NAG Toolbox for MATLAB

Chapter Introduction

G13 – Time Series Analysis

Contents

1 Scope of the Chapter

This chapter provides facilities for investigating and modelling the statistical structure of series of observations collected at equally spaced points in time. The models may then be used to forecast the series.

The chapter covers the following models and approaches.

1. Univariate time series analysis, including autocorrelation functions and autoregressive moving average (ARMA) models.
2. Univariate spectral analysis.
3. Transfer function (multi-input) modelling, in which one time series is dependent on other time series.
4. Bivarate spectral methods including coherency, gain and input response functions.
5. Vector ARMA models for multivariate time series.
6. Kalman filter models.
7. GARCH models for volatility.

2 Background to the Problems

2.1 Univariate Analysis

Let the given time series be $x_1, x_2, \ldots, x_n$, where n is its length. The structure which is intended to be investigated, and which may be most evident to the eye in a graph of the series, can be broadly described as:

(a) trends, linear or possibly higher-order polynomial;

(b) seasonal patterns, associated with fixed integer seasonal periods. The presence of such seasonality and the period will normally be known *a priori*. The pattern may be fixed, or slowly varying from one season to another;

(c) cycles or waves of stable amplitude and period p (from peak to peak). The period is not necessarily integer, the corresponding absolute frequency (cycles/time unit) being $f = 1/p$ and angular frequency $\omega = 2\pi f$. The cycle may be of pure sinusoidal form like $\sin(\omega t)$, or the presence of higher harmonic terms may be indicated, e.g., by asymmetry in the wave form;

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

(e) irregular statistical fluctuations and swings about the overall mean or trend.

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 a finite data set it is not always easy to discriminate between these two types, and a common description using the class of autoregressive integrated moving-average (ARIMA) models is now widely used. The form of these models is that of difference equations (or recurrence relations) relating present and past values of the series. You are referred to Box and Jenkins 1976 for a thorough account of these models and how to use them. We follow their notation and outline the recommended steps in ARIMA model building for which functions are available.

2.1.1 Transformations

If the variance of the observations in the series is not constant across the range of observations it may be useful to apply a variance-stabilizing transformation to the series. A common situation is for the variance to increase with the magnitude of the observations and in this case typical transformations used are the log or square root transformation. A range-mean or standard deviation-mean plot provides a quick and easy way of detecting non-constant variance and of choosing, if required, a suitable transformation. This is a plot of the range or standard deviation of successive groups of observations against their means.

2.1.2 Differencing operations

These may be used to simplify the structure of a time series.

First-order differencing, i.e., forming the new series

$$\nabla x_t = x_t - x_{t-1}$$

will remove a linear trend. 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. Differencing is also therefore appropriate when the series has something of the nature of a random walk, which is by definition the accumulation of independent changes.

Differencing may be applied repeatedly to a series, giving

$$w_t = \nabla^d \nabla_s^D x_t$$

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 $t = 1 + d + s \times D, \ldots, n$.

2.1.3 Sample autocorrelations

Given that a series has (possibly as a result of simplifying by differencing operations) a homogeneous appearance throughout its length, fluctuating with approximately constant variance about an overall mean level, it is appropriate to assume that its statistical properties are stationary. For most purposes the correlations ρ_k between terms x_t, x_{t+k} or w_t, w_{t+k} separated by lag k give an adequate description of the statistical structure and are estimated by the sample ACF r_k, for $k = 1, 2, \ldots$.

As described by Box and Jenkins 1976, these may be used to indicate which particular ARIMA model may be appropriate.

2.1.4 Partial autocorrelations

The information in the autocorrelations, ρ_k, may be presented in a different light by deriving from them the coefficients of the partial autocorrelation function (PACF) $\phi_{k,k}$, for $k = 1, 2, \ldots$. $\phi_{k,k}$ measures the correlation between x_t and x_{t+k} conditional upon the intermediate values $x_{t+1}, x_{t+2}, \ldots, x_{t+k-1}$. The corresponding sample values $\hat{\phi}_{k,k}$ give further assistance in the selection of ARIMA models.

Both ACF and PACF may be rapidly computed, particularly in comparison with the time taken to estimate ARIMA models.

2.1.5 Finite lag predictor coefficients and error variances

The partial autocorrelation coefficient $\phi_{k,k}$ is determined as the final parameter in the minimum variance predictor of x_t in terms of $x_{t-1}, x_{t-2}, \ldots, x_{t-k}$,

$$x_t = \phi_{k,1} x_{t-1} + \phi_{k,2} x_{t-2} + \cdots + \phi_{k,k} x_{t-k} + e_{k,t}$$

where $e_{k,t}$ is the prediction error, and the first subscript k of $\phi_{k,i}$ and $e_{k,t}$ emphasises the fact that the parameters will alter as k increases. Moderately good estimates $\hat{\phi}_{k,i}$ of $\phi_{k,i}$ are obtained from the sample ACF, and after calculating the PACF up to lag L, the successive values $v_1, v_2, \ldots, v_L$ of the prediction error variance estimates, $v_k = \mathrm{var}\left(e_{k,t}\right)$, are available, together with the final values of the coefficients $\hat{\phi}_{k,1}, \hat{\phi}_{k,2}, \ldots, \hat{\phi}_{k,L}$. If x_t has nonzero mean, $\bar{x}$, it is adequate to use $x_t - \bar{x}$ in place of x_t in the prediction equation.

Although Box and Jenkins 1976 do not place great emphasis on these prediction coefficients, their use is advocated for example by Akaike 1971, who recommends selecting an optimal order of the predictor as the lag for which the final prediction error (FPE) criterion $(1 + k/n)(1 - k/n)^{-1} v_k$ is a minimum.

2.1.6 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. If the stationary series w_t has been derived by differencing from 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

$$w_t = \phi_1 w_{t-1} + \cdots + \phi_p w_{t-p} + a_t - \theta_1 a_{t-1} - \cdots - \theta_q a_{t-q}, \tag{1}$$

where a_t is an uncorrelated series (white noise) with mean 0 and constant variance σ_a^2. If w_t has a nonzero mean c, then this is allowed for by replacing $w_t, w_{t-1}, \ldots$ by $w_t - c, w_{t-1} - c, \ldots$ 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 also be usefully interpreted as the linear innovations in x_t (and in w_t), i.e., the error if x_t were to be predicted using the information in all past values $x_{t-1}, x_{t-2}, \ldots$, provided also that $\theta_1, \theta_2, \ldots, \theta_q$ satisfy invertibility constraints. This allows the series a_t to be regenerated by rewriting the model equation as

$$a_t = w_t - \phi_1 w_{t-1} - \cdots - \phi_p w_{t-p} + \theta_1 a_{t-1} + \cdots + \theta_q a_{t-q}. \tag{2}$$

For a series with short-term correlation only, i.e., r_k is not significant beyond some low lag q (see Box and Jenkins 1976 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 $\hat{\phi}_{k,k}$ 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 $w_t = \nabla^d \nabla_s^D x_t$, the series is represented in a two-stage manner via an intermediate series e_t:

$$w_t = \Phi_1 w_{t-s} + \cdots + \Phi_P w_{t-s\times P} + e_t - \Theta_1 e_{t-s} - \cdots - \Theta_Q e_{t-s\times Q} \tag{3}$$

$$e_t = \phi_1 e_{t-1} + \cdots + \phi_p e_{t-p} + a_t - \theta_1 a_{t-1} - \cdots - \theta_q a_{t-q} \tag{4}$$

where Φ_i, Θ_i are the seasonal parameters and P, Q are the corresponding orders. Again, w_t may be replaced by $w_t - c$.

