Intermediate Econometrics

Homeworks

Máté Kormos

Compiled on September 28, 2017

This note is written to cover some of the homeworks without the aim of being comprehensive. Some comments are included as well.

1. Week 1, lecture 1 (Sep 19)

Topic: OLS basics

1.1. Comments

Model identification.

A structure is a function which assigns numerical values to the model parameters, formally $g : \theta \in \Theta \mapsto \mathbb{R}^{dim(\theta)}$. Then a model is identified iff there is no structure $h \neq g$ such that $f_g(\boldsymbol{y} \mid \boldsymbol{X}) = f_h(\boldsymbol{y} \mid \boldsymbol{X})$. That is to say, for the same $\boldsymbol{X}$, there are no two different structures which produce the same conditional density for $\boldsymbol{y}$.

E.g. probit model: $y_i := \mathbb{1}(y_{i1} - y_{i0} > 0) = \mathbb{1}(\boldsymbol{x_i'}\boldsymbol{\beta} - u_i > 0)$, $u_i \sim \mathcal{N}(0,1)$, where we only observe y_i and $\boldsymbol{x_i}$. Altenatively consider $y_i := \mathbb{1}(z_{i1} - z_{i0} > 0) = \mathbb{1}(\sigma(\boldsymbol{x_i'}\boldsymbol{\beta} - u_i) > 0) := \mathbb{1}(\boldsymbol{x_i'}\boldsymbol{\delta} + v_i > 0) = \mathbb{1}(\boldsymbol{x_i'}\boldsymbol{\beta} - u_i > 0)$ for $\sigma > 0$ and $v_i \sim \mathcal{N}(0, \sigma^2)$. Given the same $\boldsymbol{x_i}$, the two model $(\boldsymbol{\beta}, \mathbf{Var}\,[\text{disturbance}] = 1)$ and $(\boldsymbol{\delta}, \mathbf{Var}\,[\text{disturbance}] = \sigma^2)$ will generate the same y_i, so we cannot tell apart the two models, cannot separate the scale.

1.2. Tasks

1. Show that on violation of the homoskedasticity assumption, OLS stops being BLUE inasmuch as it won't have the smallest variance.

   We present another estimator, GLS, which is linear, unbiased, and has a samller covariance matrix than OLS. The GLS first transforms the model from $\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{u}$ to $\boldsymbol{Wy} = \boldsymbol{WX\beta} + \boldsymbol{Wu}$, or with another notation $\tilde{\boldsymbol{y}} = \tilde{\boldsymbol{X}}\boldsymbol{\beta} + \tilde{\boldsymbol{u}}$ (note that $\boldsymbol{W}$ is invertible, hence the transformation is valid). The weight matrix $\boldsymbol{W}$ is constructed to eliminate the heteroskedasticity from the model:

$$\boldsymbol{W} = \begin{bmatrix} \frac{1}{\sigma_1} & & & \\ & \frac{1}{\sigma_2} & & \\ & & \ddots & \\ & & & \frac{1}{\sigma_n} \end{bmatrix}, \tag{1}$$

$$\sigma_i = \mathbb{E}\left[u_i^2\right]^{1/2} = \mathbf{Var}\left[u_i\right]^{1/2}. \tag{2}$$

Consequently, the new error term will exhibit unit variance:

