

3a. Unbiased estimator of the residual variance I: Model without exogenous variables

- ▶ Model: $y = \beta_0 + \epsilon$, $\epsilon \sim \text{i.i.d.}$
- ▶ OLS estimator: $\hat{\beta}_0 = \bar{y} = 1/n \sum_i y_i$
- ▶ Minimized SSE: $S(\hat{\beta}_0) = \sum_i (y_i - \bar{y})^2$

Expectation value of $S(\hat{\beta}_0)$ in normal sum notation:

$$\begin{aligned}
 E(S) &= \sum_i E(y_i - \bar{y})^2 \stackrel{\text{all terms equal}}{=} nE(y_1 - \bar{y})^2 \\
 &= nE \left(y_1 - \frac{1}{n} \sum_j y_j \right)^2 \\
 &= nE(y_1^2) - 2 \sum_j E(y_j y_1) + \frac{1}{n} \sum_j \sum_k E(y_j y_k)
 \end{aligned}$$

Residual variance w/o exogenous variables (ctned)

With (remember the i.i.d. property and $E(\epsilon_k) = 0$)

$$E(y_j y_k) = E(\beta_0 + \epsilon_j)(\beta_0 + \epsilon_k) = \beta_0^2 + \sigma^2 \delta_{jk}, \quad \delta_{jk} = \begin{cases} 1 & j = k \\ 0 & j \neq k \end{cases},$$

we have

$$\begin{aligned} E(S) &= nE(y_1^2) - 2 \sum_j E(y_j y_1) + \frac{1}{n} \sum_j \sum_k E(y_j y_k) \\ &= n(\beta_0^2 + \sigma^2) - 2(n\beta_0^2 + \sigma^2) + \frac{1}{n}(n^2\beta_0^2 + n\sigma^2) \\ &= (n - 2n + n)\beta_0^2 + (n - 2 + 1)\sigma^2 \\ &= (n - 1)\sigma^2. \end{aligned}$$

We conclude that $\hat{\sigma}^2 = \frac{1}{n-1} \sum_i (y_i - \hat{y})^2 = \frac{S}{n-1}$ is unbiased, i.e.,
 $E(\hat{\sigma}^2) = \frac{E(S)}{n-1} = \sigma^2$.

2. Residual variance for the multivariate linear model (matrix-vector notation)

- ▶ Model: $y = \sum_{j=0}^{p+1} \beta_j x_j + \epsilon = \hat{y}(\mathbf{x}) + \epsilon$, $\epsilon \sim \text{i.i.d.}$
- ▶ OLS estimator: $\hat{\beta} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$
- ▶ Minimized SSE:

$$\begin{aligned}
 S(\hat{\beta}) &= \sum_i (y_i - \hat{y}(x_i))^2 = (\hat{\mathbf{y}} - \mathbf{y})' (\hat{\mathbf{y}} - \mathbf{y}) \\
 &= (\mathbf{X} \hat{\beta} - \mathbf{y})' (\mathbf{X} \hat{\beta} - \mathbf{y}) \\
 &= (\mathbf{X} \hat{\beta})' (\mathbf{X} \hat{\beta}) - (\mathbf{X} \hat{\beta})' \mathbf{y} - \mathbf{y}' (\mathbf{X} \hat{\beta}) + \mathbf{y}' \mathbf{y} \\
 &= (\mathbf{X} \hat{\beta})' (\mathbf{X} \hat{\beta}) - 2(\mathbf{X} \hat{\beta})' \mathbf{y} + \mathbf{y}' \mathbf{y}
 \end{aligned}$$

Replace $\hat{\beta} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$:

$$S(\hat{\beta}) = \mathbf{y}' (\mathbf{I} - \mathbf{X} (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}') \mathbf{y}$$

Estimation of the multivariate residual variance (ctned)

Replace the observed endogeneous data vector $\mathbf{y}$ by the model

$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$ Notice: $\boldsymbol{\beta}$ is the true and immutable parameter vector $\boldsymbol{\beta}$:

$$\begin{aligned} S(\hat{\boldsymbol{\beta}}) &= (\mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon})' (\mathbf{1} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}') (\mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}) \\ &= \boldsymbol{\epsilon}'(\mathbf{1} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}')\boldsymbol{\epsilon} \\ &\quad + 2(\mathbf{X}\boldsymbol{\beta})'(\mathbf{1} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}')\boldsymbol{\epsilon} \\ &\quad + \boldsymbol{\beta}'\mathbf{X}'(\mathbf{1} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}')\mathbf{X}\boldsymbol{\beta} \end{aligned}$$

