7	0.607
1-18	13.07	13	0.443
1-24	16.34	19	0.635
1-30	18.31	25	0.829

Model comparison statistics : AIC = 578.016 SBC = 593.635

Test set forecasting results : M.S.E.(F) = 5.20 M.A.P.E.(F) = 8.58%

3.4 STATE SPACE MODELS

Two basic structural models were fitted to each of the time series in this study, one with dummy seasonal components and the other with trigonometric seasonal components. For each model various approaches were taken to find optimal estimates of the state vector α_t, t = 1,...,T. The simplest of these was to assume starting values of μ = 0 and C_0 = 100 000 I, where I is the identity matrix, to fix the parameters as σ_ε^2 = 5, $\sigma_\eta^2 = \sigma_\zeta^2 = \sigma_\omega^2$ = 0.1 and to apply the Kalman filtering equations to find a minimum mean square estimate of α_t. The results of this method are denoted by $KF^{(1)}$ in the ensuing tables. In a second approach, the starting values of μ = 0 and C_0 = 100 000 I were held fixed and maximum likelihood estimates of the parameters σ_ε^2, σ_η^2, σ_ζ^2 and σ_ω^2 were derived using the Kalman filter. Two different techniques for obtaining these estimates, the one involving direct maximisation, and the other the EM algorithm were used and the results of these methods are denoted by $KF^{(2)}$ and EM respectively in the later tables. A further enhancement was the inclusion of a maximum likelihood estimate of $\hat{\alpha}_0$ and the results for this are denoted by $KF^{(3)}$.

The procedures described above were implemented using programs written in the GAUSS language. The GAUSS function OPTMUM was invoked in the direct maximisation calculations. This routine uses a convergence criterion based on the change of gradients, whereas convergence within the EM algorithm was assumed when changes in the likelihood function with each iteration were less than 0.0001. The first iteration of the Kalman Filter was ignored in all calculations of the likelihood function.

The fitted models were used to forecast the observations of the test set and the results were compared using the criteria

$$\text{M.S.E.(F)} = \left(\frac{1}{12}\right)\sum_{t=T-1}^{T-12}\left(Y_t - Y_{t:T}\right)^2$$

and $\quad$ M.A.P.E.(F) = $\left(\frac{1}{12}\right)\sum_{t=T+1}^{T+12}\frac{\left|Y_t-\hat{Y}_{t|T}\right|}{Y_t}$ $\quad$ as defined previously.

3.4.1 MUNICIPALITY A

Basic structural models with dummy and also with trigonometric seasonal components were fitted to the time series of monthly electricity consumption for Municipality A and the results, including estimates of the unknown parameters, are summarised in Table 3. 4. 1.

Table 3. 4. 1 Municipality A : Results for BSMs fitted to the complete time series

	$-\ln L(\theta\|Y_1,\ldots Y_T)$	θ	$\hat{\sigma}_\varepsilon^2$	$\hat{\sigma}_\eta^2$	$\hat{\sigma}_\varsigma^2$	$\hat{\sigma}_\omega^2$	M.S.E.(F)	M.A.P.E.(F)
BSM with dummy seasonality :								
KF $^{(1)}$	573.902		5.000	0.100	0.100	0.100	6.614	1.92%
KF $^{(2)}$	**546.693**	$\sigma_\varepsilon^2,\Sigma$	**8.778**	**0.460**	**0.000**	**0.251**	**6.520**	**1.90%**
KF $^{(3)}$	546.693	$\sigma_\varepsilon^2,\Sigma,\mu$	8.778	0.460	0.000	0.251	6.520	1.90%
EM	546.793	$\sigma_\varepsilon^2,\Sigma$	8.767	0.466	0.000	0.251	6.514	1.90%
BSM with trigonometric seasonality :								
KF $^{(1)}$	581.760		5.000	0.100	0.100	0.100	11.376	2.54%
KF $^{(2)}$	**556.085**	$\sigma_\varepsilon^2,\Sigma$	**8.729**	**0.413**	**0.000**	**0.008**	**6.723**	**1.81%**
KF $^{(3)}$	556.085	$\sigma_\varepsilon^2,\Sigma,\mu$	8.729	0.413	0.000	0.008	6.723	1.81%
EM	556.612	$\sigma_\varepsilon^2,\Sigma$	8.360	0.427	0.000	0.013	6.849	1.84%

The likelihood function converged more quickly when maximising directly as opposed to using the EM algorithm and in general provided smaller values of the likelihood function indicating that better estimates of the unknown parameters were derived. The value of μ had very little effect on the Kalman filtering results unless it was taken to be extremely large, thus KF $^{(2)}$ and KF $^{(3)}$ give identical results throughout this study.

Comparing the results of the BSM with dummy and trigonometric seasonal components, where the parameters were derived using the method of direct maximisation, the former model was found to be a better fit according to the criteria M.S.E.(F) whereas the latter model performed better when using the criteria M.A.P.E.(F). Obviously this indicates that there is not much difference between the models, and either would be acceptable. For the purposes of this study, the former model which is simpler was adopted. From the final estimate of the state vector derived using this model, the linear trend is given by $\mu_T = 103.888$, the slope is $\beta_T = 0.284$ and the seasonal components are given by

Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
-11.095	-0.011	3.858	0.258	7.041	8.391	6.509	1.314	-5.815	1.211	-7.310	-4.351

where the seasonal component for December is calculated using $\gamma_{12} = -\sum_{j=1}^{11} \gamma_{T-j}$. The large negative seasonal component in January reflects the annual closure during the festive season of many factories within the municipal boundaries. These results are similar to those of the Holt-Winters method where the level and trend components were found to be $L_T =$ 102.305 and T_T = 0.270 respectively. It is interesting to note that even though the parameters derived from the EM algorithm resulted in a larger likelihood function than when using parameters derived using direct maximisation, it was purely by chance that the BSM with dummy seasonal components with these parameters resulted in the smallest criterion value M.S.E.(F).

The time series Y_t can be decomposed into the four component series of level, trend, seasonality and error for t = 1,...,T. The decomposition for the BSM with dummy seasonal components and parameter estimates $\hat{\sigma}_\varepsilon^2$ = 8.778 , $\hat{\sigma}_\eta^2$ = 0.460, $\hat{\sigma}_\varsigma^2$ = 0 and $\hat{\sigma}_\omega^2$ = 0.251 is illustrated in Figure 3. 4. 1 and the residual series is shown in Figure 3. 4. 2. The high residual value in January 1989 is, as mentioned previously, due to an unusually long billing month of 34 days and the low value associated with January 1990 coincides with the installation of an electronic metering system which resulted in a short billing month. Otherwise the residuals appear to be random indicating that the BSM has captured the systematic variation of the original time series.

Figure 3. 4. 1 Municipality A : Decomposition of the time series for BSM with dummy seasonal components

