INSTITUTE OF ACTUARIAL & QUANTITATIVE STUDIES

Class: MSc

Subject : Statistical and Risk Modelling - 4

Chapter: Unit 4 (Entire)

Chapter Name: Times series 2

Today’s Agenda

1. Achieving Stationarity
2. Identification of Time Series Processes
3. Fitting a Time series model using Box-Jenkin’s Methodology
4. Forecasting
5. Other Time series processes

1.1 Compensating for trend and seasonality

In this section, we deal with possible sources of non-stationarity and how to compensate for them.

Lack of stationarity may be caused by the presence of deterministic effects in the quantity being observed.

We can identify three possible causes of non-stationarity:

1. a deterministic trend (eg exponential or linear growth)
2. a deterministic cycle (eg seasonal effect)
3. the time series is integrated.

It is worth pointing out that this list is not exhaustive.

1.2 Detecting non-stationary series

- The most useful tools in identifying non-stationarity are the simplest: a plot of the series against t , and the sample ACF.

- Plotting the series will highlight any obvious trends in the mean and will show up any cyclic variation which could also form evidence of non-stationarity. This should always be the first step in any practical time series analysis.

- The sample ACF should, in the case of a stationary time series, ultimately converge towards zero exponentially fast,

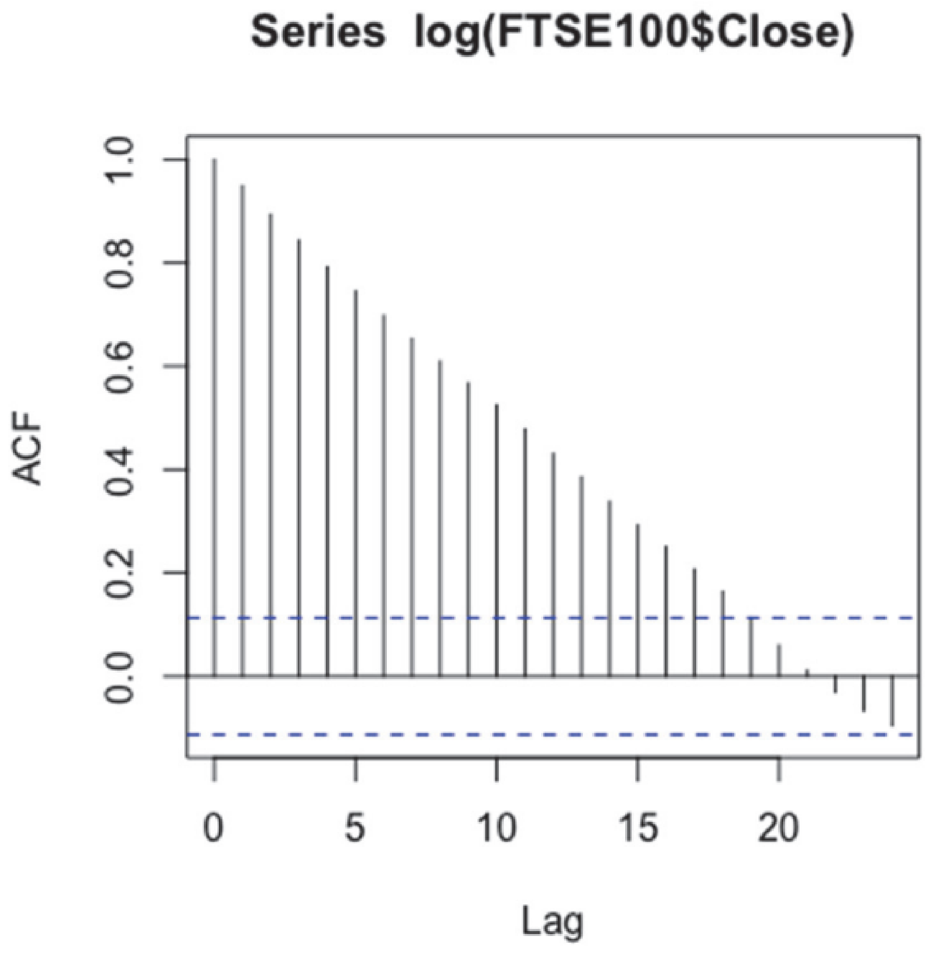
- If the sample ACF decreases slowly but steadily from a value near 1, we would conclude that the data need to be differenced before fitting the model. If the sample ACF exhibits a periodic oscillation, however, it would be reasonable to conclude that there is some underlying cause of the variation.

