BHARATHIDASAN UNIVERSITY

Tiruchirappalli- 620024

Tamil Nadu, India.

Programme: M.Sc. Statistics

Course Title: Stochastic Processes and Time Series Analysis

Course Code: 23ST02DEC

Unit-V

Forecasting

Dr. T. Jai Sankar
Associate Professor and Head
Department of Statistics

Ms. I. Angel Agnes Mary
Guest Faculty
Department of Statistics

UNIT - V

FORECASTING

Auto-correlation and auto-covariance functions

If a time series is stationary this means that the joint probability distribution of any two observations, say, y_t and y_{t+k}, is the same for any two time periods t and t + k that are separated by the same interval k. Useful information about this joint distribution and hence about the nature of the time series, can be obtained by plotting a scatter diagram of all of the data pairs y_t, y_{t+k} that are separated by the same interval k. The interval k is called the lag.

The covariance between y_t and its value at another time period, say, y_{t+k} is called the **auto-covariance** at lag k, defined by

$$\gamma_k = \text{Cov}(y_t, y_{t+k}) = E[(y_t - \mu)(y_{t+k} - \mu)]$$

The collection of the values of γ_k, k = 0, 1, 2, ... is called the **auto-covariance function**. Note that the auto-covariance at lag k = 0 is just the variance of the time series; that is, $\gamma_0 = \sigma_y^2$.

The **auto-correlation coefficient** at lag k is

$$\rho_k = \frac{E[(y_t - \mu)(y_{t+k} - \mu)]}{\sqrt{E[(y_t - \mu)^2]\,E[(y_{t+k} - \mu)^2]}} = \frac{Cov(y_t, y_{t+k})}{Var(y_t)} = \frac{\gamma_k}{\gamma_0}$$

Example

Calculate autocorrelation and auto-covariance function for the following data: 23.2, 23.6, 25.3, 25.2, 25.1, 25.6, 24.6, 24.6, 23.9 and 24.1

Procedure

- To calculate the mean
- To calculate the autocorrelation function

$$corr(X, X) = \frac{cov(X, X)}{\sigma_X \sigma_X}$$

for ACF(l_1)

$$corr(X, l_1) = \frac{cov(X, l_1)}{\sigma_X \sigma_{l_1}} = \frac{\frac{1}{n-1}\sum_{i=1}^{n-1}(X_{i+1} - \bar{X})(l_{1i} - \bar{X})}{\sigma_X \sigma_{l_1}}$$

for ACF(l_k)

$$corr(X, l_k) = \frac{cov(X, l_k)}{\sigma_X \sigma_{l_k}} = \frac{\frac{1}{n-k}\sum_{i=1}^{n-k}(X_{i+k} - \bar{X})(l_{ki} - \bar{X})}{\sigma_X \sigma_{l_k}}$$

where, l_1 = lag 1 of X time series, the length of l_1 series will be n-1.

Calculation

The autocorrelation function of lag 1 is,

$$ACF(l_1) = \frac{\frac{1}{n-1}\sum_{i=1}^{n-1}(X_{i+1} - \bar{X})(l_{1i} - \bar{X})}{\sigma_X \sigma_{l_1}}$$

X	$(X - \bar{X})$	$(X - \bar{X})^2$	Lag 1	Lag 1 - $\bar{X}$	$(X - \bar{X})(l_1 - \bar{X})$
23.2	-1.32	1.74			
23.6	-0.92	0.85	23.2	-1.32	1.2144
25.3	0.78	0.61	23.6	-0.92	-0.7176
25.2	0.68	0.46	25.3	0.78	0.5304
25.1	0.58	0.34	25.2	0.68	0.3944
25.6	1.08	1.17	25.1	0.58	0.6264
24.6	0.08	0.01	25.6	1.08	0.0864
24.6	0.08	0.01	24.6	0.08	0.0064
23.9	-0.62	0.38	24.6	0.08	-0.0496
24.1	-0.42	0.18	23.9	-0.62	0.2604
245.2		**5.75**			**2.3516**