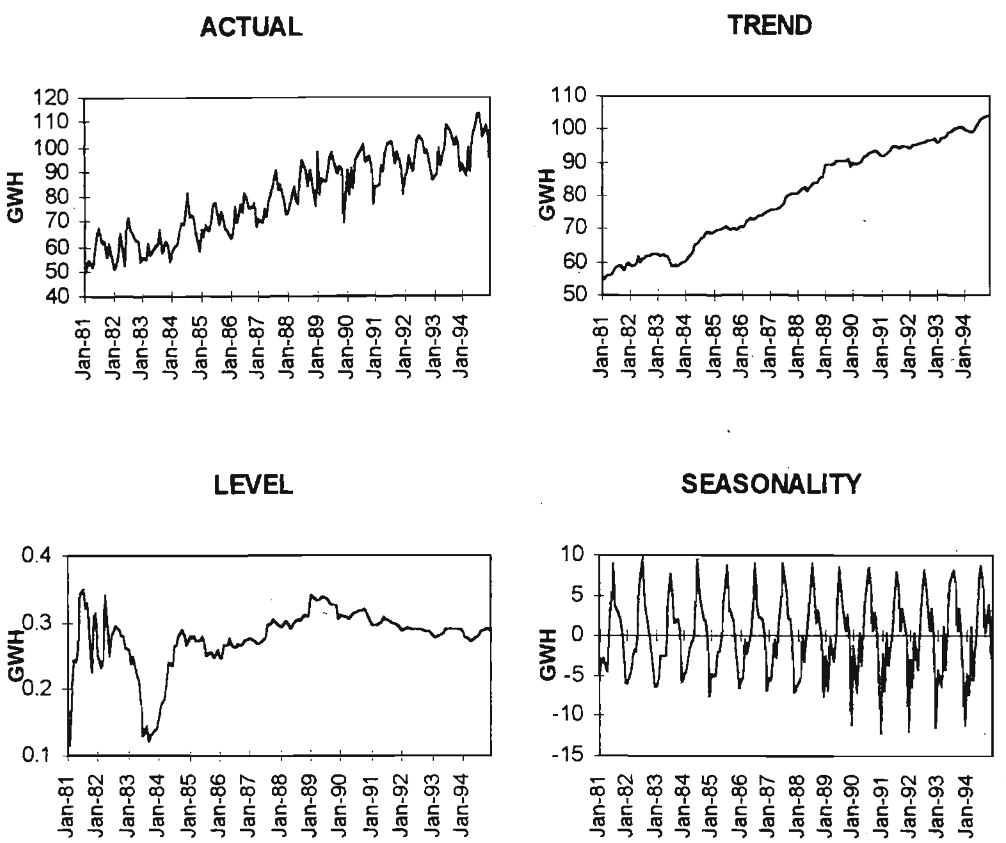
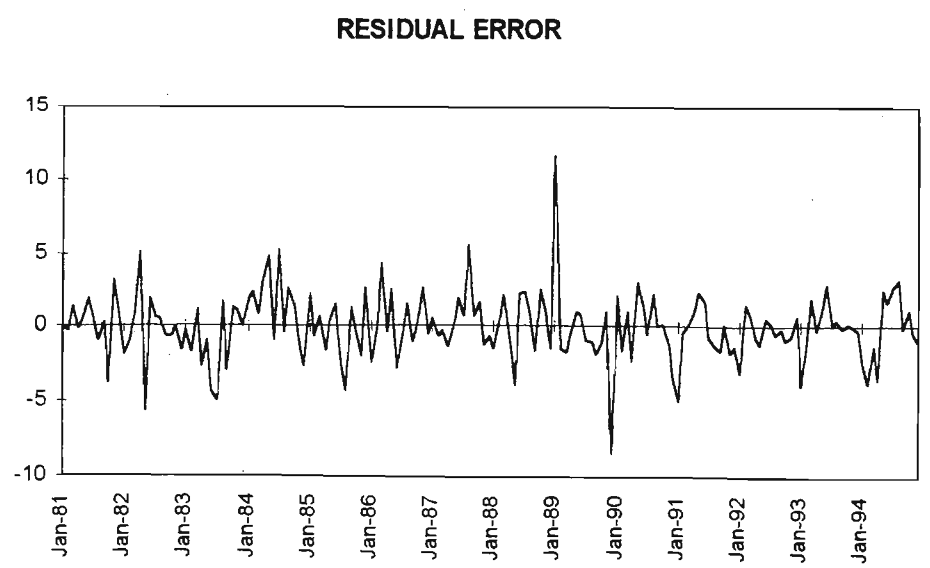


Figure 3. 4. 2 Municipality A : Residual errors for BSM with dummy seasonal components



The sub-series of monthly electricity consumption of Municipality A between January 1990 and December 1994, when the meters were read electronically, was again modelled separately to investigate whether or not this would improve the forecasting results. The results given in Table 3. 4. 2 as compared with those of Table 3. 4. 1 indicate that overall better forecasts were derived using the whole time series. However, it is interesting to observe that the estimated variances for the shorter series are generally smaller than those obtained for the full series, indicating that regular metering periods have a stabilising effect on the time series.

Table 3. 4. 2 Municipality A : Results for BSMs fitted to the electronically metered time series

	$-\ln L(\theta \vert Y_1,...Y_T)$	θ	$\hat{\sigma}_\varepsilon^2$	$\hat{\sigma}_\eta^2$	$\hat{\sigma}_\varsigma^2$	$\hat{\sigma}_\omega^2$	M.S.E.(F)	M.A.P.E.(F)
BSM with dummy seasonality :								
KF $^{(1)}$	203.971		5.000	0.100	0.100	0.100	9.555	2.01%
KF $^{(2)}$	**198.242**	$\sigma_\varepsilon^2, \Sigma$	**4.347**	**0.000**	**0.001**	**0.000**	**6.745**	**1.84%**
KF $^{(3)}$	198.242	$\sigma_\varepsilon^2, \Sigma, \mu$	4.347	0.000	0.001	0.000	6.745	1.84%
EM	198.296	$\sigma_\varepsilon^2, \Sigma$	4.288	0.012	0.001	0.022	6.736	1.85%
BSM with trigonometric seasonality :								
KF $^{(1)}$	220.708		5.000	0.100	0.100	0.100	11.387	2.56%
KF $^{(2)}$	**207.135**	$\sigma_\varepsilon^2, \Sigma$	**2.590**	**0.000**	**0.000**	**0.029**	**11.085**	**2.59%**
KF $^{(3)}$	207.135	$\sigma_\varepsilon^2, \Sigma, \mu$	2.590	0.000	0.000	0.029	11.085	2.59%
EM	212.763	$\sigma_\varepsilon^2, \Sigma$	0.062	0.011	0.000	0.194	19.697	3.34%

3.4.2 MUNICIPALITY B

The results of modelling the time series of monthly electricity consumption for Municipality B are summarised in Table 3. 4. 3. In contrast to the results for Municipality A, the BSM with trigonometric seasonality provided better forecasts than the BSM with dummy seasonal components, as measured by the criteria of M.S.E.(F) and M.A.P.E.(F).

Table 3. 4. 3 Municipality B : Results for BSM fitted to the complete time series

	$-\ln L(\theta \mid Y_1,\ldots Y_T)$	θ	$\hat{\sigma}_\varepsilon^2$	$\hat{\sigma}_\eta^2$	$\hat{\sigma}_\varsigma^2$	$\hat{\sigma}_\omega^2$	M.S.E.(F)	M.A.P.E.(F)
BSM with dummy seasonality :								
KF (1)	450.666		5.000	0.100	0.100	0.100	2.077	6.58%
KF (2)	**331.605**	$\sigma_\varepsilon^2, \Sigma$	**0.544**	**0.028**	**0.000**	**0.072**	**0.405**	**2.78%**
KF (3)	331.605	$\sigma_\varepsilon^2, \Sigma, \mu$	0.544	0.028	0.000	0.072	0.405	2.78%
EM	331.898	$\sigma_\varepsilon^2, \Sigma$	0.548	0.025	0.000	0.071	0.408	2.79%
BSM with trigonometric seasonality :								
KF (1)	519.653		5.000	0.100	0.100	0.100	0.491	2.96%
KF (2)	**343.81**	$\sigma_\varepsilon^2, \Sigma$	**0.614**	**0.026**	**0.000**	**0.001**	**0.385**	**2.70%**
KF (3)	343.81	$\sigma_\varepsilon^2, \Sigma, \mu$	0.614	0.026	0.000	0.001	0.385	2.70%
EM	344.697	$\sigma_\varepsilon^2, \Sigma$	0.628	0.014	0.000	0.001	0.393	2.72%

Again the sub-series of electricity consumption for Municipality B, when the meters were read electronically, was modelled separately to ascertain whether or not this would result in better forecasts. It is clear from Table 3. 4. 4 that better forecasts were not obtained. It is again interesting to observe that all the estimated variances decreased for this more regular time series.

Table 3. 4. 4 Municipality B : Results for BSMs fitted to the electronically metered time series

	$-\ln L(\theta \mid Y_1,\ldots Y_T)$	θ	$\hat{\sigma}_{\varepsilon}^2$	$\hat{\sigma}_{\eta}^2$	$\hat{\sigma}_{\varsigma}^2$	$\hat{\sigma}_{\omega}^2$	M.S.E.(F)	M.A.P.E.(F)
BSM with dummy seasonality :								
KF $^{(1)}$	189.258		5.000	0.100	0.100	0.100	0.505	2.61%
KF $^{(2)}$	**142.316**	$\sigma_{\varepsilon}^2, \Sigma$	**0.324**	**0.000**	**0.000**	**0.036**	**0.444**	**2.82%**
KF $^{(3)}$	142.316	$\sigma_{\varepsilon}^2, \Sigma, \mu$	0.324	0.000	0.000	0.036	0.444	2.82%
EM	142.396	$\sigma_{\varepsilon}^2, \Sigma$	0.301	0.006	0.000	0.042	0.411	2.76%
BSM with trigonometric seasonality :								
KF $^{(1)}$	213.080		5.000	0.100	0.100	0.100	0.490	2.96%
KF $^{(2)}$	**151.673**	$\sigma_{\varepsilon}^2, \Sigma$	**0.343**	**0.000**	**0.000**	**0.001**	**0.471**	**2.93%**
KF $^{(3)}$	151.673	$\sigma_{\varepsilon}^2, \Sigma, \mu$	0.343	0.000	0.000	0.001	0.471	2.93%
EM	151.969	$\sigma_{\varepsilon}^2, \Sigma$	0.288	0.006	0.000	0.002	0.448	2.88%

3.4.3 MUNICIPALITY C

The time series of monthly electricity consumption for Municipality C was clearly affected by a number of intervention events as described earlier. To monitor the improvements gained by including these intervention events into the modelling process, the time series was firstly modelled using the BSM with dummy seasonal components and excluding intervention events and the results are summarised in Table 3. 4. 5. Thereafter, the time series was modelled incorporating the intervention events of water restriction periods between January 1983 and March 1984 and again between August 1993 and January 1994, the permanent closure of a mine on the outskirts of the municipality's supply area, and a period of 40 days between meter readings in July 1991. These interventions and the associated parameters are summarised in Table 3. 4. 6.