2.1.7 ARIMA model estimation

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. This is based on the least-squares principle by which parameters are chosen to minimize the sum of squares of the innovations a_t, which are regenerated from the data using (2), or the reverse of (3) and (4) in the case of seasonal models.

Lack of knowledge of terms on the right-hand side of (2), when $t = 1, 2, \ldots, \max(p, q)$, is overcome by introducing q unknown series values $w_0, w_1, \ldots, w_{1-q}$ which are estimated as nuisance parameters, and using correction for transient errors due to the autoregressive terms. If the data $w_1, w_2, \ldots, w_N = w$ is viewed as a single sample from a multivariate Normal density whose covariance matrix V is a function of the ARIMA model parameters, then the exact likelihood of the parameters is

$$-\tfrac{1}{2}\log|V| - \tfrac{1}{2}w^{\mathrm{T}} V^{-1} w.$$

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

$$QF = w^{\mathrm{T}} V^{-1} w$$

as an objective function to be minimized. Neglecting the term $-\frac{1}{2}\log|V|$ 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.

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

The model residuals, $\hat{a}_t$, are the innovations resulting from the estimation and are usually examined for the presence of autocorrelation as a check on the adequacy of the model.

2.1.8 ARIMA model 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.

This process requires knowledge only of the model orders and parameters together with a limited set of the terms $a_{t-i}, e_{t-i}, w_{t-i}, x_{t-i}$ which appear on the right-hand side of the models (3) and (4) (and the differencing equations) when $t = n$. It does not require knowledge of the whole series.

We call this the state set. It is conveniently constituted after model estimation. Moreover, if new observations $x_{n+1}, x_{n+2}, \ldots$ come to hand, then the model equations can easily be used to update the state set before constructing forecasts from the end of the new observations. This is particularly useful when forecasts are constructed on a regular basis. The new innovations $a_{n+1}, a_{n+2}, \ldots$ may be compared with the residual standard deviation, σ_a, of the model used for forecasting, as a check that the model is continuing to forecast adequately.

2.2 Univariate 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 ω, $0 \leq \omega \leq \pi$, is specified by its amplitude $R > 0$ and phase ϕ, $0 \leq \phi < 2\pi$. 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 the 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 ω is given to order $1/n$ by

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

2.2.1 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 (σ_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 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 π/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 γ_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 that 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.

2.2.2 Spectral smoothing by lag window

The estimate is calculated directly from the sample covariances c_k as

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

the smoothing being induced by the lag window weights w_k which extend up to a truncation lag M which is generally much less than n. The smaller the value of M, the greater the degree of smoothing, the spectrum estimates being independent only at a wider frequency separation indicated by the bandwidth b which is proportional to $1/M$. It is wise, however, to calculate the spectrum at intervals appreciably less than this. Although greater smoothing narrows the error limits, it can also distort the spectrum, particularly by flattening peaks and filling in troughs.

2.2.3 Direct spectral smoothing

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.

2.3 Linear Lagged Relationships Between Time Series

We now consider the context in which one time series, called the dependent or output series, $y_1, y_2, \ldots, y_n$, is believed to depend on one or more explanatory or input series, e.g., $x_1, x_2, \ldots, x_n$. This dependency may follow a simple linear regression, e.g.,

$$y_t = vx_t + n_t$$

or more generally may involve lagged values of the input

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

The sequence $v_0, v_1, v_2, \ldots$ is called the impulse response function (IRF) of the relationship. The term n_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 n_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, n_t, are independent.

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

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

so $y_t = z_t + n_t$.

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

$$y_t = z_{1,t} + z_{2,t} + n_t.$$

2.3.1 Transfer function models

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 $\omega_0, \omega_1, \ldots, \omega_q$:

$$z_t = \delta_1 z_{t-1} + \delta_2 z_{t-2} + \cdots + \delta_p z_{t-p} + \omega_0 x_{t-b} - \omega_1 x_{t-b-1} - \cdots - \omega_q x_{t-b-q}. \tag{5}$$

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 $\omega_0, \omega_1, \ldots, \omega_q$.

2.3.2 Cross-correlations

An important tool for investigating how an input series x_t affects an output series y_t is the sample cross-correlation function (CCF) $r_{xy}(k)$, for $k = 0, 1, 2, \ldots,$ between the series. If x_t and y_t are (jointly) stationary time series this is an estimator of the theoretical quantity

$$\rho_{xy}(k) = \mathrm{corr}\left(x_t, y_{t+k}\right).$$

The sequence $r_{yx}(k)$, for $k = 0, 1, 2, \ldots$, is distinct from $r_{xy}(k)$, though it is possible to interpret

$$r_{yx}(k) = r_{xy}(-k).$$

When the series y_t and x_t are believed to be related by a TF model, the CCF is determined by the IRF $v_0, v_1, v_2, \ldots$ and the ACF of the input x_t.

In the particular case when x_t is an uncorrelated series or white noise (and is uncorrelated with any other inputs):

$$\rho_{xy}(k) \propto v_k$$

and the sample CCF can provide an estimate of v_k:

$$\tilde{v}_k = \left(s_y/s_x\right) r_{xy}(k)$$

where s_y and s_x are the sample standard deviations of y_t and x_t, respectively.

In theory the IRF coefficients $v_b, \ldots, v_{b+p+q}$ determine the parameters in the TF model, and using $\tilde{v}_k$ to estimate $\tilde{v}_k$ it is possible to solve for preliminary estimates of $\delta_1, \delta_2, \ldots, \delta_p$, $\omega_0, \omega_1, \ldots, \omega_q$.

2.3.3 Prewhitening or filtering by an ARIMA model

In general an input series x_t is not white noise, but may be represented by an ARIMA model with innovations or residuals a_t which are white noise. If precisely the same operations by which a_t is generated from x_t are applied to the output y_t to produce a series b_t, then the transfer function relationship between y_t and x_t is preserved between b_t and a_t. It is then possible to estimate

$$\tilde{v}_k = (s_b/s_a)r_{ab}(k).$$

The procedure of generating a_t from x_t (and b_t from y_t) is called prewhitening or filtering by an ARIMA model. Although a_t is necessarily white noise, this is not generally true of b_t.