$$\text{Mean} = \frac{245.2}{10} = 24.52$$

The autocorrelation function of lag 2 to lag 9 is

$(X - \bar{X})$	Lag 2	$(X - \bar{X}) * lag\ 2$	Lag 3	$(X - \bar{X}) * lag\ 3$	Lag 4	$(X - \bar{X}) * lag\ 4$	Lag 5	$(X - \bar{X}) * lag\ 5$
-1.32								
-0.92								
0.78	-1.32	-1.0296						
0.68	-0.92	-0.6256	-1.32	-0.8976				
0.58	0.78	0.4524	-0.92	-0.5336	-1.32	-0.7656		
1.08	0.68	0.7344	0.78	0.8424	-0.92	-0.9936	-1.32	-1.4256
0.08	0.58	0.0464	0.68	0.0544	0.78	0.0624	-0.92	-0.0736
0.08	1.08	0.0864	0.58	0.0464	0.68	0.0544	0.78	0.0624
-0.62	0.08	-0.0496	1.08	-0.6696	0.58	-0.3596	0.68	-0.4216
-0.42	0.08	-0.0336	0.08	-0.0336	1.08	-0.4536	0.58	-0.2936
		-0.4188		**-1.1912**		**-2.4556**		**-2.102**

$(X - \bar{X})$	Lag 6	$(X - \bar{X}) * lag\ 6$	Lag 7	$(X - \bar{X}) * lag\ 7$	Lag 8	$(X - \bar{X}) * lag\ 8$	Lag 9	$(X - \bar{X}) * lag\ 9$
-1.32								
-0.92								
0.78								
0.68								
0.58								
1.08								
0.08	-1.32	-0.1056						
0.08	-0.92	-0.0736	-1.32	-0.1056				
-0.62	0.78	-0.4836	-0.92	0.5704	-1.32	0.8184		
-0.42	0.68	-0.2856	0.78	-0.3276	-0.92	0.3864	-1.32	0.5544
		-0.9484		**0.1372**		**1.2048**		**0.5544**

$$\mathrm{ACF}(l_1) = \frac{2.3516}{5.75} = 0.409$$

$$\mathrm{ACF}(l_2) = \frac{-0.4188}{5.75} = -0.073$$

$$\mathrm{ACF}(l_3) = \frac{-1.1912}{5.75} = -0.207$$

$$\mathrm{ACF}(l_4) = \frac{-204556}{5.75} = -0.427$$

$$\mathrm{ACF}(l_5) = \frac{-2.102}{5.75} = -0.366$$

$$\mathrm{ACF}(l_6) = \frac{-0.9484}{5.75} = -0.165$$

$$\mathrm{ACF}(l_7) = \frac{0.1372}{5.75} = -0.024$$

$$\mathrm{ACF}(l_8) = \frac{1.2048}{5.75} = 0.210$$

$$\mathrm{ACF}(l_9) = \frac{0.5544}{5.75} = 0.096$$

Result

The autocorrelation and Auto-covariance function is,

Lag	Autocorrelation	Auto-covariance
0	1	5.75
1	0.410	2.3516
2	-0.073	-0.4188
3	-0.208	-1.1912
4	-0.428	-2.4556
5	-0.366	-2.102
6	-0.165	-0.9484
7	-0.024	0.1372
8	0.210	1.2048
9	0.097	0.5544

Linear stationary models

A time series has stationarity if a shift in time doesn't cause a change in the shape of the distribution. Basic properties of the distribution like the mean, variance and covariance are constant over time. Most forecasting methods assume that a distribution has stationarity. For example, Autocovariance and autocorrelations rely on the assumption of stationarity. An absence of stationarity can cause unexpected or bizarre behaviors, like t-ratios not following a t-distribution or high r-squared values assigned to variables that aren't correlated at all. There are three types of stationary models.

- Autoregressive
- Moving Average
- Autoregressive Moving Average

Autoregressive

In a multiple regression model, we forecast the variable of interest using a linear combination of predictors. In an autoregression model, we forecast the variable of interest using a linear combination of past values of the variable. The term autoregression indicates that it is a regression of the variable against itself.