Table 3. 4. 5 Municipality C : Results for BSMs

	$-\ln L(\theta \vert Y_1,...Y_T)$	θ	$\hat{\sigma}_{\varepsilon}^2$	$\hat{\sigma}_{\eta}^2$	$\hat{\sigma}_{\varsigma}^2$	$\hat{\sigma}_{\omega}^2$	M.S.E.(F)	M.A.P.E.(F)
BSM with dummy seasonality :								
KF [(1)]	442.108		5.000	0.100	0.100	0.100	0.533	7.78%
KF [(2)]	**210.948**	$\sigma_{\varepsilon}^2, \Sigma$	**0.181**	**0.004**	**0.000**	**0.002**	**0.161**	**3.82%**
KF [(3)]	210.948	$\sigma_{\varepsilon}^2, \Sigma, \mu$	0.181	0.004	0.000	0.002	0.161	3.82%
EM	210.978	$\sigma_{\varepsilon}^2, \Sigma$	0.180	0.005	0.000	0.002	0.162	3.86%

Table 3. 4. 6 Municipality C : Summary of intervention events

INTERVENTION SERIES	PARAMETER	DESCRIPTION
$I_{1,t} = \begin{cases} 1 & t = Jan'83 -> Mar'84 \\ 0 & all\ \ other\ \ months \end{cases}$	λ_1	Water restrictions between January 1983 and March 1984.
$I_{2,t} = \begin{cases} 1 & t = Jul'91 \\ 0 & all\ \ other\ \ months \end{cases}$	λ_2	There was a long billing month of 40 days in July 1991 when the meter reading system changed from manual to electronic.
$I_{3,t} = \begin{cases} 1 & t = Jan'80 - Jul'93 \\ 0 & all\ \ other\ \ months \end{cases}$	λ_3	In August 1993 a large mine just outside the municipality's area of supply closed down permanently.
$I_{4,t} = \begin{cases} 1 & t = Aug'93 - Jan'94 \\ 0 & all\ \ other\ \ months \end{cases}$	λ_4	Water restrictions between August 1993 and January 1994.

Estimates of the intervention parameters, together with the t-ratios for testing whether or not the corresponding true parameters are equal to zero, are given in Table 3.4.7.

Table 3. 4. 7 Municipality C : Estimates of the intervention parameters

PARAMETER	VALUE	T-RATIO
λ_1	-0.364	-2.401
λ_2	2.850	6.943
λ_3	-0.280	-1.262
λ_4	-0.426	-2.044

Clearly λ_3, the intervention parameter associated with the mine closure, is again negative and has a non-significant t-ratio suggesting that this parameter can be dropped from the model. The results excluding this intervention are given in Table 3. 4. 8. Overall, it is clear that the BSM with dummy seasonal components and including the intervention events is the best model and that satisfactory estimates of the variance parameters are derived using direct maximisation. From the final state vector, the trend is given by μ_T = 7.807, the slope is β_T = 0.014 and the seasonal components are given by

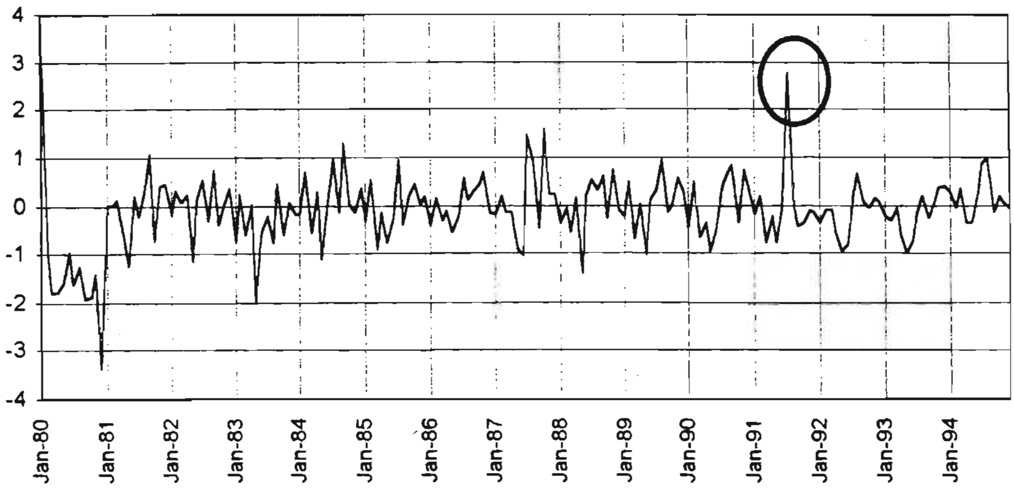
Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
-0.517	-0.175	0.079	0.082	0.927	1.038	0.976	-0.033	-0.581	-0.507	-0.753	-0.536

The last three values in the state vector $\hat{\alpha}_T$ pertain to the intervention events and indicate that the two water restriction periods had the effect of reducing electricity consumption by 0.356 and 0.359 GWh respectively and that the longer billing period in July 1992 increased the consumption by 2.939 GWh. A comparison of plots of the residual errors for the BSM excluding and including intervention events is given in Figure 3. 4. 3 and illustrates the improvement derived from including these interventions in the model.

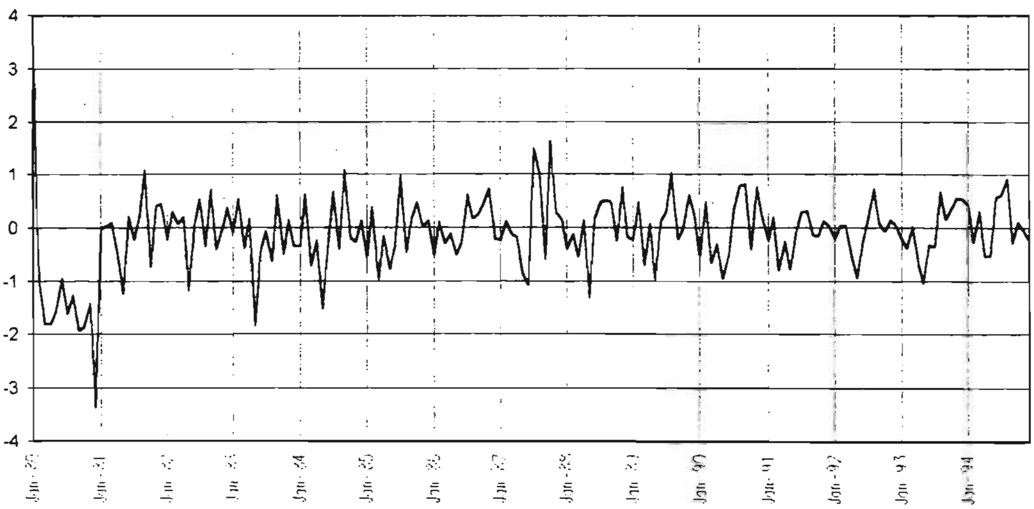
Table 3. 4. 8 Municipality C : Results for BSMs including intervention events

