Durham University

Time Series and Forecasting, including Kalman Filtering

Mark Pilling, Trevelyan College.

SEPTEMBER 1993
Contents

§1. Introduction page 1
§2. Preliminary Data Analysis page 2
§3. Models for Stationary Data page 8
§4. Estimating Model Parameters page 13
§5. Forecasting page 19
§6. State-Space Models and Kalman Filtering page 21
§7. Worked Example on Wolfer Sunspots page 28
§8. Epilogue page 37
§9. Appendix page 38
§10. References page 39
§1. Introduction

What I hope to do with this essay is to first give an explanation of some of the terminology used, then build up a theoretical knowledge of some of the methods used to analyse data (and what these methods achieve), until I can finally give a practical example of how these methods can draw realistic conclusions from a real-world data set, by using EXCEL to process information, and MINITAB8 to perform matrix operations.

Time series are just a set of observations of a situation, which are ordered with respect to a single variable. This variable could be anything, but is conventionally referred to as "time", and usually the time values are equally spaced for ease of later calculation. A time series does not have to involve just two variables (e.g. a spatial series with points in the plane) but methods for analysing simple time series can be readily extended to more complex situations. At this stage it needs to be stated that the main difference between time series and other probability distributions that we have studied is that previously, we were dealing with independent observations, whilst time series deal with successive observations dependant on time.
§2. Preliminary Data Analysis

Given a continuous time series we can "read off" (i.e. digitise) values at equal intervals of time to gain a sampled series of observations (i.e. discrete). If a variable does not have an instantaneous value (e.g. monthly unemployment figures), we can aggregate the values over equal intervals of time to gain the discrete form we can use.

**Terminology**

*Deterministic* means a time series which can be predicted exactly from past observations.

*Stochastic* means the predicted values of the time series can only be estimated from past observations.

Initially, in order to construct a sensible model for the information, we must plot the data so that we can recognise;

a) Trend: "Long term change in the mean level". This is a subjective assessment. (e.g. possible to misinterpret long term cyclic variation in the data as trend.)

b) Seasonal Variation: Annual in period (e.g. unemployment figures)

c) Cyclic Variations: This includes seasonal variations, but they usually have different periodicities due to regular physical effects on the data. (i.e. some oscillations have no fixed period, but are still predictable, to a degree)

d) Others: Such factor can be shown to exist by subtracting all the previous terms from the data, and plotting the residuals.

Most models for analysing and predicting time series are designed to work on stationary time series.

**Definition**

A time series is stationary if there is;

(i) no systematic change in mean (trend),

(ii) no systematic change in variance, and

(iii) strictly periodic variations are removed.

More rigorously, a time series is strictly stationary if, for a series of observations \( (X_t) \), the joint distribution \( (X_1, X_2, X_3, ..., X_n) \) is the same as the joint distribution of \( (X_{1+T}, X_{2+T}, X_{3+T}, ..., X_{n+T}) \) for all constant \( T \in \mathbb{Integers} \). (i.e. not dependant on the origin, just the intervals between successive observations.)

However, this definition is unwieldy too work with, so instead we will use the following definition, called second order (weakly) stationary;

(i) \( \text{E}(X_t) = \mu \) independent of time.

(ii) \( \text{Cov}(X_t, X_{t+T}) = \gamma(T) \) independent of time, but not lag \( T \).

Hence, we need to remove the effects of trend etc. from the data to see if there is an underlying (i.e. not immediately apparent) model behind the observations.

(Note: Sometimes these periodic variations are the only things we are looking for, so further analysis is unnecessary)
Analysing Time Series which contain a Trend

The simplest trend is given by the following model:

\[ X_t = \alpha + \beta t + \varepsilon_t \]

where \( X_t \) is the observation at time \( t \), \( \varepsilon_t \) is the random error at time \( t \), and \( \alpha, \beta \) are constants.

The trend \( \beta \), is a deterministic function of time. In this model it is the global trend. However, it is usually unrealistic to assume there is an overall trend, and so only local linear trends are estimated. Apart from models, other methods of interpreting the data are;

a) Curve Fitting:
This is normally just fitting a sample function (e.g. a polynomial curve) to non-seasonal data which contains a trend (e.g. fitting a quartic to data coming from a model which has a quartic trend component. i.e.

\[ X_t = \alpha + \beta t + \gamma t^2 + \delta t^3 + \sigma t^4 + \varepsilon_t \]

b) Filtering:
Filters are used to produce an output which emphasises the variations at particular frequencies. A linear filter deals with a time series \( (X_t) \) by converting it into another time series \( (Y_t) \) by

\[ Y_t = \sum_{r=-q}^{s} a_r X_{r+t} = Sm(X_t) \]

where \( (a_r) \) are the weights.

This is a low pass filter removing local fluctuations.
To help smooth out the local fluctuations and estimate the local mean, we should choose the weights so that they sum to 1, and now the process is termed a Moving Average process. Finally, after estimating the trend with \( Sm(X_t) \) we may want to consider how much of an influence the local fluctuations have on the data by calculating,

\[ \text{Res}(X_t) = (X_t) - Sm(X_t) = \sum_{r=-q}^{s} b_r X_{r+t} \]

i.e. This is another linear filter, called a high pass filter which removes long term, low frequency fluctuations. If \( \sum a_r = 1 \), this implies that \( \sum b_r = 0 \). Therefore, \( b_0 = 1 - a_0 \) and \( b_r = -a_r \forall r \neq 0 \) (by comparing coefficients).

At this point it should also be noted that other filters can be used (e.g. exponential smoothing), and that filters can be used in series.

**Example**

Let \( (a_1, a_2, a_3) \) represent the weights of a linear filter. Hence, a filter may be represented as a convolution of two other filters. e.g.

\[ \frac{1}{4} \cdot \frac{1}{2} \cdot \frac{1}{4} * \frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{2} \]

i.e. given a time series \( (X_1, X_2, X_3, ..., X_n) \), smoothing is done in three stages, by adding two successive pairs of observations together, and the dividing by four.
c) Differencing:
Differencing a time series \( (X_t) \) is done by creating a new time series \( (Y_t) \) from the difference of successive observations.
First Order differencing:
\[
Y_t = \nabla X_t = X_t - X_{t+1}
\]
Second differencing:
\[
Y_t = \nabla^2 X_t = X_t - 2X_{t+1} + X_{t+2}
\]
For non-seasonal data, differencing is sufficient to obtain apparent stationarity (i.e. removal of linear trend).
Second Order differencing:
\[
Y_t = \nabla^2 X_t = \nabla (X_t - \nabla X_t) = X_t - 2X_{t+1} + X_{t+2}
\]

Analysing Time Series which contain a Seasonal Variation

Some possible seasonal models are:

1. \( X_t = m_t + S_t + \varepsilon_t \)
\[ \text{where } m_t \text{ is the deseasonalised mean level,} \]
2. \( X_t = m_t S_t + \varepsilon_t \)
\[ S_t \text{ is the seasonal effect, and } \varepsilon_t \text{ is random error.} \]
3. \( X_t = m_t S_t \varepsilon_t \)

(Note: by random error, I do not mean that it follows a distribution [e.g. \( N(0,1) \)], but that it is purely random.)

Model (1) shows additive seasonality, while models (2)&(3) show multiplicative seasonality, but it is always possible that a more complex variation on these is needed. With the additive model, the seasonal effect can be eliminated using differencing and other more specific techniques. Therefore, in the cases of models (2)&(3) we should transform the data using logarithms so that these standard techniques can be employed on them as well.

Autocorrelation

A guide to the properties of time series is provided by the sample autocorrelation coefficients, which simply measures the correlation between observations at different distances apart, and we can use them as a guide for later models when we are using stationary data. If the data is not stationary, just use the techniques mentioned above to make sure it is stationary.
Given N observations $X_1,X_2,X_3,\ldots,X_n$ from a discrete time series we can form $(N-1)$ pairs of observations $(X_1,X_2),(X_2,X_3),\ldots,(X_{n-1},X_n)$. Regarding the first observation in each pair as one variable, and the second observation as another, the correlation between $X_t$ and $X_{t+1}$ is \textit{approximated} by;

$$\text{Sample Autocorrelation Coefficient (a.c.v.f.)} = r_t = \frac{\sum_{t=1}^{N-1} (X_t - \bar{X})(X_{t+1} - \bar{X})}{\sum_{t=1}^{N} (X_t - \bar{X})^2} \approx \gamma(1)$$

= Autocorrelation coefficient

Similarly, we can find the correlation between observations k apart.

$$\text{Sample Autocorrelation Coefficient (a.c.v.f.)} = r_t = \frac{\sum_{t=1}^{N-k} (X_t - \bar{X})(X_{t+k} - \bar{X})}{\sum_{t=1}^{N} (X_t - \bar{X})^2} \approx \gamma(k)$$

= Autocorrelation coefficient at lag k

These approximations are based on the assumption that N is large.

Similarly, we can calculate the covariance between terms k steps apart.

$$\text{Sample Covariance Coefficient (a.c.f.)} = \frac{r_k}{\gamma_0} \approx \rho_k = \rho(k) = \text{Covariance at lag K}$$

(k \in \text{integers})

\textbf{A Correlogram} is a point graph plotting lag k against sample autocorrelation $r_k$, and is very useful as a guide to which kind of model to fit to stationary data.

Interpreting the correlogram;

(a) If a time series is completely random, then for $k \neq 0$ and N large, $r_k \approx 0$.

