Regression Analysis

Formulations and Problems

Regression Analysis

Formulations and Problems

Assumptions of a Regression Model

$$Y_i = \beta_1 + \beta_2 x_i + e_i$$

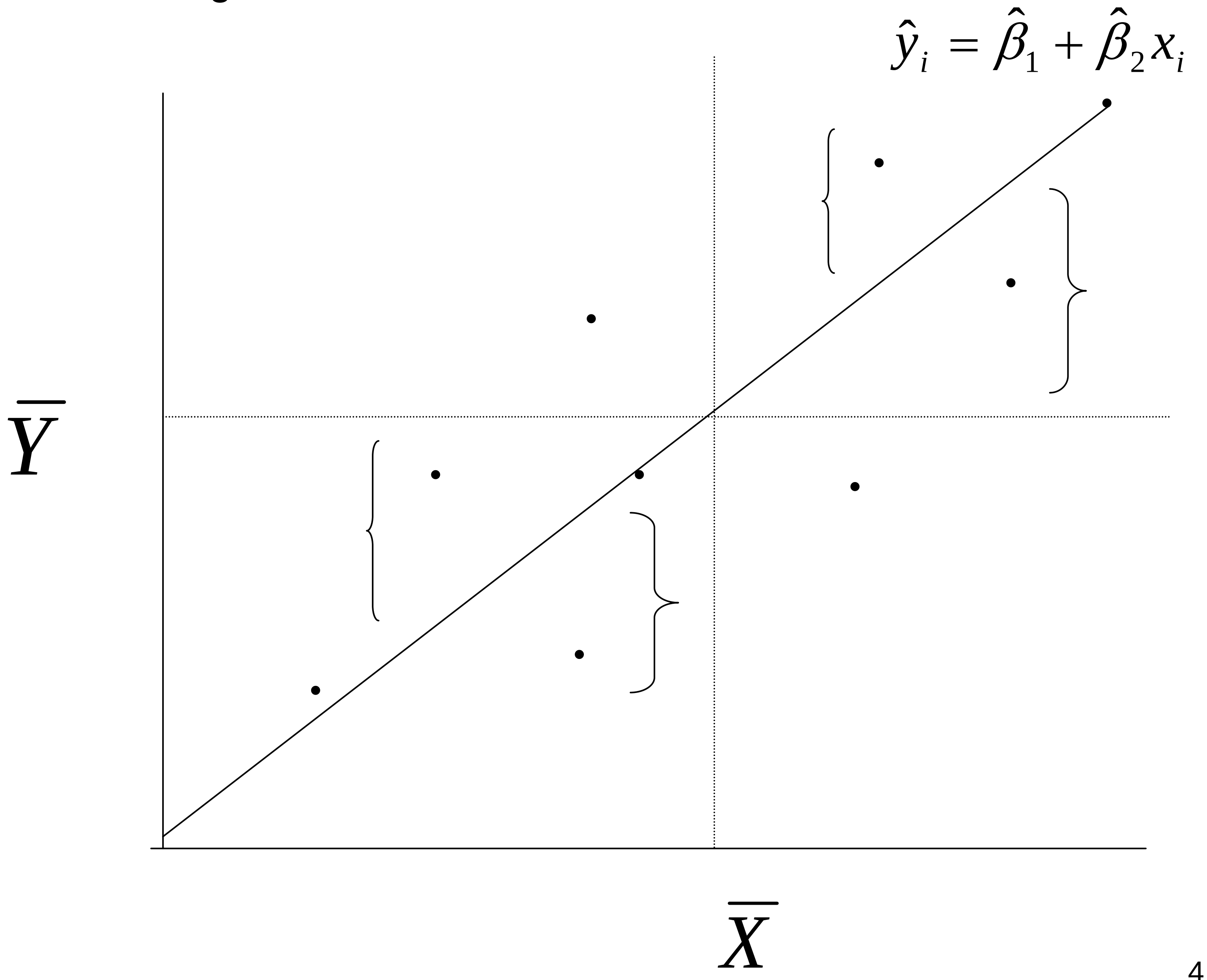
$$E[e_i] = 0 \qquad \mathrm{var}[e_i] = \sigma^2$$

$$\mathrm{cov}(e_i e_j) = 0 \quad for\ all \quad i \neq j$$

$$E[e_i x_i] = 0$$

x_i is exogenous, not random

How Regression Can do Better than Means in Prediction ?



