CHAPTER III

STATISTICAL FORECASTING AND BAYESIAN METHODS
3.1 INTRODUCTION

Most of the forecasting problems suffer from inadequate or total absence of historical (prior) information at the time the initial forecast is required. Thus, early forecasts must be based largely on subjective considerations. As and when the time series information becomes available, we must modify our subjective estimates in the light of the actual data. Bayesian methods are often useful in statistical inference problem of this type. It is assumed that the original subjective forecast can be translated into a subjective estimate of the parameters of the forecasting model. The Bayesian procedures are then used to revise the parameter estimates when time series information is available.

3.2 BAYESIAN ESTIMATION

Bayes' theorem forms the basis for estimating parameters in the time series. Let $X$ be a random variable with probability density function (pdf) $f$ characterized by an unknown parameter $\theta$. The probability function is $f(x|\theta)$. The parameter $\theta$ is assumed to be a random variable with a probability density $h_0(\theta)$, which is called the prior density of $\theta$. This probability distribution measures subjective or "degree of belief" information about $\theta$. If we are relatively confident about the value of $\theta$, we would choose a prior distribution with small variance. If we are relatively uncertain about $\theta$, we might choose a prior distribution with a larger variance.

Suppose the initial estimate of $\theta$ is given as a probability distribution $h_0(\theta)$ and the subsequent information about the time series is obtained in the form of the statistic $Y$ whose probability distribution $f(y|\theta)$ depends upon $\theta$,
where \( y \) is a sample of observations, \( x_1, x_2, \ldots, x_t \), from the time series, of some sufficient statistic computed from the sample. The new estimate of \( \theta \) will be in the form of revised distribution, \( h_1(\theta|y) \), called the posterior distribution.

Using Bayes' theorem, we have

\[
h_1(\theta|y) = \frac{h_0(\theta) f(y|\theta)}{\int h_0(\theta) f(y|\theta) \, d\theta} = \frac{h_0(\theta) f(y|\theta)}{g(y)} \quad \ldots(3.2.1)
\]

where, \( h_0(\theta) = \) prior distribution of \( \theta \) (the marginal distribution of \( \theta \))

\( f(y|\theta) = \) likelihood of \( y \), given \( \theta \) (the conditional distribution of \( y \) given \( \theta \))

\( g(y) = \) unconditional distribution of \( y \), averaged over all \( \theta \) (the marginal distribution of \( y \)).

\( h_1(\theta|y) = \) posterior distribution of \( \theta \), given the information \( y \) (the conditional distribution of \( \theta \), given \( y \)).

If \( \theta \) is a discrete variable, the integral sign in the equation (3.2.1) will be replaced by a summation sign.

The equation (3.2.1) provides a means for blending the observed data \( Y \) with the prior information \( h_0(\theta) \) to obtain a revised description of our uncertainty about \( \theta \).

The Bayes estimator of \( \theta \), denoted by \( \theta^* \), is defined as the expected value of the posterior density, or

\[
\theta^* = \int \theta \, h_1(\theta|Y) \, d\theta \quad \ldots(3.2.2)
\]
We use $\theta^*$ as an estimate of $\theta$ in the forecasting model. For certain cases, it can be shown that $\theta^*$ is optimal in the sense of minimizing the variance of forecast error.

Consider the following example of a normal sampling process and normal prior.

**Example 1:** Suppose $y$ is normally distributed with mean $\theta$ and variance $\sigma_y^2$; that is

$$f(Y|\theta) \approx (\theta, \sigma_y^2) = (2\pi\sigma_y^2)^{-1/2} \exp\left\{-1/2\left[ (y-\theta)/\sigma_y \right]^2 \right\} \quad \ldots(3.2.3)$$

where $\sigma_y^2$ is known. The prior distribution of $\theta$ is also normal with mean $\bar{\theta}'$ and variance $\nu'_{\theta}$; that is

$$h_\theta(\theta) = N(\bar{\theta}', \nu'_{\theta}) = (2\pi\nu'_{\theta})^{-1/2} \exp\left\{- (\theta - \bar{\theta}')^2 / 2\nu'_{\theta} \right\} \quad \ldots(3.2.4)$$

