Chapter g13 – Time Series Analysis

1. **Scope of the Chapter**

This chapter provides routines for investigating and modelling the statistical structure of series of observations collected at equally spaced points in time. The models may also be used to forecast values for the series. The routines for model fitting and forecasting cover both univariate time series and the analysis of a single output series depending on a number of independent input series. This chapter also contains routines for the construction of Kalman filters for state space models. The chapter only includes time domain methods; no frequency domain or spectral methods are included.

2. **Background**

A time series is a set of measurements of a variable recorded over that time. We assume that measurements are obtained at equally spaced points in time to provide a sample \( x_1, x_2, \ldots, x_n \) of length \( n \).

Interest usually focuses on describing or modelling the structure of the series. The aim of the analysis may be to forecast or extrapolate the sample into the future. Probability limits or estimated errors should also be computed to reflect the uncertainty associated with the forecasts. Other areas of application include decomposition of the series into various components, simulation of series based on the model, and control of a series.

The first step in the analysis of time series is to graph the series. This may reveal the structure of the series in terms of various components or features, which are evident to the eye. The graph may show:

(a) trends – linear or higher order polynomial.

(b) seasonal patterns associated with fixed integer seasonal periods. The presence of seasonality and the period is usually known a priori. The pattern may be fixed or slowly varying from season to season.

(c) cycles, or waves of stable amplitude and period.

(d) quasi-cycles, i.e., waves of fluctuating period and amplitude.

(e) noise – irregular statistical fluctuations and swings about the overall mean or trend, due to observational error or sampling variability.

Trends, seasonal patterns, and cycles might be regarded as deterministic components following fixed mathematical equations, and the quasi-cycles and other statistical fluctuations as stochastic and describable by short term correlation structure. For finite data sets it is not always easy to distinguish between deterministic and stochastic components. The class of autoregressive integrated moving-average (ARIMA) models is widely used to describe the structure; these models are in the form of difference equations (or recurrence relations) relating past and present values of the series. For a thorough account of these models, see Box and Jenkins (1976).

Another, more general, class of models are the state space models. The models can be fitted using the Kalman filter.

Both ARIMA and state space models are part of the time domain analysis of time series in which the state of the series at time \( t \) is related to the state at time \( t - 1, \ t - 2, \ldots \). An alternative approach is frequency domain or spectral analysis which interpret the time series in terms of component sine waves of various frequencies.

Model building first involves selecting a small number of models on the basis of inspecting the data and various statistical summaries. The parameters of the models are then estimated and the best model selected. This model is tested and checked in various ways, to see if there is any structure
in the data that has not been adequately accounted for. If necessary the model is revised and re-estimated before being used, say for example, in forecasting.

2.1. Univariate time series

If the analysis of a single series is done without reference to other variables it is called univariate analysis.

2.1.1. ARIMA Models

The correlation structure in stationary time series may often be represented by a model with a small number of parameters belonging to the autoregressive moving average (ARMA) class. Given knowledge of the characteristic acf (autocorrelation function) and pacf (partial autocorrelation function) patterns of ARMA models, these sample values may be used to select ARMA models for the series. If the stationary series \( w_t \) has been derived by differencing the original series \( x_t \), then \( x_t \) is said to follow an ARIMA model.

Taking \( w_t = \nabla^d x_t \) the (non-seasonal) ARIMA \((p, d, q)\) model with \( p \) autoregressive parameters \( \phi_1, \phi_2, \ldots, \phi_p \) and \( q \) moving average parameters \( \theta_1, \theta_2, \ldots, \theta_q \), represents the structure of \( w_t \) by the equation

\[
w_t = \phi_1 w_{t-1} + \cdots + \phi_p w_{t-p} + a_t - \theta_1 a_{t-1} - \cdots - \theta_q a_{t-q}
\]

(1)

where \( a_t \) is an uncorrelated series (often called white noise) with mean 0 and constant variance \( \sigma^2 \).

If \( w_t \) has a non-zero mean \( c \), then this is allowed for by replacing \( w_t \) by \( w_t - c \), \( w_{t-1} - c \), etc. in the model. Although \( c \) is often estimated by the sample mean of \( w_t \), this is not always optimal.

A series generated by this model will only be stationary provided restrictions are placed on \( \phi_1, \phi_2, \ldots, \phi_p \) to avoid unstable growth of \( w_t \). These are called stationarity constraints. The series \( a_t \) may be regenerated by rewriting the model equation as

\[
a_t = w_t - \phi_1 w_{t-1} - \cdots - \phi_p w_{t-p} + \theta_1 a_{t-1} + \cdots + \theta_q a_{t-q}
\]

(2)

provided also that the parameters \( \theta_1, \theta_2, \ldots, \theta_q \) satisfy the invertibility constraints. The series \( a_t \) may also be interpreted as the linear innovations in \( x_t \) (and in \( w_t \)); the innovations being the errors if \( x_t \) were to be predicted using the information in all past values \( x_{t-1}, x_{t-2}, \ldots \).

For a series with short term correlation only, i.e., \( r_k \) is not significant beyond some low lag \( q \) (see Box and Jenkins for the statistical test), then the pure moving average model MA\((q)\) is appropriate, with no autoregressive parameters, i.e., \( p = 0 \).

Autoregressive parameters are appropriate when the acf pattern decays geometrically, or with a damped sinusoidal pattern which is associated with quasi-periodic behaviour in the series. If the sample pacf \( \phi_{k,p} \) is significant only up to some low lag \( p \), then a pure autoregressive model AR\((p)\), is appropriate with \( q = 0 \). Otherwise moving average terms will need to be introduced, as well as autoregressive terms.