- The least square line is the line best fits the data set.

- The least square line passes through the average values of variables X and Y;.

- Differences between each observation and the line is represented by error terms .

  As some of them are above the line and others below the lin positive errors cancel out with the negative errors.
- Each dot in the above graph represents an observation.
- Some observations lie above the least square line and other observations lie below it.

These errors represent missing elements from this relationship.

Omitted variables
Measurement errors
Misspecification

Choose $\beta_1 \beta_2$ that Minimised the Sum of Error Square Or best fits the data

$$e_i = Y_i - \beta_1 - \beta_2 x_i$$

$$S = \sum_i e_i^2 = \sum_i \left(Y_i - \beta_1 - \beta_2 x_i\right)^2$$

$$\left(-e_i^2\right) > 0$$

$$\frac{\partial S}{\partial \beta_1} = 2\sum \left(Y_i - \beta_1 - \beta_2 x_i\right)\left(-1\right) = 0$$

$$\frac{\partial S}{\partial \beta_2} = 2\sum \left(Y_i - \beta_1 - \beta_2 x_i\right)\left(-x_i\right) = 0$$

Normal Equations and Estimators

$$\sum_i y_i = N\beta_1 + \beta_2 \sum_i x_i$$

$$\sum_i x_i y_i = \beta_1 \sum_i x_i + \beta_2 \sum_i x_i^2$$

$$\hat{\beta}_2 = \frac{N\sum_i x_i y_i - \sum_i x_i \sum_i y_i}{N\sum_i x_i^2 - \left(\sum_i x_i\right)^2}$$

$$\hat{\beta}_1 = \bar{y} - \hat{\beta}_2 \bar{x}$$

[Total variation] = [Explained variation] + [Residual variation]

df = T-1 K-1 T-K-1

$$Var\left(y_i\right)=\sum_i\left[y_i-\bar{y}\right]^2=\sum_i\left[\left(y_i-\hat{y}_i+\hat{e}_i\right)\right]^2$$

$$=\sum_i\left(y_i-\hat{y}_i\right)^2+\sum_i\hat{e}_i^2+2\sum_i\left(y_i-\hat{y}_i\right)\hat{e}_i$$

$$Var\left(y_i\right)=\sum_i\left(\hat{y}_i-\bar{y}\right)^2+\sum_i\hat{e}_i^2$$

$$\because\sum_i\left(y_i-\hat{y}_i\right)\hat{e}_i=0$$

Coefficient of determination

$$1 = \frac{\sum (y_i - \bar{y})^2}{\sum (y_i - \bar{y})^2} = \frac{\sum (\hat{y}_i - \bar{y})^2}{\sum (y_i - \bar{y})^2} + \frac{\sum \hat{e}_i^2}{\sum (y_i - \bar{y})^2} = R^2 + (1 - R^2)$$

$$R^2 = \frac{\sum (\hat{y}_i - \bar{y})^2}{\sum (y_i - \bar{y})^2} \qquad 0 \le R^2 \le 1$$

Like Parameters R-Square is also a random variable.

