Econometrics 1

Lecture 7

Multicollinearity

What is multicollinearity

Multiple regression model:

$$Y_i = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + ... + \beta_k X_k + e_i$$

Exact or perfect linear relation among the independent variables:

$$\lambda_1 X_1 + \lambda_2 X_2 + \lambda_3 X_3 + \lambda_4 X_4 + ... + \lambda_k X_k = 0$$

Inexact or imperfect linear relation among the independent variables:

$$\lambda_1 X_1 + \lambda_2 X_2 + \lambda_3 X_3 + \lambda_4 X_4 + ... + \lambda_k X_k + v_i = 0$$

X1	10	15	18	24	30
X2	50	75	90	120	150

OLS Estimates in a Multiple Regression Model

$$Y_i = \beta_1 + \beta_2 X_{1,i} + \beta_3 X_{2,i} + e_i \quad (7.4)$$

$$\hat{\beta}_2 = \frac{\sum_i y_i x_{1,i} \sum_i x_{2,i}^2 - \sum_i x_{1,i} x_{2,i} \sum_i y_i x_{2,i}}{\sum_i x_{1,i}^2 \sum_i x_{2,i}^2 - \left(\sum_i x_{1,i} x_{2,i} \right)^2} \quad (7.10)$$

$$\hat{\beta}_3 = \frac{\sum_i y_i x_{2,i} \sum_i x_{1,i}^2 - \sum_i x_{1,i} x_{2,i} \sum_i y_i x_{1,i}}{\sum_i x_{1,i}^2 \sum_i x_{2,i}^2 - \left(\sum_i x_{1,i} x_{2,i} \right)^2} \quad (7.11)$$

$$\hat{\beta}_1 = \bar{Y} - \hat{\beta}_2 \bar{X}_1 - \hat{\beta}_3 \bar{X}_2$$

Mutlicollinearity

$$\hat{\beta}_3 = \frac{\sum_i y_i x_{2,i} \sum_i x_{1,i}^2 - \sum_i x_{1,i} x_{2,i} \sum_i y_i x_{1,i}}{\sum_i x_{1,i}^2 \sum_i x_{2,i}^2 - \left(\sum_i x_{1,i} x_{2,i} \right)^2} \qquad (7.11)$$

- consequences:

Here x1 is constant, x2 and x3 are explanatory variables. Further assume that x2 and x3 are perfectly correlated: $x_{2,i} = \lambda x_{1,i}$. Then (7.10) and (7.11) become as following:

$$\hat{\beta}_2 = \frac{\lambda^2 \sum_i y_i x_{1,i} \sum_i x_{1,i}^2 - \lambda^2 \sum_i x_{1,i}^2 \sum_i y_i x_{1,i}}{\lambda^2 \left(\sum_i x_{1,i}^2 \right)^2 - \lambda^2 \left(\sum_i x_{1,i}^2 \right)^2} = \frac{0}{0} = \infty \qquad (7.12)$$

$$\hat{\beta}_3 = \frac{\lambda \sum_i y_i x_{1,i} \sum_i x_{1,i}^2 - \lambda \sum_i x_{1,i}^2 \sum_i y_i x_{1,i}}{\lambda^2 \left(\sum_i x_{1,i}^2 \right)^2 - \lambda^2 \left(\sum_i x_{1,i}^2 \right)^2} = \frac{0}{0} = \infty \qquad (7.13)$$

This is the proof of the fact that when two variables are exactly correlated to each other the least square procedure completely breaks down.

Multicollinearity: A Numerical Example

Observations	x1	x2	x1x2	x1sq	x2sq
1	1	5	5	1	25
2	1.5	7.5	11.25	2.25	56.25
3	1.8	9	16.2	3.24	81
4	2.4	12	28.8	5.76	144
5	3	15	45	9	225
Total	9.7	48.5	106.25	21.25	531.25
	1.94	9.7			

$$\sum x_1^2 = \sum \left(X_1 - \bar{X}_1\right)^2 = \sum X_1^2 - N\bar{X}_1^2 = 21.25 - 5 \times (1.94)^2 = 2.432$$

$$\sum x_2^2 = \sum \left(X_2 - \bar{X}_2\right)^2 = \sum X_2^2 - N\bar{X}_2^2 = 531.25 - 5 \times (9.7)^2 = 60.8$$

$$\sum x_1 x_2 = \sum \left(X_1 - \bar{X}_1\right)\left(X_2 - \bar{X}_2\right) = \sum X_1 X_2 - N\bar{X}_1\bar{X}_2 = 106.25 - 5(1.94)(9.7) = 12.16$$