Thus, an autoregressive model of order *p* can be written as

$$Y_t = \mu + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + ... + \phi_p Y_{t-p} + \varepsilon_t$$

where, y is the prediction outcome of multiple orders of previous results multiplied by their respective coefficients, ϕ. μ is the constant coefficient and ε_t is white noise.

An **autoregressive model** is when a value from a time series is regressed on previous values from that same time series. for example, y_t on y_{t-1}:

$$y_t = c + \phi_1 y_{t-1} + \varepsilon_t$$

In this regression model, the response variable in the previous time period has become the predictor and the errors have our usual assumptions about errors in a simple linear regression model. The **order** of an autoregression is the number of immediately preceding values in the series that are used to predict the value at the present time. So, the preceding model is a first-order autoregression, written as AR(1).

If we want to predict y this year (y_t) using measurements of global temperature in the previous two years (y_{t-1}, y_{t-2}), then the autoregressive model for doing so would be:

$$y_t = c + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \varepsilon_t$$

This model is a second-order autoregression, written as AR(2) since the value at time *t* is predicted from the values at times *t−1* and *t−2*. More generally, a k^{th}-order autoregression, written as AR(*k*), is a multiple linear regression in which the value of the series at any time *t* is a (linear) function of the values at times *t−1, t−2,…, t−p*.

Moving Average

The notation MA(*q*) refers to the moving average model of order *q*:

Moving Average process of order (q) is, $Y_t = \mu - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - .... - \theta_q \varepsilon_{t-q} + \varepsilon_t$

Where Y_t is Yield, ε_t's are independently and normally distributed with zero mean and constant variance σ^2 for t = 1,2,..., n; θs are coefficients to be estimated.

Autoregressive Moving Average

This is a model that is combined from the AR and MA models. In this model, the impact of previous lags along with the residuals is considered for forecasting the future values of the time series. Here β represents the coefficients of the AR model and α represents the coefficients of the MA model.

Autoregressive process of order (p) is, $Y_t = \mu + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + ... + \phi_p Y_{t-p} + \varepsilon_t$;

Moving Average process of order (q) is, $Y_t = \mu - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - .... - \theta_q \varepsilon_{t-q} + \varepsilon_t$; and

The general form of ARMA model of order (p, q) is

$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + ... + \phi_p Y_{t-p} + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + ... + \theta_q \varepsilon_{t-q} + \varepsilon_t$$

Where Y_t is Yield, ε_t's are independently and normally distributed with zero mean and constant variance σ^2 for t = 1,2,..., n; and ϕs and θs are coefficients to be estimated.

Differencing:

The difference operator ∇ is given by,

$$\nabla X_t = X_t - X_{t-1}$$

These differences form a new time series ∇X (of length $n-1$ if the original series had length n). Similarly

$$\nabla^2 X_t = \nabla(\nabla X_t) = X_t - 2X_{t-1} + X_{t-2}$$

and so on.

If our original time series is not stationary, we can look at the first order difference process ∇X, or second order differences ∇^2X, and so on. If we find that a differenced process is a stationary process, we can look for an ARMA model of that differenced process.

In practice if differencing is used, usually d = 1, or maybe d = 2, is enough.

The Box-Jenkins approach

The Box-Jenkins approach to fitting ARIMA models can be divided into three parts:

- Identification
- Estimation
- Verification

Identification

This refers to initial preprocessing of the data to make it stationary and choosing plausible values of p and q (which can of course be adjusted as model fitting progresses).

To assess whether the data come from a stationary process we can

- Look at the data: e.g. a time plot as we looked at for the l h series;
- Consider transforming it (e.g. by taking logs;)
- Consider if we need to difference the series to make it stationary.

For stationarity the ACF should decay to zero fairly rapidly. If this is not true, then try differencing the series, and maybe a second time if necessary. (In practice it is rare to go beyond d = 2 stages of differencing.)

The next step is initial identification of p and q. For this we use the ACF and the PACF, recalling that

- For an MA(q) series, the ACF is zero beyond lag q;
- for an AR(p) series, the PACF is zero beyond lag p.

We can use plots of the ACF/PACF and the approximate $\pm\frac{2}{\sqrt{n}}$ confidence bounds.