(it does not matter if there are a few non-zero, but not too many)

(b) Short term correlation:-

Stationary series will exhibit short term correlation, which means that initially (i.e. k small) there will be a large value of $r_k \approx 1$, followed by terms tending to zero

e.g.

We can interpret the correlogram as follows. A large correlation means that an observation above the mean tends to be followed by another above the mean, and vice-versa.
(c) Alternating Series:-
If a time series has a tendency to oscillate about it's overall mean, the correlogram also tends to oscillate.

(d) Non Stationary Time Series:-
Little can be inferred from these correlograms since any trend in the time series will effect the correlogram. Therefore we should use techniques, such as filtering and differencing etc., to make the time series is stationary before plotting any correlograms.

(e) Seasonal Fluctuations:-
If the time series has a seasonal fluctuation, the correlogram will have oscillations of the same frequency.

(f) Outliers:-
An outlier in the time series will give two "extreme" points in the correlation, since if \( X_T \) is an outlier then \( Corr(X_T, X_{T+k}) \approx 0 \), which will depress the sample correlation towards zero.

Tests of Randomness

A visual inspection is usually enough to determine if a data set is random. However, it is also possible to take a sample which is known to be random (i.e. a random sample from an unknown population) and compare this to the original data set to see if the two distributions have roughly the same number of maxima and minima.
Stochastic Processes

This refers to time series which continue to evolve over time. Although it is only possible to make one observation at time \( t \) and so gain one time series, we can regard this as one of an infinite set of time series called an ensemble, so that every member of the ensemble is a possible realisation of a stochastic process. One way of describing a stochastic process is by giving the moments of the process;

1st Moment - Mean and Variance functions.
2nd Moment - Autocovariance function.

\[
\text{Mean} = \mu(t) = E[X(t)]
\]
\[
\text{Variance} = \sigma^2(t)
\]
\[
\text{Autocovariance} = \gamma(t_1, t_2) = E[(X(t_1) - \mu(t_1))(X(t_2) - \mu(t_2))]
\]

From this second moment, it is easy to gain the a.c.f. =
\[
\rho(t_1 - t_2) = \rho(\tau) = \frac{\gamma(\tau)}{\gamma(0)}
\]

Properties of the a.c.f.

The (theoretical) a.c.f. of a stationary stochastic process is a tool for assessing which process is underlying the data. Suppose a stationary stochastic process \( X_t \) has mean \( \mu \), variance \( \sigma^2 \), a.c.f. \( \rho(\tau) \), and a.c.v.f. \( \gamma(\tau) \). Then;
(i) The a.c.f. is an even function (i.e. \( \rho(\tau) = \rho(-\tau) \)).
(ii) \( |\rho(\tau)| \leq 1 \)
(iii) The function is not unique. Although every stochastic process has a unique a.c.v.f., if we are given a a.c.v.f., we cannot find a unique model.

[Example
Random walk:-
Suppose \( \{Z_t\} \) is a discrete purely random process (i.e. \( Z_t \in H(\mu, \sigma^2_z) \) where \( H \) is some distribution with mean \( \mu \) and variance \( \sigma^2_z \)). A process \( \{X_t\} \) is said to be a random walk if it can be written as
\[
X_t = X_{t-1} + Z_t
\]
The process is usually started at \( t = 0 \), so this can be rewritten as
\[
X_t = \sum_{i=1}^t Z_i,
\]
\[
\therefore E[X_t] = t\mu, \text{Var}[X_t] = t\sigma^2_z
\]
We can note that since the mean and variance of the process \( \{X_t\} \) are dependant on time, the process is not stationary.]

7
§3. Models for Stationary Data

Moving Average Processes MA(q)

Suppose we have \( \{Z_t\} \) as before, then a process \( \{X_t\} \) is said to be a **moving average** process of order q (MA(q)), if it can be written in the form:

\[
X_t = \beta_0 Z_t + \beta_1 Z_{t-1} + \beta_2 Z_{t-2} + \ldots + \beta_q Z_{t-q}
\]

where \( \{\beta_i\} \) are constants, and the \( Z_i \) are usually scaled so that \( \beta_0 = 1 \).

\[\Rightarrow E[X_t] = 0, \text{Var}[X_t] = \sigma^2 \sum_{i=1}^{q} \beta_i^2\]

Since the \( Z_i \)'s are independently distributed,

\[
\gamma(k) = \text{Cov}(X_t, X_{t+k}) = \text{Cov}(\beta_0 Z_t + \ldots + \beta_q Z_{t-q}, \beta_0 Z_{t+k} + \ldots + \beta_q Z_{t+k-q})
\]

\[\approx \begin{cases} 
0 & \text{if } k > q \\
\sigma^2 \sum_{i=0}^{q-k} \beta_i \beta_{i+k} & k = 0, 1, \ldots, q \\
\gamma(-k) & k < 0
\end{cases}\]

Since the mean is constant and the a.c.v.f. is not dependant on time, the process is second-order stationary. Using these results we can evaluate;

\[
a.c.f. = \rho(k) = \frac{\gamma(k)}{\gamma(0)} \approx \begin{cases} 
1 & \text{if } k = 0 \\
\sum_{i=0}^{q-k} \beta_i \beta_{i+k} / \sum_{i=0}^{q} \beta_i^2 & k = 1, 2, \ldots, q \\
\rho(-k) & k < 0
\end{cases}
\]

[Example

An MA(1) process with \( \beta_0 = 1 \). In this case \( \rho(k) \approx \begin{cases} 
1 & \text{if } k = 0 \\
\beta_1 / (1 + \beta_1^2) & \text{if } k = \pm 1 \\
0 & \text{otherwise}
\end{cases}\)

Autoregressive Processes AR(p)

Suppose \( \{Z_t\} \) is a purely random process as before. A process is called **autoregressive** AR(p) if it can be written in the form,

\[
X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \ldots + \phi_p X_{t-p} + Z_t
\]

where \( \phi_1, \phi_2, \ldots, \phi_p \) are constants.

It we wish to fit this process to a given set of data, we need to estimate the parameters \( \phi_1, \phi_2, \ldots, \phi_p \) by least squares. I will address the specifics of the method later in §4.

The reason for the name of the process is that \( X_t \) is being regressed on past values of \( X_k \) where \( k < t \).

Notation
Let \( B^j X_t = X_{t-j} \quad \forall \ j \in \text{(non-negative integers)} \)
where B is called the **backward shift operator**.

[e.g. \( X_t = \beta_0 Z_t + \beta_1 Z_{t-1} + \ldots + \beta_q Z_{t-q} = Z_t (\beta_0 + \beta_1 B + \ldots + \beta_q B^q) = \Theta(B) Z_t \)]

where \( \Theta(B) \) is a polynomial of order q in B.

**AR(p)** \( X_t = \phi_1 X_{t-1} + \ldots + \phi_p X_{t-p} + Z_t \)
\[\Rightarrow Z_t = X_t (1 - \phi_1 B - \ldots - \phi_p B^p) = \Phi(B) X_t \]
where \( \Phi(B) \) is a polynomial of order p in B.

Therefore, any general process which can be written in the form
\( X_t = \phi_1 X_{t-1} + \ldots + \phi_p X_{t-p} + Z_t + \theta_1 Z_{t-1} + \ldots + \theta_q Z_{t-q} \)
can also be written as,
\[ X_t \Phi(B) = Z_t \Theta(B) \]
\[ \Phi(B) = 1 - \phi_1 B - \ldots - \phi_p B^p \]
where \( \Theta(B) = 1 + \theta_1 B + \ldots + \theta_q B^q \)

For an **AR(p)** process it is also possible to use the backwards shift operator to express
the equation as an **MA** process of infinite order.

i.e. \[ X_t = \frac{Z_t}{(1 - \phi_1 B - \ldots - \phi_p B^p)} = Z_t f(B) \]

[e.g. \( X_t = \phi_1 X_{t-1} + Z_t \Rightarrow X_t (1 - \phi_1 B) = Z_t \)]
\[\Rightarrow X_t = \frac{Z_t}{(1 - \phi_1 B)} = Z_t \sum_{j=0}^{\infty} (\phi_1 B)^j \text{(MA(\infty))} \]
by using the Binomial expansion (assuming \(|\phi_1| < 1\)).]

It should be noted that although the models we have been considering are applicable
to zero mean stationary data only, they can be easily adapted to apply to stationary
data with mean \( \mu \).

i.e. \( (X_t - \mu) = \phi_1 (X_{t-1} - \mu) + \ldots + \phi_p (X_{t-p} - \mu) + Z_t \)

**Mixed MA and AR modelling [ARMA(p,q)]**

These are just an additive combination of the **MA(q)** and **AR(p)** models.

i.e. \( X_t = \phi_1 X_{t-1} + \ldots + \phi_p X_{t-p} + Z_t + \theta_1 Z_{t-1} + \ldots + \theta_q Z_{t-q} \) and therefore can be
written as \( X_t \Phi(B) = Z_t \Theta(B) \) using the backwards shift operator.
General properties of AR and MA models

If we have a standard mixed ARMA(p,q) process,

\[ X_t \Phi(B) = Z_t \Theta(B) \]

where \( \Phi(B) = 1 - \phi_1 B - \ldots - \phi_p B^p \)
\( \Theta(B) = 1 + \theta_1 B + \ldots + \theta_q B^q \)

Theorem
Suppose the polynomials \( \Phi(y) = 0, \Theta(y) = 0 \) have no common roots \( \forall y \in \) (complex no.).
The ARMA(p,q) process is stationary and causal \( \iff \) The solutions to the polynomial \( \Phi(y) = 0 \) lie in \( \{ y \in \text{(complex no.)} : |y| > 1 \} \)
(i.e. outside the unit circle)

Proof
At this time, I will just show that the theorem is true in the case of the AR(p) model.
(i.e. \( \Theta(y) \equiv 1 \))

\( \Rightarrow \) Assuming the R.H.S. is true, we can factorise \( \Phi \) as,

\[ \Phi(y) = 1 - \phi_1 y - \ldots - \phi_p y^p = (1 - \lambda_1 y)\ldots(1 - \lambda_p y) \]
with \( |\lambda_i| < 1, \ 1 \leq i \leq p \) and the \( \lambda_i \) are distinct.