1.2 Detecting non-stationary series

Example

Below is the ACF plotted for FTSE 100 Index.

The sample ACF is clearly non-stationary as the **values decrease in some linear fashion;** differencing is therefore required before fitting a stationary model.



1.3 Least squares trend removal

The simplest way to remove a linear trend is by ordinary least squares. This is equivalent to fitting the model:

$$\boldsymbol{x_t = a + bt + y_t}$$

where a and b are constants and y is a zero-mean stationary process. The parameters a and b can be estimated by linear regression prior to fitting a stationary model to the residuals y_t.

Differencing

Differencing may well be beneficial if the sample ACF decreases slowly from a value near 1 but has useful effects in other instances as well. If, for instance, $x_t = a + bt + y_t$, then:

$$\boldsymbol{\nabla x_t = b + \nabla y_t}$$

so that the differencing has removed the trend in the mean.

1.4 Seasonal Differencing

Where seasonal variation is present in the data, one way of removing it is to take a seasonal difference.

Example 1
Suppose that the time series x records the monthly average temperature in London. A model of the form:

$$x_t = \mu + \theta_t + y_t$$

might be applied, where θ is a periodic function with period 12 and y is a stationary series.

Then the seasonal difference of x is defined as $(\nabla_{12}x)_t = x_t - x_{t-12}$ and we see that:

$$(\nabla_{12}x)_t = x_t - x_{t-12} = (\mu + \theta_t + y_t) - (\mu + \theta_{t-12} + y_{t-12}) = y_t - y_{t-12}$$

is a stationary process.

We can then model the seasonal difference of x as a stationary process and reconstruct the original process x itself afterwards.

1.5 Method of moving averages

The method of moving averages makes use of a simple linear filter to eliminate the effects of periodic variation.

A linear filter is a transformation of a time series x (the input series) to create an output series y that satisfies:

$$y_t = \sum_{k=-\infty}^{\infty} a_k \, x_{t-k}$$

1.6 Method of seasonal means

The simplest method for removing seasonal variation is to subtract from each observation the estimated mean for that period, obtained by simply averaging the corresponding observations in the sample.

Suppose that the time series x records the monthly average temperature in London. A model of the form:

$$x_t = \mu + \theta_t + y_t$$

might be applied, where θ is a periodic function with period 12 and y is a stationary series. The term θ_t contains the deviation of the model at time t due to the seasonal effect. So:

$$y_t = x_t - \mu - \theta_t$$

When fitting the model to a monthly time series x extending over 10 years from January 1990 the estimate for μ is $\bar{x}$ and the estimate for θ_{January} is:

$$\hat{\theta}_{\text{January}} = \frac{1}{10}(x_1 + x_{13} + x_{25} + \cdots + x_{109}) - \hat{\mu}$$

So, in this case, we can remove the seasonal variation by deducting the January average, $\bar{x}_{\text{January}} = \frac{1}{10}(x_1 + x_{13} + x_{25} + \cdots + x_{109})$, from all the January values.

1.7 Transformation of the data

Diagnostic procedures such as an inspection of a plot of the residuals may suggest that even the best-fitting standard linear time series model is failing to provide an adequate fit to the data. Before attempting to use more advanced non-linear models, it is often worth attempting to transform the data in some straightforward way in an attempt to find a data set on which the linear theory will work properly.

1.7 Transformation of the data

Variance-stabilising transformations

Transformations are most commonly used when a dependence is suspected between the variance of the residuals and the size of the fitted values. If, for example, the standard deviation of $X_{t+1} - X_t$ appears to be proportional to X_t, then it would be appropriate to use the logarithmic transformation, to work on the time series $Y = \ln X$.

1.7 Transformation of the data

Transformations to increase normality

In certain applications it may be found that most residuals are small and negative, with a few large positive values to offset them. This may be taken to indicate that the distribution of the error terms is non-normal, leading to doubts as to whether the standard time series procedures, designed for normal errors, are applicable. It may be possible to find a transformation which will improve the normality of the error terms of the transformed process, but care should be taken that this does not lead to instability in the variance. A further caution when using transformed data involves the final step of turning forecasts for the transformed process into forecasts for the original process, as some transformations introduce a systematic bias.