The seasonal ARIMA \((p, d, q, P, D, Q, s)\) model allows for correlation at lags which are multiples of the seasonal period \( s \). Taking \( w_t = \nabla^d \nabla^D w_t \), the series is represented in a two-stage manner via an intermediate series \( \epsilon_t \)

\[
w_t = \Phi_1 w_{t-s} + \cdots + \Phi_p w_{t-s \times p} + \epsilon_t - \Theta_1 \epsilon_{t-s} - \cdots - \Theta_Q \epsilon_{t-s \times Q}
\]

(3)

\[
\epsilon_t = \phi_1 \epsilon_{t-1} + \cdots + \phi_p \epsilon_{t-p} + a_t - \theta_1 a_{t-1} - \cdots - \theta_q a_{t-q}
\]

(4)

where \( \Phi_1, \ldots, \Phi_p \) and \( \Theta_1, \ldots, \Theta_Q \) are the seasonal parameters, and \( P \) and \( Q \) are the corresponding orders. Again, \( w_t \) may be replaced by \( w_t - c \).

2.1.2. Model identification
As already mentioned it is usually advisable to examine a graph of the series as a first step. If the variance is not constant across the range of observations it may be useful to apply a variance stabilising transformation to the series.

Differencing operations may be used to simplify the structure of a time series. For example, first order differencing, that is forming the series \( \nabla x_t = x_t - x_{t-1} \) will remove a linear trend and first order seasonal differencing \( \nabla_s x_t = x_t - x_{t-s} \) eliminates a fixed seasonal pattern.

These operations reflect the fact that it is often appropriate to model a time series in terms of changes from one value to another. For example, if a series has something of a nature of a random walk, for which successive steps from one value to another are random, then differencing may be appropriate since the random walk is by definition the accumulation of independent changes.

Differencing may be applied repeatedly to a series giving \( w_t = \nabla^d \nabla^D x_t \) where \( d \) and \( D \) are the orders of differencing. The derived series \( w_t \) will be shorter, of length \( N = n - d - s \times D \) and extend for \( t = 1 + d + s \times D, \ldots, n \).

The selection of the appropriate order of differencing is largely based on the visual inspection of the series and its differences, and of the sample autocorrelations and partial autocorrelations of the series.

The term stationarity describes series which appear to have reached statistical equilibrium about a constant mean level, and for which the other statistical properties such as variance and correlation do not vary with time.

Sample autocorrelations and partial autocorrelations are useful aids in the process of model identification. Given that a series is stationary (possibly after a transformation and/or differencing), the correlations, \( \rho_k \), between terms \( x_t \) and \( x_{t+k} \) separated by lag \( k \) give an adequate description of the statistical structure and are estimated by the sample autocorrelations \( r_k \). These may be computed using the routine nag_tsa_auto_corr (g13abc).

The information in the autocorrelations may also be presented by deriving the coefficients of the partial autocorrelations which measures the correlations between \( x_t \) and \( x_{t+k} \) conditional upon the intermediate values \( x_{t+1}, x_{t+2}, \ldots, x_{t+k-1} \). The sample partial autocorrelations may be computed by nag_tsa_auto_corr_part (g13acc).

### 2.1.3. ARIMA model estimation

The autocorrelations properties of the model define the likelihood for any finite data set; thus in theory the parameters of an ARIMA model are determined by a sufficient number of autocorrelations \( \rho_1, \rho_2, \ldots \). Using the sample values \( r_1, r_2, \ldots \) in their place it is usually (but not always) possible to solve for the corresponding ARIMA parameters. These are rapidly computed but are not fully efficient estimates, particularly if moving average parameters are present. They do provide useful preliminary values for an efficient but relatively slow iterative method of estimation. Estimation of the model essentially means searching for parameter values which minimise an estimation criterion. For the purpose of defining an estimation criterion it is assumed that the series \( a_t \) is a sequence of independent Normal variates having mean 0 and variance \( \sigma_a^2 \).

The innovations are regenerated from the data for any prescribed parameter values of the model. The early values have little or no previous data from which the predictions, and thereby the errors \( a_t \), can be calculated. This lack of knowledge about the terms on the right hand side of equation (2) for \( t = 1, 2, \ldots, \max(p, q) \) is overcome using some form of backforecasting.

In all the estimation criteria the main term is a sum of squares function

\[
S = \sum a_t^2
\]

which is a direct measure of the accuracy of prediction of the model when applied to the given data, since the innovations \( a_t \) are the one step ahead prediction errors.
The ideal criterion is the exact likelihood of the data assuming independent Normal \((0, \sigma_a^2)\) innovations. If the differenced data \(w_1, w_2, \ldots, w_N = w\) is viewed as a single sample from a multivariate Normal density whose covariance matrix \(\sigma_a^2 V\) is a function of the ARIMA model parameters, then the exact log-likelihood of the parameters is

\[-\frac{n}{2} \log(2\pi \sigma_a^2) - \frac{1}{2} \log |V| - \frac{1}{2\sigma_a^2} w'V^{-1}w.\]

This exact likelihood criterion considers the set of past values \(a_0, a_{-1}, a_{-2}, \ldots\) as unobserved random variables with known distribution. The calculation of the likelihood requires theoretical integration over the range of this past set. This yields a criteria of the form \(D = |V|^{-1/2} S/N\). Use of this quantity as an objective function is preferable to the use of \(S\) alone, on the grounds that it avoids appreciable bias in the ARIMA model parameter estimates and yields a better conditioned estimation problem. However there is a moderate computational penalty in calculating \(D\).