Hence \( (\Phi(y))^{-1} = \frac{1}{(1 - \lambda_1 y)\ldots(1 - \lambda_p y)} = \frac{A_1}{(1 - \lambda_1 y)} + \ldots + \frac{A_n}{(1 - \lambda_n y)} \)
where the \( A_i \) are constants, by using partial fractions.

[Note: it can be seen that \( \sum_{i=1}^{p} A_i = 1 \) by putting \( y = 0 \) on the above equation.]

By the formula for geometric series, \( \frac{1}{(1 - \lambda_i y)} = \sum_{j=0}^{\infty} (\lambda_i y)^j \) since \( |\lambda_i| < 1 \).

Hence, we can write \( (\Phi(y))^{-1} = \sum_{j=0}^{\infty} a_j y^j \) where \( a_j = A_1 \lambda_1^j + \ldots + A_p \lambda_p^j \)
\( a_0 = 1 \), and the \( a_j \) decrease geometrically.

Therefore we can see that \( \Phi(y)(\Phi(y))^{-1} = (1 - \phi_1 y - \ldots - \phi_p y^p)(1 + a_1 y + \ldots) = 1 \)
i.e. we can write the process as an MA(\( \infty \)), which is therefore stationary
[since \( E(Z_t) = 0 \ \forall t \implies E(X_t) \) is independent of \( t \), and we can similarly show
that \( E(X_t X_{t+k}) \) is also independent of time.]

\( \Rightarrow \) Assuming the L.H.S. is true, set \( \Phi'(y) = (1 - \lambda_2 y)\ldots(1 - \lambda_p y) \)
so that \( X_t(1 - \lambda_1 B) \Phi'(B) = Z_t \), and let \( Y_t = \frac{Z_t}{(1 - \lambda_1 B)} = \frac{X_t \Phi(B)}{(1 - \lambda_1 B)} \).

\[ \iff Y_t(1 - \lambda_1 B) = Z_t \]
Therefore, \( \iff Y_t = \lambda_1 Y_{t-1} + Z_t \) which is an AR(1) process.

Since it is possible for \( \lambda_i \) to be complex, this means \( Y_t \) could also be complex.
Therefore we consider, \( E(Y_tY_t) = E[(Z_t + \lambda_1 Y_{t-1})(Z_t + \lambda_1 Y_{t-1})] \)
\[ \Rightarrow E(Y_tY_t) = E(Z_tZ_t) + E(Z_t)E(\lambda_1 Y_{t-1}) + E(Z_t)E(\lambda_1 Y_{t-1}) + \lambda_1 \lambda_1 E(Y_{t-1}Y_{t-1}) \]
Since \( Z_t \) and \( Y_t \) are independent, \( \lambda_1 \) and \( \bar{\lambda}_1 \) are constants.
\[ \Rightarrow E(|Y_t|^2) = \sigma^2 + |\lambda_1|^2 E(|Y_{t-1}|^2) = \sigma^2 + |\lambda_1|^2 E(|Y_t|^2), \text{ since } Y_t \text{ is stationary.} \]
\[ \Rightarrow E(|Y_t|^2) = \frac{\sigma^2}{(1-|\lambda_1|^2)} \text{ and therefore we must have } |\lambda_1| < 1 \text{ otherwise L.H.S. < R.H.S.} \]

Finally, it should be noted that \( \lambda_1 \) is not special in this set of equations. By factorising differently at the start, and consequently choosing a different \( Y_t \), we can gain this condition for any of the \( \lambda_i \). Hence, we need \( |\lambda_i| < 1 \) \( \forall \ i = 1, \ldots, p \). Q.E.D.

Although no restrictions are placed on the \( \{\theta_i\} \) if an MA(q) process is to be invertible, it is sometimes useful to impose a condition to ensure \textbf{invertibility}. This is the analogous condition that the roots of \( \Theta(y) = 0 \) lie outside the unit circle.

Hence, this means that given a MA(q) process \( X_t = Z_t + \theta_1 Z_{t-1} + \ldots + \theta_q Z_{t-q} = Z_t\Theta(B) \)
we can write it in the form \( Z_t = \frac{X_t}{\Theta(B)} = X_tg(B) \) an AR(\( \infty \)) process.
\[ \text{[e.g. } X_t = Z_t + \theta_1 Z_{t-1} \]
\[ \Rightarrow Z_t = \frac{X_t}{(1 + \theta_1 B)} = X_t \sum_{j=0}^{\infty} (-\theta_1 B)^j. \]

\textbf{Integrated ARIMA models}

To use an ARMA(p,q) model on a set of data we first have to make sure that non-stationary sources of variation had to be removed from the data. ARIMA(p,d,q) models are the same as ARMA models, except that they incorporate differencing into their structure, then fit an ARMA model, and finally sum or "integrate" to give a final model which applies to the original data which was non-stationary.

i.e. Let \( w_t = \nabla^d X_t = (1-B)^d X_t \). Hence, we can write the ARMA(p,q) model as,
\[ w_t = \left( \phi_1 X_t + \ldots + \phi_p X_{t-p} + Z_t \right) + \left( \theta_1 Z_{t-1} + \ldots + \theta_q Z_{t-q} \right) \]
\[ \Rightarrow w_t \Phi(B) = X_t(1-B)^d \Phi(B) = Z_t \Theta(B) \]
We have gained an ARMA(p,q) model for \( w_t \).
Overall, the two equations represent an ARIMA(p,d,q) process.
Definition
The infinite order MA process with non-zero mean given by,

\[(X_t - \mu) = \sum_{j=0}^{\infty} \beta_j Z_{t-j}\]

is called a general linear process since it can be obtained by passing a purely random process through a linear filter.

[Note: all the models we have considered so far only apply to discrete data, but once the models are understood it is relatively easy to generalise them to encompass continuous processes.]
The Yule-Walker Equations

To begin with, I will consider the set of formulae called the Yule-Walker equations which are used in calculating the a.c.f of any stationary causal AR(p) process. However, we will see later on that these equations can be useful in many other areas beside this specific case.

Given a stationary causal AR(p) process,

$$X_t - \phi_1 X_{t-1} - \ldots - \phi_p X_{t-p} = Z_t$$

multiply by $X_{t-k}$ (k≠0), take expectations, and divide by $\gamma(0)$ to gain;

$$\frac{E[X_t X_{t-k} - \phi_1 X_{t-1} X_{t-k} - \ldots - \phi_p X_{t-p} X_{t-k}]}{\gamma(0)} = \frac{E[Z_t X_{t-k}]}{\gamma(0)} = \frac{E[Z_t]E[X_{t-k}]}{\gamma(0)} = 0$$

since $X_{t-k}$ is independent of $Z_t$, and $E[Z_t] = 0 \ \forall \ t$.

Hence,

$$\frac{\gamma(k) - \phi_1 \gamma(k-1) - \ldots - \phi_p \gamma(k-p)}{\gamma(0)} = 0$$

and this implies;

$$\rho(k) - \phi_1 \rho(k-1) - \ldots - \phi_p \rho(k-p) = 0 \ \text{The Y.W. equations}$$

This is valid for K=1,2,3,...

Solutions to the Y.W. equations are of the form,

$$\rho(j) = A_1 \Pi_1^j + A_2 \Pi_2^j + \ldots + A_p \Pi_p^j$$

where the $A_i$ are constants and,

$$y = \Pi_j$$

are solutions to

$$y^p - \phi_1 y^{p-1} - \ldots - \phi_{p-1} y - \phi_p = 0$$

[Note: If all $|\Pi_j| < 1$ then $\rho(j)$ decays geometrically, but does not cut off.]
[Example]

Calculate $\rho(j)$ for the process $X_t -\frac{1}{3}X_{t-1} -\frac{2}{9}X_{t-2} = Z_t$ using the Y.W. equations.

Hence, consider;

$$\rho(j) - \frac{1}{3}\rho(j-1) - \frac{2}{9}\rho(j-2) = 0$$

which has solutions of the form,

$$\rho(j) = A\Pi_1^j + B\Pi_2^j$$

where $\Pi_j$ are solutions to

$$y^2 - \frac{1}{3}y - \frac{2}{9} = (y - \frac{\sqrt{3}}{3})(y + \frac{\sqrt{3}}{3}) = 0,$$

therefore, $\Pi_1 = -\frac{\sqrt{3}}{3}, \Pi_2 = \frac{\sqrt{3}}{3}$.

To find A, B we use Y.W.(1) and Y.W.(0).

$$\rho(1) - \frac{1}{3}\rho(0) - \frac{2}{9}\rho(-1) = 0,$$

and $\rho(0) = 1, \rho(1) = \rho(-1)$, so $\rho(1) = \frac{3}{7}$.

Also,

$$\rho(1) = A\left(\frac{-1}{3}\right)^1 + B\left(\frac{2}{3}\right)^1 = \frac{3}{7}$$

and $\rho(0) = 1 = A + B$.

