Class V

Multiple Linear Regressions

December 4, 2024

Theorem

Assume that the data generating process of (x_j, y_j) *is*

$$y_j = \beta_0 + \beta_1 x_j + \varepsilon_j, \quad \mathbb{E}\left[\varepsilon_j\right] = 0, \quad \mathbb{E}\left[\varepsilon_j \mid X\right] = \mathbb{E}\left[\varepsilon_j\right] = 0.$$

and that ε_j *are independent. Assume further that*

$$\mathbb{V}\left[\varepsilon_j \mid x\right] = \mathbb{E}\left[\varepsilon_j^2 \mid x\right] = \mathbb{E}\left[\varepsilon_j^2\right] = \sigma_\varepsilon^2.$$

Therefore

$$\begin{aligned}
\mathbb{V}\left[\widehat{\beta}_1 \,\middle|\, x\right] &= \frac{\sigma_\varepsilon^2}{\sum_{j=1}^n \left(x_j - \overline{x}_n\right)^2} = \frac{\sigma_\varepsilon^2}{\mathsf{S}_2\left(x\right)} \\
\mathbb{V}\left[\widehat{\beta}_0 \,\middle|\, x\right] &= \frac{\sigma_\varepsilon^2}{n} \frac{\sum_{j=1}^n x_j^2}{\sum_{j=1}^n \left(x_j - \overline{x}_n\right)^2} = \frac{\sigma_\varepsilon^2}{n} \frac{\sum_{j=1}^n x_j^2}{\mathsf{S}_2\left(x\right)}.
\end{aligned}$$

$$
\begin{aligned}
\widehat{\beta}_1 &= \beta_1 + \frac{\sum_{j=1}^{n} \overbrace{\left(x_j - \overline{x}_n\right)}^{w_j} \varepsilon_j}{S_2(x)} = \beta_1 + \frac{\sum_{j=1}^{n} w_j \, \varepsilon_j}{S_2(x)} \Rightarrow \\
\widehat{\beta}_1^2 &= \beta_1^2 + \frac{\left(\sum_{j=1}^{n} w_j \, \varepsilon_j\right)^2}{S_2(x)^2} + 2\,\beta_1 \frac{\sum_{j=1}^{n} w_j \, \varepsilon_j}{S_2(x)} \\
&= \beta_1^2 + \frac{\sum_{j=1}^{n} w_j^2 \, \varepsilon_j^2 + \sum_{\ell \neq j} w_j \, w_\ell \, \varepsilon_j \, \varepsilon_\ell}{S_2(x)^2} + 2\,\beta_1 \frac{\sum_{j=1}^{n} w_j \, \varepsilon_j}{S_2(x)} \Rightarrow \\
\mathbb{E}\left[\widehat{\beta}_1^2 \,\middle|\, x\right] &= \beta_1^2 + \frac{\sum_{j=1}^{n} w_j^2 \, \mathbb{E}\left[\varepsilon_j^2 \mid x\right] + \sum_{\ell \neq j} w_j \, w_\ell \, \overbrace{\mathbb{E}\left[\varepsilon_j \, \varepsilon_\ell \mid x\right]}^{=0 \text{ by indep.}}}{S_2(x)^2} + \\
&\quad +2\,\beta_1 \frac{\sum_{j=1}^{n} w_j \, \overbrace{\mathbb{E}\left[\varepsilon_j \mid x\right]}^{=0}}{S_2(x)} = \beta_1^2 + \sigma_\varepsilon^2 \frac{\sum_{j=1}^{n} w_j^2}{S_2(x)^2} \\
&= \beta_1^2 + \sigma_\varepsilon^2 \frac{\sum_{j=1}^{n} \left(x_j - \overline{x}_n\right)^2}{S_2(x)^2} = \beta_1^2 + \sigma_\varepsilon^2 \frac{S_2(x)}{S_2(x)^2} = \beta_1^2 + \frac{\sigma_\varepsilon^2}{S_2(x)}.
\end{aligned}
$$

$$\mathbb{E}\left[\widehat{\beta}_1^2 \,\middle|\, x\right] = \beta_1^2 + \frac{\sigma_\varepsilon^2}{\mathsf{S}_2(x)}.$$

Since $\mathbb{E}\left[\widehat{\beta}_1 \,\middle|\, x\right] = \beta_1$ we get

$$\mathbb{V}\left[\widehat{\beta}_1 \,\middle|\, x\right] = \mathbb{E}\left[\widehat{\beta}_1^2 \,\middle|\, x\right] - \left(\mathbb{E}\left[\widehat{\beta}_1\right] \,\middle|\, x\right)^2 = \frac{\sigma_\varepsilon^2}{\mathsf{S}_2(x)}.$$

Remarks. Note that

I) As $\sigma_\varepsilon \to 0$ we have $\mathbb{V}\left[\widehat{\beta}_1 \,\middle|\, x\right] \to 0$: the estimator gets more and more precise.