$$\begin{aligned} \mathbf{Var}\left[\tilde{\boldsymbol{u}}\right] &= \mathbb{E}\left[\tilde{\boldsymbol{u}}\tilde{\boldsymbol{u}}'\right] && (3)\\ &= \mathbb{E}\left[\boldsymbol{W}\boldsymbol{u}\boldsymbol{u}'\boldsymbol{W}'\right] && (4)\\ &= \boldsymbol{W}\,\mathbb{E}\left[\boldsymbol{u}\boldsymbol{u}'\right]\boldsymbol{W}' && (5)\\ &= \boldsymbol{W}\boldsymbol{\Sigma}\boldsymbol{W}' && (6)\\ &= \begin{bmatrix} \frac{1}{\sigma_1} & & & \\ & \frac{1}{\sigma_2} & & \\ & & \ddots & \\ & & & \frac{1}{\sigma_n} \end{bmatrix} \begin{bmatrix} \sigma_1^2 & & & \\ & \sigma_2^2 & & \\ & & \ddots & \\ & & & \sigma_n^2 \end{bmatrix} \begin{bmatrix} \frac{1}{\sigma_1} & & & \\ & \frac{1}{\sigma_2} & & \\ & & \ddots & \\ & & & \frac{1}{\sigma_n} \end{bmatrix} && (7)\\ &= \boldsymbol{I} && (8)\\ \boldsymbol{W} &= \boldsymbol{\Sigma}^{-1/2} = \boldsymbol{W}', \qquad \boldsymbol{W}\boldsymbol{W}' = \boldsymbol{\Sigma}^{-1}. && (9) \end{aligned}$$

The GLS estimates the tranformed model with OLS, which makes it a linear estimator. The unbiasedness is not proven here but the proof is pretty much the same as for the OLS, but this with $\tilde{\boldsymbol{y}} = \tilde{\boldsymbol{X}}\boldsymbol{\beta} + \tilde{\boldsymbol{u}}$. Now, let's compare the covariance matrices. First, the OLS estimator has

$$\begin{aligned} \mathbf{Var}\left[\hat{\boldsymbol{\beta}}^{OLS}\right] &= \mathbb{E}\left[\left(\hat{\boldsymbol{\beta}}^{OLS} - \mathbb{E}\left[\hat{\boldsymbol{\beta}}^{OLS}\right]\right)\left(\hat{\boldsymbol{\beta}}^{OLS} - \mathbb{E}\left[\hat{\boldsymbol{\beta}}^{OLS}\right]\right)'\right] && (10)\\ &= \mathbb{E}\left[(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{u}\boldsymbol{u}'\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\right] \quad \text{OLS still unbaised} && (11)\\ &\overset{L.I.E.}{=} \mathbb{E}\left[(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\,\mathbb{E}\left[\boldsymbol{u}\boldsymbol{u}' \mid \boldsymbol{X}\right]\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\right] && (12)\\ &= \mathbb{E}\left[(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{\Sigma}\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\right]. && (13) \end{aligned}$$

Second, the GLS, the estimate of which is $\hat{\boldsymbol{\beta}}^{GLS} = \boldsymbol{\beta} + (\boldsymbol{X'W'WX})^{-1}\boldsymbol{X'W'Wu}$, has

$$\begin{aligned}
\mathbf{Var}\left[\hat{\boldsymbol{\beta}}^{GLS}\right] &= \mathbb{E}\left[\left(\hat{\boldsymbol{\beta}}^{GLS} - \mathbb{E}\left[\hat{\boldsymbol{\beta}}^{GLS}\right]\right)\left(\hat{\boldsymbol{\beta}}^{GLS} - \mathbb{E}\left[\hat{\boldsymbol{\beta}}^{GLS}\right]\right)'\right] && (14)\\
&= \mathbb{E}\left[(\boldsymbol{X'W'WX})^{-1}\boldsymbol{X'W'Wuu'W'WX}(\boldsymbol{X'W'WX'})^{-1}\right] && (15)\\
&= \mathbb{E}\left[(\boldsymbol{X'W'WX})^{-1}\boldsymbol{X'W'}\,\mathbb{E}\left[W'uu'W \mid \boldsymbol{X}\right]'\boldsymbol{WX}(\boldsymbol{X'W'WX'})^{-1}\right] && (16)\\
&= \mathbb{E}\left[(\boldsymbol{X'W'WX})^{-1}\right] && (17)\\
&= \mathbb{E}\left[(\boldsymbol{X'\Sigma^{-1}X})^{-1}\right]. && (18)
\end{aligned}$$

Then we compare the covariance matrices: if $\mathbf{Var}\left[\hat{\boldsymbol{\beta}}^{OLS}\right] - \mathbf{Var}\left[\hat{\boldsymbol{\beta}}^{GLS}\right]$ is positive definite, we showed GLS to be more efficient.