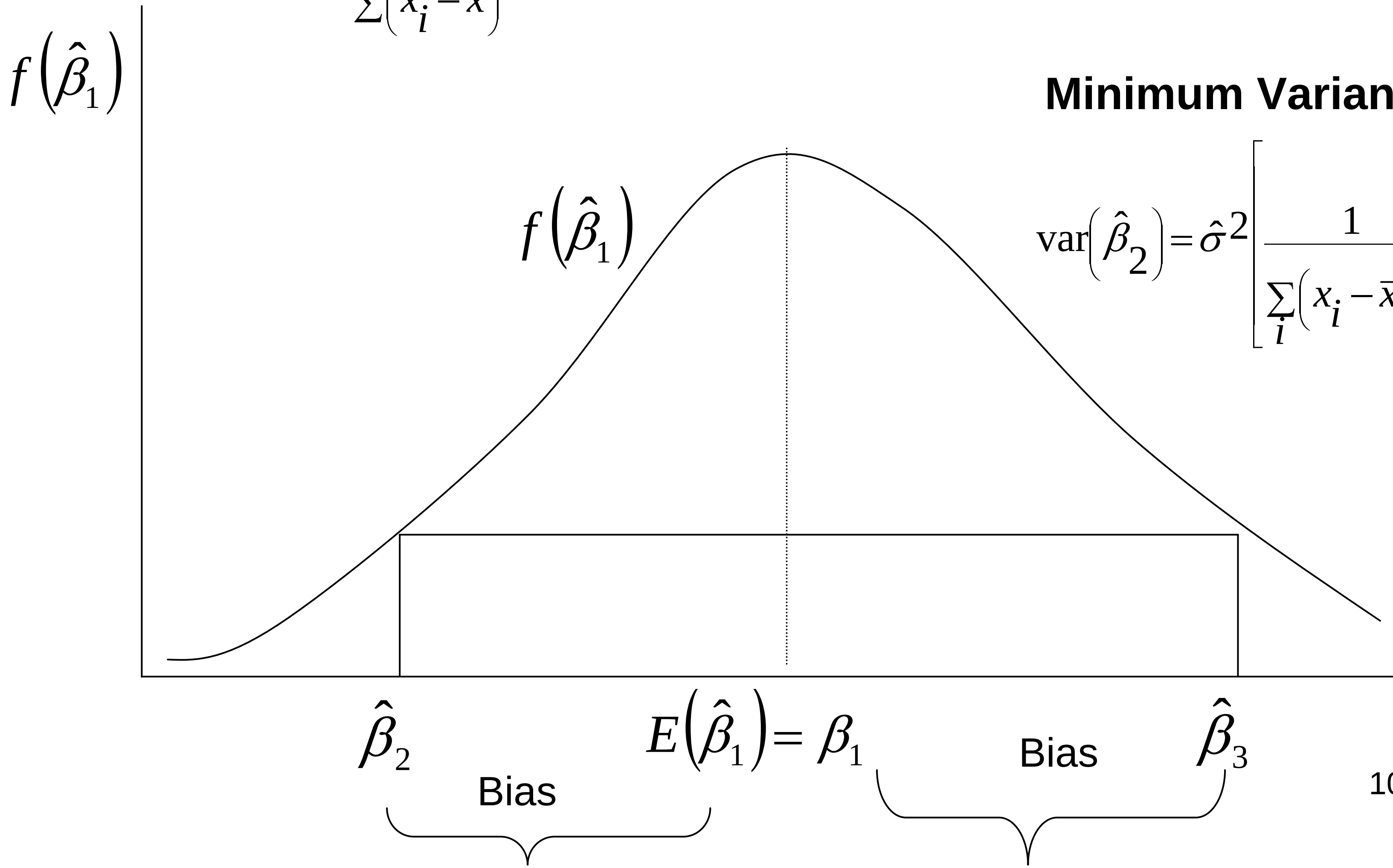
Linear, Unbiasedness and Minimum Variance Properties of an Estimator (BLUE Property)

Linearity: $\hat{\beta}_2 = \dfrac{\sum (x_i - \bar{x}) Y_i}{\sum (x_i - \bar{x})^2} = \sum w_i Y_i$

Unbiasedness $E(\hat{\beta}_1) = \beta_1$

Minimum Variance

$$\operatorname{var}(\hat{\beta}_2) = \hat{\sigma}^2 \left[\frac{1}{\sum_i (x_i - \bar{x})^2} \right]$$



