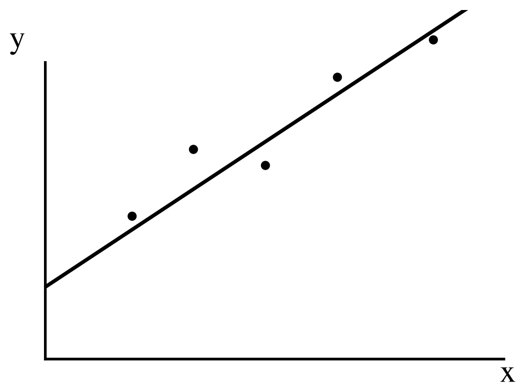
Chapter 2 Autoregressive Moving Average Models

Simple Linear Regression Models

Idea: to express the dependence of one set of observations y_t on another set x_t

Assumption: y_t's are independent or uncorrelated.



y_1	y_2	y_3	... y_n	response variable
x_1	x_2	x_3	... x_n	predictor variable

The dependence of y on x can be expressed by a linear regression model:

$$y_t = \beta_0 + \beta_1 x_t + \varepsilon_t \qquad t = 1, 2, \dots, n$$

After removing the mean from data:

$$y_t = \beta_1 x_t + \varepsilon_t \qquad t = 1, 2, \dots, n,$$

Above model can "best fit" the given data sets in the sense of minimizing the sum of squares of the residuals ε_t: i.e.,

$$\sum \varepsilon_t^2 = \sum (\dot{y}_t - \beta_0 - \beta_1 \dot{x}_t)^2$$

By differentiating with respect to β_0, β_1,

$$\frac{\partial \Sigma\ \varepsilon_t^2}{\partial \beta_0} = 0 \quad \text{and} \quad \frac{\partial \Sigma\ \varepsilon_t^2}{\partial \beta_1} = 0$$

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$$

$$\hat{\beta}_1 = \frac{\sum_{t=1}^{n} (\dot{y}_t - \bar{y})(\dot{x}_t - \bar{x})}{\sum_{t=1}^{n} (\dot{x}_t - \bar{x})^2} \qquad \bar{x} = \frac{1}{n}\sum_{t=1}^{n} \dot{x}_t \qquad \bar{y} = \frac{1}{n}\sum_{t=1}^{n} \dot{y}_t$$

$$\hat{\beta}_1 = \frac{\sum_{t=1}^{n} y_t\, x_t}{\sum_{t=1}^{n} x_t^2} \qquad \text{and} \quad \varepsilon_t \text{ -- NID}(o,\ \sigma_e^2) \qquad \sigma_\varepsilon^2 = \frac{1}{n-1}\sum_{t=1}^{n} (y_t - \hat{\beta}_1 x_t)^2$$

Remarks:

- y_t is decomposed into two parts:

  $\beta_1 x_t$ and ε_t
- When y_t is not known, ε_t is a random variable and is assumed to be normally independently distributed.
- The linear regression model can be used to predict the unknown value of y_t from x_t if a causation is assumed between y and x. (e.g., pressure vs. temp)

  $$\hat{y}_t = \beta_1\, x_t$$

  The prediction is subjected to errors which will be bounded by certain probability limits.

  e.g., 95% probability limits for will be

  $$\hat{y}_t \pm 1.96\, \sigma_\varepsilon$$
- Under the assumptions of the standard statistical model, the least squares estimates are unbiased: $E(\hat{\beta}_j) = \beta_j$, for j=0,1
- An unbiased estimate of σ^2 = Var(y)

  $$s^2 = \frac{RSS}{n-2} = \frac{\sum_{i=1}^{n} (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2}{n-2}$$

Regression Regarding Multiple Variables

<u>Idea</u>: to express the dependence of one variable y_t on several variables x_{t1}, x_{t2},..x_{tn}

<u>Assumption</u>: y_t's are independent or uncorrelated.

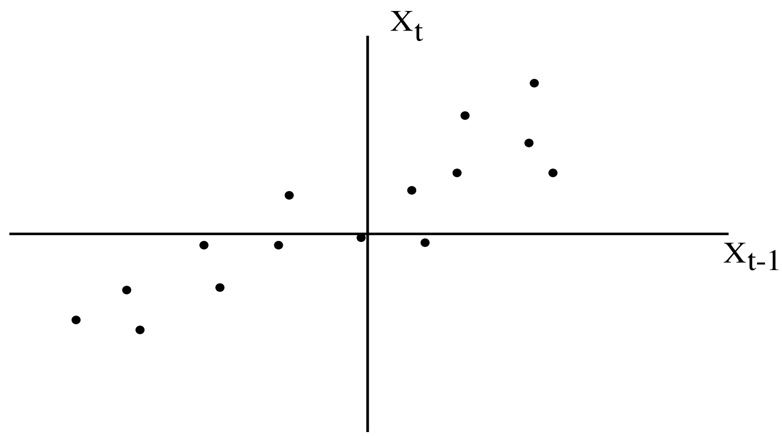
$$y_t = \beta_1 x_{t1} + \beta_2 x_{t2} + \ldots + \beta_n x_{tn} + \varepsilon_t \qquad t = 1, 2, \ldots, N$$