II) As $n \to \infty$ we (typically) have

$$\mathsf{S}_2(x)^2 = \sum_{j=1}^{n} (x_j - \overline{x}_n)^2 \to \infty \text{ (in some sense, e.g. almost surely)},$$

and so, again,

$$\mathbb{V}\left[\widehat{\beta}_1 \,\middle|\, x\right] \to 0$$

the estimator gets more and more precise.

Concerning the variance of $\widehat{\beta}_0$ recall that

$$\begin{aligned}
\widehat{\beta}_0 &= \overline{y}_n - \widehat{\beta}_1\,\overline{x}_n = \beta_0 + \beta_1\,\overline{x}_n + \overline{\varepsilon}_n - \widehat{\beta}_1\,\overline{x}_n = \beta_0 + \left(\beta_1 - \widehat{\beta}_1\right)\overline{x}_n + \overline{\varepsilon}_n \\
\widehat{\beta}_0^2 &= \beta_0^2 + \left(\beta_1 - \widehat{\beta}_1\right)^2 \overline{x}_n^2 + \overline{\varepsilon}_n^2 + 2\,\beta_0\left(\beta_1 - \widehat{\beta}_1\right) x_n + 2\,\beta_0\,\overline{\varepsilon}_n \\
&\quad +2\left(\beta_1 - \widehat{\beta}_1\right)\overline{x}_n\,\overline{\varepsilon}_n \Rightarrow \\
\mathbb{E}\left[\widehat{\beta}_0^2 \,\middle|\, X\right] &= \beta_0^2 + \mathbb{V}\left[\widehat{\beta}_1 \,\middle|\, X\right]\overline{x}_n^2 + \mathbb{E}\left[\overline{\varepsilon}_n^2 \,\middle|\, X\right] + 0 + 0 \\
&\quad +2\,\overline{x}_n\,\mathbb{E}\left[\left(\beta_1 - \widehat{\beta}_1\right)\overline{\varepsilon}_n \,\middle|\, x\right].
\end{aligned}$$

Recall that (define $w_j' = (x_j - \overline{x}_n)/\mathsf{S}_2\,(x)$)

$$\begin{aligned}
\widehat{\beta}_1 &= \beta_1 + \frac{\sum_{j=1}^n (x_j - \overline{x}_n)\,\varepsilon_j}{\mathsf{S}_2\,(x)} \Rightarrow \widehat{\beta}_1 - \beta_1 = \frac{\sum_{j=1}^n (x_j - \overline{x}_n)\,\varepsilon_j}{\mathsf{S}_2\,(x)} = \sum_{j=1}^n w_j'\,\varepsilon_j. \\
\mathbb{E}\left[\left(\beta_1 - \widehat{\beta}_1\right)\overline{\varepsilon}_n \,\middle|\, x\right] &= -\mathbb{E}\left[\sum_{j=1}^n w_j'\,\varepsilon_j\,\frac{1}{n}\sum_{\ell=1}^n \varepsilon_\ell \,\middle|\, x\right] = -\frac{1}{n}\sum_{j=1}^n\sum_{\ell=1}^n w_j'\,\mathbb{E}\left[\varepsilon_j\varepsilon_\ell\right] \\
&= -\frac{\sigma_\varepsilon^2}{n}\sum_{j=1}^n w_j' = 0.
\end{aligned}$$