The least-squares criterion is equivalent to using the quadratic form

\[w'V^{-1}w = S\]

as an objective function to be minimized. The algorithm used for the least squares criterion is equivalent to taking the infinite set of past values \(a_0, a_{-1}, a_{-2}, \ldots\) as nuisance parameters. Neglecting the term \(-\frac{1}{2} \log |V|\) when using the least squares criteria \(S\) yields estimates which differ very little from the exact likelihood except in small samples, or in seasonal models with a small number of whole seasons contained in the data. In these cases bias in moving average parameters may cause them to stick at the boundary of their constraint region, resulting in failure of the estimation method.

The third criterion available is marginal likelihood. This is only distinct from exact likelihood in models with simple input series, see Section 2.2, or in univariate models if a constant term is fitted, because this is treated as a simple input. The constant term is effectively treated as a random variable and the marginal likelihood is the exact likelihood for those components of the data which give no information about \(c\). This may be expressed as the likelihood for \(x_t - 2\) for \(t = 1, 2, \ldots, n\) or for \(\nabla x_t\) for \(t = 1, 2, \ldots, n\). The marginal likelihood approach is useful for further reducing the bias in the ARMA model parameters, particularly when modelling a large number of simple inputs.

Approximate standard errors of the parameter estimates and the correlations between them are available after estimation.

### 2.1.4. Model checking

The model residuals \(\hat{a}_t\) are the innovations resulting from the estimation, and are the main focus of model checking. A good model should have no significant residual autocorrelations; it is recommended that both the residual autocorrelations and partial autocorrelations are calculated and graphed. The portmanteau statistic may be used to get an overall assessment of the size of the acf. It is also useful to see how well the model forecasts. This may be done by leaving out a few data values from the end of the series and checking how well the model forecasts these values. There are at present no routines available for residual checking.

### 2.1.5. Forecasting

An ARIMA model is particularly suited to extrapolation of a time series. The model equations are simply used for \(t = n + 1, n + 2, \ldots\) replacing the unknown future values of \(a_t\) by zero. This produces future values of \(w_t\), and if differencing has been used, this process is reversed (the so-called integration part of ARIMA models) to construct future values of \(x_t\). Forecast error limits are easily deduced. These forecasts may be computed using the routine \texttt{nag_jsa_multiinput_forecast}\(\texttt{(g13bjc)}\).

### 2.2. Transfer function modelling
We now consider multivariate time series in which a single output (dependent) series \( y_t \) is believed to depend on a number of input (explanatory) series \( x_t \). This dependency may follow a simple linear regression, e.g.,

\[
y_t = \alpha x_t + \eta_t
\]

or more generally may involve lagged values of the input

\[
y_t = \alpha_0 x_t + \alpha_1 x_{t-1} + \alpha_2 x_{t-2} + \ldots + \eta_t.
\]

The sequence \( \eta_0, \eta_1, \eta_2, \ldots \) is called the impulse response function (IRF) of the relationship. The term \( \eta_t \) represents that part of \( y_t \) which cannot be explained by the input, and it is assumed to follow a univariate ARIMA model. We call \( \eta_t \) the (output) noise component of \( y_t \), and it includes any constant term in the relationship. It is assumed that the input series \( x_t \) and the noise component \( \eta_t \) are independent.

The part of \( y_t \) which is explained by the input is called the input component \( z_t \):

\[
z_t = \alpha_0 x_t + \alpha_1 x_{t-1} + \alpha_2 x_{t-2} + \ldots
\]

so that \( y_t = z_t + \eta_t \).

The eventual aim is to model both these components of \( y_t \) on the basis of observations of \( y_1, y_2, \ldots, y_t \) and \( x_1, x_2, \ldots, x_p \). In applications to forecasting or control, both the input and noise components are important. In general there may be more than one input series, e.g., \( x_{1t} \) and \( x_{2t} \), which are assumed to be independent and corresponding components \( z_{1t} \) and \( z_{2t} \), so that

\[
y_t = z_{1t} + z_{2t} + \eta_t.
\]

In a similar manner to that in which the structure of a univariate series may be represented by a finite parameter ARIMA model, the structure of an input component may be represented by a transfer function (TF) model with delay time \( b \), \( p \) autoregressive-like parameters \( \delta_1, \delta_2, \ldots, \delta_p \) and \( q + 1 \) moving average-like parameters \( \omega_0, \omega_1, \ldots, \omega_q \):

\[
z_t = \delta_1 z_{t-1} + \delta_2 z_{t-2} + \ldots + \delta_p z_{t-p} + \omega_0 x_{t-b} - \omega_1 x_{t-b-1} - \ldots - \omega_q x_{t-b-q}.
\]

If \( p > 0 \) this represents an IRF which is infinite in extent and decays with geometric and/or sinusoidal behaviour. The parameters \( \delta_1, \delta_2, \ldots, \delta_p \) are constrained to satisfy a stability condition identical to the stationarity condition of autoregressive models. There is no constraint on \( \omega_0, \omega_1, \ldots, \omega_q \).

### 2.2.1. Model estimation

Given that the orders of all the transfer function models and the ARIMA model of a multi-input model have been specified, the various parameters in those models may be (simultaneously) estimated.

The innovations are derived for any proposed set of parameter values, by calculating the response of each input to the transfer functions then evaluating the noise \( \eta_t \) as the difference between this response (combined for all the inputs) and the output. The innovations are derived from the noise using the ARIMA model in the same manner as for a univariate series.

As in univariate modelling, in estimating the parameters consideration has to be given to the lagged terms in the various model equations which are associated with times prior to the observation period, and are therefore unknown. The subroutine descriptions provide the necessary detail as to how this problem is treated.

As described in Section 2.1.3 three estimation criterion are available: least squares, exact likelihood and marginal likelihood which is similar to exact likelihood but can counteract bias in the ARIMA model due to the fitting of a large number of simple inputs.