The posterior distribution of $\theta$, given $y$, is

$$h_\theta(\theta|y) = \frac{2\pi(\nu\sigma_y^2)^{-1/2} \exp\left\{-1/2\left[ (\theta - \bar{\theta}')/\nu_{\theta} + (y - \theta)^2 / \sigma_y^2 \right] \right\}}{\int_0^\infty 2\pi(\nu\sigma_y^2)^{-1/2} \exp\left\{-1/2\left[ (\theta - \bar{\theta}')/\nu_{\theta} + (y - \theta)^2 / \sigma_y^2 \right] \right\} d\theta}$$

on simplification, we get

$$h_\theta(\theta|y) = \left[ 2\pi(\nu_{\theta}$$

$$\sigma_y^2)^{-1/2} \right]$$
\[
\exp \left( -\frac{1}{2} \left[ \theta - (Y \nu' \theta + \delta \sigma^2) \right]^2 \right) \\
\frac{1}{\sigma^2 (\nu' + \sigma^2)}
\]

which it is recognized as a normal density with mean

\[
\bar{\theta} = E(\theta|Y) = \frac{Y \nu' + \theta' \sigma^2}{\nu' + \sigma^2} \quad \ldots(3.2.5)
\]

and variance \(\nu^2 = \nu(\theta|y) = \frac{\nu \sigma^2}{\nu' + \sigma^2} \quad \ldots(3.2.6)\)

Therefore, it is found that the posterior distribution is normally distributed if the prior distribution is normal and the sampling process is normal with known variance.

If we believe that the time series is adequately described by the constant model: \(x_t = b + \varepsilon_t\), where \(b\) is the unknown mean and \(\varepsilon_t\) is the random component assumed to be normally distributed with mean \(\theta\) and known variance.

i.e., \(\varepsilon_t \sim N(\theta, \sigma^2)\). Thus, we are assuming that the mortality rate in any period \(t\) has probability distribution \(f(x_t/b) = N(b, \sigma^2)\). Since the variance is assumed to be known, the problem is to estimate \(b\).
3.2.1 ESTIMATION

Assume that at the start of the forecasting process (time zero) we estimate the mean mortality rate to \( \hat{\theta}_0 \). Assume that the uncertainty about this estimate is such that the true mean \( b \) is then assigned the following probability distribution:

\[
\theta_0(b) \sim N(\hat{\theta}_0, \nu_0')
\]

where the variance \( \nu_0' \) is a measure of the uncertainty in the estimate.

After one period, the observation \( x_1 \) is known. The problem is to modify the estimate \( \hat{\theta}_0 \) and the measure of uncertainty \( \nu_0' \) in the light of this information. This can be done by using the results of the above example as follows:

\[
\theta_1(\theta/x_1) \sim N[\hat{\theta}^*(1), \nu_1^*(1)],
\]

where

\[
\hat{\theta}^*(1) = E(b/x_1) = \frac{x_1 \nu_0' + \hat{\theta}_0 \sigma_0^2}{\nu_0' + \sigma_0^2}
\]

\[
\nu_1^*(1) = \text{Var}(b/X_1) = \frac{\nu_0' \sigma_0^2}{\nu_0' + \sigma_0^2}
\]

At the end of period 2, when \( x_2 \) is known, the Bayesian updating process transforms \( h_1(b/x) \) into \( h_2(b/x_1, x_2) \) in the following way:
\[ h_2(b/x_1, x_2) = \frac{h_1(b/x_1)f(x_2/b)}{\int h_1(b/x_1)f(x_2/b)db} \]

Hence \( h_1 \) is treated as a prior and, together with the likelihood of \( x_2 \), it is used to obtain the probability distribution of \( b \) at the end of period 2. Again using the results of the above example, we get

\[ h_2(b/x_1, x_2) \sim N(\delta^*(2), \nu^*(2)) \]

where

\[ \delta^*(2) = E(b/x_1, x_2) = \frac{\bar{x} \nu + \delta'(\sigma^2_2/2)}{\nu + (\sigma^2_2/2)} \]

and

\[ \bar{x} = (x_1 + x_2)/2 \]

It may be easily verified that \( h_2(b/x_1, x_2) = \bar{x} h_2(b/\bar{x}) \); that is the same posterior is obtained using \( \bar{x} \) as from using \( x_1 \) and \( x_2 \) sequentially (or jointly). This is because \( \bar{x} \) is a sufficient statistic for estimating \( b \).