2.3.4 Multi-input model estimation

The term multi-input model is used for the situation when one output series y_t is related to one or more input series $x_{j,t}$, as described in Section 2.3. If for a given input the relationship is a simple linear regression, it is called a simple input; otherwise it is a transfer function input. The error or noise term follows an ARIMA model.

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 procedure used is closely related to the least-squares principle applied to the innovations in the ARIMA noise model.

The innovations are derived for any proposed set of parameter values by calculating the response of each input to the transfer functions and then evaluating the noise n_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, and as described in Section 2.1.5.

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 (sub)program descriptions provide the necessary detail as to how this problem is treated.

Also, as described in Section 2.1.6 the sum of squares criterion

$$S = \sum a_t^2$$

is related to the quadratic form in the exact log-likelihood of the parameters:

$$-\tfrac{1}{2}\log|V| - \tfrac{1}{2}w^{\mathrm{T}}V^{-1}w.$$

Here w is the vector of appropriately differenced noise terms, and

$$w^{\mathrm{T}}V^{-1}w = S/\sigma_a^2,$$

where σ_a^2 is the innovation variance parameter.

The least-squares criterion is therefore identical to minimization of the quadratic form, but is not identical to exact likelihood. Because V may be expressed as $M\sigma_a^2$, where M is a function of the ARIMA model parameters, substitution of σ_a^2 by its maximum likelihood (ML) estimator yields a concentrated (or profile) likelihood which is a function of

$$|M|^{1/N}S.$$

N is the length of the differenced noise series w, and $|M| = \det M$.

Use of the above quantity, called the deviance, D, as an objective function is preferable to the use of S alone, on the grounds that it is equivalent to exact likelihood, and yields estimates with better properties. However, there is an appreciable computational penalty in calculating D, and in large samples it differs very little from S, except in the important case of seasonal ARIMA models where the number of whole seasons within the data length must also be large.

You are given the option of taking the objective function to be either S or D, or a third possibility, the marginal likelihood. This 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.

The model residuals $\hat{a}_t$ are the innovations resulting from the estimation, and they are usually examined for the presence of either autocorrelation or cross-correlation with the inputs. Absence of such correlation provides some confirmation of the adequacy of the model.

2.3.5 Multi-input model forecasting

A multi-input model may be used to forecast the output series provided future values (possibly forecasts) of the input series are supplied.

Construction of the forecasts requires knowledge only of the model orders and parameters, together with a limited set of the most recent variables which appear in the model equations. This is called the state set. It is conveniently constituted after model estimation. Moreover, if new observations $y_{n+1}, y_{n+2}, \ldots$ of the output series and $x_{n+1}, x_{n+2}, \ldots$ of (all) the independent input series become available, then the model equations can easily be used to update the state set before constructing forecasts from the end of the new observations. The new innovations $a_{n+1}, a_{n+2}, \ldots$ generated in this updating may be used to monitor the continuing adequacy of the model.

2.3.6 Transfer function model filtering

In many time series applications it is desired to calculate the response (or output) of a TF model for a given input series.

Smoothing, detrending, and seasonal adjustment are typical applications. You must specify the orders and parameters of a TF model for the purpose being considered. This may then be applied to the input series.

Again, problems may arise due to ignorance of the input series values prior to the observation period. The transient errors which can arise from this cause may be substantially reduced by using ‘backforecasts’ of these unknown observations.

2.4 Multivariate Time Series

Multi-input modelling represents one output time series in terms of one or more input series. Although there are circumstances in which it may be more appropriate to analyse a set of time series by modelling each one in turn as the output series with the remainder as inputs, there is a more symmetric approach in such a context. These models are known as vector autoregressive moving-average (VARMA) models.

2.4.1 Differencing and transforming a multivariate time series

As in the case of a univariate time series, it may be useful to simplify the series by differencing operations which may be used to remove linear or seasonal trend, thus ensuring that the resulting series to be used in the model estimation is stationary. It may also be necessary to apply transformations to the individual components of the multivariate series in order to stabilize the variance. Commonly used transformations are the log and square root transformations.

2.4.2 Model identification for a multivariate time series

Multivariate analogues of the autocorrelation and partial autocorrelation functions are available for analysing a set of k time series, $x_{i,1}, x_{i,2}, \ldots, x_{i,n}$, for $i = 1, 2, \ldots, k$, thereby making it possible to obtain some understanding of a suitable VARMA model for the observed series.

It is assumed that the time series have been differenced if necessary, and that they are jointly stationary. The lagged correlations between all possible pairs of series, i.e.,

$$\rho_{ijl} = \operatorname{corr}\left(x_{i,t}, x_{j,t+l}\right)$$

are then taken to provide an adequate description of the statistical relationships between the series. These quantities are estimated by sample auto- and cross-correlations r_{ijl}. For each l these may be viewed as elements of a (lagged) autocorrelation matrix.

Thus consider the **vector process** x_t (with elements x_{it}) and lagged autocovariance matrices Γ_l with elements of $\sigma_i \sigma_j \rho_{\mathrm{ijl}}$ where $\sigma_i^2 = \operatorname{var}\left(x_{i,t}\right)$. Correspondingly, Γ_l is estimated by the matrix C_l with elements $s_i s_j r_{\mathrm{ijl}}$ where s_i^2 is the sample variance of x_{it}.

For a series with short-term cross-correlation only, i.e., r_{ijl} is not significant beyond some low lag q, then the pure vector MA(q) model, with no autoregressive parameters, i.e., $p = 0$, is appropriate.

The correlation matrices provide a description of the joint statistical properties of the series. It is also possible to calculate matrix quantities which are closely analogous to the partial autocorrelations of univariate series (see Section 2.1.3). Wei 1990 discusses both the partial autoregression matrices proposed by Tiao and Box 1981 and partial lag correlation matrices.

In the univariate case the PACF between x_t and x_{t+l} is the correlation coefficient between the two after removing the linear dependence on each of the intervening variables $x_{t+1}, x_{t+2}, \ldots, x_{t+l-1}$. This partial autocorrelation may also be obtained as the last regression coefficient associated with x_t when regressing x_{t+l} on its l lagged variables $x_{t+l-1}, x_{t+l-2}, \ldots, x_t$. Tiao and Box 1981 extended this method to the multivariate case to define the partial autoregression matrix. Heyse and Wei 1985 also extended the univariate definition of the PACF to derive the correlation matrix between the vectors x_t and x_{t+l} after removing the linear dependence on each of the intervening vectors $x_{t+1}, x_{t+2}, \ldots, x_{t+l-1}$, the partial lag correlation matrix.