The numberator of the above formula collapses to zeoro:

$$\sum x_1^2 \sum x_2^2 - \left(\sum x_1 x_2\right)^2 = 2.432(60.8) - (12.16)^2 = 147.87 - 147.87 = 0$$

Mutlicollinearity: In Matrix Notation

$$Y_i = \beta_1 + \beta_2 X_{1,i} + \beta_3 X_{2,i} + e_i \qquad (7.4)$$

$$\hat{\beta} = \begin{bmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \\ \hat{\beta}_3 \end{bmatrix} = \underbrace{\begin{bmatrix} N & \sum_i X_{1,i} & \sum_i X_{2,i} \\ \sum_i X_{1,i} & \sum_i X_{1,i}^2 & \sum_i X_{1,i} X_{2,i} \\ \sum_i X_{2,i} & \sum_i X_{1,i} X_{2,i} & \sum_i X_{2,i}^2 \end{bmatrix}^{-1}}_{\text{X'X}} \underbrace{\begin{bmatrix} \sum_i Y_i \\ \sum_i Y_i X_{1,i} \\ \sum_i Y_i X_{2,i} \end{bmatrix}}_{\text{X'Y}} \qquad (7.5)$$

$$\beta = \left(X' X \right)^{-1} X' Y \qquad (7.6)$$

We can solve this system using the Cramer’s Rule. First write in the deviation form.

$$\begin{bmatrix} \hat{\beta}_2 \\ \hat{\beta}_3 \end{bmatrix} = \begin{bmatrix} \sum_i x_{1,i}^2 & \sum_i x_{1,i} x_{2,i} \\ \sum_i x_{1,i} x_{2,i} & \sum_i x_{2,i}^2 \end{bmatrix}^{-1} \begin{bmatrix} \sum_i y_i x_{1,i} \\ \sum_i y_i x_{2,i} \end{bmatrix} \qquad (7.7)$$

Mutlicollinearity: Cramer's Rule for Estimation OLS Parameters

$$\hat{\beta}_2 = \frac{\begin{vmatrix} \sum_i y_i x_{1,i} & \sum_i x_{1,i} x_{2,i} \\ \sum_i y_i x_{2,i} & \sum_i x_{2,i}^2 \end{vmatrix}}{\begin{vmatrix} \sum_i x_{1,i}^2 & \sum_i x_{1,i} x_{2,i} \\ \sum_i x_{1,i} x_{2,i} & \sum_i x_{2,i}^2 \end{vmatrix}} = \frac{\sum_i y_i x_{1,i} \sum_i x_{2,i}^2 - \sum_i x_{1,i} x_{2,i} \sum_i y_i x_{2,i}}{\sum_i x_{1,i}^2 \sum_i x_{2,i}^2 - \left(\sum_i x_{1,i} x_{2,i} \right)^2} \quad (7.8)$$

$$\hat{\beta}_3 = \frac{\begin{vmatrix} \sum_i x_{1,i}^2 & \sum_i y_i x_{1,i} \\ \sum_i x_{1,i} x_{2,i} & \sum_i y_i x_{2,i} \end{vmatrix}}{\begin{vmatrix} \sum_i x_{1,i}^2 & \sum_i x_{1,i} x_{2,i} \\ \sum_i x_{1,i} x_{2,i} & \sum_i x_{1,i}^2 \end{vmatrix}} = \frac{\sum_i y_i x_{2,i} \sum_i x_{1,i}^2 - \sum_i x_{1,i} x_{2,i} \sum_i y_i x_{1,i}}{\sum_i x_{1,i}^2 \sum_i x_{2,i}^2 - \left(\sum_i x_{1,i} x_{2,i} \right)^2} \quad (7.9)$$

$$\hat{\beta}_2 = \frac{\sum_i y_i x_{1,i} \sum_i x_{2,i}^2 - \sum_i x_{1,i} x_{2,i} \sum_i y_i x_{2,i}}{\sum_i x_{1,i}^2 \sum_i x_{2,i}^2 - \left(\sum_i x_{1,i} x_{2,i} \right)^2} \quad (7.10)$$

