REGRESSION WITH CORRELATED ERRORS

C.A. GLASBEY

SYSTEMATIC RESIDUALS

When data exhibit systematic departures from a fitted regression line (see for example Figs 1 and 2), either the regression function is inappropriate, or the errors are correlated, or both. In most cases it is assumed that the function is deficient, and it is changed. But there are situations where the assumption of independent errors is not wholly plausible. For example, some sources of error will persist over several observations when repeated measurements are made on a single experimental unit.

Systematic departures may be modelled either by another regression function, or by correlated errors. To illustrate, consider

$$y_i = a + bx_i + c \sin x_i + e_i \qquad i=1, \ldots, n,$$

where a, b, $x_1, \ldots, x_n$ are constants, c is normally distributed with mean 0, variance τ^2, and $e_1, \ldots, e_n$ are independently normally distributed with means 0, variances σ^2. Two models are equally valid for the y's, either they are independently normally distributed with means $a+bx_i + c\sin x_i$, variances σ^2, or they are correlated, with means $a+bx_i$, variances $\tau^2 \sin^2 x_i + \sigma^2$ and covariances $\tau^2 \sin x_i \sin x_j$ between y_i and y_j.

In general the modelling objective determines the choice: for a simple summary it may be preferable for the regression function to explain all systematic variability, whereas a correlated stochastic component may be of more assistance in understanding the data generating mechanism. A succinct summary of data is often achieved by using the regression function to describe the long-term trends and the correlations the short-term fluctuations. This will be the assumed case from now on.

CORRELATED ERRORS

In the presence of correlated errors, ordinary least squares regression parameter estimators remain unbiased, but they may be inefficient, and the conventional estimators of the variances of