Solving this simultaneous equation, we gain $A = \frac{5}{21}, B = \frac{16}{21}$, and therefore,

$$\rho(j) = \frac{1}{21}\left[5\left(\frac{-1}{3}\right)^j + 16\left(\frac{2}{3}\right)^j\right].$$

Estimating a.c.v.f.’s, a.c.f.’s and **partial** a.c.f.’s

Using the Yule-Walker equations if we know the situation is from a stationary causal AR(p) model, or more generally, the formula from §2, it is simply a matter of number-crunching to calculate the a.c.v.f. $\gamma(j)$ and the a.c.f. $\rho(j)$.

Partial autocorrelation functions are evaluated when we believe the underlying model for given stationary data is an AR(p) process, and their use is explained in the next section, but this section will just deal with how they are defined in terms the Yule-Walker equations.

Say we observe $x_1, \ldots, x_n$ from $(X_t)$ where $Y_t = X_t - \mu$ satisfies

$$Y_t = \phi_1Y_{t-1} + \ldots + \phi_pY_{t-p} + Z_t$$

(i.e. an AR(p) process.)

Using least squares, and assuming that $n$ is large $\bar{x} = \hat{\mu}$ and $\hat{\phi}_1, \ldots, \hat{\phi}_p$ are the estimates for the parameters.

In this situation the Yule-Walker equations are

$$\rho(k) = \phi_1\rho(k-1) + \ldots + \phi_p\rho(k-p)$$
Therefore, the first \( p \) Y.W. equations in matrix form are:

\[
\begin{pmatrix}
\rho(1) \\
\rho(2) \\
\vdots \\
\rho(p)
\end{pmatrix} = \begin{pmatrix}
\rho(0) & \rho(1) & \cdots & \rho(p-1) \\
\rho(1) & \rho(0) & & \\
\vdots & & & \ddots \\
\rho(p-1) & \cdots & \rho(1) & \rho(0)
\end{pmatrix} \begin{pmatrix}
\phi_1 \\
\phi_2 \\
\vdots \\
\phi_p
\end{pmatrix}
\]

Using the estimates \( r_k \) for the theoretical \( \rho(j) \), we can calculate the estimates \( \hat{\phi}_1 \) for the \( \phi_j \).

\[
\begin{pmatrix}
r_1 \\
r_2 \\
\vdots \\
r_p
\end{pmatrix} = \begin{pmatrix}
\hat{\phi}_1 \\
\hat{\phi}_2 \\
\vdots \\
\hat{\phi}_p
\end{pmatrix}
\]

where \( R^{(p)} \) is a \((p \times p)\) matrix with \((R^{(p)})_{ij} = r_{i-j}\).

The \textbf{Kth partial sample a.c.f.} is the last entry \( \hat{\phi}_k^{(k)} \) in the solution to,

\[
R^{(k)} = \begin{pmatrix}
\hat{\phi}_1^{(k)} \\
\hat{\phi}_2^{(k)} \\
\vdots \\
\hat{\phi}_k^{(k)}
\end{pmatrix} \begin{pmatrix}
r_1 \\
r_2 \\
\vdots \\
r_k
\end{pmatrix}
\]

\textbf{Example}

If \( k = 2 \) then,

\[
\begin{pmatrix}
r_1 \\
r_2
\end{pmatrix} = \begin{pmatrix}
1 & r_1 \\
r_1 & 1
\end{pmatrix} \begin{pmatrix}
\hat{\phi}_1 \\
\hat{\phi}_2
\end{pmatrix} \Rightarrow \begin{pmatrix}
\hat{\phi}_1 \\
\hat{\phi}_2
\end{pmatrix} = \frac{1}{1 - r_1^2} \begin{pmatrix}
1 & -r_1 \\
-1 & 1
\end{pmatrix} \begin{pmatrix}
r_1 \\
r_2
\end{pmatrix}.
\]

Now, consider the properties of the Kth partial sample a.c.f.

(i) If \( k = p \) then \( \hat{\phi}_1^{(k)}, \ldots, \hat{\phi}_p^{(k)} \) are estimates for \( \phi_1, \ldots, \phi_p \). Hence,

\( \hat{\phi}_k^{(k)} \approx \phi_k \neq 0 \).

(ii) If \( k > p \) then the same is true as before, except that \( \hat{\phi}_{p+1}, \ldots, \hat{\phi}_k \) are all zero.

Hence, \( \hat{\phi}_k^{(k)} \approx \phi_k = 0 \).

In fact, for \( n \) large, \( \hat{\phi}_k^{(k)} \sim \text{N}(0, 1/n) \) \( \Rightarrow P\left( \left| \hat{\phi}_k^{(k)} \right| > 1.96/\sqrt{n} \right) \approx 0.05 \).

(iii) If \( k < p \) then we still expect \( \hat{\phi}_k^{(k)} \approx \phi_k \neq 0 \).

[C.F. to the a.c.f. of an MA(q) process]
Interpreting the Correlogram, with regard to MA and AR models

From §2, we defined the correlogram as the plot of the sample autocorrelation function against lag k. If we are given a set of stationary data and we need to decide whether to fit an MA(q) or AR(p) process, then we can use the correlogram to justify fitting one over the other.

(A) Suppose the following correlogram is gained:

The a.c.f. cuts off after $k = 3$, and for $k > 3$ it remains within the bounds of $\pm 1.96/\sqrt{n}$ (a 5% significance interval). Since the a.c.f. has all the characteristics of the theoretical a.c.f. of an MA(3) process, this is evidence for fitting an MA(3) model, and against fitting an AR(p) model.

[Note: even if the a.c.f. deviated slight outside the $\pm 1.96/\sqrt{n}$ interval, only if there were "significant" deviation would this cast doubt on the wisdom of fitting the MA(q) model]

(B) Instead, suppose the following correlogram is gained:

The a.c.f. does not appear to "cut-off", but instead decays (possibly geometrically) so that this is evidence against the data being produced by a MA(q) process. In this situation it is prudent to examine the Kth partial sample a.c.f. ($k$ being determined using expert opinion or other methods).
Suppose the following plot is gained when \( k=3 \);

![Plot showing partial sample ACF](image)

The p.a.c.f. cuts off after \( k=3 \), and for \( k>3 \) it remains within the bounds of \( \pm 1.96/\sqrt{n} \) (a 5% significance interval). Since the p.a.c.f. has all the characteristics of the theoretical p.a.c.f. of an AR(3) process, this is evidence for fitting an AR(3) model.

However, if the plot of the p.a.c.f. is also inconclusive, then the only thing we can conclude is that these plots give evidence against the stationary data being produced by MA or AR models, and so we should not attempt to fit such models to the data.

In the case where we believe a mixed ARMA(p,q) model should be used, the correlogram and plot of the partial a.c.f. provide us with no clear evidence for its use, since it is very difficult to calculate the behaviour of the theoretical a.c.f. and p.a.c.f.

**Fitting an MA(q), AR(p), or ARMA(p,q) model**

By fitting a model to stationary data, I mean that I am estimating values of the parameters for each specific model. At this stage I will just consider the combined ARMA(p,q) model;

\[
X_t = \phi_1 X_{t-1} + \ldots + \phi_p X_{t-p} + Z_t + \theta_1 Z_{t-1} + \ldots + \theta_q Z_{t-q}
\]

since the other two cases are included within this general case.

Now this equation can be rewritten as

\[
Z_t = X_t - \sum_{j=1}^{p} \phi_j X_{t-j} - \sum_{j=1}^{q} \theta_j Z_{t-j}
\]

Now, we can gain the residuals \( z_t \) by recursion, using the past values of \( X_t = x_t \).

Since, \( E(X_t) = E(Z_t) = 0 \quad \forall \ t \leq 0 \), we set \( x_t = z_t = 0 \quad \forall \ t \leq 0 \) (i.e. the "observed" values from the data are given in small script, by convention.)
Hence, we gain;

\[ z_1 = x_1 \]
\[ z_2 = x_2 - \phi_1 x_1 - \theta_1 z_1 = x_2 - (\phi_1 + \theta_1) x_1 \]
\[ z_3 = x_3 - \phi_1 x_2 - \phi_2 x_1 - \theta_1 z_2 - \theta_2 z_1 \]

[Note: in fact these are the same terms we would gain if we had calculated the
Taylors expansion of \( Z = \Phi(B)X \), to gain the AR(\( \infty \)) process \( Z_t = \sum_{j=0}^{\infty} b_j X_{t-j}, \)
and then used \( z_t = \sum_{j=0}^{\infty} b_j x_{t-j} \) to estimate the residuals.]

Therefore, we have \( z_t = z_t(\Phi, \Theta) \) for \( 1 \leq t \leq n \).

The least squares estimates \( \hat{\Phi}, \hat{\Theta} \) for \( \Phi, \Theta \) are chosen to minimise
\[ \sum_{t=1}^{n} [z_t(\Phi, \Theta)]^2. \]

If the parameters were for an AR(p) process, we can use calculus and the Yule-Walker equations as before to estimate the parameters. However, for all other cases, "brute force" is needed. By this I mean that a grid of values of \( \Phi, \Theta \) is used and the values which minimise the residual sum of squares are used as the least squares estimates.

It is also possible to use the Kalman filter to approximate the ARMA model parameters.