Breakdown of OLS Estimation In Case of Multicollinearity in Matrix

$$\sum x_1^2 = \sum \left(X_1 - \overline{X}_1\right)^2 = \sum X_1^2 - N\overline{X}_1^2 = 21.25 - 5 \times (1.94)^2 = 2.432$$

$$\sum x_2^2 = \sum \left(X_2 - \overline{X}_2\right)^2 = \sum X_2^2 - N\overline{X}_2^2 = 531.25 - 5 \times (9.7)^2 = 60.8$$

$$\sum x_1 x_2 = \sum \left(X_1 - \overline{X}_1\right)\left(X_2 - \overline{X}_2\right) = \sum X_1 X_2 - N\overline{X}_1 \overline{X}_2 = 106.25 - 5(1.94)(9.7) = 12.16$$

In the presence of multicollinearity $\left|X'X\right| = 0$

$$\begin{vmatrix} \sum_i x_{1,i}^2 & \sum_i x_{1,i}x_{2,i} \\ \sum_i x_{1,i}x_{2,i} & \sum_i x_{2,i}^2 \end{vmatrix} = \begin{vmatrix} 2.432 & 12.16 \\ 12.16 & 60.8 \end{vmatrix} = 147.86 - 147.68 = 0$$

General Formula for the OLS Parameters in Matrix Notation

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

or,

$$\begin{bmatrix} \beta_0 \\ \beta_1 \\ . \\ . \\ \beta_k \end{bmatrix} = \begin{bmatrix} N & \sum X_{1,i} & . & . & \sum X_{k,i} \\ \sum X_{1,i} & \sum X_{1,i}^2 & . & . & \sum X_{1,i} X_{k,i} \\ . & . & . & . & . \\ . & . & . & . & . \\ \sum X_{k,i} & \sum X_{1,i} X_{k,i} & . & . & \sum X_{k,i}^2 \end{bmatrix}^{-1} \begin{bmatrix} \sum Y_i \\ \sum Y_1 X_1 \\ . \\ . \\ \sum Y_k X_k \end{bmatrix}$$

$(X'X)^{-1}$ is singular in presence of perfect multicollinearity ; $|X'X| = 0$. Therefore $\hat{\beta}$ cannot be estimated.

Variance in Algebra and in Matrix Notation

- Variance with partial correlation relation among regressors (as in page 154 of the HGJ –equation 7.3.1) can derived from the matrix

$$Var(b_2) = \frac{\sigma^2}{\sum (X_{1,t} - \bar{X}_1)(1 - r_{2,3}^2)} \qquad (7.16)$$

Variance depends on σ^2, $(1 - r_{2,3}^2)$, variation around the mean $\sum (X_{1,t} - \bar{X}_1)$ and the number of observations. Here $\frac{1}{(1 - r_{2,3}^2)}$ is the variance inflation factor. Using the matrix method to find the variance

$$\begin{bmatrix} \hat{\beta}_2 \\ \hat{\beta}_3 \end{bmatrix} = \begin{bmatrix} \sum_i x_{1,i}^2 & \sum_i x_{1,i} x_{2,i} \\ \sum_i x_{1,i} x_{2,i} & \sum_i x_{2,i}^2 \end{bmatrix}^{-1} \begin{bmatrix} \sum_i y_i x_{1,i} \\ \sum_i y_i x_{2,i} \end{bmatrix} (7.17)$$

$$\text{var}(\hat{\beta}_2) = \frac{\sigma^2 \sum_i x_{2,i}^2}{\begin{vmatrix} \sum_i x_{1,i}^2 & \sum_i x_{1,i} x_{2,i} \\ \sum_i x_{1,i} x_{2,i} & \sum_i x_{2,i}^2 \end{vmatrix}} = \frac{\sigma^2 \sum_i x_{2,i}^2}{\sum_i x_{1,i}^2 \sum_i x_{2,i}^2 - \left(\sum_i x_{1,i} x_{2,i} \right)^2} \quad (7.18)$$

