Chapter 29
Time Series Analysis

1 Scope of the Chapter

This chapter provides procedures for the analysis of time series data.

2 Available Modules

Module 29.1: `nag_tsa_identify` — **Time series analysis**

This module contains procedures for calculating the sample autocorrelations of a univariate time series. Procedures are provided for

- calculating the autocorrelation function;
- calculating the partial autocorrelation function.

Module 29.2: `nag_tsa_kalman` — **Kalman filtering**

This module contains procedures for the Kalman filtering of time series data. The procedures are based on the square-root algorithm. Facilities are provided for

- an initial estimate for the state covariance matrix;
- a prediction step of the square-root covariance Kalman filter;
- a combined update-prediction step, using the time-varying square-root covariance Kalman filter;
- a combined update-prediction step, using the time-invariant square-root covariance Kalman filter.

Module 29.3: `nag_tsa_spectral` — **Time Series Spectral Analysis**

This module contains procedures for calculating the smoothed sample spectrum of a univariate and bivariate time series. Procedures are provided for

- calculating the smoothed sample spectrum of a univariate time series;
- calculating the smoothed sample spectrum of a univariate time series using autocovariances;
- calculating the smoothed sample cross spectrum of a bivariate time series;
- calculating the smoothed sample cross spectrum of a bivariate time series using autocovariances;
- calculating the squared coherency, the cross amplitude, the gain and the phase spectra;
- calculating the noise spectrum and the impulse response function from a linear system.

3 Background

3.1 Introduction

Time series data, x_t, for $t = 1, 2, \ldots, n$, generally consist of both *deterministic* and *stochastic* components. The deterministic component gives rise to trends, seasonal patterns and cycles, while the stochastic component causes statistical fluctuations which have a short term correlation structure.

A time series is *stationary* if the structure of the series depends only the relative position of the observations; that is, the joint distribution of x_t and x_{t+l} depends only on l not on t. If we define $E(x_t) = \mu_t$, $\text{var}(x_t) = E(x_t - \mu_t)^2$, and $\text{cov}(x_t, x_{t-l}) = \gamma(t, t-l) = E\left((x_t - \mu_t)(x_{t-l} - \mu_{t-l})\right)$, then a series is *second-order* stationary if $\mu_t = \mu$ for all t and $\gamma(t, t-l) = \gamma_l$ for all t and l.

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.

There are two basic approaches to analysing time series, the time domain and the frequency domain. In the time domain the relationship of an observation at time t to observations at previous time points is examined and modelled using, for example, an ARMA model or Kalman filter model. In the case of the frequency domain approach the sinusoidal components across the whole series are examined using spectral analysis.

3.2 ARMA Time Series Models

A stationary time series may often be modelled as an auto-regressive moving average (ARMA) process. For the univariate time series x_t, for $t = 1, 2, \ldots, n$, with mean μ, an ARMA(p, q) model is

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

where the ϕ are the p auto-regressive parameters, the θ are the q moving average parameters, $w_t = x_t - \mu$ and ϵ_t is an uncorrelated sequence of noise with zero mean and variance σ_ϵ^2.
If we define the back shift operator $B^j x_t = x_{t-j}$ then the above equation can be written more compactly as

$$\phi_p(B) w_t = \theta_q(B) \epsilon_t,$$

where the auto-regressive operator $\phi_p(B)$ is

$$\phi_p(B) = 1 - \phi_1 B - \phi_2 B^2 \cdots - \phi_p B^p,$$

and the moving average operator $\theta_q(B)$ is

$$\theta_q(B) = 1 - \theta_1 B - \theta_2 B^2 \cdots - \theta_q B^q.$$

For the process to be stationary the roots of the polynomial equation $\phi_p(z) = 0$ must lie outside the unit circle. A second important property is invertibility; this allows the process to be written as a pure autoregressive process and avoids model multiplicities. For the process to be invertible we require that the roots of the polynomial equation $\theta_q(z) = 0$ lie outside the unit circle.

The advantage of an ARMA model over a pure AR model is that a complex series can often be represented using an ARMA model with fewer parameters than would be needed by a suitable AR model.

A k-variate ARMA(p, q) can be represented using similar equations to those given above, except that w_t and ϵ_t are now vectors of length k and the ϕ and θ are now $k \times k$ matrices.

Seasonal, SARMA(p, q, P, Q, s), models also allow for correlation at lags which are multiples of seasonal period s. In this case the model can be written as

$$\Phi_P(B^s) \phi_p(B) w_t = \Theta_Q(B^s) \theta_q(B) \epsilon_t,$$

where the seasonal auto-regressive operator $\Phi_P(B)$ is

$$\Phi_P(B^s) = 1 - \Phi_1 B^s - \Phi_2 B^{2s} \cdots - \Phi_P B^{Ps},$$

and the seasonal moving average operator $\Theta_Q(B^s)$ is

$$\Theta_Q(B^s) = 1 - \Theta_1 B^s - \Theta_2 B^{2s} \cdots - \Theta_Q B^{Qs}.$$

For the SARMA process to be stationary the roots of the polynomial equation $\Phi_P(z)\phi_p(z) = 0$ must lie outside the unit circle, while for the process to be invertible the roots of the polynomial equation $\Theta_Q(z)\theta_q(B) = 0$ must lie outside the unit circle.