In summary, recalling that $\mathbb{E}\left[\widehat{\beta}_0 \,\middle|\, X\right] = \beta_0$, we have

$$\begin{aligned}
\mathbb{E}\left[\widehat{\beta}_0^2 \,\middle|\, X\right] &= \beta_0^2 + \mathbb{V}\left[\widehat{\beta}_1 \,\middle|\, X\right] \overline{x}_n^2 + \mathbb{E}\left[\overline{\varepsilon}_n^2\right] \\
&= \left(\mathbb{E}\left[\widehat{\beta}_0 \,\middle|\, X\right]\right)^2 + \mathbb{V}\left[\widehat{\beta}_1 \,\middle|\, X\right] \overline{x}_n^2 + \mathbb{E}\left[\left(\frac{1}{n}\sum_{j=1}^{n} \varepsilon_j\right)^2\right] \\
&= \left(\mathbb{E}\left[\widehat{\beta}_0 \,\middle|\, X\right]\right)^2 + \frac{\sigma_\varepsilon^2}{\mathsf{S}_2(x)} \overline{x}_n^2 + \frac{1}{n^2}\left(\sum_{j=1}^{n} \mathbb{E}\left[\varepsilon_j^2\right] + \sum_{j\neq\ell}^{n} \mathbb{E}\left[\varepsilon_j \varepsilon_\ell\right]\right) \\
&= \left(\mathbb{E}\left[\widehat{\beta}_0 \,\middle|\, X\right]\right)^2 + \frac{\sigma_\varepsilon^2}{\mathsf{S}_2(x)} \overline{x}_n^2 + \frac{1}{n^2}\sum_{j=1}^{n} \sigma_\varepsilon^2 \\
&= \left(\mathbb{E}\left[\widehat{\beta}_0 \,\middle|\, X\right]\right)^2 + \frac{\sigma_\varepsilon^2}{\mathsf{S}_2(x)} \overline{x}_n^2 + \frac{1}{n}\sigma_\varepsilon^2 \Rightarrow \\
\mathbb{V}\left[\widehat{\beta}_0^2 \,\middle|\, X\right] &= \frac{\sigma_\varepsilon^2}{\mathsf{S}_2(x)} \overline{x}_n^2 + \frac{1}{n}\sigma_\varepsilon^2 = \sigma_\varepsilon^2\left(\frac{\overline{x}_n^2}{\mathsf{S}_2(x)} + \frac{1}{n}\right).
\end{aligned}$$

In (partial) summary

$$\mathbb{V}\left[\widehat{\beta}_0^2 \,\middle|\, X\right] = \sigma_\varepsilon^2 \left(\frac{\overline{x}_n^2}{\mathsf{S}_2(x)} + \frac{1}{n}\right).$$

Recall that

$$\begin{aligned}
\mathsf{S}_2(x) &= \sum_{j=1}^{n} (x_j - \overline{x}_n)^2 = \sum_{j=1}^{n} \left(x_j^2 + \overline{x}_n^2 - 2\,x_j\,\overline{x}_n\right) \\
&= \sum_{j=1}^{n} x_j^2 + n\,\overline{x}_n^2 - 2\,\overline{x}_n \sum_{j=1}^{n} x_j = \sum_{j=1}^{n} x_j^2 + n\,\overline{x}_n^2 - 2\,n\,\overline{x}_n^2 \\
&= \sum_{j=1}^{n} x_j^2 - n\,\overline{x}_n^2.
\end{aligned}$$

When this expression is plugged into the the denominator we get

$$\begin{aligned}
\mathbb{V}\left[\widehat{\beta}_0^2 \,\middle|\, X\right] &= \sigma_\varepsilon^2 \left(\frac{\overline{x}_n^2}{\sum_{j=1}^{n} x_j^2 - n\,\overline{x}_n^2} + \frac{1}{n}\right) = \sigma_\varepsilon^2 \,\frac{n\,\overline{x}_n^2 + \sum_{j=1}^{n} x_j^2 - n\,\overline{x}_n^2}{n\left(\sum_{j=1}^{n} x_j^2 - n\,\overline{x}_n^2\right)} \\
&= \frac{\sigma_\varepsilon^2}{n}\,\frac{\sum_{j=1}^{n} x_j^2}{\mathsf{S}_2(x)}.
\end{aligned}$$

$$\mathbb{V}\left[\widehat{\beta}_0^2 \,\middle|\, X\right] = \frac{\sigma_\varepsilon^2}{n} \frac{\sum_{j=1}^n x_j^2}{\mathsf{S}_2\left(x\right)}$$

Remarks. Note that

I) As $\sigma_\varepsilon \to 0$ we have $\mathbb{V}\left[\widehat{\beta}_0 \,\middle|\, X\right] \to 0$: the estimator gets more and more precise.