$$\begin{aligned}
\mathbf{Var}\left[\hat{\boldsymbol{\beta}}^{OLS}\right] - \mathbf{Var}\left[\hat{\boldsymbol{\beta}}^{GLS}\right] &= \mathbb{E}\left[(\boldsymbol{X'X})^{-1}\boldsymbol{X'\Sigma X}(\boldsymbol{X'X})^{-1}\right] - \mathbb{E}\left[(\boldsymbol{X'\Sigma^{-1}X})^{-1}\right] && (19)\\
&\mathbb{E}\left[(\boldsymbol{X'X})^{-1}\boldsymbol{X'\Sigma X}(\boldsymbol{X'X})^{-1} - (\boldsymbol{X'\Sigma^{-1}X})^{-1}\right]. && (20)
\end{aligned}$$

To simplify our proof, let's assume that there is only one regressor. This we can do without loss of generality: the estimator's properties do not depend on the number of regressors in this setting. So with $\boldsymbol{X} = \boldsymbol{x_i} \in \mathcal{M}^{n\times 1}$

$$(\boldsymbol{X'X})^{-1}\boldsymbol{X'\Sigma X}(\boldsymbol{X'X})^{-1} = \frac{\sum_i x_i^2\sigma_i^2}{(\sum_i x_i^2)(\sum_i x_i^2)} \qquad (21)$$

$$(\boldsymbol{X'\Sigma^{-1}X})^{-1} = \left(\sum_i \frac{x_i^2}{\sigma_i^2}\right)^{-1} = \frac{1}{\sum_i \frac{x_i^2}{\sigma_i^2}}, \quad \text{so the difference} \qquad (22)$$

$$\begin{aligned}
\frac{\sum_i x_i^2\sigma_i^2}{(\sum_i x_i^2)(\sum_i x_i^2)} - \frac{1}{\sum_i \frac{x_i^2}{\sigma_i^2}} &= \frac{\left(\sum_i \frac{x_i^2}{\sigma_i^2}\right)(\sum_i x_i^2\sigma_i^2) - (\sum_i x_i^2)(\sum_i x_i^2)}{(\sum_i x_i^2)(\sum_i x_i^2)\left(\sum_i \frac{x_i^2}{\sigma_i^2}\right)} && (23)\\
&= \frac{\sum_{i=1}^{n-1}\sum_{j=i+1}^{n} x_i^2 x_j^2\left(\frac{\sigma_j^2}{\sigma_i^2} + \frac{\sigma_i^2}{\sigma_j^2} - 2\right)}{(\sum_i x_i^2)(\sum_i x_i^2)\left(\sum_i \frac{x_i^2}{\sigma_i^2}\right)}, && (24)
\end{aligned}$$

where the last result is obtained by some rather soul-destroying arithmetic. We can be reasonably happy now as $(\frac{\sigma_j^2}{\sigma_i^2} + \frac{\sigma_i^2}{\sigma_j^2}$ is bounded from below, with a global minimum of 2, reached at $\sigma_i = \sigma_j$. Therefore, as the denominator is always positive, the numerator will be larger than or equal to 0, reaching 0 only in the homoskedastic case ($\sigma_i = \sigma_j \ \forall (i,j)$), so the expectation will also be positive in the heteroskedastic case $\iff$ OLS exhibits larger variance than GLS, our claim is proved.

2. Derive the finite and the limiting distribution of $\hat{\boldsymbol{\beta}}^{OLS}$. By definition of OLS we have

$$
\begin{aligned}
\hat{\boldsymbol{\beta}}^{OLS} =& (\boldsymbol{X'X})^{-1}\boldsymbol{X'y} && (25)\\
=& \boldsymbol{\beta} + (\boldsymbol{X'X})^{-1}\boldsymbol{X'u} && (26)\\
=& \boldsymbol{\beta} + \left(\sum_{i=1}^{n} \boldsymbol{x_i x_i'}\right)^{-1} \sum_{i=1}^{n} \boldsymbol{x_i} u_i && (27)\\
=& \boldsymbol{\beta} + \left(\frac{1}{n}\sum_{i=1}^{n} \boldsymbol{x_i x_i'}\right)^{-1} \frac{1}{n}\sum_{i=1}^{n} \boldsymbol{x_i} u_i. && (28)
\end{aligned}
$$

Limiting distribution.