Question

CT6 April 2015 Q7

The following time series model is being used to model monthly data:

$$Y_t = Y_{t-1} + Y_{t-12} - Y_{t-13} + e_t + \beta_1 e_{t-1} + \beta_{12} e_{t-12} + \beta_1 \beta_{12} e_{t-13}$$

where e_t is a white noise process with variance σ^2.

(i) Perform two differencing transformations and show that the result is a moving average process which you may assume to be stationary. [3]

(ii) Explain why this transformation is called seasonal differencing. [1]

(iii) Derive the auto-correlation function of the model generated in part (i). [8]

Solution

(i) Set $X_t = (1 - B^{12})(1 - B)\ Y_t$ where B is the background shift operator

i.e. $X_t = Y_t - Y_{t-1} - Y_{t-12} + Y_{t-13}$

then we have $X_t = e_t + \beta_1 e_{t-1} + \beta_{12} e_{t-12} + \beta_1\beta_{12} e_{t-13}$

$$= (1 + \beta_1 B)(1 + \beta_{12} B^{12}) e_t$$

which is a moving average process [of order 13].

(ii) This is called seasonal differencing because it compares the monthly change in Y_t with the corresponding monthly change at the same time last year.

Solution

(iii) We can see that

$$\gamma_0 = \text{Cov}(X_t, X_t) = (1+\beta_1^2+\beta_{12}^2+\beta_1^2\beta_{12}^2)\sigma^2 = (1+\beta_1^2)(1+\beta_{12}^2)\sigma^2$$

$$\begin{aligned}\gamma_1 = \text{Cov}(X_t, X_{t-1}) &= \text{Cov}(e_t + \beta_1 e_{t-1} + \beta_{12} e_{t-12} + \beta_1\beta_{12} e_{t-13};\\ &\qquad e_{t-1} + \beta_1 e_{t-2} + \beta_{12} e_{t-13} + \beta_1\beta_{12} e_{t-14})\\ &= (\beta_1+\beta_1\beta_{12}^2)\sigma^2 = \beta_1(1+\beta_{12}^2)\sigma^2\end{aligned}$$

$$\begin{aligned}\gamma_{11} = \text{Cov}(X_t, X_{t-11}) &= \text{Cov}(e_t + \beta_1 e_{t-1} + \beta_{12} e_{t-12} + \beta_1\beta_{12} e_{t-13};\\ &\qquad e_{t-11} + \beta_1 e_{t-12} + \beta_{12} e_{t-23} + \beta_1\beta_{12} e_{t-24})\\ &= \beta_1\beta_{12}\ \sigma^2\end{aligned}$$

$$\gamma_{12} = (\beta_{12}+\beta_1^2\beta_{12})\sigma^2 = \beta_{12}(1+\beta_1^2)\sigma^2$$

$$\gamma_{13} = \beta_1\beta_{12}\ \sigma^2$$

Solution

and $\gamma_s = 0$ for all other values of s.

so $$\rho_1 = \frac{\beta_1(1+\beta_{12}^2)}{(1+\beta_1^2)(1+\beta_{12}^2)} = \frac{\beta_1}{1+\beta_1^2}$$

$$\rho_{11} = \rho_{13} = \frac{\beta_1\beta_{12}}{(1+\beta_1^2)(1+\beta_{12}^2)}$$

$$\rho_{12} = \frac{\beta_{12}(1+\beta_1^2)}{(1+\beta_1^2)(1+\beta_{12}^2)} = \frac{\beta_{12}}{1+\beta_{12}^2}$$

and $\rho_0 = 1$ and $\rho_s = 0$ for all other s.

2 Identification of Time Series Processes

2.1 Identification of MA(q)

The distinguishing characteristic of $MA(q)$ is that $\rho_k = 0$ for all $k > q$. A test for the appropriateness of an $MA(q)$ model, therefore, is that r_k is close to 0 for all $k > q$.

1. Plot r_k vs k
2. Asymptotic distribution for r_k(for $k > q$)

$$\rho_{\mathrm{k}} \sim \mathrm{N}\left(0,1/\mathrm{n}\left(1+2*\sum_{i=1}^{q}\rho_{\mathrm{i}}^{2}\right)\right)$$

Confidence Interval for r_k

$$= \pm \mathrm{z}_{\alpha/2}\sqrt{1/\mathrm{n}}\left(1+2*\sum_{i=1}^{q}\rho_{\mathrm{i}}^{2}\right.$$

2 Identification of Time Series Processes

2.2 Identification of AR(p)

The corresponding diagnostic procedure for an autoregressive model is based on the sample partial ACF, since the PACF of an AR (p) is distinctive, being equal to zero for k > p.