Estimating the Parameters of an ARIMA model

In this situation, if we believe an ARIMA(p,d,q) model should be fitted and we need to estimate the parameters in the model, then it is simplest to just difference the data, fit an ARMA(p,q) model in the manner described above, then "integrate" this model so that it is applicable to the original data set.

Residual Analysis (Diagnostics)

To check a model provides an adequate description of the data we need to consider the residuals.

\[ \text{Residual} = \text{Observed} - \text{Fitted value} \]

A "good" model has residuals which are apparently random, and close to zero. In order to determine whether this is the case we can treat the residuals \( Z_t \) in the same manner we would any other time series. Therefore we plot them on a graph, and also evaluate the correlogram of these residuals (i.e. to show gross correlations, and plotting the 5% significance level \( \pm 1.96/\sqrt{n} \) to see if significantly many points lie outside it, which is evidence that the incorrect model has been fitted.).
§5. Forecasting

Suppose we have an observed time series $X_1, X_2, X_3, \ldots, X_n$. The problem is to estimate the value of $X_{n+k}$, where $k \in \text{integers}$ is the lead time. The forecast of $X_{n+k}$, made at time $N$ for $k$ steps ahead is denoted by $\hat{X}_{N,k}$.

Forecasting methods:
(a) Subjective (i.e. intuitive)
(b) Univariate: A forecast of the variable $\hat{X}_{N,k}$ based only on past observations of the time series $X_1, X_2, X_3, \ldots, X_n$ (called naive/projection methods)
(c) Multivariate: A forecast of the variable dependant, at least partially, on another time series of predictor/explanatory values. (e.g. consumer spending dependant on economic climate, time of year etc.)

It is prudent to use a combination of these methods, and not to rely too heavily on any single one. Also, it should be noted that some forecasts may influence the future when they are used, and so are potentially self-fulfilling. However, whichever method is used in forecasting, a forecast monitoring scheme (diagnostics) is needed to tell when things start to go wrong. For this we calculate residuals, and it is easiest to understand them when they are applied to a particular situation, which is why I will consider;

Univariate Methods
(i) Extrapolation of trend curves can be used, but there cannot be any objective justification for one curve over another.
(ii) Exponential Smoothing; can only be applied to time series which have no seasonal or systematic trends. The method is that given a time series $X_1, X_2, X_3, \ldots, X_n$, we calculate an estimate of $X_{N+1}$ as the weighted sum of past observations.

\[ \hat{X}_{N+1} = c_0 X_N + c_1 X_{N-1} + \ldots + c_N X_0 \]

where \( \{c_i\} \) are the weights.

A "sensible" idea is to give more weight to the most recent results, as compared to those in the distant past, so a geometric distribution would be reasonable to use.

In order that the weights sum to 1, let

\[ c_i = \alpha (1 - \alpha)^i \text{ where } i=0,1,\ldots \text{ and } |\alpha| \leq 1 \]

Hence, equation \( \otimes \) becomes,

\[ \hat{X}_{N+1} = \alpha X_N + \alpha (1 - \alpha) X_{N-1} + \ldots \]

\[ \hat{X}_{N+1} = \alpha X_N + (1 - \alpha) \hat{X}_{N-1} \]
If we set $\hat{X}_{1,1} = X_1$, then this equation can be used recursively to compute forecasts, and also means that forecasts can easily be updated.

i.e. $\hat{X}_{N,1} = \alpha e_N + \hat{X}_{N-1,1} = \alpha \left[ X_N - \hat{X}_{N-1,1} \right] + \hat{X}_{N-1,1}$

where $e_N$ is the prediction error at time $N$.

How does the value of $\alpha$ effect the forecast?

For $\alpha \approx 1$ the forecast is heavily dependant on the most recent observations. For $0.1 \leq \alpha \leq 0.3$ the forecast is less dependant on recent observations, and we need a large value of $N$ if we are to be able to make accurate forecasts.

⊕⊕ To find the optimum value of $\alpha$, work out the sum of squares of the prediction errors (residuals), and find it's minimum when varying $\alpha$.

i.e. $\hat{X}_2 = X_1 \Rightarrow e_2 = X_2 - \hat{X}_2 = X_2 - X_1$

$\hat{X}_3 = \hat{X}_{2,1} = \alpha e_2 + \hat{X}_{1,1} \Rightarrow e_3 = X_3 - \hat{X}_3 = X_3 - \hat{X}_{2,1}$

$e_4 = X_4 - \hat{X}_{3,1}$

$\ldots$

$e_N = X_N - \hat{X}_{N-1,1}$
§6. State-Space Models and Kalman Filtering

State space models are a general class of models originally developed for engineering situations (e.g. position of a rocket), but are also useful for short term forecasting in time series problems. "Kalman Filtering is a method of signal processing giving optimal estimates of the state of a dynamic system". Essentially, it is a set of recurrence relations estimating the current state of the parameters of the system, and the variances of these estimates. Typically,

\[
\text{Observation} = \text{Signal} + \text{Noise}
\]

In state space models the signal is assumed to be a linear combination of variables \( \text{(state variables)} \) forming the state vector which describes the state of the system.

\[
X_t = h_t^T \Theta_t + n_t \quad \text{Observation equation}
\]

where \( X_t \) is the scalar variable, \( \Theta_t \) is an \( (m \times 1) \) state vector, \( h_t \) is a known vector, and \( n_t \) is the observed error.

We use the observations \( X_t \) to make inferences about \( \Theta_t \), and it is reasonable to assume that we know how \( \Theta_t \) changes/updates through time.

\[
\Theta_t = G_t \Theta_{t-1} + w_t \quad \text{Transition equation}
\]

where \( G_t \) is a known \( (m \times m) \) matrix and \( w_t \) is a vector of deviations.

These equations are the general form of a univariate state space model.

[Note: We can easily change \( X_t \) into the vector \( X_t \) by turning \( h_t \) into a matrix, and \( n_t \) into a vector in the equations above.]

An inherent feature of structural (and state space) models is that the observation equation involves a linear function of the state variables. This allows local features, such as trend and seasonality, to be updated by the transition equations. Now I will deal with more specific applications of this basic type of method.

The Steady Model

Suppose the observation equation is,

\[
X_t = \mu_t + n_t
\]

where the unobservable current level \( \mu_t \) is assumed to follow a random walk, given by the transition equation,

\[
\mu_t = \mu_{t-1} + w_t
\]

In this case, the state vector \( \Theta_t \) is just a scalar \( \mu_t \), and therefore the corresponding variables \( h_t = G_t = 1 \). We also assume that the error terms \( n_t, w_t \) are independant, with normal distributions.
Then, by definition, we have

\[ \text{signal to noise ratio} = \frac{\sigma_w^2}{\sigma_n^2} \]

This is called the steady model because there is no trend term, just a "randomly" changing mean. It is useful because we can show that exponential smoothing produces optimal forecasts of the steady model, and that it shares the same forecast function with the ARIMA(0,1,1) process.

The Linear Growth Model

This is specified by three equations;

\[ X_t = \mu_t + n_t \quad \text{Observation equation} \]

\[ \mu_t = \mu_{t-1} + \beta_{t-1} + w_{1,t} \quad \text{Transition equations} \]

\[ \beta_t = \beta_{t-1} + w_{2,t} \]

Hence, the state vector \( \theta_t \) equals \( (\mu_t, \beta_t) \), and has two components interpreted as local mean \( \mu_t \), and local trend \( \beta_t \). Therefore,

\[ X_t = h_t^T(\mu_t) + n_t \Rightarrow h_t^T = (1,0) \quad \text{so that } X_t = \mu_t + n_t \]

But,

\[ \theta_t = \begin{pmatrix} \mu_t \\ \beta_t \end{pmatrix} = G_t \begin{pmatrix} \mu_{t-1} \\ \beta_{t-1} \end{pmatrix} + \begin{pmatrix} w_{1,t} \\ w_{2,t} \end{pmatrix} \Rightarrow G_t = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \text{ so that the observation equations are satisfied.} \]

This type of model allows trend and mean to vary independently. However, if \( w_{1,t} \) and \( w_{2,t} \) have zero variance the trend will be constant and we have a **global** linear trend model. Unfortunately, it is not always reasonable to make this assumption concerning \( w_{1,t} \) and \( w_{2,t} \), so instead we use **local** linear trend models.

[Note: State space models have the advantage of covering a wider variety of models than ARIMA models can.]
The Basic Structural Model

This incorporates seasonality into the state space model;

\[ X_t = \mu_t + i_t + n_t \]
\[ \mu_t = \mu_{t-1} + \beta_{t-1} + w_{1,t} \]
\[ \beta_t = \beta_{t-1} + w_{2,t} \]

where \( i_t = w_{3,t} - \sum_{j=1}^{s-1} i_{t-j}, \mu_t \) is the local mean level,

\[ \beta_t \] is local trend, and \( i_t \) is the local seasonal index.

If there are S periods in a year/season, then this model assumes that the expectation of the "sum of seasonal effects" over the year is zero. The state vector now has \((S+1)\) components, \( \mu_t, \beta_t, i_t, i_{t-1}, \ldots, i_{t-s+1} \)

Bayesian Forecasting

This is a general approach to forecasting, using the dynamic linear model (which includes MA models). Instead of having to decompose seasonal data, it can be directly represented in a state space model, but care must be taken that the seasonal terms in the equation are additive. The Bayesian interpretation of the Kalman filter is as a way of updating the state vector \( \theta_t \) whenever a new observation becomes available.

It is also possible to study several different models so that you can either compute forecasts based on these different models and consider which is the most reasonable, or refine the number of models in the light of fresh data.