Use convergence properties (i) $\operatorname{plim} A = a,\ B \xrightarrow{d} b \implies AB \xrightarrow{d} ab$ (ii) $\operatorname{plim} \bar{c} = \mathbb{E}[c]$ (iii) continuous mapping of plim, and apply the central limit theorem:

$$
\begin{aligned}
\operatorname{plim}\left(\left(\frac{1}{n}\sum_{i=1}^{n} \boldsymbol{x_i x_i'}\right)^{-1}\right) =& \mathbb{E}\left[\boldsymbol{xx'}\right]^{-1} && (29)\\
\frac{1}{n}\sum_{i=1}^{n} \boldsymbol{x_i} u_i \xrightarrow{d}& \mathcal{N}\left(\mathbb{E}[\boldsymbol{x}u], \frac{\mathbf{Var}[\boldsymbol{x}u]}{n}\right), && (30)\\
\mathbf{Var}[\boldsymbol{x}u] =& \mathbb{E}\left[(\boldsymbol{x}u - \mathbb{E}[\boldsymbol{x}u])(\boldsymbol{x}u - \mathbb{E}[\boldsymbol{x}u])'\right] && (31)\\
=& \mathbb{E}\left[\boldsymbol{x'}uu'\boldsymbol{x'}\right] && (32)\\
=& \sigma^2\,\mathbb{E}\left[\boldsymbol{xx'}\right] \implies && (33)
\end{aligned}
$$

$$
\begin{aligned}
\hat{\boldsymbol{\beta}}^{OLS} \xrightarrow{d}& \boldsymbol{\beta} + \mathbb{E}\left[\boldsymbol{xx'}\right]^{-1}\mathcal{N}\left(\mathbf{0}, \frac{\mathbb{E}[\boldsymbol{xx'}]}{n}\right) && (34)\\
\xrightarrow{d}& \mathcal{N}\left(\boldsymbol{\beta}, \frac{\sigma^2}{n}\mathbb{E}\left[\boldsymbol{xx'}\right]^{-1}\mathbb{E}\left[\boldsymbol{xx'}\right]\mathbb{E}\left[\boldsymbol{xx'}\right]^{-1}\right) && (35)\\
\xrightarrow{d}& \mathcal{N}\left(\boldsymbol{\beta}, \frac{\sigma^2}{n}\mathbb{E}\left[\boldsymbol{xx'}\right]^{-1}\right) && (36)\\
\xrightarrow{d}& \mathcal{N}\left(\boldsymbol{\beta}, \frac{\sigma^2}{n}\boldsymbol{Q}^{-1}\right) = \mathcal{N}(\boldsymbol{\beta}, \mathbf{0}) \quad \text{because } n \to \infty. && (37)
\end{aligned}
$$

It is important to distinguish between limiting distribution and asymptotic distribution. As we see above, the limiting distribution exhibits zero variance, with which no inference can be made. Hence, we define the asymptotic distribution, where the variance does not collapse, as follows.

Asymptotic distribution.

$$\sqrt{n}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \overset{A}{\sim} \mathcal{N}(\mathbf{0}, \sigma^2 \, \mathbb{E}\left[\boldsymbol{x}\boldsymbol{x'}^{-1}\right]). \tag{38}$$

Finite distribution.