This may be generalized and it may be shown that after observing \( x_0 \), we calculate the posterior as:

\[ h_1(b/x_1, x_2, \ldots, x_n) = h_0(b/\bar{x}) \sim N(\delta^*(t), \nu^*(t)) \]

where

\[ \delta^*(t) = \frac{\bar{x} \nu + \delta'(\sigma^2_i/t)}{\nu + (\sigma^2_i/t)} \]
\[ \nu_1^2(t) = \frac{\nu_1 \sigma^2}{\nu_1 + \sigma^2} \text{ and } \bar{x} = \frac{\sum x_i}{i} \]

The Bayes estimator of 'b' after 't' periods is \( b^*(t) = \tilde{b}^*(t) \). Further, it may be written as:

\[ b^*(t) = \frac{t}{\mu + t} \bar{x} + \frac{\mu}{\mu + t} b^{-1}, \]

where \( \mu = \frac{\sigma^2}{\nu_1} \).

It may be noticed that the Bayes estimator of 'b' is just a weighted average of the sample mean and the subjective initial estimate, b. Further more, \( b(t) \) may be expressed recursively as:

\[ b^* = \alpha x_t + (1 - \alpha) b^*(t - 1) \] \( \quad \ldots \text{(3.2.8)} \)

where \( \alpha = \alpha(t) = \frac{1}{\mu + t} = \frac{\nu_1}{\nu_1 + \sigma^2} \) \( \quad \ldots \text{(3.2.9)} \)

From this equation, it may be noticed that it reveals that the estimate of the demand rate is updated each period by a form similar to exponential smoothing (exponential smoothing in demand forecasting where the demand process changes over time). However, from equation (3.2.9) we observe that \( \alpha \) is a function of 't' becoming smaller as 't' increased. Since \( \nu_1^2(t) = \alpha(t) \sigma^2 \), the uncertainty in the estimate of 'b' decreases to zero as 't' becomes infinitely large.
Also the weight given to \( \tilde{b} \), the initial estimate, decreases as \( t \) becomes large, and correspondingly more weight is given to the actual time series data.

Since the variance of the demand process, \( \sigma^2 \), usually is not known, it must be estimated. If its value is not certain, it could be treated as a random variable, just as it was done in case of \( b \), by expanding the unknown parameter set to \( \theta = \{ b, \sigma^2 \} \). Then we would define a prior joint probability distribution on \( b \) and \( \sigma^2 \) is defined. The sample mean and the sample variance of the \( t \) observations is utilized while calculating the posterior probability distribution.

### 3.2.2 Forecasting

Once the posterior distribution is determined, it may be used as the basis for forecasting. For a constant process, the forecasting equation is

\[
\hat{x}_{t+k} = \hat{b}(t) \quad \text{... (3.2.10)}
\]

Using Bayes estimator \( b^*(t) \) as estimator of \( b \), the forecast for period \( t+k \) is

\[
\hat{x}_{t+k} = b^*(t) \quad \text{... (3.2.11)}
\]

The uncertainty in the estimate of \( b \) is measured by the posterior variance, \( \nu_t^*(t) \).

From equation (3.2.11) we get the forecast variance, which is given by

\[
E[(b - b^*)^2] \text{ or } V(\hat{x}_{t+k}(t)) = \nu_t^* \quad \text{... (3.2.12)}
\]
The variance of the forecast error may be shown as:

\[ \nu[e_k (t+k)] = \nu[x_{t+k} - \hat{x}_{t+k} (t)] = \sigma^2_x + \nu^*_x \]  \hspace{1cm} (3.2.13)

It should be noted that the forecast error variance is independent of the forecast lead time \( k \) in the Bayesian case.

From the equation (3.2.13), assuming \( b \) and \( x \) are normally distributed, a 100 (1-\( \tau \)) per cent prediction interval for \( X_{t+k} \) may be computed as \( b^*(t) \pm u_{\tau/2} \sqrt{\sigma_b^2 + \nu^*_b} \) where \( u_{\tau/2} \) is the 100 \( (\tau/2) \) percentile of the standard normal distribution.