Multiple Regression Analysis

Two explanatory variables

$$Y_i = \beta_1 + \beta_2 X_{1,i} + \beta_2 X_{2,i} + e_i$$

A multiple regression Model with k explanatory variables

$$Y_i = \beta_1 + \beta_2 X_{1,i} + ... + \beta_k X_{k,i} + e_i$$

Wage determination model

$$w = f\left(S, A, A^2, T, G, L, \varepsilon\right)$$

Profit

$$\pi = \pi\left(p, a, a^2, P_c, y, i, w, \varepsilon\right)$$

Investment

$$I = I\left(r, w, sk, t, D, P, \varepsilon\right)$$

Money demand

$$\frac{M}{P} = \frac{M}{P}\left(i, Y, c, E\right)$$

Derivation of Normal Equations in a Multiple Regression Analysis

$$\frac{\partial S}{\partial \beta_1} = 2\sum\left(Y_i - \beta_1 - \beta_2 X_{1,i} - \beta_3 X_{2,i}\right)\left(-1\right) = 0 \quad \text{and}$$

$$\frac{\partial S}{\partial \beta_2} = 2\sum\left(Y_i - \beta_1 - \beta_2 X_{1,i} - \beta_3 X_{2,i}\right)\left(-X_{1,i}\right) = 0$$

$$\frac{\partial S}{\partial \beta_3} = 2\sum\left(Y_i - \beta_1 - \beta_2 X_{1,i} - \beta_3 X_{2,i}\right)\left(-X_{2,i}\right) = 0$$

Thus normal equations are

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

$$\sum_i X_{1,i} Y_i = \beta_1 \sum_i X_{1,i} + \beta_2 \sum_i X^2_{1,i} + \beta_3 \sum_i X_{1,i} X_{2,i} \qquad (3)$$

$$\sum_i X_{2,i} Y_i = \beta_1 \sum_i X_{2,i} + \beta_2 \sum_i X_{1,i} X_{2,i} + \beta_3 \sum_i X^2_{2,i} \qquad (4)$$

Algebraic Method

It is easier to solve this system in a deviation form defining deviation from the mean $\sum_i x_{i,i} = \sum_i \left(X_{1,i} - \overline{X}_1\right) = 0$; $\sum_i x_{2,i} = \sum_i \left(X_{2,i} - \overline{X}_2\right) = 0$; $\sum_i y_i = \sum_i \left(Y_i - \overline{Y}_i\right) = 0$

$$\sum_i x_{1,i} y_i = \beta_2 \sum_i x_{1,i}^2 + \beta_3 \sum_i x_{1,i} x_{2,i} \qquad (3')$$

$$\sum_i x_{2,i} y_i = \beta_2 \sum_i x_{1,i} x_{2,i} + \beta_3 \sum_i x_{2,i}^2 \qquad (4')$$

In order get value of β_3 eliminate β_2 by multiplying the (3') by $\sum x_{1,i} x_{2,i}$ and (4') by $\sum x_{1,i}^2$

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

Algebraic Derivation of Parameters

Use value of the $\hat{\beta}_3$ in equation (3') to get value of $\hat{\beta}_2$

$$\sum_i x_{1,i} y_i = \hat{\beta}_2 \sum_i x_{1,i}^2 + \hat{\beta}_3 \sum_i x_{1,i} x_{2,i} \qquad (3')$$

$$\Rightarrow \frac{\sum_i x_{1,i} y_i}{\sum_i x_{1,i}^2} = \hat{\beta}_2 + \hat{\beta}_3 \frac{\sum_i x_{1,i} x_{2,i}}{\sum_i x_{1,i}^2}$$

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

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

The values of $\hat{\beta}_3$ and $\hat{\beta}_2$ can be used to find the value of $\hat{\beta}_1$.