Approximate standard errors of the parameter estimates and the correlations between them are available after estimation.
2.2.2. Forecasting

A multi-input model may be used to forecast the output series using the routine nag__multin__model__forecast (g13bec), provided future values (possibly forecasts) of the input series are supplied. The forecasts are computed from the model equations replacing the unknown future values of \( a_t \) by zero.

2.3. Kalman Filters and State Space Models

Kalman filtering can be used for estimating or filtering a multi-dimensional stochastic process \( X_i \) on which linear observations \( Y_i \) are made. The series \( Y_i \) can be considered as a multivariate time series. The technique assumes that the processes \( X_i \) and \( Y_i \) obey the linear system

\[
X_{i+1} = A_i X_i + B_i W_i + D_i U_i
\]

and the linear observation process

\[
Y_i = C_i X_i + V_i
\]

where \( X_i \) is the state vector to be estimated, \( Y_i \) is the measurement vector, \( U_i \) is the deterministic input vector, \( W_i \) is the process noise, \( V_i \) is the measurement noise, \( A_i \) is the system state transition matrix, \( B_i \) is the system input weight matrix, \( C_i \) is the system output weight matrix and \( D_i \) is the control matrix (where the subscript \( i \) refers to the value of the appropriate quantity at time \( i \)).

The vectors \( X_i, Y_i, U_i \) and \( W_i \) are of dimension \( n, p, r \) and \( m \), respectively. The matrices \( A_i, B_i, C_i \) and \( D_i \) are of dimension \( n \times n \), \( n \times p \) by \( m \) by \( n \times m \) and \( m \times r \), respectively.

Here the process noise and the measurement noise sequences are assumed to be uncorrelated and have zero mean. This implies that:

\[
E\{W_i\} = 0 \quad E\{V_i\} = 0 \quad \text{and} \quad E\{W_i V_i^T\} = 0,
\]

and the covariance matrices are

\[
E\{W_i W_i^T\} = Q_i \quad E\{V_i V_i^T\} = R_i
\]

where the operator \( E \) denotes the “expectation value”, and \( Q_i \) and \( R_i \) are positive definite matrices.

At instant \( i \), \( Q_i \) is the process noise covariance matrix whilst \( R_i \) is the measurement noise covariance matrix. [Note that in the case of the information filter the first condition is relaxed, i.e., the mean of the process noise may be non-zero.]

If the system matrices \( A_i, B_i, C_i, D_i \) and also the covariance matrices \( Q_i, R_i \) are known then Kalman filtering can be used to compute the minimum variance estimate of the stochastic variable \( X_i \) estimated from the observed values \( Y_i \) to \( Y_j \)

\[
\hat{X}_{i|i} = \hat{X}_{i|Y_1...Y_j}
\]

When \( j = i \) the above estimate is called the filtered estimate, and when \( j = i - 1 \) it is known as the one-step predicted estimate, or simply the predicted estimate.

Kalman filtering uses a recursive method which involves computing the state covariance matrices \( P_{i|i} \) and/or \( P_{i|i-1} \) and the estimates \( \hat{X}_{i|i} \) and/or \( \hat{X}_{i|i-1} \) from their previous values for \( i = 1, 2, \ldots \)

If the covariance of the initial state \( X_0 \) (represented by \( P_{0|0} \)) is known, and the mean of the initial state \( X \) (represented by \( X_{0|0} \)) is given then the following recurrence relations provide the required estimates.
\[ H_i = R_i + C_i P_{i|i-1} C_i^T \]  
(9)
\[ K_i = P_{i|i-1} C_i^T H_i^{-1} \]  
(10)
\[ P_{i|i} = [I - K_i C_i] P_{i|i-1} \]  
(11)
\[ \hat{X}_{i|i} = \hat{X}_{i|i-1} + K_i \nu_i \]  
(12)

where the one step ahead prediction error is given by \( \nu_i = Y_i - C_i \hat{X}_{i|i-1} \)

\[ P_{i+1|i} = A_i P_{i|i} A_i^T + B_i Q_i B_i^T \]  
(13)
\[ \hat{X}_{i+1|i} = A_i \hat{X}_{i|i} + D_i U_i \]  
(14)

where \( K_i \) is referred to as the Kalman gain matrix and \( H_i \) contains the covariance matrix of the prediction errors \( \nu_i \). It can be seen that equations (9), (11) and (13) define the recursion involving the covariance matrices \( P_{i|i-1} \), \( P_{i|i} \) and \( P_{i+1|i} \). These matrices are positive semidefinite and can therefore be factorised into their Cholesky (“square root”) factors. Equations (11) and (12) yielding \( \hat{X}_{i|i} \) and \( P_{i|i} \) from \( \hat{X}_{i|i-1} \) and \( P_{i|i-1} \) are termed measurement-update equations, while equations (13) and (14) yielding \( \hat{X}_{i+1|i} \) and \( P_{i+1|i} \) from \( \hat{X}_{i|i} \) and \( P_{i|i} \) are known as time-update equations.

2.3.1. The information filter

An alternative set of Kalman filter equations can be constructed which use the inverse of the covariance matrices. These matrices (e.g. \( P_{i+1|i}^{-1} \)) are also positive semidefinite and are termed information matrices.

Although the information filter has the disadvantage that it requires the inverses \( A_i^{-1} \) and \( R_i^{-1} \) to be computed, it is preferable to the covariance filter in situations where there is no (very little) information concerning the initial state of the system. In these circumstances the covariance filter will fail because the initial state covariance matrix \( P_{0|0} \) is infinite (very large), whilst the corresponding information filter initial state \( P_{0|0}^{-1} = 0 \) (very small) incurs no such difficulties.