	$-\ln L(\theta\|Y_1,\dots Y_T)$	θ	σ^2_ε	σ^2_η	σ^2_ς	σ^2_ω	λ_1 (t-ratio)	λ_2 (t-ratio)	λ_4 (t-ratio)	M.S.E.(F)	M.A.P.E.(F)
BSM with dummy seasonality :											
KF $^{(2)}$	**205.917**	$\sigma^2_\varepsilon, \Sigma$	**0.125**	**0.003**	**0.000**	**0.004**	**-0.356** **(-2.445)**	**2.939** **(7.424)**	**-0.359** **(-1.934)**	**0.147**	**4.23%**
KF $^{(3)}$	205.917	$\sigma^2_\varepsilon, \Sigma, \mu$	0.125	0.003	0.000	0.004	-0.356 (-2.445)	2.939 (7.424)	-0.359 (-1.934)	0.147	4.23%
EM	213.655	$\sigma^2_\varepsilon, \Sigma$	0.213	0.001	0.000	0.001	-0.358 (-2.193)	2.903 (5.919)	-0.364 (-1.653)	0.180	4.42%
BSM with trigonometric seasonality :											
KF $^{(2)}$	**213.668**	$\sigma^2_\varepsilon, \Sigma$	**0.150**	**0.002**	**0.000**	**0.000**	**-0.364** **(-2.402)**	**2.850** **(6.934)**	**-0.333** **(-1.706)**	**0.176**	**4.32%**
KF $^{(3)}$	213.668	$\sigma^2_\varepsilon, \Sigma, \mu$	0.150	0.002	0.000	0.000	-0.364 (-2.402)	2.850 (6.934)	-0.333 (-1.706)	0.176	4.32%
EM	226.163	$\sigma^2_\varepsilon, \Sigma$	0.214	0.001	0.000	0.000	-0.326 (-1.747)	2.916 (5.790)	-0.396 (-1.738)	0.155	4.08%

Figure 3. 4. 3 Municipality C : Residual errors for the BSM with dummy seasonality



(a) Excluding intervention events



(b) Including intervention events

3.4.4 MUNICIPALITY D

Two separate time series involving the monthly electricity consumption of Municipality D, one consisting of the monthly electricity consumption of the municipality excluding that of the large factory within the municipality's area of supply and the other, the monthly electricity consumption for the factory, were considered. Basic structural models with dummy and also with trigonometric seasonal components were fitted to the former time series. Only the portion of time series of the factory's monthly electricity consumption from June 1992 onwards was used for modelling purposes, as discussed previously in Section 3. 3. 4 and, since this series displays no seasonality, the local linear trend model of Section 2. 4. 2 was invoked. The results are summarised in Tables 3. 4. 9 and 3. 4. 10. It should be noted that in November 1995, equipment failure at the factory caused an unexpected decrease in electricity consumption, resulting in a large forecasting error for that month and hence for the test set. It is thus only by coincidence that the Kalman filtering with fixed parameter values, $KF^{(1)}$, produces the best test set forecast according to the criteria M.S.E.(F), since the test set for the factory's monthly electricity consumption does not represent the usual electricity consumption pattern.

Table 3. 4. 9 Municipality D excluding factory : Results for the BSM

	$-\ln L(\theta \mid Y_1,\ldots Y_T)$	θ	$\hat{\sigma}_{\varepsilon}^{2}$	$\hat{\sigma}_{\eta}^{2}$	$\hat{\sigma}_{\varsigma}^{2}$	$\hat{\sigma}_{\omega}^{2}$	**M.S.E.(F)**	**M.A.P.E.(F)**
BSM with dummy seasonality :								
KF $^{(1)}$	457.236		5.000	0.100	0.100	0.100	10.574	12.43%
KF $^{(2)}$	**371.412**	$\sigma_{\varepsilon}^{2},\Sigma$	**0.881**	**0.000**	**0.000**	**0.136**	**5.933**	**9.24%**
KF $^{(3)}$	371.412	$\sigma_{\varepsilon}^{2},\Sigma,\mu$	0.881	0.000	0.000	0.136	5.933	9.24%
EM	371.451	$\sigma_{\varepsilon}^{2},\Sigma$	0.860	0.011	0.000	0.139	5.408	8.69%
BSM with trigonometric seasonality :								
KF $^{(1)}$	521.940		5.000	0.100	0.100	0.100	8.059	10.92%
KF $^{(2)}$	**379.196**	$\sigma_{\varepsilon}^{2},\Sigma$	**0.831**	**0.000**	**0.000**	**0.004**	**6.153**	**9.50%**
KF $^{(3)}$	379.196	$\sigma_{\varepsilon}^{2},\Sigma,\mu$	0.831	0.000	0.000	0.004	6.153	9.50%
EM	379.321	$\sigma_{\varepsilon}^{2},\Sigma$	0.843	0.00	0.000	0.004	5.885	9.21%

Table 3. 4. 10 Factory : Results for the local linear trend model

	$-\ln L(\theta \mid Y_1,\ldots Y_T)$	θ	$\hat{\sigma}_{\varepsilon}^{2}$	$\hat{\sigma}_{\eta}^{2}$	$\hat{\sigma}_{\varsigma}^{2}$	**M.S.E.(F)**	**M.A.P.E.(F)**
KF $^{(1)}$	85.277		5.000	0.100	0.100	18.883	13.11%
KF $^{(2)}$	**82.927**	$\sigma_{\varepsilon}^{2},\Sigma$	**7.670**	**0.071**	**0.000**	**27.168**	**14.68%**
KF $^{(3)}$	82.927	$\sigma_{\varepsilon}^{2},\Sigma,\mu$	7.670	0.071	0.000	27.168	14.68%
EM	82.969	$\sigma_{\varepsilon}^{2},\Sigma$	7.620	0.094	0.000	25.336	14.13%

3.4.5 SUMMARY

The BSM with dummy seasonal components resulted in better forecasts as measured by the criterion M.S.E.(F), than the BSM with trigonometric seasonal components, for every time series modelled except for the complete time series of monthly electricity consumption for Municipality B. This was also true for the criterion M.A.P.E.(F) except for the case when modelling the complete time series of monthly electricity consumption for Municipality A. The results were better for the BSM with trigonometric seasonal components according to the criterion M.A.P.E.(F), but not for the criterion M.S.E.(F), which indicates that one model is not necessarily outright better than the other.

The method of direct maximisation converged notably faster than the EM algorithm and resulted is a smaller likelihood function within a reasonable period. It was frequently the case that, even though the parameters derived using the EM algorithm resulted in a larger likelihood function than when using those derived using the method of direct maximisation, the forecasting results according to the criteria M.S.E.(F) were better. This is presumably a result of chance where the test set deviated from the usual electricity consumption pattern. Overall the preferred approach to obtaining maximum likelihood estimates of the parameters would seem to be that involving direct maximisation of the likelihood function.

It is interesting to note that unless μ was selected to be extremely large, its effect on the model was minimal. A further point of interest is that the variance σ_{ς}^{2} always tends to be close to zero indicating a small change in the level of the series over time.

3.5 COMPARISON OF RESULTS

The forecasting results for each of the best fitting exponential smoothing, ARIMA and state space models discussed in this study, as indicated by the criterion of minimum M.S.E.(F), are summarised in Table 3. 5. 1.

Table 3. 5. 1 : Summary of forecasting results for each method

METHOD	Time series	M.S.E.(F)	M.A.P.E.(F)
Exponential Smoothing	Municipality A	7.567	1.94%
ARIMA	Municipality A	6.795	1.96%
State Space Model	Municipality A	**6.520**	1.90%
Exponential Smoothing	Municipality B	0.429	2.88%
ARIMA	Municipality B	0.432	2.72%
State Space Model	Municipality B	**0.385**	2.70%
Exponential Smoothing	Municipality C	0.138	3.87%
ARIMA	Municipality C	**0.065**	2.13%
State Space Model	Municipality C	0.147	4.23%
Exponential Smoothing	Municipality D (Excluding factory)	5.427	7.85%
ARIMA	Municipality D (Excluding factory)	**5.200**	8.58%
State Space Model	Municipality D (Excluding factory)	5.933	9.24%
Exponential Smoothing	Factory	**17.048**	11.33%
ARIMA	Factory	19.780	12.37%
State Space Model	Factory	27.168	14.68%

State space models resulted in the best forecasts for both of the time series of monthly electricity consumption for Municipality A and B. However, the best results for the time series of the monthly electricity consumption for Municipality C, which was affected by the intervention events, were derived using ARIMA models which incorporate intervention events. Surprisingly the state space model including intervention events did not perform well, and in fact the results were better for the exponential smoothing method which did not include these intervention events. This is probably because the intervention events were sufficiently early in the series to have a minimal affect on the exponential smoothing parameters. The ARIMA model produced the best forecast for the time series of the monthly electricity consumption for Municipality D, excluding the factory's electricity consumption. The results for the non-seasonal time series of the monthly electricity consumption for the factory are distorted by the decrease in electricity consumption in November 1995 caused by equipment failing at the factory. Thus the test set does not reflect the usual electricity consumption pattern and it is surmised that, purely by chance, the exponential smoothing method resulted in the smallest criterion M.S.E.(F).

For all three methods the forecasting results using the complete time series of monthly electricity consumption were better than those obtained when using the shorter series of electronically metered electricity consumption. The inclusion of the intervention events when modelling the time series of the monthly electricity sales to Municipality C improved the results of both the ARIMA and state space models. It is interesting to note however that it was not necessary to include the intervention relating to the mine closure in either model.