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

Matrix Approach

$$\hat{\beta} = \begin{bmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \\ \hat{\beta}_2 \end{bmatrix} = \underbrace{\begin{bmatrix} N & \sum_i X_{1,i} & \sum_i X_{21,i} \\ \sum_i X_{1,i} & \sum_i X_{1,i}^2 & \sum_i X_{1,i} X_{2,i} \\ \sum_i X_{21,i} & \sum_i X_{1,i} X_{2,i} & \sum_i X_{1,i}^2 \end{bmatrix}^{-1}}_{\text{X'X}} \underbrace{\begin{bmatrix} \sum_i Y_i \\ \sum_i Y_i X_{1,i} \\ \sum_i Y_i X_{2,i} \end{bmatrix}}_{\text{X'Y}}$$

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

Using Cramer's Rule for Deriving the Slope Parameters

$$\begin{bmatrix} \hat{\beta}_2 \\ \hat{\beta}_3 \end{bmatrix} = \begin{bmatrix} \sum_i x_{1,i}^2 & \sum_i x_{1,i} x_{2,i} \\ \sum_i x_{1,i} x_{2,i} & \sum_i x_{2,i}^2 \end{bmatrix}^{-1} \begin{bmatrix} \sum_i y_i x_{1,i} \\ \sum_i y_i x_{2,i} \end{bmatrix} \tag{18}$$

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

(19)

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

(20)

Variance and Covariance of Estimated OLS Parameters

$$\text{cov}(\beta) = \begin{bmatrix} \text{var}(\beta_1) & Cov(\beta_1, \beta_2) & Cov(\beta_1, \beta_3) \\ Cov(\beta_1, \beta_2) & \text{var}(\beta_2) & Cov(\beta_2, \beta_3) \\ Cov(\beta_1, \beta_3) & Cov(\beta_2, \beta_3) & \text{var}(\beta_2) \end{bmatrix};$$

$$Cov(\beta) = (X'X)^{-1}\sigma^2$$

This is also true for the deviation form

$$Cov(\beta) = (x'x)^{-1}\sigma^2 .$$

In this matrix, diagonal elements are variances and off-diagonal elements are corresponding covariance.

Using the property of a symmetric matrix we can write

the variance and covariance of $\hat{\beta}_2$ and $\hat{\beta}_3$.

Estimators of Variance and Covariance of the OLS Estimators

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

$$\operatorname{var}\left(\hat{\beta}_3\right)=\frac{\left|\sum_i x_{1,i}^2\right|\hat{\sigma}^2}{\begin{vmatrix}\sum_i x_{1,i}^2 & \sum_i x_{1,i}x_{2,i}\\ \sum_i x_{1,i}x_{2,i} & \sum_i x_{1,i}^2\end{vmatrix}}=\frac{\hat{\sigma}^2\sum_i x_{2,i}^2}{\sum_i x_{1,i}^2\sum_i x_{2,i}^2-\sum_i x_{1,i}x_{2,i}\sum_i x_{1,i}x_{2,i}}$$

$$\operatorname{cov}ar\left(\hat{\beta}_{2,}\hat{\beta}_3\right)=\frac{\left|\sum_i x_{1,i}x_{2,i}\right|\hat{\sigma}^2}{\begin{vmatrix}\sum_i x_{1,i}^2 & \sum_i x_{1,i}x_{2,i}\\ \sum_i x_{1,i}x_{2,i} & \sum_i x_{1,i}^2\end{vmatrix}}=\frac{\hat{\sigma}^2\sum_i x_{1,i}x_{2,i}}{\sum_i x_{1,i}^2\sum_i x_{2,i}^2-\sum_i x_{1,i}x_{2,i}\sum_i x_{1,i}x_{2,i}}$$

In order to get the variance of $\hat{\beta}_1$ we need to solve 3 by 3 inverse for $(X'X)^{-1}$ matrix.

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.

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 $\left| X'X \right| = 0$

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

Hetrosckedasticity

- LS assumption: variance of e_i is constant $\operatorname{var}\left[e_i\right] = \sigma^2$ for every ith observation, $\operatorname{var}\left(\hat{\beta}_2\right) = \hat{\sigma}^2\left[\frac{1}{\sum_i\left(x_i - \bar{x}\right)^2}\right]$ but it is possible that

$$\sigma_i^2 = \sigma^2 x_i$$

Causes: Learning, growth, improved data collection, outliers, omitted variables;