Variance in Algebra and in Matrix Notation

or by dividing both numerator and denominator by $\sum_i x_{2,i}^2$

$$\operatorname{var}\left(\hat{\beta}_2\right) = \frac{\sigma^2}{\dfrac{\sum x_{1,i}^2 \sum_i x_{2,i}^2}{\sum_i x_{2,i}^2} - \dfrac{\left(\sum_i x_{1,i} x_{2,i}\right)^2}{\sum_i x_{2,i}^2}} \quad (7.19)$$

$$\operatorname{var}\left(\hat{\beta}_2\right) = \frac{\sigma^2}{\sum x_{1,i}^2 \left[1 - \dfrac{\left(\sum_i x_{1,i} x_{2,i}\right)^2}{\sum x_{1,i}^2 \sum_i x_{2,i}^2}\right]} = \frac{\sigma^2}{\sum x_{1,i}^2 \left[1 - r_{1,3}^2\right]} \quad (7.20)$$

Obviously if $r_{1,3}^2$ =1, the $\operatorname{var}\left(\hat{\beta}_2\right) = \infty$. Covariannce cannot be estimated because $\left| X'X \right| = 0$

Consequences of Multicollinearity

Variance is infinite. This results in insignificant t-ratios.

$$t^* = \frac{\hat{\beta} - \beta}{SE\left(\hat{\beta}\right)} = \frac{\hat{\beta} - \beta}{\sqrt{\hat{\sigma}2}}$$

Siginificant (high) R^2 but insignificant t-ratios,

OLS estimators have larger variance and covariances

Wider confidence intervals

variance inflation factor

Detection of Multicollinearity

Detection: examination of partial correlations, auxiliary regressions,

Klien's rule of thumb

$$R_i = \frac{R^2_{x_1.x_2..X_k} \Big/ (k-2)}{\left(1 - R^2_{x_1.x_2..X_k}\right) \Big/ (n-k+1)}$$

Multicollinearity is problem only if R^2 from the auxiliary regression is bigger than the R^2 from the overall model.

Remedial Measures

- A-priori information, ie. $\beta_1 = 0.1\beta_2$ to eliminate a perticular variable.

- Pooling cross section and time series (Tobins method)

If a consumption demand panel data model is

$$Y_t = \beta_0 + \beta_1 \ln P_t + \beta_2 \ln I_t + e_i$$

If income and prices are correlated over time, first estimate relation $Y_t = \beta_0 \beta_2 \ln I_t + e_i$ for a cross section then use predicted values to estimate relation between income and prices.

$Y^*_t = \beta_0 + \beta_1 \ln P_t + e_i$ where $Y^*_t = \ln Y_t - \beta_2 \ln I_t + u_t$

dropping variables or specification bias

Remedial Measure

transformation of variables (i.e. first difference)

$$Y_t = \beta_0 + \beta_1 \ln P_t + \beta_2 \ln I_t + e_i$$

Iterate back

$$Y_{t-1} = \beta_0 + \beta_1 \ln P_{t-1} + \beta_2 \ln I_{t-1} + e_{t-1}$$

Then take the difference

$$Y_t - Y_{t-1} = \beta_0 + \beta_1 \ln\left(P_t - P_{t-1}\right) + \beta_2\left(I_t - I_{t-1}\right) + u_{t-1}$$

The first difference of the prices and investment may not be correlated.