Making use of associativity, we realize that the second and third term are each equal to zero, so

$$S(\hat{\boldsymbol{\beta}}) = \boldsymbol{\epsilon}'(\mathbf{1} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}')\boldsymbol{\epsilon}.$$

Excursion: trace of a matrix

$S(\hat{\beta})$ is a scalar made up of many vector and matrix products. To simplify, we use the fact that the **trace** of a square matrix (sum of its diagonal elements) allows additional identity operations, commutativity and cyclic permutation, whenever the corresponding products of generally non-square matrices are defined:

$$\text{tr}(\mathbf{A}\mathbf{B}) = \text{tr}(\mathbf{B}\mathbf{A}), \quad \text{tr}(\mathbf{A}\mathbf{B}\mathbf{C}) = \text{tr}(\mathbf{B}\mathbf{C}\mathbf{A}) = \text{tr}(\mathbf{C}\mathbf{A}\mathbf{B})$$

- ▶ Also applies to degenerate matrices (vectors, scalars)
- ▶ The permutations may even change the matrix dimensions: if ϵ is a n -vector and $\mathbf{M}$ a $n \times n$ matrix, then $\epsilon'\mathbf{M}\epsilon$ is a number and $\mathbf{M}\epsilon\epsilon'$ a $n \times n$ matrix, still $\text{tr}(\epsilon'\mathbf{M}\epsilon) = \text{tr}(\mathbf{M}\epsilon\epsilon')$
- ▶ For a scalar S , we trivially have $\text{tr}(S) = S \Rightarrow$ apply to $S(\beta)$

Estimation of the multivariate residual variance (ctned)

$$\begin{aligned}
 S(\hat{\beta}) &= \epsilon'(\mathbf{1} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}')\epsilon \\
 &= \text{tr}(\epsilon'(\mathbf{1} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}')\epsilon) \\
 &= \text{tr}(\epsilon'\epsilon) - \text{tr}(\epsilon'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\epsilon) \\
 &= \text{tr}(\epsilon'\epsilon) - \text{tr}(\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\epsilon\epsilon')
 \end{aligned}$$

Expectation: of course, $E(\cdot)$ and $\text{tr}(\cdot)$ commute. Furthermore (statistical Gauß Markow assumptions) $E(\epsilon_i\epsilon_j) = \sigma^2\delta_{ij}$:

$$\begin{aligned}
 E(S) &= \text{tr}(E(\epsilon'\epsilon)) - \text{tr}(\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'E(\epsilon\epsilon')) \\
 &= n\sigma^2 - \text{tr}(\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}')\sigma^2 \\
 &= n\sigma^2 - \text{tr}((\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X})\sigma^2 \\
 &= n\sigma^2 - p\sigma^2 = \underline{\underline{(n-p)\sigma^2}}
 \end{aligned}$$

(notice that $(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X}$ is a $p \times p$ identity matrix) $\Rightarrow \hat{\sigma}^2 = \frac{S}{n-p}$ is an unbiased estimator for σ^2 .

Distribution function of $\hat{\sigma}^2$

$S(\hat{\beta}) = \epsilon'(\mathbf{1} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}')\epsilon$ is a quadratic form of ϵ where also, although uncorrelated, the distribution function of the nondiagonal $\epsilon_i\epsilon_j$ is nontrivial. It is beyond this course to show that

$\mathbf{1} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$ is a projection matrix with rank $n - p$, so all the nondiagonals of $\epsilon_i\epsilon_j$ vanish, so $S(\hat{\beta})/\sigma^2$ is a sum of squared independent standardnormal distributed random variables Z defining the χ^2 distribution:

$$\frac{S(\hat{\beta})}{\sigma^2} = \frac{(n - p)\hat{\sigma}^2}{\sigma^2} = \sum_{i=1}^{n-p} Z_i^2 \sim \chi^2(n - p)$$