4. CONCLUSION

The aim of this thesis was to identify and study appropriate methods of forecasting by month, one year ahead, the electricity consumption for selected municipalities in Kwa-Zulu Natal. In general the time series of monthly electricity consumption for these municipalities displayed a trend and, except for the time series of monthly electricity consumption of the factory within Municipality D's area of supply, seasonality. The exponential smoothing method and ARIMA and state space modelling were identified as appropriate approaches for forecasting and were compared and contrasted.

In summary, the exponential smoothing method is simple, robust and easy to implement. It can be fully automated and requires limited calculations and data storage space. The ARIMA methodology requires the time series to be stationary, and if it is not, the trend and seasonality to be removed by differencing which is not always acceptable. Furthermore the model identification stage is often difficult, and can be subjective and time consuming and if the model is incorrectly identified, the resulting forecasts can be very unsatisfactory. State space models on the other hand incorporate the trend and seasonality, and as with exponential smoothing, the time series can be expressed in terms of the trend, level, seasonal and error components. An added advantage of state space modelling over exponential smoothing is that it is a formal modelling technique. Once a model is expressed in state space form, Kalman filtering is easily applied with pleasing results. Unfortunately state space models and Kalman filtering are not included in the majority of forecasting packages. For example SAS invokes state space models to determine the maximum likelihood estimates for ARIMA models but does not include basic structural models.

For cases in which a time series is affected by intervention events and these are not included in the modelling process, the forecasting results are often unsatisfactory. This is particularly true if the event occurs towards the latter part of the time series. ARIMA and state space models allow the incorporation of intervention events and this can greatly enhance the forecasting results and decrease the residual errors.

Further areas of interest are the application of the above methods to the time series of monthly electricity consumption for other groups of Eskom customers whose electricity consumption patterns differ from those of the municipal customers, such as the various railway lines, coal mines and industries within Kwa-Zulu Natal. There are also other forecasting methods and techniques available which need to be investigated, one of these being neural networks which is reported to give good results for less regular time series.

APPENDIX A

A.1 : Conditional expectations of terms in the log likelihood function for a state space model

(i) Since $E(\alpha_0|Y_1,...Y_T)=\hat{\alpha}_{0|T}$ and $Var(\alpha_0|Y_1,.........Y_T)=C_{0|T}$,

$$E\{(\alpha_0-\mu)^T\Sigma^{-1}(\alpha_0-\mu)|Y_1,.......Y_T\}$$

$$=E\{tr[(\alpha_0-\mu)'\Sigma^{-1}(\alpha_0-\mu)]\}$$

$$=E\{tr[\Sigma^{-1}(\alpha_0-\mu)(\alpha_0-\mu)^T]\}$$

$$=tr[\Sigma^{-1}E\{(\alpha_0-\mu)(\alpha_0-\mu)^T\}]$$

$$=tr[\Sigma^{-1}\{(\hat{\alpha}_{0|T}-\mu)(\hat{\alpha}_{0|T}-\mu)^T+C_{0|T}\}]$$

(ii) Since $\begin{pmatrix}\alpha_t\\ \alpha_{t-1}\end{pmatrix}|Y_1,..........Y_T\sim N\left[\begin{pmatrix}\hat{\alpha}_{t|T}\\ \hat{\alpha}_{t-1|T}\end{pmatrix},\begin{pmatrix}C_{t|T} & C_{t,t-1|T}\\ C_{t,t-1|T}^T & C_{t-1|T}\end{pmatrix}\right]$,

$$\alpha_t-\Phi\alpha_{t-1}|Y_1,...Y_T\sim N(\hat{\alpha}_{t|T}-\Phi\hat{\alpha}_{t-1|T},C_{t|T}-C_{t,t-1|T}\Phi^T-\Phi C_{t,t-1|T}^T+\Phi C_{t-1|T}\Phi^T).$$

Thus $E\{(\alpha_t-\Phi\alpha_{t-1})^T\Sigma^{-1}(\alpha_t-\Phi\alpha_{t-1})\}$

$$=-\frac{1}{2}tr\{\Sigma^{-1}[\sum_{t=1}^{T}(C_{t|T}+\hat{\alpha}_{t|T}\hat{\alpha}_{t|T}^T)-\sum_{t=1}^{T}(C_{t,t-1|T}+\hat{\alpha}_{t|T}\hat{\alpha}_{t-1|T}^T)\Phi^T$$

$$-\Phi\sum_{t=1}^{T}(C_{t,t-1|T}+\hat{\alpha}_{t|T}\hat{\alpha}_{t-1|T}^T)^T+\Phi\sum_{t=1}^{T}(C_{t-1|T}+\hat{\alpha}_{t-1|T}\hat{\alpha}_{t-1|T}^T)\Phi^T]\}$$

(iii) Since $\alpha_t|Y_1,......Y_T\sim N(\hat{\alpha}_{t|T},C_{t|T})$, it follows that

$$E[\frac{(Y_t-h^T\alpha_t)^2}{\sigma_\varepsilon^2}|Y_1,.......Y_T)=\frac{(Y_t-h^T\hat{\alpha}_{t|T})^2+h'C_{t|T}h}{\sigma_\varepsilon^2}$$

(Shumway and Stoffer, 1982).

A.2 : Maximum likelihood estimates of σ_ε^2 and Σ^{-1} in a state space model

The expectation $E[\ln L(\sigma_\varepsilon^2, \Sigma | Y_1, \ldots Y_T)]$ is maximised by setting the derivatives with respect to σ_ε^2 and Σ^{-1} equal to zero and solving for σ_ε^2 and Σ^{-1} (Shumway and Stoffer, 1982). In particular let

$$A = \sum_{t=1}^{T} (C_{t|T} + \hat{\alpha}_{t|T} \hat{\alpha}_{t|T}^{T}) - \sum_{t=1}^{T} (C_{t,t-1|T} + \hat{\alpha}_{t|T} \hat{\alpha}_{t-1|T}^{T}) \Phi^T$$

$$- \Phi \sum_{t=1}^{T} (C_{t,t-1|T} + \hat{\alpha}_{t|T} \hat{\alpha}_{t-1|T}^{T})^T + \Phi \sum_{t=1}^{T} (C_{t-1|T} + \hat{\alpha}_{t-1|T} \hat{\alpha}_{t-1|T}^{T}) \Phi^T$$

Then consider the terms in $E[\ln L(\sigma_\varepsilon^2, \Sigma | Y_1, \ldots Y_T)]$ involving Σ ,written as

$$f(\Sigma) = -\frac{T}{2} \log|\Sigma| - \frac{1}{2} tr[\Sigma^{-1} A]$$

$$= \frac{T}{2} \log|\Sigma^{-1}| - \frac{1}{2} tr[\Sigma^{-1} A].$$

From the results of Mardia, Kent and Bibby (1979; appendix A 9.3 and A 9.4), and defining $diag(A)$ as the matrix containing only the diagonal elements of A along its own diagonal, it follows that

$$\frac{\partial f(\Sigma)}{\partial \Sigma} = \det(\Sigma^{-1}) \frac{T}{2} [2(\Sigma^{-1})^{-1} - diag(\Sigma)] - \frac{\partial(\Sigma^{-1})}{\partial \Sigma} \frac{1}{2} [(2A) - diag(A)]$$

and this derivative equals zero when $\hat{\Sigma} = \frac{1}{T} A$.

Similarly, let $B = \sum_{t=1}^{T} [(Y_t - h^T \hat{\alpha}_{t|T})^2 + h^T C_{t|T} h]$.

Then the term involving σ_ε^2 is given by

$$f(\sigma_\varepsilon^2) = -\frac{T}{2} \ln \sigma_\varepsilon^2 - \frac{B}{2\sigma_\varepsilon^2}.$$

Thus $\frac{\partial f(\sigma_\varepsilon^2)}{\partial \sigma_\varepsilon^2} = -\frac{T}{2\sigma_\varepsilon^2} + \frac{B}{2(\sigma_\varepsilon^2)^2}$ equals zero when $\hat{\sigma}_\varepsilon^2 = \frac{1}{T} B$.