- Enlarging the sample by adding new data

  Larger sample means more efficient estimates

  Other methods (factor analysis, principal component and ridge)

AKAIKE (1969) FINAL PREDICTION ERROR - FPE = 6.2051

(FPE IS ALSO KNOWN AS AMEMIYA PREDICTION CRITERION - PC)

$$\hat{\sigma}^2 \left(\frac{N + K}{N - K} \right)$$

AKAIKE (1974) INFORMATION CRITERION - AIC = 6.1713

$$\hat{\sigma}^2 \exp \frac{2K}{N}$$

AKAIKE (1973) INFORMATION CRITERION - LOG AIC = 1.8199

$$\ln \left(\hat{\sigma}^2 \right) + \frac{2K}{N}$$

SCHWARZ (1978) CRITERION - LOG SC = 1.8804

$$\ln \left(\hat{\sigma}^2 \right) + \frac{K \ln N}{N}$$

CRAVEN-WAHBA (1979) GENERALIZED CROSS VALIDATION - GCV = 6.4637

$$\hat{\sigma}^2 \left(1 - \frac{K}{N} \right)^{-2}$$

HANNAN AND QUINN (1979) CRITERION = 5.7749

$$\hat{\sigma}^2 (\ln N)^{\frac{2K}{N}}$$

RICE (1984) CRITERION = 6.8946

$$\hat{\sigma}^2 \left(1 - \frac{2K}{N} \right)^{-1}$$

SHIBATA (1981) CRITERION = 5.7915

$$\hat{\sigma}^2 \left(\frac{N + 2K}{N} \right)$$

SCHWARZ (1978) CRITERION - SC = 6.5563

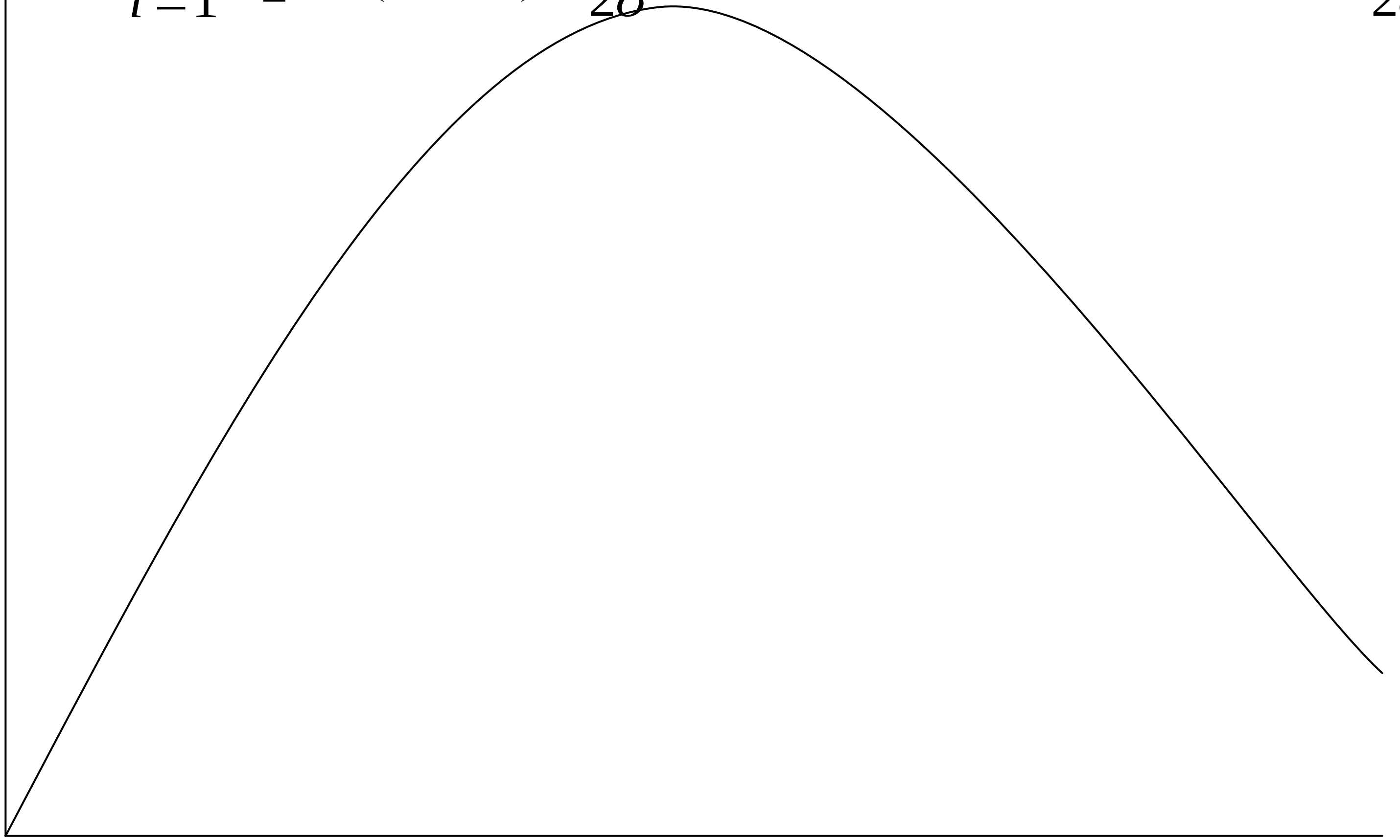
$$\hat{\sigma}^2 N^{\frac{K}{N}}$$

Given everything else a model with smaller values of these tests are preferable than with larger values.

Likelihood and Log-Likelihood Functions

$$Log(\alpha,\beta,\sigma)=\sum_{i=1}^{n}-\frac{1}{2}\log\left(2\pi\sigma^{2}\right)-\frac{1}{2\sigma^{2}}\left(y_i-\alpha-\beta x_i\right)^{2}=c-\frac{1}{2}\log\left(\sigma^{2}\right)-\frac{Q}{2\sigma^{2}}$$

Value of Likelihood function



Parameter Set, α,β,σ

$$L(\alpha,\beta,\sigma)=f(y_1,y_1,.....y_n)=\prod_{i=1}^{n}\left(\frac{1}{2\pi\sigma^{2}}\right)^{\frac{1}{2}}\exp\left[-\frac{1}{2\sigma^{2}}\left(y_i-\alpha-\beta x_i\right)^{2}\right]$$