Note that the partial lag correlation matrix is a correlation coefficient matrix since each of its elements is a properly normalized correlation coefficient. This is not true of the partial autoregression matrices (except in the univariate case for which the two types of matrix are the same). The partial lag correlation matrix at lag 1 also reduces to the regular correlation matrix at lag 1; this is not true of the partial autoregression matrices (again except in the univariate case).

Both the above share the same cut-off property for autoregressive processes; that is for an autoregressive process of order p, the terms of the matrix at lags $p + 1$ and greater are zero. Thus if the sample partial cross-correlations are significant only up to some low lag p then a pure vector AR(p) model is appropriate with $q = 0$. Otherwise moving-average terms will need to be introduced as well as autoregressive terms.

Under the hypothesis that x_t is an autoregressive process of order $l - 1$, n times the sum of the squared elements of the partial lag correlation matrix at lag l is asymptotically distributed as a χ^2 variable with k^2 degrees of freedom where k is the dimension of the multivariate time series. This provides a diagnostic aid for determining the order of an autoregressive model.

The partial autoregression matrices may be found by solving a multivariate version of the Yule–Walker equations to find the autoregression matrices, using the final regression matrix coefficient as the partial autoregression matrix at that particular lag.

The basis of these calculations is a multivariate autoregressive model:

$$x_t = \phi_{l,1} x_{t-1} + \cdots + \phi_{l,l} x_{t-l} + e_{l,t}$$

where $\phi_{l,1}, \phi_{l,2}, \ldots, \phi_{l,l}$ are matrix coefficients, and $e_{l,t}$ is the vector of errors in the prediction. These coefficients may be rapidly computed using a recursive technique which requires, and simultaneously furnishes, a backward prediction equation:

$$x_{t-l-1} = \psi_{l,1} x_{t-l} + \psi_{l,2} x_{t-l+1} + \cdots + \psi_{l,l} x_{t-1} + f_{l,t}$$

(in the univariate case $\psi_{l,i} = \phi_{l,i}$).

The forward prediction equation coefficients, $\phi_{l,i}$, are of direct interest, together with the covariance matrix D_l of the prediction errors $e_{l,t}$. The calculation of these quantities for a particular maximum equation lag $l = L$ involves calculation of the same quantities for increasing values of $l = 1, 2, \ldots, L$.

The quantities $v_l = \det D_l / \det \Gamma_0$ may be viewed as generalized variance ratios, and provide a measure of the efficiency of prediction (the smaller the better). The reduction from v_{l-1} to v_l which occurs on extending the order of the predictor to l may be represented as

$$v_l = v_{l-1} \left(1 - \rho_l^2\right)$$

where ρ_l^2 is a multiple squared partial autocorrelation coefficient associated with k^2 degrees of freedom.

Sample estimates of all the above quantities may be derived by using the series covariance matrices C_l, for $l = 1, 2, \ldots, L$, in place of Γ_l. The best lag for prediction purposes may be chosen as that which yields the minimum FPE criterion:

$$\text{FPE}(l) = v_l \times \frac{\left(1 + lk^2/n\right)}{\left(1 - lk^2/n\right)}.$$

An alternative method of estimating the sample partial autoregression matrices is by using multivariate least-squares to fit a series of multivariate autoregressive models of increasing order.

2.4.3 VARMA model estimation

The cross-correlation structure of a stationary multivariate time series may often be represented by a model with a small number of parameters belonging to the VARMA class. If the stationary series w_t has been derived by transforming and/or differencing the original series x_t, then w_t is said to follow the VARMA model:

$$w_t = \phi_1 w_{t-1} + \cdots + \phi_p w_{t-p} + \epsilon_t - \theta_1 \epsilon_{t-1} - \cdots - \theta_q \epsilon_{t-q},$$

where ϵ_t is a vector of uncorrelated residual series (white noise) with zero mean and constant covariance matrix Σ, $\phi_1, \phi_2, \ldots, \phi_p$ are the p autoregressive (AR) parameter matrices and $\theta_1, \theta_2, \ldots, \theta_q$ are the q moving-average (MA) parameter matrices. If w_t has a nonzero mean μ, then this can be allowed for by replacing $w_t, w_{t-1}, \ldots$ by $w_t - \mu, w_{t-1} - \mu, \ldots$ in the model.

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 stationarity constraints. The series ϵ_t may also be usefully interpreted as the linear innovations in w_t, i.e., the error if w_t were to be predicted using the information in all past values $w_{t-1}, w_{t-2}, \ldots$, provided also that $\theta_1, \theta_2, \ldots, \theta_q$ satisfy what are known as invertibility constraints. This allows the series ϵ_t to be generated by rewriting the model equation as

$$\epsilon_t = w_t - \phi_1 w_{t-1} - \cdots - \phi_p w_{t-p} + \theta_1 \epsilon_{t-1} + \cdots + \theta_q \epsilon_{t-q}.$$

The method of ML may be used to estimate the parameters of a specified VARMA model from the observed multivariate time series together with their standard errors and correlations.

The residuals from the model may be examined for the presence of autocorrelations as a check on the adequacy of the fitted model.

2.4.4 VARMA model forecasting

Forecasts of the series may be constructed using a multivariate version of the univariate method. Efficient methods are available for updating the forecasts each time new observations become available.

2.5 Cross-spectral Analysis

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

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

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

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

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

2.5.1 The sample cross-spectrum

This 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

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

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

2.5.2 The amplitude and phase spectrum

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:

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

If the series x_t and y_t contain deterministic sinusoidal components of amplitudes R_y, R_x and phases ϕ_y, ϕ_x at frequency ω, then $A^*(\omega)$ will have a peak of approximate width π/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

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

where $\gamma_{xy}(k) = \mathrm{cov}\left(x_t, y_{t+k}\right)$. 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 $\hat{f}_{xy}(\omega)$.

2.5.3 The coherency spectrum

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

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

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,

$$\left|f^*_{xy}(\omega)\right|^2 = f^*_{xx}(\omega) f^*_{yy}(\omega),$$

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.

2.5.4 The gain and noise spectrum

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 ω (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 recognized and selected. The noise spectrum may also be estimated as

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

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 $\hat{v}_k$. It is assumed that there is no anticipatory response between y_t and x_t, i.e., no coefficients v_k with $k = -1$ or -2 are needed (their presence might indicate feedback between the series).

2.5.5 Cross-spectrum smoothing by lag window

The estimate of the cross-spectrum is calculated from the sample cross-variances as

$$\hat{f}_{xy}(\omega) = \frac{1}{2\pi} \sum_{-M+S}^{M+S} w_{k-S} c_{xy}(k) e^{i\omega k}.$$

The lag window w_k extends up to a truncation lag M as in the univariate case, but its centre is shifted by an alignment lag S usually chosen to coincide with the peak cross-correlation. This is equivalent to an alignment of the series for peak cross-correlation at lag 0, and reduces bias in the phase estimation.