1. Plot $\phi_k^{\wedge}$ vs k
2. Asymptotic distribution for $\phi_k^{\wedge}$(for $k > p$)

$$\phi_\mathrm{k} \sim \mathrm{N}(0,1/\mathrm{n})$$

Confidence Interval for ϕ_k

$$= \pm \mathrm{z}_{\alpha/2}\sqrt{1/n} \approx \pm 2/\sqrt{n}$$

2 Identification of Time Series Processes

2.3 Identification of WNP (White Noise Process) (Method 1)

1. Calculate e_t values using Backward forecasting
2. Plot e_t against time
3. Asymptotic distribution for

$$\rho_{\mathrm{k}} \sim N(0,1/n)$$
$$\phi_{\mathrm{k}} \sim N(0,1/n)$$

Confidence Interval for both

$$= \pm z_{\alpha/2}\sqrt{1/n} \approx \pm 2/\sqrt{n}$$

2 Identification of Time Series Processes

2.3 Identification of WNP (White Noise Process) (Method 2)

Portmanteau test by Ljung & Box

- Overall goodness of fit test
- $\mathrm{H_0}$ - residuals are WNP v/s $\mathrm{H_1} -$ not $\mathrm{H_0}$
- Model for WNP

$$X_t = \mu + e_t$$

$$\text{Test static} = n(n+2)\sum_{k=1}^{m} \frac{r_k^2}{n-k} \sim \chi_m^2$$

for each m

- Decision criteria : Reject $\mathrm{H_0}$ at $\alpha\%$ level of significance if

$$\chi^2{}_{\mathrm{cal}} > \chi^2{}_{\mathrm{tab}}$$

This is a one-sided test. A large test statistic indicates that the data do not confirm to a white noise process.

3 Fitting a time series model using the Box-Jenkins methodology

The Box-Jenkins approach allows one to find an ARIMA model which is reasonably simple and provides a sufficiently accurate description of the behavior of the historical data.

Main steps in the Box-Jenkins approach to modelling

The main steps of the approach are:

- tentative identification of a model from the ARIMA class
- estimation of parameters in the identified model
- diagnostic checks.

If the tentatively identified model passes the diagnostic tests, the model is ready to be used for forecasting. If it does not, the diagnostic tests should indicate how the model ought to be modified, and a new cycle of identification, estimation and diagnosis is performed.

3 Fitting a time series model using the Box-Jenkins methodology

1. Tentative identification of a model from the ARIMA class

An ARIMA (p,d,q) model is completely identified by the choice of non-negative integer values for the parameters p , d and q . The parameter d is the number of times we have to difference the time series x to convert it to some stationary level.

3 Fitting a time series model using the Box-Jenkins methodology

The following principles can be used to choose the appropriate value of d :

1. A time series x can be modelled by a stationary ARMA model if the sample autocorrelation function r_k decays rapidly to zero with k. If, on the other hand, a slowly decaying positive sample autocorrelation function r_k is observed, this should be taken to indicate that the time series needs to be differenced to convert it into a likely realisation of a stationary random process.

2. Let $\hat{\sigma}_d^2$ denote the sample variance of the process $z^{(d)} = \nabla^d x$, ie the sample variance of the data values after they have been differenced d times. It is normally the case that $\hat{\sigma}_d^2$ first decreases with d until stationarity is achieved and then starts to increase. Therefore, d can be set to the value which minimises $\hat{\sigma}_d^2$. This could be $d = 0$ if the original time series x is already stationary.

3 Fitting a time series model using the Box-Jenkins methodology

Fitting an ARMA(p, q) model

Suppose now that the appropriate value for the parameter d has been found, and the time series $\{z_{d+1}, z_{d+2}, \dots, z_n\}$ is adequately stationary. (Notice that a differenced series has d fewer observation than the original series.) We shall assume throughout this section that the sample mean of the z sequence is zero; if this is not the case, obtain a new sequence by subtracting $\hat{\mu} = \bar{\mathbf{z}}$ from each value in the sequence. We shall also assume, for the sake of simplicity in setting down the lower and upper limits of sums, that $d = 0$.

In the framework of the Box-Jenkins approach we try to find an $\mathrm{ARMA}(p, q)$ model which fits the data z.