Model Building

The problem in finding a suitable state space model is that initially, we do not know the structure of the time series. We need to know this in order to guess;

(i) \( h_t \) and \( G_t \),

(ii) the variances and covariances of the disturbance terms (i.e. \( \sigma_n^2 \) and \( \sigma_w^2 \))

The only way to do this is by studying graphs of the original data.
A Regression Model with Time-varying Coefficients

Suppose that an observed variable \( X_t \) is known to be linearly independant related to an explanatory variable \( \mu_t \) by;

\[
X_t = a_t + b_t u_t + n_t
\]

where the regression coefficients \( a_t, b_t \) are allowed to evolve through time according to a random walk. Let,

\[
\theta_t^T = (a_t, b_t), h_t^T = (1, u_t)
\]

Now we can write this model in the state space form;

\[
X_t = h_t^T \theta_t + n_t
\]

\[
\theta_t = \theta_{t-1} + w_t
\]

The advantage of this form of equation is that we can now consider a more general class of models, which includes simple regression, and find their properties.
The Kalman Filter

The Kalman filter is a general way of aiding state space modelling by estimating the state vector $\theta_t$, and is useful if the signal being monitored is swapped by noise. It is a two-stage process using equations to update and predict the state vector. Suppose we have observed a time series up to time (t-1), and $\hat{\theta}_{t-1}$ is the best estimator for $\theta_{t-1}$ (i.e. minimum least squares estimator).

Let $P_{t-1}$ be the variance/covariance matrix of $\hat{\theta}_{t-1}$.

(A) The prediction stage;

We forecast $\theta_t$ from time (t-1) and denote the resulting estimator as $\hat{\theta}_{t,t-1}$.

The transition equation is $\hat{\theta}_{t,t-1} = G_t \hat{\theta}_{t-1}$

so an obvious estimators are $\hat{\theta}_{t,t-1} = G_t \hat{\theta}_{t-1}$

Prediction equations

$P_{t,t-1} = G_t P_{t-1} G_t^T + w_t$

where $P_{t,t-1}$ is the standard variance/covariance matrix.

(B) When the new observation $X_t$ at time t occurs, the current best estimator of $\theta_t$ can be corrected.

Prediction error $e_t = X_t - h_t^T \hat{\theta}_{t,t-1}$

We can show that,

$\hat{\theta}_t = \hat{\theta}_{t,t-1} + k_t e_t$

$P_t = P_{t,t-1} - k_t h_t^T P_{t,t-1}$

Updating equations

where $k_t =$ Kalman gain matrix $= \left( h_t^T P_{t,t-1} h_t + \sigma_n^2 \right)^{-1},$

A major advantage of the Kalman filter is that the calculations are recursive, so we don't have to re-evaluate $\hat{\theta}_t$ using every past observation when a piece of new data becomes available, just update $\hat{\theta}_{t-1}$. Also, it converges quickly when there is a constant underlying model, whilst still being flexible enough to follow an evolving situation.
Examples

(i) Consider the steady model where $\theta_t$ is just a single variable, the current level $\mu_t$.
It is possible to show that, as $t \to \infty$, the Kalman filter reduces to the recurrence relation,
$$\hat{\mu}_t = \hat{\mu}_{t-1} + \alpha e_t$$
where the smoothing constant $\alpha$ is a function of $\frac{\sigma_w^2}{\sigma_n^2}$.
This is a simple exponential smoothing

(ii) Consider the linear regression model with varying time coefficients.
We can demonstrate that if $w_t$ is zero, the regression coefficients are constant, and $G_t$ is the identity matrix and invariant through time, so that; $P_{t,t-1} = P_{t-1}$
and the Kalman filter reduces to:-
$$\theta_t = \theta_{t-1} + k_t e_t$$
$$P_t = P_{t-1} - k_t h_t^T P_{t-1}$$
(from the prediction & updating eq.)
and, $e_t = X_t - h_t^T \theta_{t-1}$, $k_t = \frac{P_{t,t-1} h_t}{(h_t^T P_{t,t-1} h_t + \sigma_n^2)}$ as before.
"In order to initialise the Kalman filter, we need values of $\theta_t$ and $P_t$ at the start of the series. This can be done by initial guesswork, relying on the fact that the Kalman filter with rapidly update these estimates."

Another advantage of using a Kalman filter is that it is simple to obtain forecasts
At time $t$, the k-step ahead forecast is given by,
$$\hat{X}_{t,k} = \hat{X}(t, k) = h_{t+k}^T \theta_{t+k} \tag{from prediction equations}$$
Also, if $G_t$ is constant (say G), than,
$$\hat{X}_{t,k} = h_{t+k}^T G^k \hat{\theta}_t$$
[Note: the Kalman filter can only be applied to state space models with linear parameters.]
The Linear Growth Model

If we want to apply the Kalman filter to a linear growth model, we look at the data up to time \( t-1 \) and estimate the trend and mean level, denoting these by \( \hat{\beta}_{t-1} \) and \( \hat{\mu}_{t-1} \). At time \( t-1 \), the forecasts of \( w_{1,t} \) and \( w_{2,t} \) will be zero, so that,

\[
\hat{\mu}_{t,t-1} = \hat{\mu}_{t-1} + \hat{\beta}_{t-1}, \quad \hat{\beta}_{t,t-1} = \hat{\beta}_{t-1}
\]

These are consistent with the prediction equations of the Kalman filter. When \( X_t \) is observed, we can evaluate the error term \( e_t = X_t - \hat{X}_{t,t-1} \).

Hence, using the update equations;

\[
\hat{\theta}_t = \begin{pmatrix} \hat{\mu}_t \\ \hat{\beta}_t \end{pmatrix} = \begin{pmatrix} \hat{\mu}_{t,t-1} \\ \hat{\beta}_{t,t-1} \end{pmatrix} + k_t e_t = \hat{\theta}_{t,t-1} + \begin{pmatrix} c_{1,t} \\ c_{1,t} \end{pmatrix} e_t
\]

where \( c_{1,t} \) and \( c_{2,t} \) are elements of the Kalman gain matrix \( k_t \) (in this case a \( 2 \times 1 \) vector).

\[
\Rightarrow \hat{\mu}_t = \hat{\mu}_{t,t-1} + c_{1,t} e_t = \hat{\mu}_{t-1} + c_{1,t} e_t \\
and \hat{\beta}_t = \hat{\beta}_{t,t-1} + c_{2,t} e_t = \hat{\beta}_{t-1} + c_{2,t} e_t
\]

[Note: An intuitive way to initialise the two state variable from the first two observations is to take \( \hat{\mu}_2 = x_2, \hat{\beta}_2 = x_2 - x_1 \).]

Finally, after all this theory, I am giving a practical example of how it applies to a real situation, and use it to make a real prediction.
§7. An Example of using these Techniques on Wolfer Sunspot Data

First, I will plot the data in order to see if there are any obvious trends or cyclic variations, so that I can try to evaluate what kind of underlying model might produce this scenario.