II) As $n \to \infty$ we (typically) have

$$\begin{aligned} \mathsf{S}_2\left(x\right)^2 &= \sum_{j=1}^n \left(x_j - \overline{x}_n\right)^2 \to \infty \\ & \sum_{j=1}^n x_j^2 \to \infty, \end{aligned}$$

and so, again,

$$\mathbb{V}\left[\widehat{\beta}_1 \,\middle|\, X\right] \to 0$$

the estimator gets more and more precise.

Problem

Suppose we have collected data $(x_1, y_1), \cdots, (x_n, y_n)$ and we want to estimate the regression

$$y_j = \beta_0 + \beta_1 x_j + \varepsilon_j.$$

How can we compute

$$\begin{aligned} \mathbb{V}\left[\widehat{\beta}_1 \,\middle|\, X\right] &= \frac{\sigma_\varepsilon^2}{\mathsf{S}_2(x)} \\ \mathbb{V}\left[\widehat{\beta}_0^2 \,\middle|\, X\right] &= \frac{\sigma_\varepsilon^2}{n} \frac{\sum_{j=1}^n x_j^2}{\mathsf{S}_2(x)} \quad ? \end{aligned}$$

The variance σ_ε is not observed!

Theorem

Assume that the data generating process of (x_j, y_j) *is*

$$y_j = \beta_0 + \beta_1 x_j + \varepsilon_j, \quad \mathbb{E}[\varepsilon_j] = 0, \quad \mathbb{E}[\varepsilon_j \mid x_1, \cdots, x_n] = \mathbb{E}[\varepsilon_j] = 0.$$

and that ε_j *are independent. Assume further that*

$$\mathbb{V}[\varepsilon_j \mid x_1, ..., x_n] = \mathbb{E}\left[\varepsilon_j^2 \mid x_1, ..., x_n\right] = \mathbb{E}\left[\varepsilon_j^2\right] = \sigma_\varepsilon^2.$$

Therefore the random variable

$$\widehat{\sigma_\varepsilon^2} \doteq \frac{1}{n-2} \sum_{j=1}^{n} \widehat{\varepsilon}_j^2 = \frac{\mathrm{SSR}}{n-2},$$

is an unbiased estimator of σ_ε^2*, in the sense that*

$$\mathbb{E}\left[\widehat{\sigma_\varepsilon^2}\right] = \sigma_\varepsilon^2.$$

skip proof

Difference between errors and residuals

Recall the definition of residuals

$$\begin{aligned}
\widehat{\varepsilon}_j &= y_j - \widehat{Y}_j \\
&= y_j - \left(\widehat{\beta}_0 + \widehat{\beta}_1 x_j\right) \\
&= \beta_0 + \beta_1 x_j + \varepsilon_j - \widehat{\beta}_0 - \widehat{\beta}_1 x_j \\
&= \beta_0 - \widehat{\beta}_0 + \left(\beta_1 - \widehat{\beta}_1\right) x_j + \varepsilon_j .
\end{aligned}$$

In summary

$$\underbrace{\widehat{\varepsilon}_j}_{\text{Residuals}} = \underbrace{\varepsilon_j}_{\text{Errors}} - \left(\widehat{\beta}_0 - \beta_0\right) - \left(\widehat{\beta}_1 - \beta_1\right) x_j .$$

So that

$$\begin{aligned}
\mathbb{E}\left[\widehat{\varepsilon}_j\right] &= \mathbb{E}\left[\mathbb{E}\left[\widehat{\varepsilon}_j \mid X\right]\right] = \mathbb{E}\left[\mathbb{E}\left[\varepsilon_j - \left(\widehat{\beta}_0 - \beta_0\right) - \left(\widehat{\beta}_1 - \beta_1\right) x_j \,\middle|\, X\right]\right] \\
&= \mathbb{E}\left[\mathbb{E}\left[\varepsilon_j \mid X\right]\right] = \mathbb{E}\left[\varepsilon_j\right] = 0.
\end{aligned}$$

Recall that the first of the FOC's is

$$\begin{aligned} 0 &= \sum_{j=1}^{n} \widehat{\varepsilon}_j \Leftrightarrow \\ 0 &= \sum_{j=1}^{n} \left(\beta_0 - \widehat{\beta}_0 + \left(\beta_1 - \widehat{\beta}_1 \right) x_j + \varepsilon_j \right) \Leftrightarrow \\ 0 &= \overline{\varepsilon}_n - \left(\widehat{\beta}_0 - \beta_0 \right) - \left(\widehat{\beta}_1 - \beta_1 \right) \overline{x}_n . \end{aligned}$$