Estimation: AR processes

For the *AR(p)* process

$$X_t = \sum_{i=1}^{p} \alpha_i X_{t-i} + \epsilon_t$$

We have the Yule-Walker equations

$$\rho_k = \sum_{i=1}^{p} \alpha_i \rho_{|i-k|},$$

for k > 0.

We fit the parameters $\alpha_1, \ldots, \alpha_p$ by solving

$$r_k = \sum_{i=1}^{p} \alpha_i r_{|i-k|}, \qquad k = 1, \ldots, p$$

These are p equations for the p unknowns $\alpha_1, \ldots, \alpha_p$ which, as before, can be solved using a Levinson-Durbin recursion.

The Levinson-Durbin recursion gives the residual variance

$$\widehat{\sigma}_p^2 = \frac{1}{n} \sum_{t=p+1}^{n} \left(X_t - \sum_{j=1}^{p} \widehat{\alpha}_j X_{t-j} \right)^2.$$

This can be used to guide our selection of the appropriate order p. Define approximate log likelihood by

$$-2 \log L = n \log(\widehat{\sigma}_p^2).$$

Then this can be used for likelihood ratio tests.

Alternatively, p can be chosen by minimizing AIC where

$$AIC = -2 \log L + 2k$$

and k = p is the number of unknown parameters in the model.

If $(X_t)_t$ is a causal AR(p) process with i.i.d. with N(0, σ^2_ε) then the Yule-Walker estimator $\hat{\underline{\alpha}}$ is optimal with respect to the normal distribution.

Moreover for the PACF of a causal AR(p) process we have that, for m > p,

$$\sqrt{n}\hat{\alpha}_{mm}$$

is asymptotically standard normal. However, the elements of the vector

$\hat{\alpha}_m = (\hat{\alpha}_{1m}, \ldots, \hat{\alpha}_{mm})$ are in general not asymptotically uncorrelated.

Estimation: ARMA processes

Now we consider an *ARMA(p, q)* process. If we assume a parametric model for the white noise – this parametric model will be that of Gaussian white noise – we can use maximum likelihood.

We rely on the prediction error decomposition. That is, $X_1, \ldots, X_n$ have joint density

$$f(X_1, \ldots, X_n) = f(X_1) \prod_{t=2}^{n} f(X_t \mid X_1, \ldots, X_{t-1}).$$

Suppose the conditional distribution of X_t given $X_1, \ldots, X_{t-1}$ is normal with mean $\widehat{X}_t$ and variance P_{t-1}, and suppose that $X_1 \sim N(\widehat{X}_1, P_0)$.

Then for the log likelihood we obtain

$$-2 \log L = \sum_{t=1}^{n} \left\{ \log(2\pi) + \log P_{t-1} + \frac{(X_t - \widehat{X}_t)^2}{P_{t-1}} \right\}.$$

Here $\widehat{X}_t$ and P_{t-1} are functions of the parameters $\alpha_1, \ldots, \alpha_p, \beta_1, \ldots, \beta_q$, and so maximum likelihood estimators can be found (numerically) by minimising $-2 \log L$ with respect to these parameters.

The matrix of second derivatives of $-2 \log L$, evaluated at the MLE, is the observed information matrix, and its inverse is an approximation to the covariance matrix of the estimators. Hence we can obtain approximate standard errors for the parameters from this matrix.

In practice, for AR(p) for example, the calculation is often simplified if we condition on the first m values of the series for some small m. That is, we use a conditional likelihood, and so the sum in the expression for $-2 \log L$ is taken over $t = m + 1$ to n.

For an AR(p) we would use some small value of m, $m > p$.

When comparing models with different numbers of parameters, it is important to use the same value of m, in particular when minimising $AIC = -2 \log L + 2(p + q)$.

Verification

The third step is to check whether the model fits the data. Two main techniques for model verification are

- *Over fitting:* add extra parameters to the model and use likelihood ratio or t tests to check that they are not significant.
- *Residual analysis:* calculate residuals from the fitted model and plot their ACF, PACF, 'spectral density estimates', etc, to check that they are consistent with white noise.