**Example 2**: We wish to forecast the demand for a new product (time is measured in weeks). We believe that demand is normally distributed and that a constant model is appropriate, but no prior information is available. A reasonable prior density for \( b \) is thought to be \( N(100, 25) \), and is estimated to be 150.

The forecast for the period 1 is \( x(0) = 100 \). The variance of the forecast error is \( 150 + 25 = 175 \). Therefore, a 95% prediction interval for \( x_1 \) is 100 + (1.960).

Suppose the actual demand in period 1 is 86.

Then using the equation (3.2.8) with \( \mu = \sigma_b^2 / \nu^*_b = 150/25 = 6 \)

And \( \sigma(1) = 1/(6 + 1) = 0.143 \), we have \( b^*(1) = \hat{b}^*(1) \)

34
\[ = \frac{1}{7} (86) + \frac{6}{7} (100) = 98.0 \text{ and} \]

\[ \nu_2^*(1) = \frac{1}{7} (150) = 21.4 \]

Therefore, the forecast for period 2 is \( \hat{x}_2(1) = 98.0 \) and a 95% prediction interval for \( x_2 \) is

\[ 98.0 \pm (1.960) \sqrt{150 + 21.4} \text{ or } [72.3, 123.7] \]

In period 2, the actual demand is 94. Then \( \alpha(2) = \frac{1}{6 + 2} = 0.125 \) and

\[ b^*(2) = \frac{1}{8} (94) + \frac{7}{8} (98.0) = 97.5 \]

\[ \nu_2^*(2) = \frac{1}{8} (150) = 18.8 \]

The 95% prediction interval is

\[ 97.5 \pm (1.960) \sqrt{150 + 18.8} \text{ or } [72.0, 123.0] \]

This procedure is continued until it is thought appropriate to switch over to a more permanent forecasting system.
3.2.3 PROCESS MODEL

Bayesian approach to parameter estimation and forecasting for time series models that are linear in the unknown parameters $b_1, b_2, \ldots, b_k$ is discussed below:

Consider the time series model:

$$X_t = b_1 z_1(t) + b_2 z_2(t) + \ldots + b_k z_k(t) + \varepsilon_t,$$

$$= \sum_{i=1}^{k} b_i z_i(t) + \varepsilon_t, \quad \ldots (3.2.14)$$

Where the $\{b_i\}$ are constants, the $\{z_i(t)\}$ are mathematical functions of $t$ and are independent variables in the model, and $\varepsilon_t$ is $N(0, \sigma^2_{\varepsilon})$.

We assume $\varepsilon_t$ is independent of $\varepsilon_{t-j}$ for all $j$. Often $z_i(t)=1$, so that $b_i$ is the constant term in the model.

In matrix notation, it is defined as

$$Z(t) = \begin{bmatrix} z_1(t) & z_2(t) & \ldots & z_k(t) \end{bmatrix}$$

$$b = \begin{bmatrix} b_1 & b_2 & \ldots & b_k \end{bmatrix}^T$$

Where the superscript $^T$ indicates transpose (all vectors are column vectors) then equation (3.2.14) can be written as:

$$x_t = b' Z(t) + \varepsilon_t, \quad \ldots (3.2.15)$$
The probability distribution of $x_t$ is

$$f\left( \frac{x_t}{b}, \sigma_t^2 \right) = N\left( b^T z(t), \sigma_t^2 \right) \quad \ldots (3.2.16)$$

'b' has to be estimated to forecast this process. It is assumed that $\sigma_t^2$ is known.

### 3.2.4 PRIOR DISTRIBUTION

Before observing the time series, the $b_i$ are assumed to be jointly normally distributed with $E(b_i) = \delta_i$, $\text{var}(b) = \nu_0$, and $\text{Cov}(b_i, b_j) = \nu_j$. That is $b'$ follows the multivariate normal distribution:

$$h_0(b) = (2\pi)^{-n/2} \nu^{-1} \exp \left\{ -1/2 (b - b')^T \nu^{-1} (b - b') \right\} \quad \ldots (3.2.17)$$

$$h_0(b) = N(\bar{b}', \nu'), \text{ where } b \text{ and } \nu \text{ are vectors.}$$

$\bar{b}' = E(b)$ and $\nu'$ is the variance-covariance matrix of the prior distribution given in equation in the light of these actual data.