A.3 : The exponential smoothing method and ARIMA models

Forecasting approach : Simple exponential smoothing and ARIMA(0,1,1) models

The one-step-ahead forecasts derived for a time series using the simple exponential smoothing method are the same as those obtained when using an ARIMA(0,1,1) model. In particular, the one-step-ahead forecast when using simple exponential smoothing is given by

$$\hat{Y}_{t+1|t} = \alpha Y_t + (1-\alpha)\hat{Y}_{t|t-1} \qquad \text{(A.1)}$$

and the one-step-ahead forecast when applying the model ARIMA(0,1,1) to a time series is given by

$$\begin{aligned}\hat{Y}_{t+1|t} &= E(Y_{t+1}|Y_t, Y_{t-1}, \ldots Y_1)\\ &= E(Y_t + Z_{t+1} - \theta Z_t | Y_t, Y_{t-1} \ldots Y_1)\\ &= Y_t - \theta Z_t \ .\end{aligned}$$

However $$Y_t - \hat{Y}_{t|t-1} = Y_{t-1} + Z_t - \theta Z_{t-1} - Y_{t-1} + \theta Z_{t-1} = Z_t$$

and thus $$\begin{aligned}\hat{Y}_{t+1|t} &= Y_t - \theta(Y_t - \hat{Y}_{t|t-1})\\ &= (1-\theta)Y_t - \theta\hat{Y}_{t|t-1} \qquad \text{(A.2)}\end{aligned}$$

On setting $1-\theta=\alpha$, it is clear that equations (A.1) and (A.2) are equivalent.

Similarly, Holt Winter's two parameter smoothing method, which incorporates trend and level components but no seasonal component, is equivalent to an ARIMA(0,2,2) process.

Firstly consider the double exponential smoothing method defined by

$$\begin{aligned}L_t &= \alpha Y_t + (1-\alpha)(L_{t-1} + T_{t-1})\\ &= \alpha Y_t + (1-\alpha)\hat{Y}_{t|t-1}\end{aligned}$$

and $$T_t = \gamma(L_t - L_{t-1}) + (1-\gamma)T_{t-1}$$

$$= \gamma L_t - \gamma(L_{t-1} + T_{t-1}) + T_{t-1}$$

$$= \gamma(\alpha Y_t + (1-\alpha)\hat{Y}_{t|t-1}) - \gamma \hat{Y}_{t|t-1} + T_{t-1}$$

$$= \gamma\alpha(Y_t - \hat{Y}_{t|t-1}) + T_{t-1}$$

Then the one-step-ahead forecast is given by

$$\hat{Y}_{t+1|t} = L_t + T_t$$

$$= \alpha Y_t + (1-\alpha)\hat{Y}_{t|t-1} + \gamma\alpha(Y_t - \hat{Y}_{t|t-1}) + T_{t-1}$$

$$= \alpha Y_t + (1-\alpha)\hat{Y}_{t|t-1} + \gamma\alpha(Y_t - \hat{Y}_{t|t-1}) + \hat{Y}_{t|t-1} - L_{t-1}$$

$$= \alpha Y_t + (1-\alpha)\hat{Y}_{t|t-1} + \gamma\alpha(Y_t - \hat{Y}_{t|t-1}) + \hat{Y}_{t|t-1} - (\alpha Y_{t-1} + (1-\alpha)\hat{Y}_{t-1|t-2})$$

$$= (\alpha + \alpha\gamma)Y_t + (2 - \alpha - \alpha\gamma)\hat{Y}_{t|t-1} - \alpha Y_{t-1} + (\alpha - 1)\hat{Y}_{t-1|t-2} \qquad \text{(A.3)}$$

The forecast $\hat{Y}_{t+1|t}$ using an ARIMA(0,2,2) model is calculated as

$$\hat{Y}_{t+1|t} = E(Y_{t+1}|Y_t, Y_{t-1}, \ldots Y_1)$$

$$= E(2Y_t - Y_{t-1} + Z_{t+1} - \theta_1 Z_t - \theta_2 Z_{t-1}|Y_t, Y_{t-1} \ldots Y_1)$$

$$= 2Y_t - Y_{t-1} - \theta_1 Z_t - \theta_2 Z_{t-1}$$

and, since

$$Y_t - \hat{Y}_{t|t-1} = 2Y_{t-1} - Y_{t-2} + Z_t - \theta_1 Z_{t-1} - \theta_2 Z_{t-1} - [2Y_{t-1} - Y_{t-2} - \theta_1 Z_{t-2} - \theta_2 Z_{t-2}] = Z_t \qquad \text{(A.4)}$$

it follows that $\hat{Y}_{t+1|t} = 2Y_t - Y_{t-1} - \theta_1[Y_t - \hat{Y}_{t|t-1}] - \theta_2[Y_{t-1} - \hat{Y}_{t-1|t-2}]$.

$$= (2 - \theta_1)Y_t - (1 + \theta_2)Y_{t-1} + \theta_1\hat{Y}_{t|t-1} + \theta_2\hat{Y}_{t-1|t-2}\ .$$

It is clear that by writing

$\theta_2 = \alpha - 1$ and $\theta_1 = 2 - \alpha - \gamma\alpha$,

equations (A.3) and (A.4) are equivalent.

Conditional least squares : Simple exponential smoothing and ARIMA(0,1,1) models

If the parameters of the ARIMA model are derived using conditional least squares, the forecast estimates derived from simple exponential smoothing and ARIMA(0,1,1) models are the same. This is readily demonstrated as follows.

Assume that the ARIMA(0,1,1) model given by $Y_t = Y_{t-1} + Z_t - \theta Z_{t-1}$ has the realisation, $y_t = y_{t-1} + z_t - \theta z_{t-1}$ and that $z_1 = E(z_1) = 0$. Then clearly the residuals are given by

$$z_2 = y_2 - y_1$$

$$\begin{aligned} z_3 &= y_3 - y_2 + \theta(y_2 - y_1) \\ &= y_3 + y_2(\theta - 1) - \theta y_1 \\ &= y_3 - \alpha y_2 + (\alpha - 1)y_1 \qquad \text{where } \theta = 1 - \alpha \end{aligned}$$

and generally,

$$z_t = y_t - \alpha y_{t-1} - \alpha(1-\alpha)y_{t-2} - \ldots\ldots - (1-\alpha)^{t-2} y_1 .$$

The conditional least squares estimates of the unknown parameters are then derived by minimising $\sum z_t^2$ with respect to α .

Similarly, using the exponential smoothing approach, and assuming $\hat{y}_{2|1} = y_1$, the forecasts are derived by

$$\hat{y}_{3|2} = \alpha y_2 + (1-\alpha)\hat{y}_{2|1} = \alpha y_2 + (1-\alpha)y_1$$

$$\hat{y}_{4|3} = \alpha y_3 + (1-\alpha)\hat{y}_{3|2} = \alpha y_3 + (1-\alpha)\alpha y_2 + (1-\alpha)^2 y_1$$

and the residuals are calculated as

$$e_2 = y_2 - \hat{y}_{2|1} = y_2 - y_1$$

$$e_3 = y_3 - \hat{y}_{3|2} = y_3 - \alpha y_2 - (1-\alpha)y_1 ,$$

and generally as

$$e_t = y_t - y_{t|t-1} = y_t - \alpha y_{t-1} - \alpha(1-\alpha) y_{t-2} - \ldots - (1-\alpha)^{t-2} y_1 .$$

Since the residuals $\sum e_t^2$ are minimised using the smoothing approach, it is clear that the estimates from ARIMA(0,1,1) and simple exponential smoothing are equivalent.

APPENDIX B : Time Series

MUNICIPALITY A ELECTRICITY CONSUMPTION IN GWH																
	1980	1981	1982	1983	1984	1985	1986	1987	1988	1989	1990	1991	1992	1993	1994	1995
Jan	48.345	50.100	51.010	55.705	57.615	66.830	62.510	70.815	73.435	98.210	90.458	84.269	87.780	87.971	93.344	97.997
Feb	50.140	51.175	53.365	54.565	59.705	63.660	65.645	69.575	76.410	80.950	80.821	84.400	89.856	88.642	88.412	96.605
Mar	48.610	54.560	57.420	61.045	61.455	68.625	76.115	75.225	84.390	86.946	91.225	92.566	96.306	99.993	99.236	107.990
Apr	48.790	51.610	65.095	56.495	65.755	65.965	69.760	71.685	78.450	85.378	83.467	89.927	90.190	92.593	89.785	100.143
May	54.760	58.910	52.089	58.180	69.695	70.240	76.655	77.305	77.010	90.303	94.383	97.002	94.615	99.972	103.826	111.047
Jun	56.115	64.770	69.065	60.475	68.415	76.610	73.685	83.135	89.300	95.749	98.048	101.611	102.055	108.492	109.161	115.323
Jul	62.630	67.870	71.524	61.545	81.565	77.165	81.720	87.275	94.690	98.037	99.730	101.925	104.387	107.444	113.125	115.034
Aug	58.830	61.065	66.660	66.660	71.845	68.550	79.175	90.535	90.535	94.726	101.243	99.500	101.871	106.492	113.152	107.693
Sep	56.810	61.700	63.925	57.260	73.010	74.090	75.070	82.770	83.740	88.978	93.736	93.196	96.824	100.511	103.334	105.099
Oct	60.735	55.320	62.210	62.335	72.210	71.920	75.850	84.535	90.490	91.919	96.392	98.250	98.744	103.933	108.644	110.6072
Nov	50.925	61.510	61.655	61.265	66.660	66.655	77.180	78.500	85.945	91.209	91.957	92.668	95.827	100.552	103.475	108.5253
Dec	49.540	54.060	54.005	54.030	57.975	66.910	67.705	73.020	76.310	69.060	76.949	81.159	86.284	89.459	91.716	94.26811
TOTAL	646.230	692.650	728.023	709.560	805.905	837.220	881.070	944.375	1000.705	1071.465	1098.409	1116.473	1144.740	1186.054	1217.210	1270.332