Here we are restricted to the conditional distribution $\hat{\boldsymbol{\beta}}^{OLS} \mid \boldsymbol{X}$. As $\boldsymbol{u} \sim \mathcal{N}(\mathbf{0}, \sigma^2 \boldsymbol{I})$ is assumed, from $\hat{\boldsymbol{\beta}}^{OLS} = \boldsymbol{\beta} + (\boldsymbol{X'X})^{-1}\boldsymbol{X'u}$ we have

$$\hat{\boldsymbol{\beta}}^{OLS} \mid \boldsymbol{X} \sim \mathcal{N}(\boldsymbol{\beta}, (\boldsymbol{X'X})^{-1}\boldsymbol{X'}\sigma^2\boldsymbol{I}\boldsymbol{X}(\boldsymbol{X'X})^{-1}) \tag{39}$$

$$\sim \mathcal{N}(\boldsymbol{\beta}, \sigma^2(\boldsymbol{X'X})^{-1}). \tag{40}$$

2. Week 1, lecture 2 (Sep 21)

Topic: maximum likelihood and least squares estimators

1. Maximum likelihood estimation of multivariate normal distribution parameters under homoskedasticity.

   Given $\boldsymbol{x} \in \mathcal{M}^{M\times 1} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ with $\boldsymbol{\Sigma} = \sigma^2\boldsymbol{I}$ and i.i.d. sample $\{\boldsymbol{x_i}\}_{i=1}^n$, we have

$$f(\boldsymbol{x_i}, \boldsymbol{\mu}, \sigma^2) = (2\pi)^{-\frac{M}{2}} |\sigma^2\boldsymbol{I}|^{-\frac{1}{2}} e^{-\frac{1}{2}(\boldsymbol{x_i}-\boldsymbol{\mu})'(\sigma^2\boldsymbol{I})^{-1}(\boldsymbol{x_i}-\boldsymbol{\mu})} := \mathscr{L}(\boldsymbol{x_i}, \boldsymbol{\mu}, \sigma^2) \tag{41}$$

$$\log \mathscr{L}(\boldsymbol{x_i}, \boldsymbol{\mu}, \sigma^2) = -\frac{M}{2}\log(2\pi) - \frac{1}{2}\log(|\sigma^2\boldsymbol{I}|) - \frac{1}{2}(\boldsymbol{x_i}-\boldsymbol{\mu})'(\sigma^2\boldsymbol{I})^{-1}(\boldsymbol{x_i}-\boldsymbol{\mu}) \tag{42}$$

$$= -\frac{M}{2}\log(2\pi) - \frac{M}{2}\log(\sigma^2) - \frac{1}{2\sigma^2}(\boldsymbol{x_i}-\boldsymbol{\mu})'(\boldsymbol{x_i}-\boldsymbol{\mu}) \overset{\text{i.i.id.}}{\Longrightarrow} \tag{43}$$

$$\log \mathscr{L}(\{\boldsymbol{x_i}\}_{i=1}^n, \boldsymbol{\mu}, \sigma^2) = \sum_{i=1}^n \log \mathscr{L}(\boldsymbol{x_i}, \boldsymbol{\mu}, \sigma^2) \tag{44}$$

$$= -\frac{nM}{2}\log(2\pi) - \frac{nM}{2}\log(\sigma^2) - \frac{1}{\sigma^2}\sum_{i=1}^n(\boldsymbol{x_i}-\boldsymbol{\mu})'(\boldsymbol{x_i}-\boldsymbol{\mu}) \tag{45}$$

$$= -\frac{nM}{2}\log(2\pi) - \frac{nM}{2}\log(\sigma^2) - \frac{1}{\sigma^2}\sum_{i=1}^n(\boldsymbol{x_i'x_i} - \boldsymbol{x_i'\mu} - \boldsymbol{\mu'x_i} + \boldsymbol{\mu'\mu}) \tag{46}$$

$$:= \ell(.), \tag{47}$$

then the FOCs and the estimates

$$\frac{\partial \ell(.)}{\partial \boldsymbol{\mu}} = \mathbf{0} \tag{48}$$

$$-\frac{1}{2\sigma^2}\sum_{i=1}^{n} = (-2\boldsymbol{x_i} + 2\boldsymbol{\mu}) = \mathbf{0} \implies \hat{\boldsymbol{\mu}}^{ML} = \sum_i \boldsymbol{x_i}/n = \bar{\boldsymbol{x}} \tag{49}$$