A matrix $G'$ is defined as follows:

$$G' = \sigma_t^2 \nu'^{-1} \quad \ldots (3.2.18)$$

so that $G'^{-1} = \nu'/\sigma_t^2 \quad \ldots (3.2.19)$

The formulae for calculating the parameters of the posterior distribution can be written in terms of $G'$. 

37
After \( t \) periods, it is derived to revise the prior distribution given in equation (3.2.17) in the light of the actual data.

We define \( X = \{x_1, x_2, \ldots, x_t\} \), where \( x \) is a vector, and

\[
Z = \begin{bmatrix} z'(1) \\ z'(2) \\ \vdots \\ z'(t) \end{bmatrix} = \begin{bmatrix} z_1(t) & z_2(t) & \cdots & z_s(t) \\ z_1(t) & z_2(t) & \cdots & z_s(t) \\ \vdots & \vdots & \ddots & \vdots \\ z_1(t) & z_2(t) & \cdots & z_s(t) \end{bmatrix}
\]

We can estimate \( b \) from the actual time series data by solving the least squares normal equations

\[
z'z\hat{b} = z'x \quad \ldots \quad (3.2.20)
\]

let \( G = z'z \) and \( g = z'x \), then \( G\hat{b} = g \) \quad \ldots \quad (3.2.21)

The least squares estimators are: \( \hat{b} = G^{-1}g \) \quad \ldots \quad (3.2.22)

Since this computation is made at the end of period \( t \), it is understood that \( \hat{b} = \hat{b}(t) \)

The least squares estimators are unbiased, that is

\[
E(\hat{b}/b) = b \quad \ldots \quad (3.2.23)
\]

and have variance-covariance matrix \( \text{cov} (\hat{b}) = G^{-1}G^{-1} \).

38
The elements of \( v \) are \( v_{ij} = \text{Cov} \left( \hat{b}_i; \hat{b}_j \right) \). \( v \) is symmetric. It can be shown that \( \hat{b} \) is sufficient for estimating \( b \) and \( \sigma^2 \) is known and that it follows the multivariate normal distribution.

\[
 f(\hat{b}/b:z,\sigma^2) = (2\pi)^{-d/2}e^{-1/2} \exp\left\{1/2[\hat{b} - b] \right\} v^{-1} [\hat{b} - b] \quad \text{...(3.2.24)}
\]

\[
 f(\hat{b} / b: z, \sigma^2) \sim \text{N}(b, v)
\]

If the prior distribution of \( b \), given in the equation (3.2.17) is considered, the marginal distribution of \( b \) can be computed for fixed \( z \) and, averaging over \( b \); that is,

\[
 g(b) = \int f(\hat{b}/b)h_b(b)db \sim \text{N}(\bar{b}', \nu_*) \quad \text{...(3.2.25)}
\]

where \( \nu_* = \nu' + \nu = (G^{-1} + G^{-1})\sigma^2 \) \( \quad \text{...(3.2.26)} \)

### 3.2.5 POSTERIOR DISTRIBUTION

The posterior distribution of \( \hat{b} \), given \( b \), is computed at time 't' as:

\[
 h_t(b/\hat{b}) = \frac{h_b(b)f(\hat{b}/b)}{g(\hat{b})} \quad \text{...(3.2.27)}
\]

It can be shown that \( h_t(b/\hat{b}) \) follows multivariate normal distribution i.e., \( h_t(b/\hat{b}) \sim \text{N}(\bar{b}^*, \nu^*) \) with mean \( \bar{b}^* \) and variance covariance matrix \( \nu^* \), which is determined from
\[ \nu^{*-1} = \nu^{-1} + \nu^{-1} \quad \ldots (3.2.28) \]

\[ \bar{\nu}^* = \nu^*(\nu^{-1}\bar{\nu}' + \nu^{-1}\dot{\nu}) \quad \ldots (3.2.29) \]

The equation (3.2.28) can also be written as:

\[ G^* = G' + G = \sigma^2 \nu^{*-1} + z'z \]

and equation (3.2.29) as:

\[ \bar{\nu}^* = G^{*-1}(G'\bar{\nu}' + G\dot{\nu}) = G^{*-1}(G'\bar{\nu}' + g) \]

These results are due to Raiffa and Schlafer (1961).