Similarity Between the OLS and Maximum Likelihood Estimators

$$L(\alpha,\beta,\sigma) = f(y_1, y_1, \ldots.. y_n) = \prod_{i=1}^{n} \left(\frac{1}{2\pi\sigma^2} \right)^{\frac{1}{2}} \exp\left[-\frac{1}{2\sigma^2} \left(y_i - \alpha - \beta x_i \right)^2 \right] \quad (1)$$

Take log of this function to get a log-likelihood function

$$Log(\alpha,\beta,\sigma) = \sum_{i=1}^{n} -\frac{1}{2} \log\left(2\pi\sigma^2 \right) - \frac{1}{2\sigma^2} \left(y_i - \alpha - \beta x_i \right)^2 = c - \frac{1}{2}\log\left(\sigma^2\right) - \frac{Q}{2\sigma^2} \quad (2$$

where $c = -\frac{1}{2}\log(2\pi)$ and $Q = \left(y_i - \alpha - \beta x_i \right)^2$

Maximising (2) w.r.t. α, β, σ is equivalent to

minimizing $Q = \left(y_i - \alpha - \beta x_i \right)^2$, which is equivalent to the square of the error term and is a negative term in the likelihood function. Therefore the estimators of α and β under the ML method are the same as in the OLS method.

Large Sample Tests

- Likelihood ratio test

$$\lambda = \frac{\max L(\theta)_R}{\max L(\theta)_U}; \qquad -2\ln\lambda \text{ has a } \chi^2 \text{ distribution.}$$

- Wald Test : $\lambda_W = \frac{SSE_R - SSE_U}{\hat{\sigma}^2} \approx \chi_j^2$, where $\hat{\sigma}^2$ is estimated from unrestricted model.

  Which is equal to $\hat{\sigma}^2 = \frac{ESS_U}{T-K}$

- Lagrange Multiplier test is similar to the Wald test except that the estimated variance is from the restricted model. $\lambda_{LM} = \frac{SSE_R - SSE_U}{\hat{\sigma}_*^2} \approx \chi_j^2$ where $\hat{\sigma}_*^2 = \frac{ESS_R}{T-K+J}$