The selection of the truncation lag M, which fixes the bandwidth of the estimate, is based on the same criteria as for univariate series, and the same choice of M and window shape should be used as in univariate spectrum estimation to obtain valid estimates of the coherency, gain, etc., and test statistics.

2.5.6 Direct smoothing of the cross-spectrum

The computations are exactly as for smoothing of the univariate spectrum except that allowance is made for an implicit alignment shift S between the series.

2.6 Kalman Filters

Kalman filtering provides a method for the analysis of multi-dimensional time series. The underlying model is:

$$X_{t+1} = A_t X_t + B_t W_t \tag{6}$$

$$Y_t = C_t X_t + V_t \tag{7}$$

where X_t is the unobserved state vector, Y_t is the observed measurement vector, W_t is the state noise, V_t is the measurement noise, A_t is the state transition matrix, B_t is the noise coefficient matrix and C_t is the measurement coefficient matrix at time t. The state noise and the measurement noise are assumed to be uncorrelated with zero mean and covariance matrices:

$$E\{W_t W_t^{\mathrm{T}}\} = Q_t \qquad \text{and} \qquad E\{V_t V_t^{\mathrm{T}}\} = R_t.$$

If the system matrices A_t, B_t, C_t and the covariance matrices Q_t, R_t are known then Kalman filtering can be used to compute the minimum variance estimate of the stochastic variable X_t.

The estimate of X_t given observations Y_1 to Y_{t-1} is denoted by $\hat{X}_{t|t-1}$ with state covariance matrix $E\left\{\hat{X}_{t|t-1}\hat{X}_{t|t-1}^{\mathrm{T}}\right\} = P_{t|t-1}$ while the estimate of X_t given observations Y_1 to Y_t is denoted by $\hat{X}_{t|t}$ with covariance matrix $E\left\{\hat{X}_{t|t}\hat{X}_{t|t}^{\mathrm{T}}\right\} = P_{t|t}$.

The update of the estimate, $\hat{X}_{t+1|t}$, from time t to time $t+1$, is computed in two stages.

First, the update equations are

$$\hat{X}_{t|t} = \hat{X}_{t|t-1} + K_t r_t, \qquad P_{t|t} = (I - K_t C_t)P_{t|t-1}$$

where the residual $r_t = Y_t - C_t X_{t|t-1}$ has an associated covariance matrix $H_t = C_t P_{t|t-1} C_t^{\mathrm{T}} + R_t$, and K_t is the Kalman gain matrix with

$$K_t = P_{t|t-1} C_t^{\mathrm{T}} H_t^{-1}.$$

The second stage is the one-step-ahead prediction equations given by

$$\hat{X}_{t+1|t} = A_t \hat{X}_{t|t}, \qquad P_{t+1|t} = A_t P_{t|t} A_t^{\mathrm{T}} + B_t Q_t B_t^{\mathrm{T}}.$$

These two stages can be combined to give the one-step-ahead update-prediction equations

$$\hat{X}_{t+1|t} = A_t \hat{X}_{t|t-1} + A_t K_t r_t.$$

The above equations thus provide a method for recursively calculating the estimates of the state vectors $\hat{X}_{t|t}$ and $\hat{X}_{t+1|t}$ and their covariance matrices $P_{t|t}$ and $P_{t+1|t}$ from their previous values. This recursive procedure can be viewed in a Bayesian framework as being the updating of the prior by the data Y_t.

The initial values $\hat{X}_{1|0}$ and $P_{1|0}$ are required to start the recursion. For stationary systems, $P_{1|0}$ can be computed from the following equation:

$$P_{1|0} = A_1 P_{1|0} A_1^{\mathrm{T}} + B_1 Q_1 B_1^{\mathrm{T}},$$

which can be solved by iterating on the equation. For $\hat{X}_{1|0}$ the value $E\{X\}$ can be used if it is available.

2.6.1 Computational methods

To improve the stability of the computations the square root algorithm is used. One recursion of the square root covariance filter algorithm which can be summarized as follows:

$$\begin{pmatrix} R_t^{1/2} & C_t S_t & 0 \\ 0 & A_t S_t & B_t Q_t^{1/2} \end{pmatrix} U = \begin{pmatrix} H_t^{1/2} & 0 & 0 \\ G_t & S_{t+1} & 0 \end{pmatrix}$$

where U is an orthogonal transformation triangularizing the left-hand pre-array to produce the right-hand post-array, S_t is the lower triangular Cholesky factor of the state covariance matrix $P_{t+1|t}$, $Q_t^{1/2}$ and $R_t^{1/2}$ are the lower triangular Cholesky factor of the covariance matrices Q and R and $H^{1/2}$ is the lower triangular Cholesky factor of the covariance matrix of the residuals. The relationship between the Kalman gain matrix, K_t, and G_t is given by

$$A_t K_t = G_t \left(H_t^{1/2}\right)^{-1}.$$

To improve the efficiency of the computations when the matrices A_t, B_t and C_t do not vary with time the system can be transformed to give a simpler structure. The transformed state vector is $U^* X$ where U^* is the transformation that reduces the matrix pair (A, C) to lower observer Hessenberg form. That is, the matrix U^* is computed such that the compound matrix

$$\begin{pmatrix} CU^{*T} \\ U^* A U^{*T} \end{pmatrix}$$

is a lower trapezoidal matrix. The transformations need only be computed once at the start of a series, and the covariance matrices Q_t and R_t can still be time-varying.

2.6.2 Model fitting and forecasting

If the state space model contains unknown parameters, θ, these can be estimated using ML. Assuming that W_t and V_t are normal variates the log-likelihood for observations Y_t, for $t = 1, 2, \ldots, n$, is given by

$$\text{constant} - \frac{1}{2}\sum_{t=1}^{n} \ln(\det(H_t)) - \frac{1}{2}\sum_{t=1}^{t} r_t^{\mathrm{T}} H_t^{-1} r_t.$$

Optimal estimates for the unknown model parameters θ can then be obtained by using a suitable optimizer function to maximize the likelihood function.

Once the model has been fitted forecasting can be performed by using the one-step-ahead prediction equations. The one-step-ahead prediction equations can also be used to 'jump over' any missing values in the series.