**Wolfer Sunspot Numbers (Yearly)**

<table>
<thead>
<tr>
<th>Year</th>
<th>No. Sunspots</th>
<th>Year</th>
<th>No. Sunspots</th>
<th>Year</th>
<th>No. Sunspots</th>
<th>Year</th>
<th>No. Sunspots</th>
</tr>
</thead>
<tbody>
<tr>
<td>1770</td>
<td>101</td>
<td>1795</td>
<td>21</td>
<td>1820</td>
<td>16</td>
<td>1845</td>
<td>40</td>
</tr>
<tr>
<td>1771</td>
<td>82</td>
<td>1796</td>
<td>16</td>
<td>1821</td>
<td>7</td>
<td>1846</td>
<td>62</td>
</tr>
<tr>
<td>1772</td>
<td>66</td>
<td>1797</td>
<td>6</td>
<td>1822</td>
<td>4</td>
<td>1847</td>
<td>98</td>
</tr>
<tr>
<td>1773</td>
<td>35</td>
<td>1798</td>
<td>4</td>
<td>1823</td>
<td>2</td>
<td>1848</td>
<td>124</td>
</tr>
<tr>
<td>1774</td>
<td>31</td>
<td>1799</td>
<td>7</td>
<td>1824</td>
<td>8</td>
<td>1849</td>
<td>96</td>
</tr>
<tr>
<td>1775</td>
<td>7</td>
<td>1800</td>
<td>14</td>
<td>1825</td>
<td>17</td>
<td>1850</td>
<td>66</td>
</tr>
<tr>
<td>1776</td>
<td>20</td>
<td>1801</td>
<td>34</td>
<td>1826</td>
<td>36</td>
<td>1851</td>
<td>64</td>
</tr>
<tr>
<td>1777</td>
<td>92</td>
<td>1802</td>
<td>45</td>
<td>1827</td>
<td>50</td>
<td>1852</td>
<td>54</td>
</tr>
<tr>
<td>1778</td>
<td>154</td>
<td>1803</td>
<td>43</td>
<td>1828</td>
<td>62</td>
<td>1853</td>
<td>39</td>
</tr>
<tr>
<td>1779</td>
<td>125</td>
<td>1804</td>
<td>48</td>
<td>1829</td>
<td>67</td>
<td>1854</td>
<td>21</td>
</tr>
<tr>
<td>1780</td>
<td>85</td>
<td>1805</td>
<td>4</td>
<td>1830</td>
<td>71</td>
<td>1855</td>
<td>7</td>
</tr>
<tr>
<td>1781</td>
<td>68</td>
<td>1806</td>
<td>28</td>
<td>1831</td>
<td>48</td>
<td>1856</td>
<td>4</td>
</tr>
<tr>
<td>1782</td>
<td>38</td>
<td>1807</td>
<td>10</td>
<td>1832</td>
<td>28</td>
<td>1857</td>
<td>23</td>
</tr>
<tr>
<td>1783</td>
<td>23</td>
<td>1808</td>
<td>8</td>
<td>1833</td>
<td>8</td>
<td>1858</td>
<td>55</td>
</tr>
<tr>
<td>1784</td>
<td>10</td>
<td>1809</td>
<td>2</td>
<td>1834</td>
<td>13</td>
<td>1859</td>
<td>94</td>
</tr>
<tr>
<td>1785</td>
<td>24</td>
<td>1810</td>
<td>0</td>
<td>1835</td>
<td>57</td>
<td>1860</td>
<td>96</td>
</tr>
<tr>
<td>1786</td>
<td>83</td>
<td>1811</td>
<td>1</td>
<td>1836</td>
<td>122</td>
<td>1861</td>
<td>77</td>
</tr>
<tr>
<td>1787</td>
<td>132</td>
<td>1812</td>
<td>5</td>
<td>1837</td>
<td>138</td>
<td>1862</td>
<td>59</td>
</tr>
<tr>
<td>1788</td>
<td>131</td>
<td>1813</td>
<td>12</td>
<td>1838</td>
<td>103</td>
<td>1863</td>
<td>44</td>
</tr>
<tr>
<td>1789</td>
<td>118</td>
<td>1814</td>
<td>14</td>
<td>1839</td>
<td>86</td>
<td>1864</td>
<td>47</td>
</tr>
<tr>
<td>1790</td>
<td>90</td>
<td>1815</td>
<td>35</td>
<td>1840</td>
<td>63</td>
<td>1865</td>
<td>30</td>
</tr>
<tr>
<td>1791</td>
<td>67</td>
<td>1816</td>
<td>46</td>
<td>1841</td>
<td>37</td>
<td>1866</td>
<td>16</td>
</tr>
<tr>
<td>1792</td>
<td>60</td>
<td>1817</td>
<td>41</td>
<td>1842</td>
<td>24</td>
<td>1867</td>
<td>7</td>
</tr>
<tr>
<td>1793</td>
<td>47</td>
<td>1818</td>
<td>30</td>
<td>1843</td>
<td>11</td>
<td>1868</td>
<td>37</td>
</tr>
<tr>
<td>1794</td>
<td>41</td>
<td>1819</td>
<td>24</td>
<td>1844</td>
<td>15</td>
<td>1869</td>
<td>74</td>
</tr>
</tbody>
</table>

The number of sunspots is given by $X_t = X(t)$, a function of time.
From this graph of the data it is evident that although there is no steady trend term, there is a very pronounced cyclic variation every eleven years or so (even though it also varies in scale over the 100 year period). Since Kalman filters can only be applied to state-space models which are linear in parameters, we must make sure that the model we use has the appropriate properties (i.e. for incorporating the cyclic variations). The Kalman filter could then be used to estimate the coefficients in the model, and also provide a relatively simple way of evaluating forecasts.

[Note: we cannot put much reliance on the simple technique of differencing with this data, since the variable cyclic factor will still effect it. However, I have calculated eleventh differences in an attempt to see whether this will provide a clearer picture.

\[
i.e. \ Y(t) = \nabla_{11} X_t = X_t - X_{t+11}
\]

Eleventh Differencing

At this stage, as a check, I have calculated the difference in the eleventh differences. i.e. \(Z_t = \nabla Y_t = Y_t - Y_{t+1} = X_t + X_{t+12} - (X_{t+1} + X_{t+11})\)

More Differencing
What the eleventh differencing shows is that a lot of the variability in the model can be reduced by this process, but that "significant" peaks still occur and so differencing alone is not sufficient to gain stationary data.

**Method 1**
I have considered using a regression model with time varying coefficients as this has a fairly straightforward state-space model and state variables, as well as being fairly flexible. However, I do not believe that this kind of model could cope with the very large cyclic variations shown by the graph. Therefore, I must examine other approaches.

**Method 2**
I will assume that the basic structural model can accurately approximate the process producing the data:

\[ X_t = \mu_t + i_t + n_t \]  
**Observation equation**

\[ \mu_t = \mu_{t-1} + \beta_{t-1} + w_{1,t} \]
\[ \beta_t = \beta_{t-1} + w_{2,t} \]  
**Transition equations**

where \( i_t = w_{3,t} - \sum_{j=1}^{s-1} i_{t-j} \).

\( \mu_t \) is the local level, \( \beta_t \) is the local trend, and \( i_t \) is the local seasonal index.

The error terms \( n_t, w_{1,t}, w_{2,t}, w_{3,t} \) are taken to be I.I.D. \( \mathcal{N}(0, \sigma^2) \).

In this problem I will regard one "season" as being 100 years. Also, judging by the initial plot, it is reasonable to assume that the expectation of the sum of seasonal effects over 1 season equals zero. i.e. mean over 100 years = 46.55

*Wolfer Sunspot data*

Therefore, from counting the number of cyclic changes on the first graph, \( s=11 \).

Putting the model in the state-space form, the state vector is;

\[ \theta_t = (\mu_t, \beta_t, i_t, i_{t-1}, \ldots, i_{t-10}) \] (twelve parameters)

Hence, the model can be written as;

\[ X_t = h_{11}^T \theta_t + n_t \]  
**Observation equation**
\[ h_t^T = (1, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0) = \text{Constant through time} \]

Also,

\[ \theta_t = G_t \theta_{t-1} + w_t \quad \text{Transition equation} \]

\[
\begin{pmatrix}
\mu_t \\
\beta_t \\
i_t \\
i_{t-1} \\
\vdots \\
i_{t-9}
\end{pmatrix} =
\begin{pmatrix}
\mu_{t-1} \\
\beta_{t-1} \\
i_{t-1} \\
i_{t-2} \\
\vdots \\
i_{t-10}
\end{pmatrix} + w_t
\]

Using the transition equations, this implies;

\[
\begin{pmatrix}
w_{1,t} \\
w_{2,t} \\
w_{3,t} \\
w_t
\end{pmatrix} =
\begin{pmatrix}
1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & -1 & -1 & -1 & -1 & -1 & -1 & -1 & -1 \\
0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1
\end{pmatrix} = G
\]

which is invariant through time.

(Note: \( \det(G) = -1 \).)

Firstly, I will complete one cycle of the equation to indicate what happens, then I will just include salient details from the MINITAB output of this process. For simplicities sake, I will also adopt the following convention.

<table>
<thead>
<tr>
<th>Year</th>
<th>1770</th>
<th>1771</th>
<th>1869</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time t</td>
<td>0</td>
<td>1</td>
<td>etc...</td>
</tr>
</tbody>
</table>
Relying on the fact that the Kalman filter will rapidly update the initial values of $\hat{\theta}_i$ and $P_{i}$, I am arbitrarily choosing:

$$
\begin{pmatrix}
1 \\
1 \\
1 \\
1 \\
1 \\
1 \\
1 \\
1 \\
1
\end{pmatrix} =
\begin{pmatrix}
1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1
\end{pmatrix}$$

and $P_{0} = \begin{pmatrix}
10 & 10 & 10 & 10 & 10 & 10 & 10 & 10 & 10 \\
10 & 10 & 10 & 10 & 10 & 10 & 10 & 10 & 10 \\
10 & 10 & 10 & 10 & 10 & 10 & 10 & 10 & 10 \\
0 & 10 & 10 & 10 & 10 & 10 & 10 & 10 & 10 \\
0 & 0 & 10 & 10 & 10 & 10 & 10 & 10 & 10 \\
0 & 0 & 0 & 10 & 10 & 10 & 10 & 10 & 10 \\
0 & 0 & 0 & 0 & 10 & 10 & 10 & 10 & 10 \\
0 & 0 & 0 & 0 & 0 & 10 & 10 & 10 & 10 \\
0 & 0 & 0 & 0 & 0 & 0 & 10 & 10 & 10 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 10 & 10 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 10
\end{pmatrix}$

a (12 x 1) vector, and a (12 x 12) matrix.