3.3 ARIMA Time Series Models

In many cases while a time series w_t may be non-stationary the differenced series $w_t - w_{t-1}$ may be stationary. Single differencing will remove linear trends, higher order and seasonal differencing may remove more complex trends and cycles.

If the time series w_t $(t = 1, 2, \ldots, n)$ is non-stationary and the differenced series follows an ARMA(p, q) then the series is termed an ARIMA(p, d, q) and can be written as

$$\phi_p(B)(1 - B)^d w_t = \theta_q(B)\epsilon_t,$$

where $(1 - B)^d$ is the dth-order differencing operator, i.e.,

$$(1 - B)w_t = w_t - w_{t-1}, \quad (1 - B)^2 w_t = w_t - 2w_{t-1} + w_{t-2}, \text{ etc.}$$

In the case of a seasonal series it can be reduced to a SARMA(p, q, P, Q, s) by using seasonal differencing, with the model being written as follows:

$$\Phi_P(B^s)\phi_p(B)(1 - B^s)^D(1 - B)^d w_t = \Theta_Q(B^s)\theta_q(B)\epsilon_t,$$

where $(1 - B^s)^D$ is the Dth-order seasonal differencing operator, i.e.,

$$(1 - B^s)w_t = w_t - w_{t-s}, \quad (1 - B^s)^2 w_t = w_t - 2w_{t-s} + w_{t-2s}, \text{ etc.}$$

3.4 Structural Models

Structural models represent a time series directly in terms of trend, seasonal and irregular components and thus offer an alternative to ARIMA models.

The basic structural model for the univariate time series y_t, for $t = 1, 2, \ldots, T$ has the form

$$y_t = \mu_t + \gamma_t + \epsilon_t$$

where μ_t, γ_t and ϵ_t are respectively the trend, seasonal and irregular components.
The equations for the process generating the trend are

$$\text{level:} \quad \mu_t = \mu_{t-1} + \beta_{t-1} + \eta_t, \quad t = 1, 2, \ldots, T$$

and

$$\text{slope:} \quad \beta_t = \beta_{t-1} + \zeta_t, \quad t = 1, 2, \ldots, T$$

where η_t and ζ_t are normally distributed white noise processes with zero means and variances σ_η^2 and σ_ζ^2, respectively. The effect of η_t is to allow the level of the trend to shift up and down, while ζ_t allows the slope to change.

If the seasonal components were deterministic then, at time t, they would have to satisfy

$$\sum_{j=0}^{s-1} \gamma_{t-j} = 0$$

where s is the number of 'seasons' in the 'year'. The introduction of a random component, $\omega_t = N(0, \sigma_\omega^2)$ on the right-hand side,

$$\sum_{j=0}^{s-1} \gamma_{t-j} = \omega_t,$$

allows the seasonal effect to change over time, but still ensures that the sum of any s consecutive seasonal components has an expected value of zero. A recursive formula for the seasonal components is

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

The above is called the dummy variable form of seasonality.

An alternative way of modelling seasonality is by a set of trigonometric terms at the seasonal frequencies, $\lambda_j = 2\pi j/s$, for $j = 1, 2, \ldots, s/2$. The lowest frequency, $2\pi/s$, is known as the fundamental frequency while the remaining frequencies are harmonics. If, as above, a random component is added to allow the season effects to change over time but still have an expected value of zero, the following recursive formula for the seasonal effect at time t is obtained:

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

where

$$\left.\begin{array}{rcl} \gamma_{j,t} & = & \gamma_{j,t-1}\cos\lambda_j + \gamma^*_{j,t-1}\sin\lambda_j + \omega_{j,t} \\ \gamma^*_{j,t} & = & -\gamma_{j,t-1}\sin\lambda_j + \gamma^*_{j,t-1}\cos\lambda_j + \omega^*_{j,t} \end{array}\right\} \quad j = 1, 2, \ldots, s/2,$$

where $\omega_{j,t}$ and $\omega^*_{j,t}$ are zero mean white noise processes which are uncorrelated and have a common variance σ^2_j. Assigning different variances to each harmonic allows them to evolve at varying rates. However, it is usually desirable to let $\mathrm{var}(\omega^*_{j,t}) = \mathrm{var}(\omega_{j,t}) = \sigma^2_j = \sigma^2_\omega$, for $j = 1, 2, \ldots, s/2$. When s is even and $j = s/2$, the sine term vanishes giving $\gamma_{j,t} = \gamma_{j,t-1}\cos\lambda_j + \omega_{j,t}$. For even s the number of parameters, γ_j and γ^*_j, is therefore always $s - 1$, which is the same as the number of coefficients in the seasonal dummy form. Provided that the full set of trigonometric terms is included, the trigonometric and dummy variable approach should give identical results. However, the advantage of the trigonometric approach is that it allows unimportant frequency components to be dropped and thus enables more compact models to be constructed.