Thus, when a normal prior distribution is selected, the posterior also will be normal and, more importantly, the parameters of the posterior can be determined from simple algebraic combinations of the prior parameters and the results of least squares analysis of the time series data.

3.2.6 BAYESIAN FORECASTING

The posterior distribution of \( b \), computed at time \( t' \) is given by the equation: \( h_b(b/b) = N(\bar{\nu'}, \nu') \). The forecast for period \( t+k \) is to be based on the posterior distribution of \( x_{t+k} \), the mortality rate in period \( t+k \).

Considering \( E(\;X_{t+k}) = \sum b_i z_i(t+k) \) and the \( (b_i) \) are jointly distributed.

\[ E(x_{t+k}) \] is a normally distributed random variable we set its mean as:
\[ \sum_{i=1}^{\hat{b}} z_i(t+k) = z'(t+k) \hat{b} \]  

... (3.2.30)

and variance

\[ \sum_{i=1}^{\hat{b}} \sum_{j=1}^{\hat{b}} z_i(t+k)z_j(t+k)\nu'' = z'(t+k)\nu'z(t+k) \]  

... (3.2.31)

Further, since \( X_{t+k} = E(x_{t+k}) + \varepsilon_{t+k} \) and \( \varepsilon_{t+k} \) is \( N(0, \sigma_{\varepsilon}^2) \) and assumed to be independent of \( b \), it is considered that \( X_{t+k} \) is a normally distributed random variable having posterior mean given in the above equation (3.2.30) and variance

\[ z'(t+k)\nu'Z(t+k) + \sigma_{\varepsilon}^2 \]  

... (3.2.32)

Therefore, the forecast error for period \( t+k \) is:

\[ e_{t+k} = x_{t+k} - \hat{x}_{t+k}(t). \]

is normally distributed with mean 0 and variance given in the equation (3.2.32) by defining

\[ m(t+k) = \sum_{i=1}^{\hat{b}} \hat{b}z_i(t+k) = z'(t+k)\hat{b} \]

and \[ s(t+k) = \sqrt{z'(t+k)\nu'Z(t+k) + \sigma_{\varepsilon}^2} \],

the 100 \((1-r)\)% prediction interval for \( x_{t+k} \) is given as:

\[ m(t+k) \pm u_{r/2}s(t+k) \]  

... (3.2.33)
3.3 Autoregressive Moving Average Model (ARMA)

Given a time series data $X_t$, the ARMA model is a total for understanding and predicting future values in the series. The model consists of two parts - an autoregressive (AR) part and a moving average (MA) part. The model is usually referred to as the ARMA $(p,q)$ model where 'p' is the order of the autoregressive part and 'q' is the order of the moving average part.

Autoregressive Model

The notation AR $(p)$ refers to the autoregressive model of order $p$. The AR $(p)$ model is written

$$X_t = \sum_{i=1}^{p} \phi X_{t-i} + \epsilon_t$$  \hspace{1cm} (3.3.1)

where the independent chance elements $\epsilon$ are identically distributed with zero mean and variance $\sigma^2$; and $\phi$ is a parameter.

This type of process is called autoregressive process since its form represents a regression of $X_t$ on $X_{t-1}$.

Moving Average Model:

The notation MA $(q)$ refers to the moving average model of order $q$.

$$X_t = \epsilon_t + \sum_{i=1}^{q} \theta_i \epsilon_{t-i}$$  \hspace{1cm} (3.3.2)

where $\theta_1, \theta_2, ..., \theta_q$ are the parameters of the model, and $\epsilon_t, \epsilon_{t-1}, ...$ are the error terms.

This process is called the moving average process. The general expression for such a process is
\[ X_t = \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \cdots - \theta_q \varepsilon_{t-q}, \]

where \( \theta_1, \theta_2, \ldots, \theta_q \) are constants

This is referred to as a finite moving average process of order \( q \), MA (q). This series may continue to infinity, in which case it is termed as an infinite moving average process.