Use the prediction equations;

$$
\begin{aligned}
\hat{\theta}_{i,t-1} &= G_t \hat{\theta}_{t-1} \\
P_{t,t-1} &= G_t P_{t-1} G_t^T + w_t
\end{aligned}
$$

$$
\Rightarrow P_{i,0} = GP_t G_t^T + w_0
$$

where $\hat{\theta}_0$ is the identity matrix (in this case 9 x 9 ).
\[ \hat{\theta}_{t,0} = G\theta_0 = \begin{pmatrix} 2 \\ 1 \\ -10 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{pmatrix} \]

Hence, Error = \( X_1 - h^T \hat{\theta}_{t,0} = 82 - (2 - 10) = 90. \)

Note: This is the point at which the observed values are used by the equations. Now use the updating equations of the Kalman filter;

\[ \hat{\theta}_t = \hat{\theta}_{t,t-1} + k_t e_t \]
\[ P_t = P_{t,t-1} - k_t h^T_t P_{t,t-1} \]

where \( k_t = \frac{P_{t,t-1} h_t}{h^T_t P_{t,t-1} h_t + \sigma_n^2} \), and we assume \( \sigma_n^2 \) is small.

\[ k_i = \frac{1}{12} \begin{pmatrix} 2 \\ 1 \\ 10 \\ -1 \\ -1 \\ -1 \\ -1 \\ -1 \end{pmatrix} \]

and \[ \hat{\theta}_t = \hat{\theta}_{t,0} + k_t e_t = \frac{1}{12} \begin{pmatrix} 1 \\ 1 \\ -1 \\ -1 \\ -1 \\ -1 \end{pmatrix} \]

and \( P_t = P_{t,0} - k_t h^T_t P_{t,0} \)
Therefore, we have come full circle. Continuing the analysis, we eventually gain;

\[
P_t = \begin{pmatrix}
155.762 & 175.044 & 135.637 \\
22.753 & 22.361 & 18.290 \\
-61.762 & -79.044 & -58.637 \\
15.645 & 15.298 & 10.151 \\
15.645 & 15.298 & 10.151 \\
15.645 & 15.298 & 10.151 \\
15.645 & 15.298 & 10.151 \\
15.645 & 15.298 & 10.151 \\
15.645 & 15.298 & 10.151 \\
\end{pmatrix}
\]

\[
\hat{\theta}_{99} = \begin{pmatrix}
-0.409091 \\
0.409091 \\
-0.090909 \\
-0.090909 \\
-0.090909 \\
-0.090909 \\
-0.090909 \\
-0.090909 \\
-0.090909 \\
-0.090909 \\
\end{pmatrix}
\]

\[
\hat{\theta}_{99} = \begin{pmatrix}
-0.409091 \\
0.409091 \\
-0.090909 \\
-0.090909 \\
-0.090909 \\
-0.090909 \\
-0.090909 \\
-0.090909 \\
-0.090909 \\
-0.090909 \\
\end{pmatrix}
\]

\[
\hat{\theta}_{100} = \begin{pmatrix}
-0.409091 \\
0.409091 \\
-0.090909 \\
-0.090909 \\
-0.090909 \\
-0.090909 \\
-0.090909 \\
-0.090909 \\
-0.090909 \\
-0.090909 \\
\end{pmatrix}
\]

\[
\hat{p} = \begin{pmatrix}
-0.409091 \\
0.409091 \\
-0.090909 \\
-0.090909 \\
-0.090909 \\
-0.090909 \\
-0.090909 \\
-0.090909 \\
-0.090909 \\
-0.090909 \\
\end{pmatrix}
\]

and for about the last 50 cycles I gained,

However, the error terms for the model were still significant (i.e. \( |e| > 20 \)) and the state-space vector also varies significantly during all (including the last three) observations, which is evidence against the suitability of using a basic structural model to the data. Despite this, I will use the final estimate \( \hat{\theta}_{100} \) for the state-space vector of the model to calculate predicted values of \( X_t \), using the following equation which is gained directly from the state-space form of the structural model;
The k-step ahead forecast is \( \hat{X}_{t,k} = \hat{X}(t,k) = H^T G \hat{\Theta} \).

Hence, using MINITAB, we gain:

<table>
<thead>
<tr>
<th>Year</th>
<th>Step-Ahead (k)</th>
<th>Predicted No. Sunspots</th>
</tr>
</thead>
<tbody>
<tr>
<td>1870</td>
<td>1</td>
<td>3.931</td>
</tr>
<tr>
<td>1871</td>
<td>2</td>
<td>230.854</td>
</tr>
<tr>
<td>1872</td>
<td>3</td>
<td>40.511</td>
</tr>
<tr>
<td>1873</td>
<td>4</td>
<td>267.434</td>
</tr>
<tr>
<td>1874</td>
<td>5</td>
<td>77.091</td>
</tr>
<tr>
<td>1875</td>
<td>6</td>
<td>304.014</td>
</tr>
<tr>
<td>1876</td>
<td>7</td>
<td>113.671</td>
</tr>
<tr>
<td>1877</td>
<td>8</td>
<td>340.594</td>
</tr>
<tr>
<td>1878</td>
<td>9</td>
<td>150.251</td>
</tr>
<tr>
<td>1879</td>
<td>10</td>
<td>377.174</td>
</tr>
<tr>
<td>1880</td>
<td>11</td>
<td>186.831</td>
</tr>
<tr>
<td>1983</td>
<td>114</td>
<td>2023.274</td>
</tr>
</tbody>
</table>

[Note: During the process to calculate estimates of the state-space vector, problems occurred in the form of the scalar which divides a matrix, to give the Kalman gain matrix, being close to zero. The computer rounded this off to zero, and so I had to replace this with an arbitrarily small number which the computer didn't round off, in order to continue with the procedure.]

Diagnostics

The following information was gained from "Data: A Collection of Problems", page 69, and are a sample of the averages in each year that I have calculated using the Zurich Monthly sunspot numbers 1749 - 1983.

<table>
<thead>
<tr>
<th>Year</th>
<th>No. Sunspots</th>
</tr>
</thead>
<tbody>
<tr>
<td>1869</td>
<td>73.9</td>
</tr>
<tr>
<td>1870</td>
<td>139.1</td>
</tr>
<tr>
<td>1871</td>
<td>111.2</td>
</tr>
<tr>
<td>1872</td>
<td>101.7</td>
</tr>
<tr>
<td>1873</td>
<td>66.3</td>
</tr>
<tr>
<td>1874</td>
<td>44.7</td>
</tr>
<tr>
<td>1875</td>
<td>17.1</td>
</tr>
<tr>
<td>1876</td>
<td>11.3</td>
</tr>
<tr>
<td>1877</td>
<td>12.3</td>
</tr>
<tr>
<td>1878</td>
<td>3.4</td>
</tr>
<tr>
<td>1879</td>
<td>6.0</td>
</tr>
<tr>
<td>1880</td>
<td>32.3</td>
</tr>
<tr>
<td>1983</td>
<td>66.6</td>
</tr>
</tbody>
</table>

I have included the yearly average for 1869 as a check that these numbers do correspond to the Wolfer sunspot data used in the problem.
From these graphs it is obvious that the predicted values in no way correspond to the actual number of sunspots occurring in each year. From this, the only conclusion is that either the basic structural model was inappropriate for this data, or that the periodicity ($s = 11$) was wrongly chosen.

**Method 3**

If I had more time remaining to me I would not simply attempt the same analysis on the data using the basic structural model and the Kalman filter, using a different values of $s$ (i.e. the periodicity), since I believe the data incorporates more than one seasonal effect. By this I mean that as well a peak occurring roughly every eleven years, the size of this peak varies periodically (and there may be many more subtle cyclic components hidden in the data).
§8. Epilogue

In this report I have briefly looked at the nature and mechanisms behind sunspots, but since this is a very complicated area and not covered by the remit of this essay, I have decided to omit any investigation in this area.

It should also be noted that according to Stetson, there are other sources who claim there is evidence for other periodicities in the data other than the obvious 11 year cycle. Namely, these are 37, 68, 77, 83, 126, 252, 300, so the situation may not be as simple as I have assumed when I fitted the basic structural model. Also, to a degree in the early records of sunspot activity, the number of "sunspots" recorded not an objective quantity and actually the Wolfer number is defined as follows. When "a large number of spots group themselves together ... an aggregation of disturbances was of more consequence in judging solar behaviour than the appearance of small isolated spots scattered here and there over the sun's surface. So (we) add to this count of the number of sunspots ten times the number of the groups which can be seen at any one time. This combined number, made up of both spots and groups, (is) designated to be the "sunspot number" for the day". In practice this calculation can be quite complicated in times of peak activity.

Finally, the lack of any computer packages available to me at this time was the reason for my resorting to unwieldy MINITAB operations in my worked example with Kalman filters. However, it has meant that I have had to demonstrate my understanding of the Kalman filter and not attempted to screen myself behind a professional program.
§9. Appendix

The following program was used in MINITAB to calculate values of the parameters in the structural model, for one value of $X_t$ (i.e. sunspot number). To calculate the numbers given in the example, I just ran this program 100 times with the appropriate values of $X_t$ in place.

```
copy c17 m3  \hat{\theta}_0

copy c22-c33 m4  \hat{P}_0

copy c4-c15 m1  G

copy c18 m2  \underline{h}_i = \underline{h}
```

These were to introduce the initial estimates for the parameters.

```
let k4 = X_t
```

This was the stage at which the data was introduced into the process.

```
	mult m1m3 m5
	trans m1 m6
	mult m1m4 m7
	mult m7m6 m7
	trans m2 m8

tmult m8m5 k1

print k1

let k2 = k4 - k1

print k2  \text{(Error)}

tmult m8m7 m9
	mult m9m2 m9

tcopy m9 c35

tcopy c35 k3

tprint c35

tmult m7m2 m10

tcopy m10 c36

tlet c36 = c36 * (1/k3)

tprint c36

tcopy c36 m10

tmult m10k2 m11

tadd m5m11 m12

tprint m12

tmult m10m8 m13

tmult m13m7 m13

tprint m13

tcopy m12 m3

tcopy m13 m4
```

where matrix 12 is the update for $\hat{\theta}_t$ and matrix 13 is the update for $\hat{P}_t$. 

38
§10. References

"Data: A Collection of Problems" D. F. Andrews & A. M. Herzberg

"Fourier Analysis of Time Series" (1976) P. Bloomfield

"Time Series Analysis: Forecasting and Control" G. Box & G. M. Jenkins

"The Analysis of Time Series" (Fourth Edition) C. Chatfield

"Probability and Statistics" (Second Edition) H. DeGroot

"Forecasting, structural time series models, and the Kalman filter" Andrew C. Harvey

"Sunspots and their effects" H. T. Stetson