Orthogonal Regressors, Correlated Regressors

H.M. James Hung

Without loss of generality, take a multiple linear regression model containing two regressors

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \varepsilon \qquad (1)$$

For n subjects rendering data $(y_{1i}, x_{1i}, x_{2i}), i = 1, \dots, n$, the OLS estimator for $(\beta_0, \beta_1, \beta_2)$ is

$$\begin{pmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \\ \hat{\beta}_2 \end{pmatrix} = (\mathrm{X}'\mathrm{X})^{-1} X'Y$$

Thus, $\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2$.
Note: $\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}_1 - \hat{\beta}_2 \bar{x}_2$
Hence $\hat{y} - \bar{y} = \hat{\beta}_1 (x_1 - \bar{x}_1) + \hat{\beta}_2 (x_2 - \bar{x}_2)$

Also note that from (1) in the last slide

$$\begin{aligned} y &= \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \varepsilon \\ &= (\beta_0 + \beta_1 \bar{x}_1 + \beta_2 \bar{x}_2) + \beta_1 (x_1 - \bar{x}_1) + \beta_2 (x_2 - \bar{x}_2) + \varepsilon \end{aligned}$$

Now denote $\beta_0^* = (\beta_0 + \beta_1 \bar{x}_1 + \beta_2 \bar{x}_2)$, a new intercept.

It can be shown that applying "centering regressors, x_1 and x_2, at their respective sample means" yields identical OLS estimators for β_1 and β_2 ; that is,

$$\begin{pmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \end{pmatrix} = \begin{bmatrix} \sum(x_{1i} - \bar{x}_1)^2 & \sum(x_{1i} - \bar{x}_1)(x_{2i} - \bar{x}_2) \\ \sum(x_{1i} - \bar{x}_1)(x_{2i} - \bar{x}_2) & \sum(x_{2i} - \bar{x}_2)^2 \end{bmatrix}^{-1} \begin{pmatrix} \sum(x_{1i} - \bar{x}_1)y_i \\ \sum(x_{2i} - \bar{x}_2)y_i \end{pmatrix}$$

x_1 and x_2 are **orthogonal** if $\sum(x_{1i} - \bar{x}_1)(x_{2i} - \bar{x}_2) = 0$.

When the regressors x_1 and x_2 are random variables, $\sum(x_{1i} - \bar{x}_1)(x_{2i} - \bar{x}_2)$ captures the estimated covariance between x_1 and x_2 . Correlation is covariance divided by the square root of the product of the variances of x_1 and x_2 . When the correlation between x_1 and x_2 are nonzero, they are correlated.

When the regressors x_1 and x_2 are fixed, there is no concept of correlation but "correlation" can be used loosely.