Whence

$$\begin{aligned} \widehat{\varepsilon}_j &= \varepsilon_j - \left(\widehat{\beta}_0 - \beta_0 \right) - \left(\widehat{\beta}_1 - \beta_1 \right) x_j \\ \widehat{\varepsilon}_j - 0 &= \varepsilon_j - \left(\widehat{\beta}_0 - \beta_0 \right) - \left(\widehat{\beta}_1 - \beta_1 \right) x_j - \left[\overline{\varepsilon}_n - \left(\widehat{\beta}_0 - \beta_0 \right) - \left(\widehat{\beta}_1 - \beta_1 \right) \overline{x}_n \right] \\ \widehat{\varepsilon}_j &= \varepsilon_j - \overline{\varepsilon}_n - \left(\widehat{\beta}_1 - \beta_1 \right) (x_j - \overline{x}_n) . \end{aligned}$$

From

$$\widehat{\varepsilon}_j = \varepsilon_j - \overline{\varepsilon}_n - \left(\widehat{\beta}_1 - \beta_1\right)\,(x_j - \overline{x}_n)$$

we have

$$\widehat{\varepsilon}_j^2 = (\varepsilon_j - \overline{\varepsilon}_n)^2 + \left(\widehat{\beta}_1 - \beta_1\right)^2 (x_j - \overline{x}_n)^2 - 2\,(\varepsilon_j - \overline{\varepsilon}_n)\left(\widehat{\beta}_1 - \beta_1\right)(x_j - \overline{x}_n)$$

So that ...

$$\begin{aligned}\sum_{j=1}^n \widehat{\varepsilon}_j^2 \quad = \quad & \underbrace{\sum_{j=1}^n (\varepsilon_j - \overline{\varepsilon}_n)^2}_{\mathrm{A}} \\ & + \underbrace{\sum_{j=1}^n \left(\widehat{\beta}_1 - \beta_1\right)^2 (x_j - \overline{x}_n)^2}_{\mathrm{B}} \\ & -2\underbrace{\left(\widehat{\beta}_1 - \beta_1\right)\sum_{j=1}^n (\varepsilon_j - \overline{\varepsilon}_n)\,(x_j - \overline{x}_n)}_{\mathrm{C}} = \mathrm{A} + \mathrm{B} - 2\,\mathrm{C}\end{aligned}$$

We have thus to compute $\mathbb{E}\,[\mathrm{A}]$, $\mathbb{E}\,[\mathrm{B}]$ and $\mathbb{E}\,[\mathrm{C}]$.

$$\mathbb{E}[\mathrm{A}] = \mathbb{E}\left[\sum_{j=1}^{n} (\varepsilon_j - \overline{\varepsilon}_n)^2\right] = \mathbb{E}\left[\sum_{j=1}^{n} \left(\varepsilon_j^2 + \overline{\varepsilon}_n^2 - 2\,\varepsilon_j\,\overline{\varepsilon}_n\right)\right]$$

Recall that $\overline{\varepsilon}_n = \frac{1}{n}\sum_{\ell=1}^{n} \varepsilon_\ell$ is the **sample mean of the errors** $\Rightarrow$ it contains ε_j

$$\begin{aligned}
\mathbb{E}[\mathrm{A}] &= \mathbb{E}\left[\sum_{j=1}^{n}\left(\varepsilon_j^2 + \overline{\varepsilon}_n^2 - 2\,\varepsilon_j\,\frac{\varepsilon_1 + \varepsilon_2 + \cdots + \varepsilon_j + \cdots + \varepsilon_n}{n}\right)\right] \\
&= \mathbb{E}\left[\sum_{j=1}^{n}\left(\varepsilon_j^2 + \overline{\varepsilon}_n^2 - 2\,\frac{\varepsilon_j\,\varepsilon_1 + \varepsilon_j\,\varepsilon_2 + \cdots + \varepsilon_j^2 + \cdots + \varepsilon_j\,\varepsilon_n}{n}\right)\right] \\
&= \sum_{j=1}^{n}\left(\mathbb{E}\left[\varepsilon_j^2\right] + \mathbb{E}\left[\overline{\varepsilon}_n^2\right] - 2\,\frac{\mathbb{E}[\varepsilon_j\,\varepsilon_1] + \cdots + \mathbb{E}\left[\varepsilon_j^2\right] + \cdots + \mathbb{E}[\varepsilon_j\,\varepsilon_n]}{n}\right) \\
&= \sum_{j=1}^{n}\left(\sigma_\varepsilon^2 + \frac{1}{n^2}\,\mathbb{E}\left[\sum_{\ell}\varepsilon_n^2 + \sum_{\ell\neq m}\varepsilon_\ell\,\varepsilon_m\right] - \frac{2\,\sigma_\varepsilon^2}{n}\right) = \\
&= \sum_{j=1}^{n}\left(\sigma_\varepsilon^2 + \frac{n}{n^2}\,\sigma_\varepsilon^2 - \frac{2\,\sigma_\varepsilon^2}{n}\right) = n\,\sigma_\varepsilon^2 + \sigma_\varepsilon^2 - 2\,\sigma_\varepsilon^2 = (n-1)\;\sigma_\varepsilon^2.
\end{aligned}$$