The combination of the two models results in the generalization. We take the first order moving average process as:

\[ X_t = \varepsilon_t - \theta \varepsilon_{t-1} \quad \text{...(3.3.3)} \]

and the first order autoregressive process as:

\[ X_t - \phi X_{t-1} = \varepsilon_t \quad \text{...(3.3.4)} \]

and by combining their structure, the following equations are obtained:

\[ X_t - \sum_{i=1}^p \phi_i X_{t-i} = \varepsilon_t - \sum_{i=1}^q \theta_i \varepsilon_{t-i} \]

\[ X_t = \varepsilon_t + \sum_{i=1}^p \phi_i X_{t-i} + \sum_{i=1}^q \theta_i \varepsilon_{t-i} \quad \text{...(3.3.5)} \]

This is referred to as autoregressive moving average model of order \((p,q)\), denoted by ARMA\((p,q)\).

The forecast, \( x_{t+1} \) is given by

\[ x_{t+1} = \phi x_t + \varepsilon_{t+1} - \theta \varepsilon_t \quad \text{...(3.3.6)} \]

So taking the conditional expectation at time ‘t’ give

\[ \bar{x}_t = \phi x_t - \theta \varepsilon_t , \]

the forecast error is \( \varepsilon_{t+1} = x_{t+1} - \bar{x}_t = \varepsilon_{t+1} \)
So, once again the one-step-ahead forecast errors are equal to the corresponding \( \varepsilon \) values. Thus, it may be written as \( \bar{x}_i = \phi x_i - \theta x_{i-1} \) or in terms of the previous forecast,

\[
x_i \bar{x}_i = (\phi - \theta) x_i + \theta \bar{x}_{i-1}
\]

If \( \phi = 1 \) and \( 0 < \theta < 1 \), this is identical to the forecasting formula of exponential smoothing. Thus, exponential smoothing is the minimum mean square error forecasting method for the model.

\[
x_i = x_{i-1} + \varepsilon_i - \theta \varepsilon_{i-1}
\] ... (3.3.7)

Here each value of \( x_i \) equals the previous value \( x_{i+1} \) plus a random term from a simple moving average stochastic process.

For example, consider now a prediction for a lead time of two in the original model:

\[
\bar{x}_{i,2} = E[\phi x_{i+1} + \varepsilon_i + \varepsilon_{i+1} / x_i, \ldots] = \phi E[x_{i+1} / x_i, \ldots] = \phi \bar{x}_{i,1}.
\]

The same calculation is applied to any lead time, \( k \), say \( \bar{x}_{i,k} = \phi \bar{x}_{i,k-1}(k \geq 2) \).

So, \( \bar{x}_{i,k} = \phi^{k-1} \bar{x}_{i,1} \).

Consider now the problem of updating the forecasts. By substituting for in this last expression;

\[
\bar{x}_{i,k} \text{ is given by}
\]

\[
\bar{x}_{i,k} = \phi^{k-1}(\phi - \theta)x_i + \theta \phi^{k-1} \bar{x}_{i-1,1}.
\]

Therefore, \( \bar{x}_{i,k} = \theta \bar{x}_{i,k-1} + \phi^{k-1}(\phi - \theta)x_i \).
The model may be generalized as:
\[ x_t - \phi_1 x_{t-1} - \ldots - \phi_p x_{t-p} = \epsilon_t - \theta_1 \epsilon_{t-1} - \ldots - \theta_q \epsilon_{t-q} \ldots (3.3.8) \]

Finally, summarizing the forecasting formulae for MA, AR and mixed models, the expression for \( x_{t+k} \) is given by the model:
\[ X_{t+k} = \text{terms in } X_{t+1}, \ldots, X_t, X_{t-1}, \ldots, \epsilon_{t+k}, \epsilon_t, \ldots, \epsilon_{t-1}; \]
which is a linear function in time.

3.4 Regression Method:

The regression method applied in estimating parameters of time series model is described by a simple linear function of time.

The notation is
\[ X_t = a + bt + \epsilon_t \ldots (3.4.1) \]

Consider now estimating the unknown parameters \( a \) and \( b \) by least squares regression. The only assumption required regarding the random error component is:

\[ E(\epsilon_t) = 0, V(\epsilon_t) = \sigma^2, \text{and } \text{cov}(\epsilon_i, \epsilon_j) = 0 \text{ for } i \neq j. \]
That is, the average error is zero, the error variance is \( \sigma^2 \), and the errors are uncorrelated, random variables.