2.6.3 Kalman filter and time series models

Many commonly used time series models can be written as state space models. A univariate ARMA(p, q) model can be cast into the following state space form:

$$\begin{aligned} x_t &= Ax_{t-1} + B\epsilon_t \\ w_t &= Cx_t \end{aligned}$$

where $r = \max(p, q + 1)$, the first element of the state vector x_t is w_t,

$$A = \begin{pmatrix} \phi_1 & 1 & & & & \\ \phi_2 & & 1 & & & \\ . & & & . & & \\ . & & & & . & \\ \phi_{r-1} & & & & & 1 \\ \phi_r & 0 & 0 & . & . & 0 \end{pmatrix}, \quad B = \begin{pmatrix} 1 \\ -\theta_1 \\ -\theta_2 \\ . \\ . \\ -\theta_{r-1} \end{pmatrix} \quad \text{and} \quad C^{\mathrm{T}} = \begin{pmatrix} 1 \\ 0 \\ 0 \\ . \\ . \\ 0 \end{pmatrix}.$$

The representation for a k-variate ARMA(p, q) series (VARMA) is very similar to that given above, except now the state vector is of length kr and the ϕ and θ are now $k \times k$ matrices and the 1s in A, B and C are now the identity matrix of order k. If $p < r$ or $q + 1 < r$ then the appropriate ϕ or θ matrices are set to zero, respectively.

Since the compound matrix

$$\begin{pmatrix} C \\ A \end{pmatrix}$$

is already in lower observer Hessenberg form (i.e., it is lower trapezoidal with zeros in the top right-hand triangle) the invariant Kalman filter algorithm can be used directly without the need to generate a transformation matrix U^*.

2.7 GARCH Models

2.7.1 ARCH models and their generalizations

Rather than modelling the mean (for example using regression models) or the autocorrelation (by using ARMA models) there are circumstances in which the variance of a time series needs to be modelled. This is common in financial data modelling where the variance (or standard deviation) is known as volatility. The ability to forecast volatility is a vital part in deciding the risk attached to financial decisions like portfolio selection. The basic model for relating the variance at time t to the variance at previous times is the autoregressive conditional heteroskedastic (ARCH) model. The standard ARCH model is defined as

$$y_t \mid \psi_{t-1} \sim N(0, h_t),$$

$$h_t = \alpha_0 + \sum_{i=1}^{q} \alpha_i \epsilon_{t-1}^2,$$

where ψ_t is the information up to time t and h_t is the conditional variance.

In a similar way to that in which AR models were generalized to ARMA models the ARCH models have been generalized to a GARCH model; see Engle 1982, Bollerslev 1986 and Hamilton 1994

$$h_t = \alpha_0 + \sum_{i=1}^{q} \alpha_i \epsilon_{t-1}^2 + \sum_{j=1}^{p} \beta h_{t-j}.$$

This can be combined with a regression model:

$$y_t = b_0 + \sum_{i=1}^{k} b_i x_{it} + \epsilon_t,$$

where $\epsilon_t \mid \psi_{t-1} \sim N(0, h_t)$ and where x_{it}, for $i = 1, \ldots, k$, are the exogenous variables.

The above models assume that the change in variance, h_t, is symmetric with respect to the shocks, that is, that a large negative value of ϵ_{t-1} has the same effect as a large positive value of ϵ_{t-1}. A frequently observed effect is that a large negative value ϵ_{t-1} often leads to a greater variance than a large positive value. The following three asymmetric models represent this effect in different ways using the parameter λ as a measure of the asymmetry.

Type I AGARCH(p, q)

$$h_t = \alpha_0 + \sum_{i=1}^{q} \alpha_i (\epsilon_{t-i} + \gamma)^2 + \sum_{j=1}^{p} \beta_i h_{t-j}.$$

Type II AGARCH(p, q)

$$h_t = \alpha_0 + \sum_{i=1}^{q} \alpha_i (|\epsilon_{t-i}| + \gamma \epsilon_{t-i})^2 + \sum_{j=1}^{p} \beta_i h_{t-j}.$$

GJR-GARCH(p, q), or Glosten, Jagannathan and Runkle GARCH (see Glosten *et al.* 1993)

$$h_t = \alpha_0 + \sum_{i=1}^{q} (\alpha_i + \gamma S_{t-1}) \epsilon_{t-1}^2 + \sum_{j=1}^{p} \beta_i h_{t-j},$$

where $S_t = 1$ if $\epsilon_t < 0$ and $S_t = 0$ if $\epsilon_t \geq 0$.

The first assumes that the effects of the shocks are symmetric about γ rather than zero, so that for $\gamma < 0$ the effect of negative shocks is increased and the effect of positive shocks is decreased. Both the Type II AGARCH and the GJR GARCH (see Glosten *et al.* 1993) models introduce asymmetry by increasing the value of the coefficient of ϵ_{t-1}^2 for negative values of ϵ_{t-1}. In the case of the Type II AGARCH the effect is multiplicative while for the GJR GARCH the effect is additive.

Coefficient	$\epsilon_{t-1} < 0$	$\epsilon_{t-1} > 0$
Type II AGARCH	$\alpha_i (1 - \gamma)^2$	$\alpha_i (1 + \gamma)^2$
GJR GARCH	$\alpha_i + \gamma$	α_i

(Note that in the case of GJR GARCH, γ needs to be positive to inflate variance after negative shocks while for Type I and Type II AGARCH, γ needs to be negative.)

A third type of GARCH model is the exponential GARCH (EGARCH). In this model the variance relationship is on the log scale and hence asymmetric.

$$\ln(h_t) = \alpha_0 + \sum_{i=1}^{q} \alpha_i z_{t-i} + \sum_{i=1}^{q} \phi_i(|z_{t-i}| - E[|z_{t-i}|]) + \sum_{j=1}^{p} \beta_i \ln\left(h_{t-j}\right),$$

where $z_t = \frac{\epsilon_t}{\sqrt{h_t}}$ and $E[|z_{t-i}|]$ denotes the expected value of $|z_{t-i}|$.

Note that the ϕ_i terms represent a symmetric contribution to the variance while the α_i terms give an asymmetric contribution.

Another common characteristic of financial data is that it is heavier in the tails (leptokurtic) that the Normal distribution. To model this the Normal distribution is replace by a scaled Student's t-distribution (that is a Student's t-distribution standardized to have variance h_t). The Student's t-distribution is such that the smaller the degrees of freedom the higher the kurtosis for degrees of freedom > 4.

2.7.2 Fitting GARCH models

The models are fitted by maximizing the conditional log-likelihood. For the Normal distribution the conditional log-likelihood is

$$\frac{1}{2}\sum_{i=1}^{T}\left(\log(h_i) + \frac{\epsilon_i^2}{h_i}\right).$$

For the Student's t-distribution the function is more complex. An approximation to the standard errors of the parameter estimates is computed from the Fisher information matrix.

3 Recommendations on Choice and Use of Available Functions

3.1 ARMA-type Models

ARMA-type modelling usually follows the methodology made popular by Box and Jenkins. It consists of four steps: identification, model fitting, model checking and forecasting. The availability of functions for each of these four steps is given below for the three types of modelling situation considered: univariate, input-output and multivariate.