Linear non-stationary models

A time series has non-stationarity if a shift in time does cause a change in the shape of the distribution. Basic properties of the distribution like the mean, variance and covariance aren't constant over time. There are three types of stationary models.

- Autoregressive Integrated Moving Average (ARIMA)
- Seasonal Autoregressive Integrated Moving Average (SARIMA)
- Exponential Smoothing

Autoregressive Integrated Moving Average (ARIMA)

The time series when differenced follows both AR and MA models and is known as ARIMA model. Hence, ARIMA model was used in this study, which required a sufficiently large data set and involved four steps: identification, estimation, diagnostic checking and forecasting. Model parameters were estimated to fit the ARIMA models.

Autoregressive process of order (p) is, $Y_t = \mu + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + ... + \phi_p Y_{t-p} + \varepsilon_t$;

Moving Average process of order (q) is, $Y_t = \mu - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - .... - \theta_q \varepsilon_{t-q} + \varepsilon_t$; and

The general form of ARIMA model of order (p,d,q) is

$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + ... + \phi_p Y_{t-p} + \mu - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - .... - \theta_q \varepsilon_{t-q} + \varepsilon_t$$

where Y_t is Yield, ε_t's are independently and normally distributed with zero mean and constant variance σ^2 for t = 1,2,..., n; d is the fraction differenced while interpreting AR and MA and ϕs and θs are coefficients to be estimated.

Trend Fitting

The Box-Ljung Q statistics was used to transform the non-stationary data into Stationarity data and also to check the adequacy for the residuals. For evaluating the adequacy of AR, MA and ARIMA processes, various reliability statistics like R^2, Stationary R^2, RMSE, MAPE, and BIC as suggested by Gideon Schwartz (1978) were used computed as below:

$$RMSE = \left[\frac{1}{n} \sum_{i=1}^{n} (Y_i - \hat{Y}_i)^2 \right]^{1/2}; \quad MAPE = \frac{1}{n} \sum_{i=1}^{n} \left| \frac{(Y_i - \hat{Y}_i)}{Y_i} \right| \text{ and } BIC(p,q) = \ln v^*(p,q) + (p+q) [\ln(n) / n]$$

where, p and q are the order of AR and MA processes respectively and n is the number of observations in the time series and v* is the estimate of white noise variance σ^2.

Seasonal Autoregressive Integrated Moving Average (SARIMA)

The seasonal ARIMA model incorporates both non-seasonal and seasonal factors in a multiplicative model. One shorthand notation for the model is

$$\text{ARIMA (p, d, q)} \times \text{(P, D, Q)}_S$$

Where, p = non-seasonal AR order, d = non-seasonal differencing, q = non-seasonal MA order, P = seasonal AR order, D = seasonal differencing, Q = seasonal MA order and S = time span of repeating seasonal pattern, Φ = Autoregressive polynomial Order and Θ = Moving average polynomial Order.

Without differencing operations, the model could be written more formally as

$$\Phi(B^S)\phi(B^p) \qquad \text{---------- (1)}$$

The non-seasonal components are:

$$\text{AR: } \phi(B) = 1 - \phi_1 B - \ldots\ldots\ldots - \phi_p B^p$$
$$\text{MA: } \theta(B) = 1 + \theta_1 B + \ldots\ldots\ldots + \theta_q B^q$$

The seasonal components are:

$$\text{Seasonal AR: } \Phi(B^S) = 1 - \Phi_1 B^S - \ldots\ldots\ldots - \Phi_{\text{P}} B^{PS}$$
$$\text{Seasonal MA: } \Theta(B^S) = 1 + \Theta_1 B^S + \ldots\ldots\ldots + \Theta_{\text{Q}} B^{QS}$$

On the left side of equation (1) the seasonal and non-seasonal AR components multiply each other, and on the right side of equation (1) the seasonal and non-seasonal MA components multiply each other.

Exponential Smoothing

Exponential smoothing is a time series method for forecasting univariate time series data. Time series methods work on the principle that a prediction is a weighted linear sum of past observations or lags. The Exponential Smoothing time series method works by assigning exponentially decreasing weights for past observations. It is called so because the weight assigned to each demand observation is exponentially decreased.