$$
\begin{aligned}
\mathbb{E}\left[\mathrm{B}\right] &= \mathbb{E}\left[\sum_{j=1}^{n}\left(\widehat{\beta}_1-\beta_1\right)^2\left(x_j-\overline{x}_n\right)^2\right] \\
&= \sum_{j=1}^{n}\mathbb{E}\left[\left(\widehat{\beta}_1-\beta_1\right)^2\left(x_j-\overline{x}_n\right)^2\right] \\
&= \sum_{j=1}^{n}\mathbb{E}\left[\mathbb{E}\left[\left(\widehat{\beta}_1-\beta_1\right)^2\left(x_j-\overline{x}_n\right)^2 \,\middle|\, X\right]\right] \\
&= \sum_{j=1}^{n}\mathbb{E}\left[\left(x_j-\overline{x}_n\right)^2\mathbb{E}\left[\left(\widehat{\beta}_1-\beta_1\right)^2 \,\middle|\, X\right]\right] \\
&= \mathbb{E}\left[\sum_{j=1}^{n}\left(x_j-\overline{x}_n\right)^2\mathbb{E}\left[\left(\widehat{\beta}_1-\beta_1\right)^2 \,\middle|\, X\right]\right] \\
&= \mathbb{E}\left[\mathrm{S}_2\left(x\right)\mathbb{V}\left[\widehat{\beta}_1 \,\middle|\, X\right]\right] \\
&= \mathbb{E}\left[\mathrm{S}_2\left(x\right)\frac{\sigma_\varepsilon^2}{\mathrm{S}_2\left(x\right)}\right] = \sigma_\varepsilon^2.
\end{aligned}
$$

$$\mathbb{E}\left[C\right] = \mathbb{E}\left[\left(\widehat{\beta}_1 - \beta_1\right) \sum_{j=1}^{n} \left(\varepsilon_j - \overline{\varepsilon}_n\right) \left(x_j - \overline{x}_n\right)\right].$$

Recall that

$$\begin{aligned}
\widehat{\beta}_1 - \beta_1 &= \frac{\sum_{j=1}^{n} \left(x_j - \overline{x}_n\right) \varepsilon_j}{S_2\left(x\right)} = \frac{\sum_{j=1}^{n} \left(x_j - \overline{x}_n\right) \left(\varepsilon_j - \overline{\varepsilon}_n + \overline{\varepsilon}_n\right)}{S_2\left(x\right)} \\
&= \frac{\sum_{j=1}^{n} \left(x_j - \overline{x}_n\right) \left(\varepsilon_j - \overline{\varepsilon}_n\right)}{S_2\left(x\right)} + \overline{\varepsilon}_n \frac{\sum_{j=1}^{n} \left(x_j - \overline{x}_n\right)}{S_2\left(x\right)} \\
&= \frac{\sum_{j=1}^{n} \left(x_j - \overline{x}_n\right) \left(\varepsilon_j - \overline{\varepsilon}_n\right)}{S_2\left(x\right)}.
\end{aligned}$$

Thus

$$\mathbb{E}\left[C\right] = \mathbb{E}\left[\left(\widehat{\beta}_1 - \beta_1\right)^2 S_2\left(x\right)\right] = \mathbb{E}\left[\mathbb{E}\left[\left(\widehat{\beta}_1 - \beta_1\right)^2 S_2\left(x\right) \,\middle|\, X\right]\right] = \sigma_\varepsilon^2.$$