3.1.1 Univariate series

(a) Model identification

The function g13au may be used in obtaining either a range-mean or standard deviation-mean plot for a series of observations, which may be useful in detecting the need for a variance-stabilizing transformation. g13au computes the range or standard deviation and the mean for successive groups of observations and g01ag may then be used to produce a scatter plot of range against mean or of standard deviation against mean.

The function g13aa may be used to difference a time series. The $N = n - d - s \times D$ values of the differenced time series which extends for $t = 1 + d + s \times D, \ldots, n$ are stored in the first N elements of the output array.

The function g13ab may be used for direct computation of the autocorrelations. It requires the time series as input, after optional differencing by g13aa.

An alternative is to use g13ca, which uses the FFT to carry out the convolution for computing the autocovariances. Circumstances in which this is recommended are

(i) if the main aim is to calculate the smoothed sample spectrum;

(ii) if the series length and maximum lag for the autocorrelations are both very large, in which case appreciable computing time may be saved.

For more precise recommendations, see Gentleman and Sande 1966. In this case the autocorrelations r_k need to be obtained from the autocovariances c_k by $r_k = c_k/c_0$.

The function g13ac computes the PACF and prediction error variance estimates from an input ACF. Note that g13dn, which is designed for multivariate time series, may also be used to compute the PACF together with χ^2 statistics and their significance levels.

Finite lag predictor coefficients are also computed by the function g13ac. It may have to be used twice, firstly with a large value for the maximum lag L in order to locate the optimum FPE lag, then again with L reset to this lag.

The function g13dx may be used to check that the AR part of the model is stationary and that the MA part is invertible.

(b) Model estimation

The function g13ad is used to compute preliminary estimates of the ARIMA model parameters, the sample autocorrelations of the appropriately differenced series being input. The model orders are required.

The main function for parameter estimation for ARIMA models is g13ae, and an easy-to-use version is g13af. Both these functions use the least-squares criterion of estimation.

In some circumstances the use of g13be or g13dc, which use ML, is recommended.

The functions require the time series values to be input, together with the ARIMA orders. Any differencing implied by the model is carried out internally. They also require the maximum number of iterations to be specified, and return the state set for use in forecasting.

g13ae should be preferred to g13af for:

(i) more information about the differenced series, its backforecasts and the intermediate series;

(ii) greater control over the output at successive iterations;

(iii) more detailed control over the search policy of the non-linear least-squares algorithm;

(iv) more information about the first and second derivatives of the objective function during and upon completion of the iterations.

g13be is primarily designed for estimating relationships between time series. It is, however, easily used in a univariate mode for ARIMA model estimation. The advantage is that it allows (optional) use of the exact likelihood estimation criterion, which is not available in g13ae or g13af. This is particularly recommended for models which have seasonal parameters, because it reduces the tendency of parameter estimates to become stuck at points on the parameter space boundary. The model parameters estimated in this function should be passed over to g13aj for use in univariate forecasting.

The function g13dc is primarily designed for fitting vector ARMA models to multivariate time series but may also be used in a univariate mode. It allows the use of either the exact or conditional likelihood estimation criterion, and allows you to fit non-multiplicative seasonal models which are not available in g13ae, g13af or g13be.

(c) Model checking

g13as calculates the correlations in the residuals from a model fitted by either g13ae or g13af. In addition the standard errors and correlations of the residual autocorrelations are computed along with a portmanteau test for model adequacy. g13as can be used after a univariate model has been fitted by g13be, but care must be taken in selecting the correct inputs to g13as. Note that if g13dc has been used to fit a non-multiplicative seasonal model to a univariate series then g13ds may be used to check the adequacy of the model.

(d) Forecasting using an ARIMA model

Given that the state set produced on estimation of the ARIMA model by either g13ae or g13af has been retained, g13ah can be used directly to construct forecasts for $x_{n+1}, x_{n+2}, \ldots$, together with probability limits. If some further observations $x_{n+1}, x_{n+2}, \ldots$ have come to hand since model estimation (and there is no desire to re-estimate the model using the extended series), then g13ag can be used to update the state set using the new observations, prior to forecasting from the end of the extended series. The original series is not required.

The function g13aj is provided for forecasting when the ARIMA model is known but the state set is unknown. For example, the model may have been estimated by a procedure other than the use of g13ae or g13af, such as g13be. g13aj constructs the state set and optionally constructs forecasts with probability limits. It is equivalent to a call to g13ae with zero iterations requested, followed by an optional call to g13ah, but it is much more efficient.

3.1.2 Input-output/transfer function modelling

(a) Model identification

Normally use g13bc for direct computation of cross-correlations, from which cross-covariances may be obtained by multiplying by $s_y s_x$, and impulse response estimates (after prewhitening) by multiplying by s_y/s_x, where s_y, s_x are the sample standard deviations of the series.

An alternative is to use g13cc, which exploits the FFT to carry out the convolution for computing cross-covariances. The criteria for this are the same as given in Section 3.1.1 for calculation of autocorrelations. The impulse response function may also be computed by spectral methods without prewhitening using g13cg.

g13ba may be used to prewhiten or filter a series by an ARIMA model.

g13bb may be used to filter a time series using a transfer function model.

(b) Estimation of input-output model parameters

The function g13bd is used to obtain preliminary estimates of transfer function model parameters. The model orders and an estimate of the impulse response function (see Section 3.2.1) are required.

The simultaneous estimation of the transfer function model parameters for the inputs, and ARIMA model parameters for the output, is carried out by g13be.

This function requires values of the output and input series, and the orders of all the models. Any differencing implied by the model is carried out internally.

The function also requires the maximum number of iterations to be specified, and returns the state set for use in forecasting.

(c) Input-output model checking

The function g13as, primarily designed for univariate time series, can be used to test the residuals from an input-output model.

(d) Forecasting using an input-output model

Given that the state set produced on estimation of the model by g13be has been retained, the function g13bh can be used directly to construct forecasts of the output series. Future values of the input series (possibly forecasts previously obtained using g13ah) are required.

If further observations of the output and input series have become available since model estimation (and there is no desire to re-estimate the model using the extended series) then g13bg can be used to update the state set using the new observations prior to forecasting from the end of the extended series. The original series are not required.

The function g13bj is provided for forecasting when the multi-input model is known, but the state set is unknown. The set of output and input series must be supplied to the function which then constructs the state set (for future use with g13bg and/or g13bh) and also optionally constructs forecasts of the output series in a similar manner to g13bh.

In constructing probability limits for the forecasts, it is possible to allow for the fact that future input series values may themselves have been calculated as forecasts using ARIMA models. Use of this option requires that these ARIMA models be supplied to the function.

(e) Filtering a time series using a transfer function model

The function for this purpose is g13bb.