Autocorrelation

Assumption behind the OLS

$$\text{cov}(e_i e_j) = 0 \quad \textit{for all} \quad i \neq j$$

Autocorrelation exists when

$$\text{cov}(e_i e_j) \neq 0 \quad \textit{for all} \quad i \neq j$$

$$e_i = \rho e_{i-1} + v_i \qquad\qquad v_i \sim N\left(0, \sigma^2\right)$$

ρ **correlation coefficient between –1 and 1**

Linearity of the OLS Estimators

Linear regression model: $Y_i = \beta_1 + \beta_2 x_i + e_i$

Linearity of estimator $\hat{\beta}_2 = \dfrac{\sum\left(x_i - \bar{x}\right)\left(y_i - \bar{y}\right)}{\sum\left(x_i - \bar{x}\right)^2}$

Proof: $\hat{\beta}_2 = \dfrac{\sum\left(x_i - \bar{x}\right)Y_i}{\sum\left(x_i - \bar{x}\right)^2} = \sum w_i Y_i$;

$w_i = \dfrac{\left(x_i - \bar{x}\right)}{\sum_i\left(x_i - \bar{x}\right)^2} = a$ is a constant Y is the only random variable.

Since β_2 is linear the β_1 is also linear. $\hat{\beta}_1 = \bar{y} - \hat{\beta}_2\bar{x}$.

Unbiasedness of the Intercept Parameter

- $E\left(\hat{\beta}_1\right) = \beta_1$

Proofs:

$$E\left(\hat{\beta}_1\right) = E\left(\bar{y} - \hat{\beta}_2 \bar{x}\right) = E\left[\frac{\sum_i y_i}{n} - \frac{\sum_i \left(x_i - \bar{x}\right)\left(y_i - \bar{y}\right)}{\sum_i \left(x_i - \bar{x}\right)^2} \bar{x}\right]$$

$$E\left(\hat{\beta}_1\right) = \sum\left[\frac{1}{n} - \bar{x} w_i\right] E\left(y_i\right) = \sum\left[\frac{1}{n} - \bar{x} w_i\right] E\left(\beta_1 + \beta_2 x_i + e_i\right)$$

or

$$E\left(\hat{\beta}_1\right) = \sum_i \left[\frac{1}{n} - \bar{x} w_i\right]\left(\beta_1 + \beta_2 x_i\right) = \left[\beta_1 - \beta_1 \bar{x} \sum_i w_i + \beta_2 \frac{\sum_i x_i}{n} - \right.$$

$= \beta_1$

where $w_i = \dfrac{\left(x_i - \bar{x}\right)}{\sum_i \left(x_i - \bar{x}\right)^2}$

Unbiasedness of the Slope Parameter

- $E\left(\hat{\beta}_2\right) = \beta_2$

Proof:

$$\hat{\beta}_2 = \frac{\sum\left(x_i - \bar{x}\right)\left(y_i - \bar{y}\right)}{\sum\left(x_i - \bar{x}\right)^2} = \frac{\sum\left(x_i - \bar{x}\right)y_i}{\sum\left(x_i - \bar{x}\right)^2} = \sum w_i y_i = \sum w_i\left(\beta_1 + \beta\right.$$

- $$E\left(\hat{\beta}_2\right) = E\left[\sum w_i y_i\right] = E\left[\sum w_i\left(\beta_1 + \beta_2 x_i + e_i\right)\right]$$ Since

$$= E\left[\sum w_i \beta_1 + \beta_2 \sum w_i x_i + \sum w_i e_i\right] = \beta_2$$

$$\sum w_i \beta_1 = \beta_1 \sum w_i = 0; \quad \sum w_i e_i = 0$$

Variance of the Intercept Parameter

$$\hat{\beta}_1 = \bar{y} - \hat{\beta}_2 \bar{x}$$

$$\hat{\beta}_1 - E\left(\hat{\beta}_1\right) = \left(\bar{y} - \hat{\beta}_2 \bar{x}\right) - \left(\bar{y} - \beta_2 \bar{x}\right)$$

$$\hat{\beta}_1 - \beta_1 = \left(\bar{y} - \hat{\beta}_2 \bar{x}\right) - \left(\bar{y} - \beta_2 \bar{x}\right) \Rightarrow \hat{\beta}_1 - \beta_1 = -\bar{x}\left(\hat{\beta}_2 - \beta_2\right)$$

$\mathrm{var}\left(\hat{\beta}_1\right) = \bar{x}^2 \,\mathrm{var}\left(\hat{\beta}_2\right)$: Since $\bar{x}^2$ is a constant the estimate of variance of β_1 has the minimum variance if β_2 has the minimum variance.