MUNICIPALITY B ELECTRICITY CONSUMPTION IN GWH																
	1980	1981	1982	1983	1984	1985	1986	1987	1988	1989	1990	1991	1992	1993	1994	1995
Jan	8.971	11.064	11.071	10.805	12.067	13.023	12.043	12.888	12.638	14.108	17.942	16.851	16.630	16.661	17.073	17.660
Feb	14.431	13.685	15.228	14.899	15.115	15.975	15.920	16.645	17.260	23.407	18.446	16.760	18.138	17.540	17.813	18.756
Mar	13.966	15.130	16.764	16.757	17.383	17.688	17.290	19.088	19.818	18.815	20.015	18.013	19.883	19.130	20.060	20.737
Apr	13.207	13.123	15.101	14.690	15.758	16.390	16.345	16.185	17.660	19.736	17.929	18.200	17.746	17.160	18.525	18.614
May	14.962	16.169	16.692	16.126	17.820	17.698	17.960	18.385	20.553	20.620	19.854	19.409	19.473	19.840	20.140	21.334
Jun	15.247	17.239	17.258	16.865	19.283	19.465	18.438	19.533	19.100	22.246	21.778	20.392	20.951	22.080	22.193	22.090
Jul	17.453	16.555	18.977	18.199	20.753	20.068	20.870	20.063	22.450	22.815	18.760	21.277	21.235	21.506	22.774	21.881
Aug	16.186	17.196	18.650	18.110	18.910	19.408	19.983	22.750	21.063	21.693	21.690	20.600	20.038	20.993	21.791	21.038
Sep	15.233	17.402	16.375	16.805	17.535	16.860	18.670	18.333	20.038	20.081	19.020	19.279	18.695	19.727	19.904	19.326
Oct	14.897	16.726	16.390	16.363	17.508	18.135	18.040	19.990	19.480	20.613	19.565	20.010	18.665	19.295	20.124	19.909
Nov	14.724	15.468	17.294	16.939	17.138	17.238	18.278	19.898	20.010	20.112	18.852	18.319	18.608	20.055	19.966	21.284
Dec	13.778	16.034	13.462	15.108	14.575	15.748	16.215	16.295	17.375	14.653	13.939	14.154	15.612	15.301	15.507	16.705
Total	173.054	185.791	193.262	191.666	203.842	207.693	210.050	220.050	227.443	238.900	227.788	223.264	225.674	229.289	235.870	239.333

MUNICIPALITY C
ELECTRICITY CONSUMPTION IN GWH

	1980	1981	1982	1983	1984	1985	1986	1987	1988	1989	1990	1991	1992	1993	1994	1995
Jan	4.717	5.178	4.853	4.805	4.938	5.237	5.280	5.664	5.832	6.096	6.144	6.744	6.723	6.688	6.990	6.990
Feb	4.733	4.690	5.458	5.686	5.625	5.717	5.280	5.616	5.832	6.600	6.816	6.770	6.529	6.330	6.605	6.733
Mar	4.448	5.012	5.126	5.054	4.818	4.802	5.400	5.808	5.712	5.952	6.264	6.240	7.089	7.000	7.332	7.668
Apr	4.589	4.707	6.049	6.000	5.618	5.414	5.664	6.024	6.720	6.672	6.528	6.888	6.757	6.670	6.938	7.419
May	4.984	5.514	5.749	4.950	5.490	6.082	6.384	6.288	5.904	6.720	7.140	7.440	7.375	7.250	8.004	8.720
Jun	6.432	6.284	6.910	6.430	6.748	6.761	7.032	6.336	8.136	8.352	7.752	8.496	8.248	8.146	9.264	9.360
Jul	5.839	6.582	7.454	6.350	7.286	7.481	7.154	8.304	7.920	7.944	8.280	11.284	8.399	7.884	9.335	9.190
Aug	6.490	6.051	6.484	5.715	6.288	6.283	6.790	7.848	7.272	8.352	8.136	8.120	8.437	7.710	8.999	8.549
Sep	5.582	6.800	6.650	5.958	6.816	5.870	6.288	5.808	7.656	7.008	8.184	7.225	7.533	6.937	7.668	7.704
Oct	5.380	5.302	5.642	4.961	5.885	6.274	6.384	8.064	6.360	7.008	6.840	7.280	7.364	7.198	7.892	7.749
Nov	5.736	5.294	5.803	5.543	5.366	5.640	6.552	6.360	7.080	7.104	7.536	6.948	7.046	7.029	7.356	7.599
Dec	4.686	5.870	5.966	4.626	5.527	5.472	5.376	6.192	6.024	6.696	6.816	6.677	6.828	6.995	7.056	7.205
TOTAL	63.614	67.283	72.145	66.079	70.406	71.033	73.584	78.312	80.448	84.504	86.436	90.113	88.327	85.838	93.438	94.885

MUNICIPALITY D
TOTAL ELECTRICITY CONSUMPTION IN GWH

	1980	1981	1982	1983	1984	1985	1986	1987	1988	1989	1990	1991	1992	1993	1994	1995
Jan	10.130	12.520	13.220	12.820	21.205	45.945	55.985	18.089	19.207	52.861	54.914	20.180	20.400	39.655	41.039	50.660
Feb	13.290	15.325	16.950	16.515	37.260	42.925	20.450	21.987	27.708	62.161	51.053	20.400	20.598	43.130	44.659	50.784
Mar	12.890	14.420	15.545	15.555	39.765	43.595	25.515	46.672	55.494	59.993	59.592	22.724	21.642	42.730	52.585	53.436
Apr	13.840	18.680	17.955	16.005	39.590	36.890	27.841	30.392	51.979	59.401	47.940	22.491	28.291	45.650	49.977	52.757
May	15.845	16.580	19.305	18.760	30.065	21.440	49.395	52.053	60.148	54.744	29.870	24.612	38.880	43.180	53.206	56.249
Jun	17.870	21.050	20.610	21.920	20.735	29.035	54.540	58.758	56.907	63.996	25.820	26.416	47.105	49.792	53.427	56.627
Jul	18.685	20.315	21.965	26.025	35.450	51.395	33.685	28.522	67.503	68.418	25.690	25.501	47.906	49.666	58.590	58.177
Aug	18.090	21.320	20.615	29.315	49.725	58.425	35.555	31.514	60.979	63.336	24.260	22.220	43.603	50.533	54.392	55.216
Sep	16.895	21.445	19.605	38.890	39.125	55.500	30.765	52.051	59.838	46.129	23.144	21.528	45.449	48.619	50.389	51.857
Oct	14.835	17.985	18.435	19.320	48.585	60.075	57.500	58.917	47.449	63.076	23.273	22.660	47.378	50.385	52.122	53.068
Nov	16.485	16.730	19.285	17.320	50.615	54.205	54.895	55.174	54.252	23.554	22.466	21.392	43.626	50.115	50.594	38.212
Dec	15.345	15.660	16.465	17.065	25.670	46.955	26.233	51.462	63.721	50.567	18.646	18.859	31.992	48.284	50.218	50.249
TOTAL	184.200	212.030	219.955	249.510	437.790	546.385	472.358	505.589	625.182	668.235	406.668	268.983	436.868	561.741	611.199	627.292