$$\frac{\partial \ell(.)}{\partial (\sigma^2)} = 0 \tag{50}$$

$$-\frac{nM}{2\sigma^2} + \frac{1}{2\sigma^4}\sum_i (\boldsymbol{x_i} - \boldsymbol{\mu})'(\boldsymbol{x_i} - \boldsymbol{\mu}) = 0 \implies \widehat{\sigma^2}^{ML} = \sum_i (\boldsymbol{x_i} - \bar{\boldsymbol{x}})'(\boldsymbol{x_i} - \bar{\boldsymbol{x}})/(nM), \tag{51}$$

where we substituted in $\hat{\boldsymbol{\mu}}^{ML}$ while solving for $\widehat{\sigma^2}^{ML}$. Hence the Hessian,

$$\boldsymbol{H} = \frac{\partial^2 \ell(.)}{\partial \begin{bmatrix}\boldsymbol{\mu}' & \sigma^2\end{bmatrix}' \partial \begin{bmatrix}\boldsymbol{\mu}' & \sigma^2\end{bmatrix}} = \begin{bmatrix} -\frac{n}{\sigma^2}\boldsymbol{I}_M & -\frac{1}{\sigma^4}\sum_i (\boldsymbol{x_i} - \boldsymbol{\mu}) \\ -\frac{1}{\sigma^4}\sum_i (\boldsymbol{x_i} - \boldsymbol{\mu}) & \frac{nM}{2\sigma^4} - \frac{\sum_i (\boldsymbol{x_i}-\boldsymbol{\mu})'(\boldsymbol{x_i}-\boldsymbol{\mu})}{6} \end{bmatrix} \tag{52}$$

$$\mathbb{E}\left[\boldsymbol{H}\right] = \begin{bmatrix} -\frac{n}{\sigma^2}\boldsymbol{I}_M & \mathbf{0}_M \\ \mathbf{0}'_M & \frac{nM}{2\sigma^4} - \frac{nM\sigma^2}{\sigma^6} \end{bmatrix} \tag{53}$$

$$= \begin{bmatrix} -\frac{n}{\sigma^2}\boldsymbol{I}_M & \mathbf{0}_M \\ \mathbf{0}'_M & -\frac{nM}{2\sigma^4} \end{bmatrix} \tag{54}$$

$$(\mathbb{E}\left[\boldsymbol{H}\right])^{-1} = \begin{bmatrix} -\frac{\sigma^2}{n}\boldsymbol{I}_M & \mathbf{0}_M \\ \mathbf{0}'_M & -\frac{2\sigma^4}{nM} \end{bmatrix} \tag{55}$$

$$\mathbf{Var}\left[\begin{bmatrix}\hat{\boldsymbol{\mu}}' & \widehat{\sigma^2}^{ML}\end{bmatrix}'\right] = -(\mathbb{E}\left[\boldsymbol{H}\right])^{-1} \tag{56}$$

$$= \begin{bmatrix} \frac{\sigma^2}{n}\boldsymbol{I}_M & \mathbf{0}_M \\ \mathbf{0}'_M & \frac{2\sigma^4}{nM} \end{bmatrix} \tag{57}$$

$$= -(\boldsymbol{I}(\Theta))^{-1}, \quad \text{the inverse information matrix} \tag{58}$$

$$= \text{Cramer-Rao lower bound (CRLB)} \tag{59}$$

2. Least squares estimation of univariate normal parameters. I.i.d. sample: $\{x_i\}_{i=1}^n$, $x \sim \mathcal{N}(\mu, \sigma^2)$.

   First, μ.