If either the correlogram or the partial correlogram appears to be close to zero for sufficiently large k , an MA (q) or AR (p) model is indicated. Otherwise, we should look for an ARMA(p, q) model with non-zero values of p and q .

3 Fitting a time series model using the Box-Jenkins methodology

Fitting an ARMA(p, q) model

A good indicator for possible values of p and q in an ARMA p q (,) is the number of spikes in the ACF and PACF until some geometrical decay to zero is observed.

For more complex models, we perform a trail and error method:

Eg - ARMA(1,1)
- ARMA(2,1)
- ARMA(1,2)

Every additional parameter improves the fit of the model by reducing the residual sum of squares. Taking this to extremes, a model with n parameters could be found to fit the data exactly.

? How will take care of overfitting the model?

3 Fitting a time series model using the Box-Jenkins methodology

Fitting an ARMA(p, q) model

The question of when to stop adding new parameters is addressed by Akaike's information criterion (AIC), which states that we should only consider adding an extra parameter if this results in a reduction of the residu sum of squares by a factor of at least $e^{-2/n}$, or alternatively, one can evaluate for each possible model the value of:

$$AIC(\text{ model }) = \log(\hat{\sigma}^2) + 2 \times \frac{\text{number of parameters}}{n}$$

and choose as the most appropriate the one corresponding to the lowest such value.

3 Fitting a time series model using the Box-Jenkins methodology

2. Parameter Estimation

Once the values of p and q have been identified, the problem becomes to estimate the values of parameters $\alpha_1, \alpha_2, \ldots, \alpha_p$ and $\beta_1, \beta_2, \ldots, \beta_q$ for the $\mathrm{ARMA}(p, q)$ model:

$$z_t = \alpha_1 z_{t-1} + \alpha_2 z_{t-2} + \cdots + \alpha_p z_{t-p} + e_t + \beta_1 e_{t-1} + \beta_2 e_{t-2} + \cdots + \beta_q e_{t-q}$$

Least squares estimation suggests itself; this is equivalent to maximum likelihood estimation if the e_t may be assumed normally distributed.

3 Fitting a time series model using the Box-Jenkins methodology

2. Parameter Estimation

In the case of a more general ARMA process we encounter the difficulty that the e_t cannot be deduced from the z_t. For example, in the case of $\mathrm{ARMA}(1,1)$ we have:

$$e_t = z_t - \alpha_1 z_{t-1} - \beta_1 e_{t-1}$$

an equation which can be solved iteratively for e_t as long as some starting value e_0 is assumed. For an $\mathrm{ARMA}(p, q)$ the list of starting values is $(e_0, \dots, e_{q-1})$.

The starting values need to be estimated, which is usually carried out by a recursive technique. First assume they are all equal to zero and estimate the α_i and β_j on that basis, then use standard forecasting techniques on the time-reversed process $\{z_n, \dots, z_1\}$ to obtain predicted values for $(e_0, \dots, e_{q-1})$, a method known as backforecasting. These new values can be used as the starting point for another application of the estimation procedure; this continues until the estimates have converged.

This is done by software packages.

3 Fitting a time series model using the Box-Jenkins methodology

2. Parameter Estimation

The final parameter of the model is σ^2, the variance of the e_t, which may be estimated using:

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{t=p+1}^{n} \hat{e}_t^2 = \frac{1}{n}\sum_{t=p+1}^{n} \left(z_t - \hat{\alpha}_1 z_{t-1} - \cdots - \hat{\alpha}_p z_{t-p} - \hat{\beta}_1 \hat{e}_{t-1} - \cdots - \hat{\beta}_q \hat{e}_{t-q}\right)^2$$

3

Fitting a time series model using the Box-Jenkins methodology

3. Diagnostic Checking

After the tentative identification of an $\mathrm{ARIMA}(p, d, q)$ model and calculation of the estimates $\hat{\mu}, \hat{\sigma}, \hat{\alpha}_1, \dots, \hat{\alpha}_p, \hat{\beta}_1, \dots, \hat{\beta}_q$ we have to perform diagnostic checking. The principle of this is that, if the $\mathrm{ARMA}(p, q)$ model is a good approximation to the underlying time series process, then the residuals $\hat{e}_t$ will form a good approximation to a white noise process.