There are \( T \) periods of data available, say \( x_1, x_2, \ldots, x_T \) to estimate \( a \) and \( b \) say by \( \hat{a}, \hat{b} \) such that the sum of squares of residuals.
\[ SS_E = \sum_{i=1}^{T} (x_i - \hat{\alpha} - \hat{\beta}t)^2 \] 

...(3.4.2)

is minimized and also satisfy that \( \hat{\alpha} \) and \( \hat{\beta} \) are derived as:

\[ \frac{\partial SS_E}{\partial \hat{\alpha}} = -2 \sum_{i=1}^{T} [x_i - \hat{\alpha} - \hat{\beta}t] = 0 \] 

...(3.4.3)

and

\[ \frac{\partial SS_E}{\partial \hat{\beta}} = -2 \sum_{i=1}^{T} [x_i - \hat{\alpha} - \hat{\beta}t] = 0 \] 

...(3.4.4)

equation (3.4.3) and (3.4.4) it may written as

\[ \hat{\alpha} \sum_{i=1}^{T} (1) + \hat{\beta} \sum_{i=1}^{T} t = \sum_{i=1}^{T} x_i \] 

...(3.4.5)

\[ \hat{\alpha} \sum_{i=1}^{T} t + \hat{\beta} \sum_{i=1}^{T} t^2 = \sum_{i=1}^{T} x_i \] 

...(3.4.6)

Equations (3.4.5) and (3.4.6) are called the least squares normal equations. Using \( \sum_{i=1}^{T} t = T(T+1)/2 \) and \( \sum_{i=1}^{T} t^2 = T(T+1)(2T+1)/6 \), the solution to the normal equations is found as:

\[ \hat{\alpha} = -\frac{2(T+1)}{T(T+1)} \sum_{i=1}^{T} x_i - \frac{6}{T(T-1)} \sum_{i=1}^{T} x_i \] 

...(3.4.7)

\[ \hat{\beta} = \frac{12}{T(T+1)} \sum_{i=1}^{T} x_i - \frac{6}{T(T-1)} \sum_{i=1}^{T} x_i \] 

...(3.4.8)

\( \hat{\alpha}(t), \hat{\beta}(t) \) are the unknown parameters as a function of time, where \( T \) is the time at which the estimates are computed. The forecast for some future period, say \( T+t \) is

\[ \hat{x}_{T+t} = \hat{\alpha}(T) + \hat{\beta}(T)(T + t) \] 

...(3.4.9)
Example 3: The weekly demand record following the introduction of a new drug is available. Use this data to estimate the parameters in the linear trend model.

<table>
<thead>
<tr>
<th>Week (t)</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Drug demand</td>
<td>10</td>
<td>12</td>
<td>15</td>
<td>18</td>
<td>20</td>
</tr>
</tbody>
</table>

Since \( \sum_{i=1}^{2} x_i = 75, \sum_{i=1}^{5} x_i = 251. \)

\[
\hat{a}(5) = \frac{2(11)(75) - 6(251)}{5(4)} = 7.2
\]

\[
\hat{b}(5) = \frac{12}{5(24)}(251) - \frac{6}{5(4)}(75) = 2.6
\]

the forecast of drug demand for the next week, that is \( r = 1 \) is

\[
x \hat{x}_r = 7.2 + 2.6(6) = 22.8 \approx 23
\]

So, finally considered a simple linear relationship provides a better model for a time series data than other models.

### 3.5 COMPARISON OF SLOPES AND INTERCEPTS

Kleibaum and Kupper (1978) have suggested statistical tests for testing the equality between slopes and intercepts. For testing the null hypothesis that the correlation between X and Y in one population is the same as correlation between X and Y in another population, the procedure developed by R.A.Fisher in 1921 is used.

For comparing correlation coefficients, one should also compare slopes. It is possible that the slope for predicting Y from X to be different in one population than in the another while the correlation between X and Y is identical in the two populations, and it is also quite possible for the correlation between X and Y to be different in one population than in the other while the slopes are identical.
The test on slopes uses a pooled error term. If the variance in the dependent variable is much greater in one group than in the other, alternative methods are suggested by Kleiassum and Kupper.