The information filter recursion (with \( D_i = 0 \)) is described by the following equations

\[ P_{i+1|i}^{-1} = [I - N_i B_i^T] M_i \]  
(15)
\[ P_{i+1|i+1}^{-1} = P_{i+1|i}^{-1} + C_i^T R_i^{-1} C_i \]  
(16)
where \( M_i = (A_i^{-1})^T P_{i|i}^{-1} A_i^{-1} \)
and \( N_i = M_i B_i (Q_i^{-1} + B_i^T M_i B_i)^{-1} \)
\[ \hat{a}_{i+1|i} = [I - N_i B_i^T] (A_i^{-1})^T \hat{a}_{i|i} \]  
(17)
\[ \hat{a}_{i+1|i+1} = \hat{a}_{i+1|i} + C_i^T R_i^{-1} Y_{i+1} \]  
(18)
where \( \hat{a}_{i+1|i} = P_{i+1|i}^{-1} \hat{X}_{i+1|i} \)
and \( \hat{a}_{i+1|i+1} = P_{i+1|i+1}^{-1} \hat{X}_{i+1|i+1} \)  
(19)

2.3.2. Square root filters

The use of the Kalman filter equations previously given can result in covariance matrices which are not positive semidefinite. This may happen if some of the measurements are very accurate and numerical computation then involves ill-conditioned quantities. Square root filtering is a technique which overcomes this difficulty by propagating the covariance matrices in Cholesky (square root) form. This has the advantage that, despite computational errors, the product of the Cholesky factors will always yield a positive definite covariance matrix. The numerical conditioning of the Cholesky square root is also generally much better than that of its corresponding covariance matrix. Since the condition number of the Cholesky factor is the square root of the condition number of
the covariance matrix, square root filtering can avoid numerical difficulties with only half as many significant digits as the full matrix Kalman filters outlined above.

2.3.3. The square root covariance filter

The time-varying square root covariance Kalman filter (nag_kalman_sqrt_filt_cov_var (g13eac)) provided by this chapter requires the construction of the following block matrix pre-array and block matrix post-array.

$$
\begin{pmatrix}
R_{i}^{1/2} & C_{i}S_{i} & 0 \\
0 & A_{i}S_{i} & B_{i}Q_{i}^{1/2}
\end{pmatrix}
U_{1} =
\begin{pmatrix}
H_{i}^{1/2} & 0 & 0 \\
G_{i} & S_{i+1} & 0
\end{pmatrix}
$$

(21)

where $U_{1}$ is an orthogonal matrix that triangularizes the pre-array and the matrices $P_{i|i-1}$, $Q_{i}$, $H_{i}$ and $R_{i}$ have been Cholesky factorised as follows:

$$
P_{i|i-1} = S_{i}S_{i}^{T}
$$

$$
Q_{i} = Q_{i}^{1/2} \left( Q_{i}^{1/2} \right)^{T}
$$

$$
R_{i} = R_{i}^{1/2} \left( R_{i}^{1/2} \right)^{T}
$$

$$
H_{i} = H_{i}^{1/2} \left( H_{i}^{1/2} \right)^{T}
$$

where the left factors are lower triangular, and the Kalman filter gain matrix $K_{i}$ is related to $G_{i}$ by

$$
A_{i}K_{i} = G_{i} \left( H_{i}^{1/2} \right)^{-1}
$$

The new state estimate can then be calculated using

$$
\hat{X}_{i+1|i} = A_{i}\hat{X}_{i|i-1} + A_{i}K_{i}(Y_{i} - C_{i}\hat{X}_{i|i-1}) + D_{i}U_{i}
$$

(22)

where the product of the matrices $A_{i}$ and $K_{i}$ is represented as $AK_{i}$.

That this method is computationally equivalent to equations (9)-(14) can be demonstrated by “squaring” each side of equation (21) (post-multiplying each side by its transpose) and then equating block matrix elements on either side. It can similarly be shown that transposition of columns 2 and 3 of the pre-array, as occurs in routine nag_kalman_sqrt_filt_cov_invar (g13ebc), does not affect the elements in the resultant post-array.

2.3.4. The square root information filter

The time-varying square root information Kalman filter (nag_kalman_sqrt_filt_info_var (g13ecc)) provided by this chapter requires the construction of the following block matrix pre-array and block matrix post-array.

$$
\begin{pmatrix}
Q_{i}^{-1/2} & 0 & 0 \\
S_{i}^{-1}A_{i}^{-1}B_{i} & S_{i}^{-1}A_{i}^{-1} & S_{i}^{-1}X_{i|i-1} \\
0 & R_{i+1}^{-1/2}C_{i+1} & R_{i+1}^{-1/2}Y_{i+1}
\end{pmatrix}
U_{2} =
\begin{pmatrix}
F_{i+1}^{-1/2} & * & * \\
0 & S_{i+1}^{-1} & \xi_{i+1|i+1} \\
0 & 0 & E_{i+1}
\end{pmatrix}
$$

(23)

3.intro-g13.8
where the asterisk represents elements that are not required, $U_3$ is an orthogonal transformation
triangularizing the pre-array and $F_{i+1}$, the matrix containing the innovations in the process noise, is given by

$$F_{i+1}^{-1} = Q_i^{-1} + R_i^T M_i B_i$$

The matrices $P_{i|i}^{-1}$, $Q_i^{-1}$, $F_{i+1}^{-1}$ and $R_i^{-1}$ have been Cholesky factorised as follows:

$$P_{i|i}^{-1} = \left(S_i^{-1}\right)^T S_i^{-1}$$
$$Q_i^{-1} = \left(Q_i^{-1/2}\right)^T Q_i^{-1/2}$$
$$R_i^{-1} = \left(R_i^{-1/2}\right)^T R_i^{-1/2}$$
$$F_{i+1}^{-1} = \left(F_{i+1}^{-1/2}\right)^T F_{i+1}^{-1/2}$$

where the right factors are upper triangular.

