

Box-Tidwell transformation of the X s

Now we consider transformations of the X 's with the model

$$Y_i = \alpha + \beta_1 X_{i1}^{\gamma_1} + \cdots + \beta_k X_{ik}^{\gamma_k} + \varepsilon_i,$$

where $\varepsilon_i \sim N(0, \sigma_\varepsilon^2)$ and all X_{ij} are positive. Although optimal values of these parameters can be found by a general nonlinear least squares approach, Box and Tidwell (1962) present an iterative procedure based on constructed variables that can also yield diagnostic plots.

More Tests for Nonconstant Error Variance

Breusch and Pagan (1979) proposed a score test based on the model

$$\sigma_i^2 = V(\varepsilon_i) = g(\gamma_0 + \gamma_1 Z_{i1} + \cdots + \gamma_p Z_{ip}),$$

that produces a test for $H_0 : \gamma_1 = \cdots = \gamma_p = 0$ using an auxiliary-regression approach. Let $U_i = E_i^2 / \hat{\sigma}_\varepsilon^2$, where $\hat{\sigma}_\varepsilon^2 = \sum E_i^2 / n$, and fit the model

$$U_i = \eta_0 + \eta_1 Z_{i1} + \cdots + \eta_p Z_{ip} + \omega_i.$$

The test statistic is then

$$S_0^2 = \frac{\sum (\hat{U}_i - \bar{U})^2}{2}$$

and is asymptotically χ^2 distributed with p degrees of freedom under H_0 . The variables (Z_j 's) in the above regression can be covariates from the original regression (X_j 's) or fitted values from the original regression (\hat{Y}_i 's as suggested by Cook and Weisberg, 1983, in which case S_0^2 has 1 degree of freedom). Anscombe (1961) also addressed this problem and suggested a transformation to correct for heteroscedasticity via a power transformation $Y^{(\tilde{\lambda})}$ with $\tilde{\lambda} = 1 - \frac{1}{2} \hat{\eta}_1 \bar{Y}$. White (1980) also developed similar approach. Since these tests are based on fitting regression models, standard plots can be produced to diagnose potential problems.

Structural Dimension

The structural dimension of a regression problem is the smallest subspace of the X_j 's required to represent the dependency of Y on the X_j 's. This is an interesting concept that, if the covariates X_j are linearly related, also yields a diagnostic check of whether the structural dimension is 1, based on the structure of the inverse regression relationships $E(X_j|y)$ and $V(X_j|y)$.

These relationships can be routinely assessed via a scatterplot matrix of Y with the X_j 's.