So, finally, recalling that

$$\sum_{j=1}^{n} \widehat{\varepsilon}_j^2 = \mathrm{A} + \mathrm{B} - 2\,\mathrm{C}$$

we have

$$\mathbb{E}\left[\sum_{j=1}^{n} \widehat{\varepsilon}_j^2\right] = \mathbb{E}\,[\mathrm{A}] + \mathbb{E}\,[\mathrm{B}] - 2\,\mathbb{E}\,[\mathrm{C}] = (n-1)\,\sigma_\varepsilon^2 + \sigma_\varepsilon^2 - 2\,\sigma_\varepsilon^2 = (n-2)\,\sigma_\varepsilon^2,$$

which is what we wanted to prove

$$\mathbb{E}\left[\frac{1}{n-2}\sum_{j=1}^{n} \widehat{\varepsilon}_j^2\right] = \sigma_\varepsilon^2.$$

$$Y_t = \beta_0 + \beta_1 X_t + \varepsilon_t. \Rightarrow \text{Too simple!}$$

The multiple regression model: two regressors

$$Y_t = \beta_0 + \beta_1 X_{t,1} + \beta_2 X_{t,2} + \varepsilon_t. \quad (\star)$$

I) Y_t = the t-th observation of the dependent variable (e.g. the daily return of a stock).

II) $X_{t,1}$ and $X_{t,2}$ the t-th observation of the two regressors (e.g. explaining factors of the stock return).

III) ε_t unknown factors.

Assume that in $(\star)$ only $X_{t,1}$ varies of $\Delta X_{t,1}$ ...

$$\Delta Y_t = \beta_1 \Delta X_{t,1} \Rightarrow \beta_1 = \frac{\Delta Y_t}{\Delta X_{t,1}} \text{ (and similarly for } \beta_2\text{).}$$

β_j = effect of a unit variation of the j-th regressor on Y keeping everything else constant.

$$Y_t = \beta_0 + \beta_1 X_{t,1} + \beta_2 X_{t,2} + \varepsilon_t. \quad (\star)$$

we will always assume that

I) $\mathbb{E}[\varepsilon_t] = 0$, any other value will be compensated by β_0.

II) $\mathbb{E}[\varepsilon_t \mid X_1, X_2] = \mathbb{E}[\varepsilon_t] = 0$.

Dummy variables

Assume that $X_{t,1} \in \{0, 1\}$. Then

$$\begin{aligned}
\mathbb{E}[Y_t \mid X_{t,1} = 0, X_{t,2}] &= \beta_0 + \beta_1 \cdot 0 + \beta_1 X_{t,2}, \\
\mathbb{E}[Y_t \mid X_{t,1} = 1, X_{t,2}] &= \beta_0 + \beta_1 \cdot 1 + \beta_1 X_{t,2}.
\end{aligned}$$

Now subtract the first from the second ...

$$\mathbb{E}[Y_t \mid X_{t,1} = 1, X_{t,2}] - \mathbb{E}[Y_t \mid X_{t,1} = 0, X_{t,2}] = \beta_1,$$

β_1 measures the difference between the expected value of Y_t when $X_{t,1} = 1$ and the expected value of Y_t when $X_{t,1} = 0$, keeping the other regressors constant.

Collect the observations $\{Y_1, ..., Y_T\}$, $\{X_{1,1}, ..., X_{T,1}\}$ and $\{X_{1,2}, ..., X_{T,2}\}$ into vectors

$$\begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_T \end{pmatrix}, \quad \begin{pmatrix} X_{1,1} \\ X_{2,1} \\ \vdots \\ X_{T,1} \end{pmatrix}, \quad \begin{pmatrix} X_{1,2} \\ X_{2,2} \\ \vdots \\ X_{T,2} \end{pmatrix}$$

The model $Y_t = \beta_0 + \beta_1 X_{t,1} + \beta_2 X_{t,2} + \varepsilon_t$, $t = 1, ..., T$, becomes written as

$$\underbrace{\begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_T \end{pmatrix}}_{T\times 1} = \beta_0 \underbrace{\begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix}}_{T\times 1} + \underbrace{\begin{pmatrix} X_{1,1} & X_{1,2} \\ X_{2,1} & X_{2,2} \\ \vdots & \vdots \\ X_{T,1} & X_{T,2} \end{pmatrix}}_{T\times 2} \underbrace{\begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix}}_{2\times 1} + \underbrace{\begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_T \end{pmatrix}}_{T\times 1}$$