The new state estimate is computed via

$$X_{i+1|i+1} = S_{i+1} \xi_{i+1|i+1}$$

That this method is computationally equivalent to equations (15)-(20) can be demonstrated by transposing (23), “squaring” the right hand side to eliminate the orthogonal matrix $U_3$ and then, after performing a block Cholesky decomposition, equating block matrix elements on either side. It can similarly be shown that transposition of rows 2 and 3 of the pre-array, as occurs in routine nagKalmanSqrtFilterCovInvar (g13edc), does not affect the elements in the resultant post-array.

2.3.5. Time invariant condensed square root filters

When the system matrices $A$, $B$, $C$ are time invariant, it can be advantageous to perform initial unitary transformations to “condense” them (create as many zeros as possible) and thereby significantly reduce the number of floating-point operations required by the algorithm. Essentially this entails creating an appropriate unitary transformation matrix $U$ and solving for the new state vector $X_t = UX$ in the transformed reference frame. After the required number of Kalman filter iterations have been performed the back transformation $X = U^T X_t$ provides the estimated state vector in the original reference frame. It can be shown from equations (9)-(14) that the transformed system matrices for the covariance filter are given by \{ $UA^T$, $UB$, $CUT^T$ \}, which are in agreement with the arguments required by routine nagKalmanSqrtFilterCovInvar (g13edc). It can similarly be shown, from equations (15)-(20), that the system matrices describing the corresponding transformed information filter are \{ $UA^{-1}U^T$, $UB$, $CUT^T$ \}. These correspond to the arguments used by the routine nagKalmanSqrtFilterCovInvar (g13edc) \{ $UA^{-1}U^T$, $UA^{-1}B$, $CUT^T$ \}, where the second matrix is input as the product of $UA^{-1}U^T$ and $UB$. It should be noted that in the transformed frame the covariance matrix $P_{i|i}$ is related to the original covariance matrix via the similarity transformation $P_{i|i} = UP_{i|i}U^T \left( (P_{i|i})^{-1} = U (P_{i|i})^{-1} U^T \right)$. This means that, for square root Kalman filter routines, the appropriate Cholesky factor of $P_{i|i}^T$ must be input.

The condensed matrix forms used by the routines in this chapter are (nagKalmanSqrtFilterCovInvar (g13edc)) the lower observer Hessenberg form where the compound matrix

$$\begin{pmatrix} UAU^T \\ CUT \end{pmatrix}$$

is lower trapezoidal and (nagKalmanSqrtFilterInfoInvar (g13edc)) the upper controller Hessenberg form where the compound matrix $(UBU^T)$ is upper trapezoidal.

Both nagKalmanSqrtFilterCovInvar (g13edc) and nagKalmanSqrtFilterInfoInvar (g13edc) contain the block matrix

$$\begin{pmatrix} U & CUT \\ UB & UAU^T \end{pmatrix}$$
within their pre-array, and the structure of this matrix (for \( n = 6, \ m = 3 \) and \( p = 2 \)) is illustrated below for both Hessenberg forms:

**Lower observer Hessenberg**

\[
\begin{pmatrix}
  x & 0 & 0 & 0 & 0 & 0 \\
  x & x & 0 & 0 & 0 & 0 \\
  x & x & x & x & 0 & 0 \\
  x & x & x & x & x & 0 \\
  x & x & x & x & x & x \\
  x & x & x & x & x & x
\end{pmatrix}
\]

**Upper controller Hessenberg**

\[
\begin{pmatrix}
  x & x & x & x & x & x \\
  x & x & x & x & x & x \\
  x & x & x & x & x & x \\
  x & x & x & x & x & x \\
  0 & x & x & x & x & x \\
  0 & 0 & x & x & x & x \\
  0 & 0 & 0 & x & x & x \\
  0 & 0 & 0 & 0 & x & x \\
  0 & 0 & 0 & 0 & 0 & x
\end{pmatrix}
\]

2.3.6. ARMA models and the Kalman filter

Wei (1990) illustrates how an ARMA model can be written as a state space model and Harvey and Phillips (1979) show how the Kalman filter can be used to construct the likelihood for an ARMA model. Because of the updating nature of Kalman filters and the presence of a prior covariance matrix Kalman filters are suitable for the Bayesian approach to statistical inference.

The logarithm of the likelihood for observations at time \( i = 1, 2, \ldots, t \) is given by

\[
l(\theta) = \kappa - \frac{1}{2} \sum_{i=1}^{t} \ln(\det(H_i)) - \frac{1}{2} \sum_{i=1}^{t} (Y_i - C_iX_{i|i-1})^T H_i^{-1}(Y_i - C_iX_{i|i-1})
\]

where \( \kappa \) is a constant and \( H_i \) (and possibly \( C_i \)) contain unknown parameters \( \theta \). The maximum likelihood estimates of \( \theta \) can then be found by maximizing \( l(\theta) \) using a suitable optimization routine.