There is a set of checks to be performed:

- Inspection of the graph of the residuals
- Inspection of the sample autocorrelation functions of the residuals
- Counting turning points

3 Fitting a time series model using the Box-Jenkins methodology

3. Diagnostic Checking

Inspection of the graph of the residuals
The visual inspection of the graph of the residuals against t or the graph of $\hat{e}_t$ against z_t can help to highlight a poorly fitting model.
If any pattern is evident, whether in the average level of the residuals or in the magnitude of the fluctuations about 0 , this should be taken to mean that the model is inadequate.

Inspection of the sample autocorrelation functions of the residuals
If the SACF or SPACF of the sequence of residuals has too many values outside the range ▪2 N , we conclude that the fitted model does not have enough parameters and a new model with additional parameters should be fitted.

3. Diagnostic Checking

Counting turning points

If $y_1, y_2, \dots, y_N$ is a sequence of numbers, then we say that the sequence has a turning point at time k if either $y_{k-1} < y_k$ and $y_k > y_{k+1}$, or $y_{k-1} > y_k$ and $y_k < y_{k+1}$.

If $Y_1, Y_2, \dots, Y_N$ is a sequence of independent random variables with continuous distribution, then the probability of a turning point at time k is $2/3$, the expected number of turning points is $\frac{2}{3}(N-2)$, and the variance is $\frac{16N-29}{90}$.

This is a result for a sequence of independent random variables. It is therefore usually applied to the residuals of the time series, not to the original time series itself, which will not be independent.

Question

CS2A A2023 Q8

Consider the time-series model:

$$y_t = a\, y_{t-2} + e_t + b\, e_{t-1} \qquad \text{(A)}$$

where et is white noise with mean 0 and variance σ^2.

(i) Derive the possible values of a and b for which the process yt is stationary and invertible. [4]

(ii) State the values of p and q for which yt is an ARMA(p, q) process. [1]

If b = 0 the original model (A) reduces to

$$y_t = a\, y_{t-2} + e_t \qquad \text{(B)}$$

(iii) Derive the autocorrelation function for this model while stationarity is assumed to hold. [8]

An actuary attempts to fit the model (A) to some time series data but concludes that the simpler model (B) is more appropriate.

(iv) Discuss how this conclusion could have been reached. [4] [Total 17]

Solution

(i)
Using the backshift operator one can show that the corresponding polynomials are
1-a B^2 [1]
and
1+bB [1]
The roots need to be in absolute value less than 1
abs(a)<1 and abs(b)<1 [2]

(ii)
ARMA(2,1) [1]

(iii)
The Yule-Walker equations are
gamma_0=a gamma_2+sigma^2 [1]
and
gamma_k=a gamma_{k-2} for k >= 1 [1]
So

Solution

gamma_1=a gamma_1 [1]
gamma_2=a gamma_0 [1]
These imply that
gamma_1=0, gamma_2=a gamma_0 and in general [1]
gamma_k =0 for k odd [1]
gamma_k = a^{k/2} gamma_0 for k even [1]
therefore
rho_k=0 for k odd [½]
rho_k=a^(k/2) for k even [½]
(There are no marks available for deriving the Yule Walker equations from first principles)

(iv)
Sample acf of the data could have indicated insignificant spikes for odd lags as for b=0 case those values are zero [2]
AIC/BIC could have also been used to confirm the statistical preference between the two models [1]
In the parameter estimation process for model (1), some low t-values could have been produced, particularly for the parameter *b*, indicating over-parametrisation. [1]
other sensible comments contrasting the fit of the two models [1]
[Marks available 5, maximum 4]
[Total 17]

4 Forecasting

Box-Jenkins approach to forecasting stationary time series

Using the Box-Jenkins approach, forecasting is relatively straightforward. Having fitted an ARMA model to the data $\{x_1, \ldots, x_n\}$ we have the equation:

$$\boldsymbol{X_{n+k} = \mu + \alpha_1(X_{n+k-1} - \mu) + \cdots + \alpha_p\left(X_{n+k-p} - \mu\right) + e_{n+k} + \beta_1 e_{n+k-1} + \cdots + \beta_q e_{n+k-q}}$$

We will now look forward to forecasting for this series.