Types of Exponential Smoothing

There are three types of Exponential Smoothing forecasting methods.

- Simple or Single Exponential Smoothing
- Double Exponential Smoothing
- Triple Exponential Smoothing

1. Simple or Single Exponential Smoothing

Simple or single exponential smoothing (SES) is the method of time series forecasting used with univariate data with no trend and no seasonal pattern. It needs a single parameter called alpha (a), also known as the smoothing factor. Alpha controls the rate at which the influence of past observations decreases exponentially. The parameter is often set to a value between 0 and 1.

The simple exponential smoothing formula is given by:

$$s_t = \alpha x_t + (1 - \alpha)s_{t-1} = s_{t-1} + \alpha(x_t - s_{t-1})$$

here,

- s_t = smoothed statistic (simple weighted average of current observation x_t)
- s_{t-1} = previous smoothed statistic
- α = smoothing factor of data; $0 < \alpha < 1$
- t = time period

2. Double Exponential Smoothing

This method is known as Holt's trend model or second-order exponential smoothing. Double exponential smoothing is used in time-series forecasting when the data has a linear trend but no seasonal pattern. The basic idea here is to introduce a term that can consider the possibility of the series exhibiting some trend.

In addition to the alpha parameter, Double exponential smoothing needs another smoothing factor called beta (b), which controls the decay of the influence of change in trend. The method supports trends that change in additive ways (smoothing with linear trend) and trends that change in multiplicative ways (smoothing with exponential trend).

The Double exponential smoothing formulas are:

$S_1 = x_1$

$B_1 = x_1 - x_0$

For t>1,

$s_t = \alpha x_t + (1 - \alpha)(s_{t-1} + b_{t-1})$

$\beta_t = \beta(s_t - s_{t-1}) + (1 - \beta)b_{t-1}$

Here,

- s_t = smoothed statistic, it is the simple weighted average of current observation x_t
- s_{t-1} = previous smoothed statistic
- α = smoothing factor of data; $0 < \alpha < 1$
- t = time period
- b_t = best estimate of trend at time t
- β = trend smoothing factor; $0 < \beta < 1$

3. Triple Exponential Smoothing

In this method, exponential smoothing applied three times. This method is used for forecasting the time series when the data has both linear trend and seasonal pattern. This method is also called Holt-Winters exponential smoothing. The triple exponential smoothing formulas are given by:

$$\begin{aligned} s_0 &= x_0 \\ s_t &= \alpha \frac{x_t}{c_{t-L}} + (1 - \alpha)(s_{t-1} + b_{t-1}) \\ b_t &= \beta(s_t - s_{t-1}) + (1 - \beta)b_{t-1} \\ c_t &= \gamma \frac{x_t}{s_t} + (1 - \gamma)c_{t-L} \end{aligned}$$

Here,

- s_t = smoothed statistic, it is the simple weighted average of current observation x_t
- s_{t-1} = previous smoothed statistic
- α = smoothing factor of data; $0 < \alpha < 1$
- t = time period
- b_t = best estimate of a trend at time t
- β = trend smoothing factor; $0 < \beta < 1$
- c_t = sequence of seasonal correction factor at time t
- γ = seasonal change smoothing factor; $0 < \gamma < 1$

Holt-Winters Exponential Smoothing has two categories depending on the nature of the seasonal component:

- Holt-Winter's Additive Method − for seasonality that is addictive.
- Holt-Winter's Multiplicative Method – for seasonality that is multiplicative.

Example

The sales of a magazine in a stall for the previous 10 months are given below:

Month	January	February	March	April	May	June	July	August	September	October
Sales	30	25	35	25	20	30	35	40	30	45

Calculate the simple exponential smoothing taking $\alpha = 0.3$.

Solution

Month	**Sales**	**Exponential smooth $\alpha = 0.3$**
January	30	30.00
February	25	30.00
March	35	28.50
April	25	30.45
May	20	14.10
June	30	15.87
July	35	20.11
August	40	24.58
September	30	29.20
October	45	29.44
November	–	34.11