3.5 State Space Models

A general class of time series models is the state space model. This comprises two equations, the state equation that models the unobserved stochastic state vector, X_t,

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

and the measurement equation,

$$Y_t = C_t X_t + V_t$$

for the observed Y_t. 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. The state noise and the measurement noise are assumed to be uncorrelated and have zero mean, and the covariance matrices are

$$\mathrm{cov}(W_t) = Q_t \text{ and } \mathrm{cov}(V_t) = R_t.$$

With state space models the primary objective of the analysis is to estimate the value of the state vector, X_t, given the observed $Y_1 \ldots Y_t$ and to forecast X_{t+1}. This is achieved as a recursive operation using the *Kalman Filter*. The current estimate of X_t given observations $Y_1 \ldots Y_{t-1}$, denoted by $\hat{X}_{t|t-1}$, is updated by the Y_t observation to give the estimate $\hat{X}_{t|t}$. This operation can be seen as being within a Bayesian framework.

Both the ARMA and the structural models described above can be written as state space models. The matrices A_t and B_t are then functions of the parameters of the model.

3.6 Model Identification

Model identification is the stage in the modelling process in which the form of the model that is thought to be suitable is selected. Identification can either be conceptual, based on theoretical knowledge of the system, or empirical, making use of data observed on the system. Because of the generality of state space models any direct model identification would have to be, in the first instance, conceptual. For ARIMA models empirical identification is possible. The basic tools for ARIMA model identification are data plot, the *autocorrelation function (acf)* and the *partial autocorrelation function (pacf)*. A simple plot of

the series will indicate if the data has to be differenced or transformed in order to produce a stationary series. Given a stationary series the autocorrelation function is the correlation of observations l apart for $l = 1, 2, \ldots$, that is an estimate of $\rho_l = \gamma_l/\gamma_0$. For pure MA processes of order q the autocorrelations should be zero for $l > q$. The partial autocorrelation is a measure of the correlation between observations l apart having removed the effect of the correlation due to the intervening observations. For a pure AR process of order p the partial autocorrelations should be zero for $l > p$. Thus examination of plots of the acf and pacf should indicate if a MA or AR model is suitable or if a more complex ARMA model is needed.

3.7 Model Fitting

Given a specified form of a model in terms of unknown parameters, θ_i, ϕ_i, etc., the model can be fitted by estimating the parameters using maximum likelihood. If the model is written as a state space model the likelihood can be computed by making use of a Kalman filter. The Kalman filter computes an estimate of the state vector X_t and its covariance matrix P_t given observations $Y_1 \ldots Y_t$ and values of the model parameters. As a by-product one-step-ahead residuals, r_t and their covariance matrix, H_t are also computed; from these the log-likelihood for the unknown parameters can be computed as

$$\kappa - \frac{1}{2}\sum_{i=1}^{t} \ln(\det(H_i)) - \frac{1}{2}\sum_{i=1}^{t} r_i^T H_i^{-1} r_i,$$

where κ is a constant. This likelihood can then be maximized using a non-linear optimizer.

3.8 Model Checking

In addition to testing the significance of the estimated parameters by using likelihood ratio tests or z-test, the fit of the model can be checked by examining the one-step-ahead residuals. These residuals should be independently and identically distributed Normal variates. The independence can be checked by computing the acf of the residuals, and the Normality can be checked using graphical methods.

3.9 Forecasting

Given a fitted state-space model, forecasts can be computed by using

$$\hat{X}_{t+i|t} = A_{t+i-1}\hat{X}_{t+i-1|i} \qquad i = 1, 2, \ldots,$$

where $\hat{X}_{t+i|t}$ is the estimate of the state vector X_{t+i} given observation up to time t. The covariance matrix, $P_{t+i|t}$, of the forecasts can be computed from

$$P_{t+i|t} = A_{t+i-1}P_{t+i-1|t}A_{t+i-1|t}^T + B_i Q_i B_i^T,$$

where $P_{t+1|t}$ is the state covariance matrix at $t+1$ for observations up to time t and is computed by the Kalman filter along with $\hat{X}_{t+1|t}$.

3.10 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 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 ω is given to order $1/n$ by

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

For a series $x_1, x_2, \ldots, x_n$ the sample spectrum 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 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 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.
The smoothed estimate of the spectrum can be calculated from autocovariances or from the unsmoothed sample spectrum.

3.11 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

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

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 cross-covariance function (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.

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:

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

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

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

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 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

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

From these representations of the estimated frequency response $\hat{V}(\omega)$, parametric transfer function (TF) models may be recognised 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 impulse response function (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 $k = -1, -2$ are needed (their presence might indicate feedback between the series).