Variance of the Slope Parameter

- $\operatorname{var}\left(\hat{\beta}_2\right) = \hat{\sigma}^2 \left[\frac{1}{\sum_i \left(x_i - \bar{x}\right)^2} \right]$

It was proved above that

$$E\left(\hat{\beta}_2\right) = E\left[\sum w_i y_i\right] = E\left[\sum w_i \left(\beta_1 + \beta_2 x_i + e_i\right)\right]$$

$$= E\left[\sum w_i \beta_1 + \beta_2 \sum w_i x_i + \sum w_i e_i\right] = \beta_2$$

$$\operatorname{var}\left(\hat{\beta}_2\right) = E\left[E\left(\hat{\beta}_2\right) - \beta_2\right]^2 =$$

$$E\left[\sum w_i e_i\right]^2 = \left[\sum w_{ii}^2 E\left(e_i\right)^2\right] = \hat{\sigma}^2 \frac{1}{\sum_i \left(x_i - \bar{x}\right)^2}$$

Linear and Unbiased Alternative Estimator

Let there be another estimator Linearity of estimator which is also linear and unbaised

$$b_2 = \frac{\sum\left(x_i - \bar{x}\right)\left(y_i - \bar{y}\right)}{\sum\left(x_i - \bar{x}\right)^2}$$

Proof: $b_2 = \frac{\sum\left(x_i - \bar{x}\right)Y_i}{\sum\left(x_i - \bar{x}\right)^2} = \sum k_i Y_i \; ; \; k_i = w_i + c_i$

$$b_2 = \sum k_i Y_i = \sum k_i E\left(\beta_1 + \beta_2 x_i + e_i\right)$$

$$= \sum\left(w_i + c_i\right)E\left(\beta_1 + \beta_2 x_i + e_i\right)$$

$$= E\left[\beta_1 \sum w_i + \beta_2 \sum w_i x_i + \sum w_i e_i + \beta_1 \sum c_i + \beta_2 \sum c_i x_i + \sum c_i e_i\right]$$

Unbiasedness requires that $\sum c_i = 0 \;\; \sum c_i x_i = 0$

$$b_2 = \beta_2 + \sum\left(w_i + c_i\right)e_i$$

Proof of the Minimum Variance of the OLS Estimator

$$b_2 = \beta_2 + \sum\left(w_i + c_i\right)e_i$$

$$\text{var}\left(b_2\right) = E\left(b_2 - \beta_2\right)^2 = E\left[\sum\left(w_i + c_i\right)e_i\right]^2$$

$$\text{var}\left(b_2\right) = \hat{\sigma}^2\left[\frac{1}{\sum_i\left(x_i - \overline{x}\right)^2}\right] + \hat{\sigma}^2\sum c_i^2$$

Thus the variance of this alternative estimator bigger than the variance of the OLS estimator $\text{var}\left(\hat{\beta}_2\right) = \hat{\sigma}^2\left[\frac{1}{\sum_i\left(x_i - \overline{x}\right)^2}\right]$.

$$\text{var}\left(b_2\right) > \text{var}\left(\hat{\beta}_2\right)$$

References

- Bhattarai (2004) Lecture Notes in Econometrics, University of Hull.
- Gujarati DN (2003) Basic Econometrics, McGraw Hill
- Griffiths ,W.E., R.C. Hill and C. G. Judge (GHJ) Learning and Practicing Econometrics, John Wiley and Sons Inc. New York, 1993
- Stock JH and MW Watson (2003) Introduction to Econometrics, Addison Wisley.
- Hill, Griffiths and Judge (2001) Undergraduate Econometrics, Second Edition, John Willey and Sons, 2001