Autocorrelation - particularly in data taken over time

We assume errors are independent. What if they are not, such that the residual at time t is correlated to the previous residual at time t-1.

This may cause runs of +++ or $- - -$ in the residuals. This could also be due to some cyclical variable (moon phase, tidal cycles) which in some cases could be fitted with other variables,

or it could be due to serial correlation which is addressed with time series

The possible effects of autocorrelation are

1) The β_i estimates are unbiased, but not minimum variance (variance estimate is inefficient).

2) MSE may underestimate the variance

3) s_b estimates may be underestimated

4) the hypothesis tests and CI's may not be valid

How serious the problems are depends on the strength of the correlation. We need a measure of and test of this correlation.

$$Y_t = \beta_0 + \beta_1 X_t + \epsilon_t$$

where

$$\epsilon_t = \epsilon_{t-1} + \mu_t$$

the t subscript is used to show that these are serial measurements (usually time, but not necessarily) and are not taken at random

The first order autoregressive model,

$$Y_t = \beta_0 + \beta_1 X_t + \epsilon_t$$

where

$$\epsilon_t = \rho\epsilon_{t-1} + \mu_t$$

The error term is a linear combination of the of the current and preceding terms, it is cumulative such that

$$\epsilon_t = \sum_{i=0}^{\infty} \rho^s \mu_{t-s}$$

and

$$E(\epsilon_t) = 0$$

$$Var(\epsilon_t) = \frac{\sigma^2}{1-\rho^2}$$

$$Covar(\epsilon_t, \epsilon_{t-1}) = \rho\frac{\sigma^2}{1-\rho^2}$$ (did not exist in previous models)

Testing for, and estimating, autocorrelation

Durban - Watson test statistic

H_0: $\rho = 0$

H_1: $\rho > 0$

$$e = Y_t - \hat{Y}_t$$

$$D = \frac{\sum_{t=2}^{n}(e_t - e_{t-1})^2}{\sum_{t=1}^{n} e_t^2}$$

an exact test is not possible, but your test gives upper and lower bounds (A6)

if $D > d_u$ conclude H_0

$D < d_l$ conclude H_1

note that smaller D values are significant

Remedial measures

1) Transformation
a) estimate ρ as "r", and remove its effect

The transformation is based on the property that,

$$Y'_t = Y_t - \rho Y_{t-1}$$

where

$$Y_t = \beta_0 + \beta_1 X_t + \epsilon_t$$

then

$$Y'_t = (\beta_0 + \beta_1 X_t + \epsilon_t) - \rho(\beta_0 + \beta_1 X_t + \epsilon_t)$$

$$Y'_t = \beta_0(1 - \rho) + \beta_1(X_t - \rho X_{t-1}) + (\epsilon_t - \rho\epsilon_{t-1})$$

$$Y_t = \beta'_0 + \beta'_1 X'_t + \mu_t$$

where

$$Y'_t = Y_t - \rho Y_{t-1} \text{ , estimated by } Y_t - rY_{t-1}$$

$$X'_t = X_t - \rho X_{t-1} \text{ , estimated by } X_t - rX_{t-1}$$

$$\beta'_0 = \beta_0(1 - \rho)$$

$$\beta'_1 = \beta_1$$

Estimating r

Cochrane - Orcutt procedure

$\epsilon_t = \rho\epsilon_{t\text{-}1} + \mu_t$ suggests that if we have
e_t and $e_{t\text{-}1}$ we could regress through the origin to get r ,
an alternative calculation is

$$r = \frac{\sum_{t=2}^{n} e_{t\text{-}1} * e_t}{\sum_{t=2}^{n} e_{t\text{-}1}^2}$$

Hildreth - Lu procedure (like Box-Cox transformation approach)
Minimize r by some procedure, such as a numerical search,
Try values of r from 0 to 1, and use best estimate

First difference procedure
Use OLS where r = 1

SEE HANDOUTS for each technique