2.4. Spectral Analysis

In describing a time series using spectral analysis the fundamental components are taken to be sinusoidal waves of the form \( R \cos(\omega t + \phi) \), which for a given angular frequency \( \omega \), \( 0 \leq \omega \leq \pi \), is specified by its amplitude \( R > 0 \) and phase \( \phi \), \( 0 \leq \phi \leq 2\pi \). Thus in a time series of \( n \) observations it is not possible to distinguish more than \( n/2 \) independent sinusoidal components. The frequency range \( 0 \leq \omega \leq \pi \) is limited to a shortest wavelength of two sampling units because any wave of higher frequency is indistinguishable upon sampling (or is aliased with) a wave within this range.
Spectral analysis follows the idea that for a series made up of a finite number of sine waves the amplitude of any component at frequency $\omega$ is given to order $1/n$ by

$$R^2 = \left(\frac{1}{n^2}\right) \left| \sum_{t=1}^{n} x_t e^{i\omega t} \right|^2.$$

2.4.1. The sample spectrum

For a series $x_1, x_2, \ldots, x_n$ this is defined as

$$f^*(\omega) = \left(\frac{1}{2n\pi}\right) \left| \sum_{t=1}^{n} x_t e^{i\omega t} \right|^2,$$

the scaling factor now being chosen in order that

$$2 \int_{0}^{\pi} f^*(\omega) d\omega = \sigma_x^2,$$

i.e., the spectrum indicates how the sample variance ($\sigma_x^2$) of the series is distributed over components in the frequency range $0 \leq \omega \leq \pi$.

It may be demonstrated that $f^*(\omega)$ is equivalently defined in terms of the sample autocorrelation function (acf) $r_k$ of the series as

$$f^*(\omega) = \left(\frac{1}{2\pi}\right) \left( c_0 + 2 \sum_{k=1}^{n-1} c_k \cos k\omega \right)$$

where $c_k = \sigma_x^2 r_k$ are the sample autocovariance coefficients.

If the series $x_t$ does contain a deterministic sinusoidal component of amplitude $R$, this will be revealed in the sample spectrum as a sharp peak of approximate width $\pi/n$ and height $(n/2\pi)R^2$. This is called the discrete part of the spectrum, the variance $R^2$ associated with this component being in effect concentrated at a single frequency.

If the series $x_t$ has no deterministic components, i.e., is purely stochastic being stationary with acf $r_k$, then with increasing sample size the expected value of $f^*(\omega)$ converges to the theoretical spectrum – the continuous part

$$f(\omega) = \left(\frac{1}{2\pi}\right) \left( \gamma_0 + 2 \sum_{k=1}^{\infty} \gamma_k \cos(\omega k) \right)$$

where $\gamma_k$ are the theoretical autocovariances.

The sample spectrum does not however converge to this value but at each frequency point fluctuates about the theoretical spectrum with an exponential distribution, being independent at frequencies separated by an interval of $2\pi/n$ or more. Various devices are therefore employed to smooth the sample spectrum and reduce its variability. Much of the strength of spectral analysis derives from the fact that the error limits are multiplicative, so features may still show up as significant in a part of the spectrum which has a generally low level, whereas they are completely masked by other components in the original series. The spectrum can help to distinguish deterministic cyclical components from the stochastic quasi-cycle components which produce a broader peak in the spectrum. (The deterministic components can be removed by regression and the remaining part represented by an ARIMA model.)

A large discrete component in a spectrum can distort the continuous part over a large frequency range surrounding the corresponding peak. This may be alleviated at the cost of slightly broadening
the peak by tapering a portion of the data at each end of the series with weights which decay smoothly to zero. It is usual to correct for the mean of the series and for any linear trend by simple regression, since they would similarly distort the spectrum.

The unsmoothed sample spectrum is calculated for a fine division of frequencies, then averaged over intervals centred on each frequency point for which the smoothed spectrum is required. This is usually at a coarser frequency division. The bandwidth corresponds to the width of the averaging interval.

2.4.2. Cross-spectral analysis

The relationship between two time series may be investigated in terms of their sinusoidal components at different frequencies. At frequency $\omega$ a component of $y_t$ of the form

$$R_y(\omega) \cos(\omega t - \phi_y(\omega))$$

has its amplitude $R_y(\omega)$ and phase lag $\phi_y(\omega)$ estimated by

$$R_y(\omega) e^{i\phi_y(\omega)} = \frac{1}{n} \sum_{t=1}^{n} y_t e^{i\omega t}$$

and similarly for $x_t$. In the univariate analysis only the amplitude was important -- in the cross analysis the phase is important.

The sample cross-spectrum is defined by

$$f^*_{xy}(\omega) = \frac{1}{2\pi n} \left( \sum_{t=1}^{n} y_t e^{i\omega t} \right) \left( \sum_{t=1}^{n} x_t e^{-i\omega t} \right).$$

It may be demonstrated that this is equivalently defined in terms of the sample CCF, $r_{xy}(k)$, of the series as

$$f^*_{xy}(\omega) = \frac{1}{2\pi} \sum_{k=-n+1}^{n-1} c_{xy}(k) e^{i\omega k}$$

where $c_{xy}(k) = s_x \cdot s_y r_{xy}(k)$ is the cross-covariance function.

The cross-spectrum is specified by its real part or cospectrum $cf^*(\omega)$ and imaginary part or quadrature spectrum $qf^*(\omega)$, but for the purpose of interpretation the cross-amplitude spectrum and phase spectrum are useful:

$$A^*(\omega) = |f^*_{xy}(\omega)|, \quad \phi^*(\omega) = \arg(f^*_{xy}(\omega)).$$

If the series $x_t$ and $y_t$ contain deterministic sinusoidal components of amplitudes $R_y, R_x$ and phases $\phi_y, \phi_x$ at frequency $\omega$, then $A^*(\omega)$ will have a peak of approximate width $\pi/n$ and height $(n/2\pi)R_y R_x$ at that frequency, with corresponding phase $\phi^*(\omega) = \phi_y - \phi_x$. This supplies no information that cannot be obtained from the two series separately. The statistical relationship between the series is better revealed when the series are purely stochastic and jointly stationary, in which case the expected value of $f^*_{xy}(\omega)$ converges with increasing sample size to the theoretical cross-spectrum

$$f_{xy}(\omega) = \frac{1}{2\pi} \sum_{k=-\infty}^{\infty} \gamma_{xy}(k) e^{i\omega k}$$

where $\gamma_{xy}(k) = \text{cov}(x_t, y_{t+k})$.