$$\hat{\mu}^{LS} = \arg\min_\mu \sum_{i=1}^{n} (x_i - \mu)^2 \implies \text{FOC}: \tag{60}$$

$$= -2\sum_i (x_i - \mu) = 0 \implies \hat{\mu}^{LS} = \bar{x}. \tag{61}$$

Second, σ^2. Let $\hat{\mu}$ denote $\hat{\mu}^{LS}$ from now on. Then

$$\begin{aligned}
\mathbb{E}\left[\sum_{i=1}^{n}(x_i-\hat{\mu})^2\right] &= \mathbb{E}\left[\sum_i (x_i^2 - 2x_i\hat{\mu} + \hat{\mu}^2)\right] && (62)\\
&= \mathbb{E}\left[\sum_i x_i^2 - 2\hat{\mu}\sum_i x_i + n\hat{\mu}^2\right] && (63)\\
&= \mathbb{E}\left[\sum_i x_i^2\right] - 2\,\mathbb{E}\left[\hat{\mu}\sum_i x_i\right] + n\,\mathbb{E}\left[\hat{\mu}^2\right] && (64)\\
&\overset{\text{iid}}{=} n\,\mathbb{E}\left[x_i^2\right] - 2\,\mathbb{E}\left[\frac{\left(\sum_i x_i\right)}{n}\left(\sum_i x_i\right)\right] + n\,\mathbb{E}\left[\frac{\left(\sum_i x_i\right)\left(\sum_i x_i\right)}{n^2}\right] && (65)\\
&= n\,\mathbb{E}\left[x_i^2\right] - \frac{1}{n}\,\mathbb{E}\left[\left(\sum_i x_i\right)\left(\sum_i x_i\right)\right] && (66)\\
&= n\,\mathbb{E}\left[x_i^2\right] - \frac{1}{n}\left(n\,\mathbb{E}\left[x_i^2\right] + n(n-1)\,\mathbb{E}\left[x_i x_j\right]\right) \quad i \neq j && (67)\\
&= n\,\mathbb{E}\left[x_i^2\right] - \frac{1}{n}\left(n\,\mathbb{E}\left[x_i^2\right] + n(n-1)\,\mathbb{E}\left[x_i\right]\mathbb{E}\left[x_j\right]\right) \quad \text{iid sample:} x_i \perp\!\!\!\perp x_j && (68)\\
&= n\,\mathbb{E}\left[x_i^2\right] - \frac{1}{n}\left(n\,\mathbb{E}\left[x_i^2\right] + n(n-1)\mu^2\right) && (69)\\
&= n\,\mathbb{E}\left[x_i^2\right] - \mathbb{E}\left[x_i^2\right] - (n-1)\mu^2 && (70)\\
&= (n-1)\,\mathbb{E}\left[x_i^2\right] - (n-1)\mu^2 && (71)\\
&= (n-1)(\mathbb{E}\left[x_i^2\right] - \mu^2) \quad (\mathbf{Var}\,W = \mathbb{E}\left[W^2\right] - \mathbb{E}\left[W\right]^2) && (72)\\
&= (n-1)\sigma^2 \implies && (73)\\
\widehat{\sigma^2}^{LS} &= \frac{\sum_{i=1}^{n}(x_i-\hat{\mu})^2}{n-1}. && (74)
\end{aligned}$$

3. Least squares estimation of the classical regression model $\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{u}$, $\mathbb{E}\left[\boldsymbol{y} \mid \boldsymbol{X}\right] = \boldsymbol{X}\boldsymbol{\beta}$ under homoskedasticity: $\boldsymbol{u} \sim \mathcal{N}(\boldsymbol{0}, \sigma^2\boldsymbol{I})$.

   First, $\boldsymbol{\beta}$.

$$\begin{aligned}
\hat{\boldsymbol{\beta}}^{LS} &= \arg\min_{\boldsymbol{\beta}}(\boldsymbol{y}-\boldsymbol{X}\boldsymbol{\beta})'(\boldsymbol{y}-\boldsymbol{X}\boldsymbol{\beta}) \iff && (75)\\
\hat{\boldsymbol{\beta}}^{LS} &= \arg\min_{\boldsymbol{\beta}}(\boldsymbol{y}'\boldsymbol{y} - \boldsymbol{y}'\boldsymbol{X}\boldsymbol{\beta} - \boldsymbol{\beta}'\boldsymbol{X}'\boldsymbol{y} + \boldsymbol{\beta}'\boldsymbol{X}'\boldsymbol{X}\boldsymbol{\beta}) \implies \text{FOC:} && (76)
\end{aligned}$$

$$-\boldsymbol{X}'\boldsymbol{y} - \boldsymbol{X}'\boldsymbol{y} + 2\boldsymbol{X}'\boldsymbol{X}\boldsymbol{\beta} = \boldsymbol{0}_K \implies \hat{\boldsymbol{\beta}}^{LS} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y}. \quad (77)$$