$$Y = \begin{bmatrix} y_1 \\ y_2 \\ . \\ . \\ y_N \end{bmatrix} \qquad X = \begin{bmatrix} x_{11} & x_{12} & .. & x_{1n} \\ x_{21} & x_{22} & .. & x_{2n} \\ . & . & .. & . \\ x_{N1} & x_{N2} & .. & x_{Nn} \end{bmatrix} \qquad \boldsymbol{\beta} = \begin{vmatrix} \beta_1 \\ \beta_2 \\ . \\ . \\ \beta_n \end{vmatrix}$$

$$\hat{\beta} = (X'X)^{-1} X'Y$$

First Order Autoregressive Models

Idea: the dependence between X_t and X_{t-1},



Assumption: a_t's at different **t** are independent, that is, a_t is independent of a_{t-1}, a_{t-2}, etc. $a_t \sim NID(0, \sigma_a^2)$

Remarks:

- An AR(1) model can be interpreted as an orthogonal decomposition of X_t into two parts: $\phi_1 X_{t-1}$ and a_t.
- At time (t-1), when X_{t-1} is observed and known, the AR(1) model is the same as a regression model, thus called "conditional regression".
- At time (t-1), since X_t is an unknown random variable, so is a_t. But, as soon as X_t is observed, a_t is a fixed number and can be computed by
  $$a_t = X_t - \phi_1 X_{t-1}$$
- An AR(1) model can also be seen as a device to transform or reduce dependent data to independent data.
- If X_{t-1} is fixed, above eq. gives a static conditional regression model.
- If X_{t-1} is not fixed, it gives a dynamic model.
  Since the observation X_{t-1} itself depends on X_{t-2} by the same model and recursively substituting,
  $$X_t = \sum_{j=0}^{\infty} \phi^j a_{t-j}$$
  It shows how the past shocks or excitations affect the present observation or how a_{t-j} are remembered.

4. Estimation:

For a given set of data, the estimates of ϕ_1 and σ_a^2 can be obtained by conditional least squares estimates

$$\hat{\phi}_1 = \frac{\sum_{t=2}^{N} X_t X_{t-1}}{\sum_{t=2}^{N} X_{t-1}^2} \qquad \sigma_a^2 = \frac{1}{N-1}\sum_{t=2}^{N}(X_t - \hat{\phi}_1 X_{t-1})^2$$

biased

$X_{t:}$	X_2	X_3	X_4	....	X_N
$X_{t-1:}$	X_1	X_2	X_3	....	X_{N-1}

Discussion:

- The parameter ϕ_1 measures the extent of the dependence of X_t on X_{t-1}: The stronger the dependence, the larger ϕ_1 will be in magnitude.
- The estimated ϕ_1 gives an estimate of the dependence or relation between the values of X_t one "lag" apart. This is also called the estimated autocorrelation at lag one:

$$\hat{\rho}_1 = \frac{\sum_{t=2}^{N} X_t X_{t-1}}{\sum_{t=2}^{N} X_{t-1}^2}$$

5. Prediction or Forecasting using AR(1):

$$X_t = \phi_1 X_{t-1} + a_t$$

one step ahead prediction at time t-1 is:

$$\hat{X}_{t-1}(1) = \phi_1 X_{t-1} \qquad e_{t-1}(1) = X_t - \hat{X}_{t-1}(1) = a_t$$

$e_{t-1}(1)$ is prediction error and has the NID(0, σ_a^2)

95% probability limits for the prediction are:

$$\hat{X}_{t-1}(1) \pm 1.96\,\sigma_a \qquad \text{that is,} \quad \phi_1 X_{t-1} \pm 1.96\,\sigma_a$$

Random Walk as a limit of AR(1)

In general, an AR(1) model is a good approximation for many systems characterized by inertia.

e.g.

for IBM stock prices data,

/

From figs. 2-13 & 2-14, it can be seen that above AR(1) model is adequate.

$$X_t = X_{t-1} + a_t \quad \text{or} \quad X_t - X_{t-1} = a_t \quad \nabla X_t = a_t$$

Remarks:

- The system is characterized by high inertia, or strong dependence / memory.
- Its response or value remains unchanged from t-1 to t, except for an random independent increment a_t. But, $E(a_t)=0$, the system would stay in the same position indefinitely.
- $\hat{X}_{t-1}(1) = X_{t-1}$

AR(2) model

$$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + a_t$$

conditional multiple linear regression

$$Y = \begin{vmatrix} X_3 \\ X_4 \\ X_N \end{vmatrix} \qquad X = \begin{vmatrix} X_2 & X_1 \\ X_3 & X_2 \\ X_{N-1} & X_{N-2} \end{vmatrix}$$

$$\begin{bmatrix} \hat{\phi}_1 \\ \hat{\phi}_2 \end{bmatrix} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$$