The sample spectrum, as in the univariate case, does not however converge to the theoretical spectrum without some form of smoothing which either implicitly (using a lag window) or explicitly
(using a frequency window) averages the sample spectrum $f_{xy}(\omega)$ over wider bands of frequency to obtain a smoothed estimate $\hat{f}_{xy}(\omega)$.

If there is no statistical relationship between the series at a given frequency, $f_{xy}(\omega) = 0$, and the smoothed estimate $\hat{f}_{xy}(\omega)$, will be close to 0. This is assessed by the squared coherency between the series:

$$\hat{W}(\omega) = \frac{|\hat{f}_{xy}(\omega)|^2}{\hat{f}_{xx}(\omega) \hat{f}_{yy}(\omega)}$$

where $\hat{f}_{xx}(\omega)$ is the corresponding smoothed univariate spectrum estimate for $x_t$, and similarly for $y_t$. The coherency can be treated as a squared multiple correlation. It is similarly invariant in theory not only to simple scaling of $x_t$ and $y_t$, but also to filtering of the two series, and provides a useful test statistic for the relationship between autocorrelated series. Note that without smoothing

$$|f_{xy}(\omega)|^2 = f_{xx}(\omega) f_{yy}(\omega),$$

so the coherency is 1 at all frequencies, just as a correlation is 1 for a sample of size 1. Thus smoothing is essential for cross-spectrum analysis.

If $y_t$ is believed to be related to $x_t$ by a linear lagged relationship as in Section 2.3, i.e.,

$$y_t = v_0 x_t + v_1 x_{t-1} + v_2 x_{t-2} + \cdots + n_t,$$

then the theoretical cross-spectrum is

$$f_{xy}(\omega) = V(\omega) f_{xx}(\omega)$$

where

$$V(\omega) = G(\omega) e^{i\phi(\omega)} = \sum_{k=0}^{\infty} v_k e^{ik\omega}$$

is called the frequency response of the relationship.

Thus if $x_t$ were a sinusoidal wave at frequency $\omega$ (and $n_t$ were absent), $y_t$ would be similar but multiplied in amplitude by $G(\omega)$ and shifted in phase by $\phi(\omega)$. Furthermore, the theoretical univariate spectrum

$$f_{yy}(\omega) = G(\omega)^2 f_{xx}(\omega) + f_n(\omega)$$

where $n_t$, with spectrum $f_n(\omega)$, is assumed independent of the input $x_t$.

Cross-spectral analysis thus furnishes estimates of the gain

$$\hat{G}(\omega) = |\hat{f}_{xy}(\omega)| / \hat{f}_{xx}(\omega)$$

and the phase

$$\hat{\phi}(\omega) = \arg \left( \hat{f}_{xy}(\omega) \right)$$

From these representations of the estimated frequency response $\hat{V}(\omega)$, parametric TF models may be recognised and selected. The noise spectrum may also be estimated as

$$\hat{f}_{yy}(\omega) = \hat{f}_{yy}(\omega) \left( 1 - \hat{W}(\omega) \right),$$

a formula which reflects the fact that in essence a regression is being performed of the sinusoidal components of $y_t$ on those of $x_t$ over each frequency band.
Interpretation of the frequency response may be aided by extracting from $\hat{V}(\omega)$ estimates of the IRF $t_k$. It is assumed that there is no anticipatory response between $y_t$ and $x_t$, i.e., no coefficients $t_k$ with $k = -1, -2$ are needed (their presence might indicate feedback between the series).

2.5. References


3. Available Functions

The following functions are for ARIMA modelling: Calculate the auto-correlation function for a univariate time series $g_{13abc}$

Calculate the partial auto-correlation function for a univariate time series $g_{13acc}$

Fit a multi-input (transfer function) model $g_{13bec}$

Compute forecasts of a time series $g_{13bzc}$

Initialise the options structure utilised by some functions in this chapter $g_{13bxc}$

Allocate memory to the pointers in the transfer function model orders structure $g_{13byc}$

Free the memory allocated by NAG functions to the transfer function model orders structure $g_{13bzc}$

Free the memory allocated by NAG functions to the options structure $g_{13zgc}$

The following routines may be used for Kalman filter computations: Compute one iteration step of the time-varying Kalman filter recursion using the square root covariance implementation $g_{13eac}$

Compute one iteration step of the time-invariant Kalman filter recursion using the square root covariance implementation with $(A, C)$ in lower observer Hessenberg form $g_{13ebc}$

Compute one iteration step of the time-varying Kalman filter recursion using the square root information implementation $g_{13ecc}$

Compute one iteration step of the time-invariant Kalman filter recursion using the square root information implementation with $(A^{-1}, A^{-1}B)$ in upper controller Hessenberg form $g_{13edc}$

The following functions may be used to compute the required matrices for the time-invariant Kalman filter recursion:

Reduce the matrix pair $(A, C)$ to lower or upper observer Hessenberg form $g_{13ewc}$

Reduce the matrix pair $(B, A)$ to lower or upper controller Hessenberg form $g_{13exc}$

The following functions are for spectral analysis:

Calculate the smoothed sample spectrum of a univariate time series $g_{13cbc}$

Calculate the smoothed sample spectrum of a bivariate time series $g_{13cdd}$

From the computed sample spectrums the following may be computed:

- Cross amplitude spectrum and squared coherency $g_{13cecc}$
- Gain and phase $g_{13cfc}$
- Noise spectrum $g_{13cgc}$