Second, σ^2. Entirely based on Greene's treatment (Econometric Analysis, p. 51 in 6th edition).

$$\begin{aligned}
\mathbb{E}\left[\hat{\boldsymbol{u}}'\hat{\boldsymbol{u}} \mid \boldsymbol{X}\right] &= \mathbb{E}\left[\boldsymbol{u}'\boldsymbol{M}'\boldsymbol{M}\boldsymbol{u} \mid \boldsymbol{X}\right] \quad \boldsymbol{M} \text{ is symm. and idempotent } \Longrightarrow && (78)\\
&= \mathbb{E}\left[\boldsymbol{u}'\boldsymbol{M}\boldsymbol{u} \mid \boldsymbol{X}\right] && (79)\\
&= \mathbb{E}\left[\operatorname{tr}\left(\boldsymbol{u}'\boldsymbol{M}\boldsymbol{u}\right) \mid \boldsymbol{X}\right] \quad \text{trace of a scalar is the scalar itself} && (80)\\
&= \mathbb{E}\left[\operatorname{tr}\left(\boldsymbol{M}\boldsymbol{u}\boldsymbol{u}'\right) \mid X\right] \quad \text{trace property} && (81)\\
&= \operatorname{tr}\left(\mathbb{E}\left[\boldsymbol{M}\boldsymbol{u}\boldsymbol{u}' \mid \boldsymbol{X}\right]\right) \quad \text{linearity of expectation} && (82)\\
&= \operatorname{tr}\left(\boldsymbol{M}\,\mathbb{E}\left[\boldsymbol{u}\boldsymbol{u}' \mid \boldsymbol{X}\right]\right) && (83)\\
&= \operatorname{tr}\left(\boldsymbol{M}\sigma^2\boldsymbol{I}\right) && (84)\\
&= \sigma^2\operatorname{tr}\left(M\right). && (85)
\end{aligned}$$

Stop here for a minute and see where $\boldsymbol{M}$ came from. It is the so called residual-maker (as $\boldsymbol{My} = \hat{\boldsymbol{u}}$), is defined as $\boldsymbol{M} = \boldsymbol{I_n} - \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'$ and has the nice property:

$$\begin{aligned}
\hat{\boldsymbol{u}} &= \boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\beta}} && (86)\\
&= \boldsymbol{y} - \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}\boldsymbol{y} && (87)\\
&= (\boldsymbol{I_n} - \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}')\boldsymbol{y} && (88)\\
&= (\boldsymbol{I_n} - \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}')(\boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{u}) && (89)\\
&= (\boldsymbol{I_n} - \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}')\boldsymbol{u} && (90)\\
&= \boldsymbol{M}\boldsymbol{u}. && (91)
\end{aligned}$$

Therefore to pick up where we left

$$\begin{aligned}
\mathbb{E}\left[\hat{\boldsymbol{u}}'\hat{\boldsymbol{u}} \mid \boldsymbol{X}\right] &= \sigma^2\operatorname{tr}\left(M\right) && (92)\\
&= \sigma^2(\operatorname{tr}\left((\boldsymbol{I_n} - \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}')\right)) && (93)\\
&= \sigma^2\left(\operatorname{tr}\left(I_n\right) - \operatorname{tr}\left(\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}\right)\right) && (94)\\
&= \sigma^2\left(n - \operatorname{tr}\left((\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{X}\right)\right) \quad \text{trace property} && (95)\\
&= \sigma^2\left(n - \operatorname{tr}\left(\boldsymbol{I_K}\right)\right) && (96)\\
&= \sigma^2(n-K) \Longrightarrow \widehat{\sigma^2} = \frac{\hat{\boldsymbol{u}}'\hat{\boldsymbol{u}}}{n-K} && (97)
\end{aligned}$$

. . . .