3.1.3 Multivariate series

(a) Model identification

The function g13dl may be used to difference the series. You must supply the differencing parameters for each component of the multivariate series. The order of differencing for each individual component does not have to be the same. The function may also be used to apply a log or square root transformation to the components of the series.

The function g13dm may be used to calculate the sample cross-correlation or cross-covariance matrices. It requires a set of time series as input. You may request either the cross-covariances or cross-correlations.

The function g13dn computes the partial lag correlation matrices from the sample cross-correlation matrices computed by g13dm, and the function g13dp computes the least-squares estimates of the partial autoregression matrices and their standard errors. Both functions compute a series of χ^2 statistic that aid the determination of the order of a suitable autoregressive model. g13db may also be used in the identification of the order of an autoregressive model. The function computes multiple squared partial autocorrelations and predictive error variance ratios from the sample cross-correlations or cross-covariances computed by g13dm.

The function g13dx may be used to check that the autoregressive part of the model is stationary and that the moving-average part is invertible.

(b) Estimation of VARMA model parameters

The function for this purpose is g13dc. This function requires a set of time series to be input, together with values for p and q. You must also specify the maximum number of likelihood evaluations to be permitted and which parameters (if any) are to be held at their initial (user-supplied) values. The fitting criterion is either exact ML or conditional ML.

g13dc is primarily designed for estimating relationships between time series. It may, however, easily be used in univariate mode for non-seasonal and non-multiplicative seasonal ARIMA model estimation. The advantage is that it allows (optional) use of the exact ML estimation criterion, which is not available in either g13ae or g13af. The conditional likelihood option is recommended for those models in which the parameter estimates display a tendency to become stuck at points on the boundary of the parameter space. When one of the series is known to be influenced by all the others, but the others in turn are mutually independent and do not influence the output series, then g13be (the TF model fitting function) may be more appropriate to use.

(c) VARMA model checking

g13ds calculates the cross-correlation matrices of residuals for a model fitted by g13dc. In addition the standard errors and correlations of the residual correlation matrices are computed along with a portmanteau test for model adequacy.

(d) Forecasting using a VARMA model

The function g13dj may be used to construct a chosen number of forecasts using the model estimated by g13dc. The standard errors of the forecasts are also computed. A reference vector is set up by g13dj so that should any further observations become available the existing forecasts can be efficiently updated using g13dk. On a call to g13dk the reference vector itself is also updated so that g13dk may be called again each time new observations are available.

3.2 Spectral Methods

3.2.1 Univariate spectral estimation

Two functions are available, g13ca carrying out smoothing using a lag window and g13cb carrying out direct frequency domain smoothing. Both can take as input the original series, but g13ca alone can use the sample autocovariances as alternative input. This has some computational advantage if a variety of spectral estimates needs to be examined for the same series using different amounts of smoothing.

However, the real choice in most cases will be which of the four shapes of lag window in g13ca to use, or whether to use the trapezium frequency window of g13cb. The references may be consulted for advice on this, but the two most recommended lag windows are the Tukey and Parzen. The Tukey window has a very small risk of supplying negative spectrum estimates; otherwise, for the same bandwidth, both give very similar results, though the Parzen window requires a higher truncation lag (more ACF values).

The frequency window smoothing procedure of g13cb with a trapezium shape parameter $p \simeq \frac{1}{2}$ generally gives similar results for the same bandwidth as lag window methods with a slight advantage of somewhat less distortion around sharp peaks, but suffering a rather less smooth appearance in fine detail.

3.2.2 Cross-spectrum estimation

Two functions are available for the main step in cross-spectral analysis. To compute the cospectrum and quadrature spectrum estimates using smoothing by a lag window, g13cc should be used. It takes as input either the original series or cross-covariances which may be computed in a previous call of the same function or possibly using results from g13bc. As in the univariate case, this gives some advantage if estimates for the same series are to be computed with different amounts of smoothing.

The choice of window shape will be determined as the same as that which has already been used in univariate spectrum estimation for the series.

For direct frequency domain smoothing, g13cd should be used, with similar consideration for the univariate estimation in choice of degree of smoothing.

The cross-amplitude and squared coherency spectrum estimates are calculated, together with upper and lower confidence bounds, using g13ce. For input the cross-spectral estimates from either g13cc or g13cd and corresponding univariate spectra from either g13ca or g13cb are required.

The gain and phase spectrum estimates are calculated together with upper and lower confidence bounds using g13cf. The required input is as for g13ce above.

The noise spectrum estimates and impulse response function estimates are calculated together with multiplying factors for confidence limits on the former, and the standard error for the latter, using g13cg. The required input is again the same as for g13ce above.

3.3 Kalman Filtering

3.4 GARCH Models

The main choice in selecting a type of GARCH model is whether the data is symmetric or asymmetric and if asymmetric what form of asymmetry should be included in the model.

A symmetric ARCH or GARCH model can be fitted by g13fa and the volatility forecast by g13fb. For asymmetric data the choice is between the type of asymmetry as described in Section 2.7.

GARCH Type	**Fit**	**Forecast**
Type I	g13fa	g13fb
Type II	g13fc	g13fd
GJR	g13fe	g13ff
EGARCH	g13fg	g13fh

All functions allow the option of including regressor variables in the model and the choice between Normal and Student's t-distribution for the errors.

3.5 Time Series Simulation

There are functions available in Chapter G05 for generating a realisation of a time series from a specified model: g05pa for univariate time series and g05pc for multivariate time series. There is also a suite of functions for simulating GARCH models: g05hk, g05hl , g05hm and g05hn.

3.6 Summary of Recommendations

4 References

Akaike H 1971 Autoregressive model fitting for control *Ann. Inst. Statist. Math.* **23** 163–180

Bollerslev T 1986 Generalised autoregressive conditional heteroskedasticity *Journal of Econometrics* **31** 307–327

Box G E P and Jenkins G M 1976 *Time Series Analysis: Forecasting and Control* (Revised Edition) Holden–Day

Engle R 1982 Autoregressive conditional heteroskedasticity with estimates of the variance of United Kingdom inflation *Econometrica* **50** 987–1008

Gentleman W S and Sande G 1966 Fast Fourier transforms for fun and profit *Proc. Joint Computer Conference, AFIPS* **29** 563–578

Glosten L, Jagannathan R and Runkle D 1993 Relationship between the expected value and the volatility of nominal excess return on stocks *Journal of Finance* **48** 1779–1801

Hamilton J 1994 *Time Series Analysis* Princeton University Press

Heyse J F and Wei W W S 1985 The partial lag autocorrelation function *Technical Report No. 32* Department of Statistics, Temple University, Philadelphia

Tiao G C and Box G E P 1981 Modelling multiple time series with applications *J. Am. Stat. Assoc.* **76** 802–816

Wei W W S 1990 *Time Series Analysis: Univariate and Multivariate Methods* Addison–Wesley