MUNICIPALITY D
ELECTRICITY CONSUMPTION IN GWH FOR THE FACTORY

	1983	1984	1985	1986	1987	1988	1989	1990	1991	1992	1993	1994	1995
Jan		7.451	31.723	41.530	3.400	3.611	36.799	35.824	2.495	3.000	22.850	23.360	29.669
Feb		20.100	25.442	2.806	4.181	9.941	43.758	33.517	2.504	3.000	26.532	26.680	27.521
Mar		22.167	24.975	6.384	27.031	34.367	45.432	41.411	3.165	3.000	23.941	30.584	27.597
Apr		22.070	18.612	9.184	11.357	34.974	39.714	28.033	2.785	10.000	26.628	32.753	33.126
May		11.585	3.100	31.125	33.852	39.252	34.515	9.400	3.037	15.000	23.411	29.605	28.334
Jun		0.000	8.054	33.435	37.530	35.733	42.619	3.104	4.015	24.040	25.568	27.839	29.398
Jul	3.000	13.325	29.720	12.460	4.197	43.132	47.344	3.000	2.914	25.276	27.815	32.507	32.378
Aug	8.233	28.038	36.436	13.263	9.966	39.789	41.294	3.000	1.433	21.752	26.244	29.054	31.761
Sep	18.700	19.186	35.687	11.078	31.277	39.464	28.572	3.000	2.128	26.657	29.420	27.920	28.941
Oct	0.000	28.341	39.369	36.333	38.498	27.769	41.401	3.000	2.648	28.033	29.390	29.856	32.600
Nov	0.000	31.508	34.204	34.000	34.871	34.933	4.371	3.000	3.000	25.016	29.120	28.624	16.816
Dec	0.000	8.210	29.297	8.378	33.892	46.244	36.625	3.000	3.000	15.099	29.280	28.080	27.660
TOTAL		211.980	316.619	239.974	270.049	389.206	442.442	169.288	33.122	199.873	320.199	346.862	345.802

MUNICIPALITY D ELECTRICITY CONSUMPTION IN GWH EXCLUDING THE FACTORY																
	1980	1981	1982	1983	1984	1985	1986	1987	1988	1989	1990	1991	1992	1993	1994	1995
Jan	10.130	12.520	13.220	12.820	13.754	14.222	14.455	14.689	15.597	16.063	19.090	17.685	17.400	16.805	17.679	20.991
Feb	13.290	15.325	16.950	16.515	17.160	17.483	17.644	17.805	17.767	18.403	17.537	17.897	17.598	16.598	17.979	23.263
Mar	12.890	14.420	15.545	15.555	17.598	18.620	19.130	19.641	21.128	14.561	18.181	19.559	18.642	18.789	22.001	25.839
Apr	13.840	18.680	17.955	16.005	17.520	18.278	18.657	19.036	17.005	19.687	19.907	19.706	18.291	19.022	17.224	19.631
May	15.845	16.580	19.305	18.760	18.480	18.340	18.270	18.201	20.896	20.229	20.470	21.575	23.880	19.769	23.601	27.915
Jun	17.870	21.050	20.610	21.920	20.735	20.982	21.105	21.228	21.174	21.378	22.716	22.401	23.065	24.224	25.588	27.229
Jul	18.685	20.315	21.965	23.025	22.125	21.675	21.225	24.325	24.372	21.074	22.690	22.588	22.630	21.851	26.083	25.799
Aug	18.090	21.320	20.615	21.082	21.687	21.989	22.292	21.549	21.190	22.042	21.260	20.787	21.851	24.289	25.338	23.455
Sep	16.895	21.445	19.605	20.190	19.939	19.813	19.687	20.775	20.374	17.557	20.144	19.400	18.792	19.199	22.469	22.917
Oct	14.835	17.985	18.435	19.320	20.244	20.706	21.167	20.419	19.680	21.676	20.273	20.012	19.345	20.995	22.266	20.468
Nov	16.485	16.730	19.285	17.320	19.108	20.001	20.895	20.304	19.320	19.183	19.466	18.392	18.610	20.995	21.970	21.396
Dec	15.345	15.660	16.465	17.065	17.460	17.658	17.855	17.570	17.476	13.942	15.646	15.859	16.893	19.004	22.138	22.587
TOTAL	184.200	212.030	219.955	219.577	225.810	229.766	232.384	235.539	235.976	225.794	237.380	235.861	236.995	241.542	264.337	281.490

REFERENCES

Abraham, B. and Ledolter, J. (1986) "Forecast Functions Implied by Autoregressive Integrated Moving Average Models and Other related forecast Procedures", *International Statistical Review,* 54 , pp. 51 - 66.

Abraham, B. and Ledolter, J. (1983) "Statistical Methods for Forecasting", *New York, Wiley*

Anderson, T. W. (1958) "An Introduction to Multivariate statistical Analysis", *New York, Wiley*

Armstrong, J. S. and Collopy, F. (1992) "Error measures for generalizing about forecasting methods: Empirical comparisons", *International Journal of Forecasting, 8,* pp 69-80.

Box, G. E. P. and Jenkins, G. M. (1970) "Time Series Analysis : Forecasting and Control" , *San Francisco : Holden Day.*

Box, G. E. P. and Tiao, G. C. (1975) "Intervention Analysis with Applications to Economic and Environmental Problems", *Journal of the American Statistical Association March 1975, 70,* pp.70-79.

Chatfield, C. (1978) "The Holt-winters forecasting procedure", *Applied Statistics,* 27, pp. 264-279.

Chatfield, C. (1988) "What is the 'best' method of forecasting?", *Journal of Applied Statistics,* 15, pp. 19-39.

Chatfield, C. (1989) "The Analysis of Time Series", 4th Ed, *Chapman & Hall.*

Chatfield, C. and YAR, M. (1988) "Holt Winters Forecasting : some practical issues", *The Statistician*, 37, pp. 129 - 140.

Chatfield, C. and YAR, M. (1991) "Prediction intervals for multiplicative Holt-Winters", *International Journal of Forecasting*, 7, pp. 31 - 37.

Chen, C. and Liu, L. (1993) "Joint Estimation of Model Parameters and Effects in Time Series", *Journal of the American Statistical Association March 1993, 88,* pp. 284-297.

Cryer, J. (1986) "Time Series Analysis", *PWS Publishers.*

Deadman D. and Pyle D. (1989) "Homicide in England and Wales : A Time Series Analysis", *University of Leicester, Department of Economics, Discussion Paper No. 102.*

Diggle, P. L. (1990) "A Biostatistical Introduction", *Clarendon Press : Oxford.*

Fildes, R. (1992) "The evaluation of extrapolative forecasting methods", *International Journal of Forecasting, 8*, pp 81-98.

Gardner, E. S. (1985) "Exponential Smoothing : The State of the Art", *Journal of Forecasting, 4,* pp. 1-28.

Granger, C. W. J. and Newbold, P. (1977) "Forecasting Economic Time Series" (2nd Ed.), *New York : Academic Press.*

Harvey, A. C. (1989) "Forecasting, structural time series models and the Kalman filter", *Cambridge University Press.*

Harvey, A. C. and Durbin, J. (1986) "The effects of Seat Belt Legislation on British Road Casualties : A Case Study in structural Time Series Modelling", *J. R. Statistical Society A. 149* , pp. 187-227.

Harrison, P.J. and Stevens, C. F. (1976) "Bayesian forecasting" (with discussion), *Journal of Royal Statistical Society, Series B,* pp 205-247.

Janacek, G. and Swift, L. (1993) "Time Series, Forecasting, Simulation, Applications", *Ellis Horwood Limited.*

Jones, R. H. (1993) "Longitudinal Data with Serial Correlation : A State-space Approach", *Chapman and Hall.*

Ledolter, J. and Chan, K. S. (1996) "Evaluating the impact of the 65 mph Maximum Speed Limit on the Iowa Rural Interstates", *The American Statistician, February 1996, 50,* pp 79-85.

Makridakis, S., Wheelwright, S. and Mcgee, V. (1983) "Forecasting: methods and applications", 2nd ed., *New York, Wiley.*

Mardia, K. V., Kent, J. T. and Bibby, J. M. (1979) "Multivariate Analysis", *London, Academic Press Inc.*

Meinhold, R.J. and Singpurwalla, N. D. (1983) "Understanding the Kalman", *Filter The American Statistician, May 1983, 37,* pp. 123-127.

Shumway, R. H. and Stoffer, D. S. (1982) "An Approach to Time Series smoothing and Forecasting Using the EM Algorithm", *Journal of Time Series Analysis, 3,* pp. 253-263.

Shumway, R.H. (1988) "Applied statistical Time Series Analysis", pp 173-200.

Tanner, M. (1993) "Tools for Statistical Inference" (2nd Ed.), *Springer*

Vandaele, W. (1983) "Applied Time Series and Box-Jenkins Models", *USA, Academic Press.*

Yar, M. and Chatfield, C. (1990) "Prediction intervals for the Holt-Winters forecasting procedure", *International Journal of Forecasting, 6*, pp. 127-137.