4 Forecasting

Box-Jenkins approach to forecasting stationary time series
Forecasting future values of an ARMA process

The forecast value of X_{n+k} given all observations up until time n, known as the k-step ahead forecast and denoted $\hat{x}_n(k)$, is obtained from this equation by:

- replacing all (unknown) parameters by their estimated values;
- replacing the random variables $X_1, \dots, X_n$ by their observed values $x_1, \dots, x_n$;
- replacing the random variables $X_{n+1}, \dots, X_{n+k-1}$ by their forecast values $\hat{x}_n(1), \dots, \hat{x}_n(k-1)$;
- replacing the innovations $e_1, \dots, e_n$ by the residuals $\hat{e}_1, \dots, \hat{e}_n$;
- replacing the random variables $e_{n+1}, \dots, e_{n+k-1}$ by their expectations, 0

For example, the one-step ahead and two-step ahead forecasts for an $AR(2)$ are given by:

$$\hat{x}_n(1) = \hat{\mu} + \hat{\alpha}_1(x_n - \hat{\mu}) + \hat{\alpha}_2(x_{n-1} - \hat{\mu})$$
$$\hat{x}_n(2) = \hat{\mu} + \hat{\alpha}_1(\hat{x}_n(1) - \hat{\mu}) + \hat{\alpha}_2(x_n - \hat{\mu})$$

4 Forecasting

Forecasting ARIMA processes

If X is an $\mathrm{ARIMA}(p, d, q)$ process, then $Z = \nabla^d X$ is $\mathrm{ARMA}(p, q)$, so the techniques of Section 4.1 can be used to produce forecasts and confidence intervals for future values of Z.
By reversing the differencing procedure these can be translated into forecasts of future values of X.

For example, suppose that X is $\mathrm{ARIMA}(0,1,1)$.
Then $Z_n = \nabla X_n = x_n - x_{n-1}$ is $\mathrm{ARMA}(0,1)$, and $x_n = x_{n-1} + z_n$.
Hence $X_{n+1} = X_n + Z_{n+1}$, and $\hat{x}_n(1) = x_n + \hat{z}_n(1)$.

5.1 Multivariate time series models

We can write a univariate time series in multivariate (or vector) form.

For example, the time series $x_t = \alpha_1 x_{t-1} + \alpha_2 x_{t-2} + e_t + \beta e_{t-1}$ can be written as

$$\begin{pmatrix} x_t \\ x_{t-1} \\ x_{t-2} \end{pmatrix} = \begin{pmatrix} 0 & \alpha_1 & \alpha_2 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x_t \\ x_{t-1} \\ x_{t-2} \end{pmatrix} + \begin{pmatrix} 1 & \beta & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} e_t \\ e_{t-1} \\ e_{t-2} \end{pmatrix}$$

ie $\underline{X}_t = A\underline{x}_{t-1} + B\underline{e}_t$

The advantage of the vector form is that it displays the Markov property.

The vector process is stationary if the eigenvalues λ of the matrix A are all strictly less than 1 in magnitude. The eigenvalues are found by solving $\det(A - \lambda l) = 0$ where I is the identity matrix.

5.2 Cointegrated series

Two time series processes X and Y are called cointegrated if:
(i) X and Y are I(1) random processes
(ii) there exists a non-zero vector (called the cointegrating vector) (α, β) such that $\alpha X + \beta Y$ is stationary.

We might expect that two processes are cointegrated if one of the processes is driving the other or if both are being driven by the same underlying process.

5.3 Other non-linear, non-stationary time series

Other examples of time series include:

- bilinear models, which exhibit 'bursty' behavior:

$$x_n - \alpha(x_{n-1} - \mu) = \mu + e_n + \beta e_{n-1} + b(x_{n-1} - \mu)e_{n-1}$$

- threshold autoregressive models, which are used to model 'cyclical' behavior:

$$x_n = \mu + \begin{cases} \alpha_1(x_{n-1} - \mu) + e_n, & \text{if } x_{n-1} \leq d \\ \alpha_2(x_{n-1} - \mu) + e_n, & \text{if } x_{n-1} > d \end{cases}$$

- random coefficient, autogressive models:

$$x_t = \mu + \alpha_t(x_{t-1} - \mu) + e_t$$

where $\{\alpha_1, \alpha_2, \ldots, \alpha_n\}$ is a sequence of independent random variables.

5.3 Other non-linear, non-stationary time series

Other examples of time series include:

- autoregressive conditional heteroscedasticity (ARCH) models, which are used to model asset prices, where we require the volatility to depend on the size of the previous value:

$$x_t = \mu + e_t \sqrt{\alpha_0 + \sum_{k=1}^{p} \alpha_k (x_{t-k} - \mu)^2}$$

Question

CT6 S2015 Q11

Consider the following pair of equations:

$$X_t = 0.5X_{t-1} + \beta Y_t + \varepsilon_t^1$$

$$Y_t = 0.5Y_{t-1} + \beta X_t + \varepsilon_t^2$$

where ε_t^1 and ε_t^2 are independent white noise processes.

(i) (a) Show that these equations can be represented as

$$M\begin{pmatrix} X_t \\ Y_t \end{pmatrix} = N\begin{pmatrix} X_{t-1} \\ Y_{t-1} \end{pmatrix} + \begin{pmatrix} \varepsilon_t^1 \\ \varepsilon_t^2 \end{pmatrix}$$

where M and N are matrices to be determined.

Question

(b) Determine the values of β for which these equations represent a stationary bivariate time series model. [9]

(ii) Show that the following set of equations represents a VAR(p) (vector auto regressive) process, by specifying the order and the relevant parameters:[3]

$$X_t = \alpha X_{t-1} + \alpha Y_{t-1} + \varepsilon_t^1$$

$$Y_t = \beta X_{t-1} - \beta X_{t-2} + \varepsilon_t^2$$

Solution

(i) (a) It follows that

$$\begin{pmatrix} 1 & -\beta \\ -\beta & 1 \end{pmatrix}\begin{pmatrix} X_t \\ Y_t \end{pmatrix} = \begin{pmatrix} 0.5 & 0 \\ 0 & 0.5 \end{pmatrix}\begin{pmatrix} X_{t-1} \\ Y_{t-1} \end{pmatrix} + \begin{pmatrix} \varepsilon_t^1 \\ \varepsilon_t^2 \end{pmatrix}$$

(b) Multiplying both sides by

$$\begin{pmatrix} 1 & -\beta \\ -\beta & 1 \end{pmatrix}^{-1} = \frac{1}{(1-\beta^2)}\begin{pmatrix} 1 & \beta \\ \beta & 1 \end{pmatrix}$$

we then have

$$\begin{pmatrix} X_t \\ Y_t \end{pmatrix} = \frac{1}{2\left(1-\beta^2\right)}\begin{pmatrix} 1 & \beta \\ \beta & 1 \end{pmatrix}\begin{pmatrix} X_{t-1} \\ Y_{t-1} \end{pmatrix} + \frac{1}{\left(1-\beta^2\right)}\begin{pmatrix} 1 & \beta \\ \beta & 1 \end{pmatrix}\begin{pmatrix} \epsilon_t^{\ 1} \\ \epsilon_t^{\ 2} \end{pmatrix}.$$

Solution

Which is a stationary VAR(1) model if the eigenvalues of

$$\mathbf{A_1} = \frac{1}{2(1-\beta^2)}\begin{pmatrix} 1 & \beta \\ \beta & 1 \end{pmatrix}$$

are those λ such that

$$\det\begin{pmatrix} 1-\lambda & \beta \\ \beta & 1-\lambda \end{pmatrix} = 0 \text{ or } \lambda_{1,2} = 1 \pm \beta$$

then the eigenvalues of A_1 are less than one in absolute value if

Solution

$$\left|\frac{1 \pm \beta}{2(1-\beta^2)}\right| < 1 \text{ i.e.}$$

$$\left|\frac{1}{2(1-\beta)}\right| < 1$$

and

$$\left|\frac{1}{2(1+\beta)}\right| < 1$$

which implies that $|\beta| < \frac{1}{2}$ or $|\beta| > \frac{3}{2}$

Solution

(ii) Here we have a VAR(2) where

$$\boldsymbol{A}_1 = \begin{pmatrix} \alpha & \alpha \\ \beta & 0 \end{pmatrix} \qquad \boldsymbol{A}_2 = \begin{pmatrix} 0 & 0 \\ -\beta & 0 \end{pmatrix}$$

since

$$\begin{pmatrix} X_t \\ Y_t \end{pmatrix} = \begin{pmatrix} \alpha & \alpha \\ \beta & 0 \end{pmatrix}\begin{pmatrix} X_{t-1} \\ Y_{t-1} \end{pmatrix} + \begin{pmatrix} 0 & 0 \\ -\beta & 0 \end{pmatrix}\begin{pmatrix} X_{t-2} \\ Y_{t-2} \end{pmatrix} + \begin{pmatrix} \epsilon_t^{\,1} \\ \epsilon_t^{\,2} \end{pmatrix}.$$