Collect the observations $\{Y_1, ..., Y_T\}$, $\{X_{1,1}, ..., X_{T,1}\}$ and $\{X_{1,2}, ..., X_{T,2}\}$ into vectors

$$\begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_T \end{pmatrix}, \quad \begin{pmatrix} X_{1,1} \\ X_{2,1} \\ \vdots \\ X_{T,1} \end{pmatrix}, \quad \begin{pmatrix} X_{1,2} \\ X_{2,2} \\ \vdots \\ X_{T,2} \end{pmatrix}$$

The model $Y_t = \beta_0 + \beta_1 X_{t,1} + \beta_2 X_{t,2} + \varepsilon_t$, $t = 1, ..., T$, becomes written as

$$\underbrace{\begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_T \end{pmatrix}}_{T\times 1} = \underbrace{\begin{pmatrix} 1 & X_{1,1} & X_{1,2} \\ 1 & X_{2,1} & X_{2,2} \\ \vdots & \vdots & \vdots \\ 1 & X_{T,1} & X_{T,2} \end{pmatrix}}_{T\times 3} \underbrace{\begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{pmatrix}}_{3\times 1} + \underbrace{\begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_T \end{pmatrix}}_{T\times 1}$$

Collect the observations $\{Y_1, ..., Y_T\}$, $\{X_{1,1}, ..., X_{T,1}\}$ and $\{X_{1,2}, ..., X_{T,2}\}$ into vectors

$$\begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_T \end{pmatrix}, \quad \begin{pmatrix} X_{1,1} \\ X_{2,1} \\ \vdots \\ X_{T,1} \end{pmatrix}, \quad \begin{pmatrix} X_{1,2} \\ X_{2,2} \\ \vdots \\ X_{T,2} \end{pmatrix}$$

$$Y = X\beta + \varepsilon$$

Collect the observations $\{Y_1, ..., Y_T\}$, $\{X_{1,1}, ..., X_{T,1}\}$ and $\{X_{1,2}, ..., X_{T,2}\}$ into vectors

$$\begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_T \end{pmatrix}, \quad \begin{pmatrix} X_{1,1} \\ X_{2,1} \\ \vdots \\ X_{T,1} \end{pmatrix}, \quad \begin{pmatrix} X_{1,2} \\ X_{2,2} \\ \vdots \\ X_{T,2} \end{pmatrix}$$

$$\boxed{\begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_T \end{pmatrix}} \longrightarrow \boxed{Y} = X\beta + \varepsilon$$

Collect the observations $\{Y_1, \dots, Y_T\}$, $\{X_{1,1}, \dots, X_{T,1}\}$ and $\{X_{1,2}, \dots, X_{T,2}\}$ into vectors

$$\begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_T \end{pmatrix}, \quad \begin{pmatrix} X_{1,1} \\ X_{2,1} \\ \vdots \\ X_{T,1} \end{pmatrix}, \quad \begin{pmatrix} X_{1,2} \\ X_{2,2} \\ \vdots \\ X_{T,2} \end{pmatrix}$$

$$\boxed{\begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_T \end{pmatrix}}$$

$$\boxed{Y} = \boxed{X}\beta + \varepsilon$$

$$\boxed{\begin{pmatrix} 1 & X_{1,1} & X_{1,2} \\ 1 & X_{2,1} & X_{2,2} \\ \vdots & \vdots & \vdots \\ 1 & X_{T,1} & X_{T,2} \end{pmatrix}}$$

Collect the observations $\{Y_1, ..., Y_T\}$, $\{X_{1,1}, ..., X_{T,1}\}$ and $\{X_{1,2}, ..., X_{T,2}\}$ into vectors

$$\begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_T \end{pmatrix}, \quad \begin{pmatrix} X_{1,1} \\ X_{2,1} \\ \vdots \\ X_{T,1} \end{pmatrix}, \quad \begin{pmatrix} X_{1,2} \\ X_{2,2} \\ \vdots \\ X_{T,2} \end{pmatrix}$$

$$\boxed{\begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_T \end{pmatrix}}$$

$$\boxed{Y} = \boxed{X}\boxed{\beta} + \varepsilon$$

$$\boxed{\begin{pmatrix} 1 & X_{1,1} & X_{1,2} \\ 1 & X_{2,1} & X_{2,2} \\ \vdots & \vdots & \vdots \\ 1 & X_{T,1} & X_{T,2} \end{pmatrix}} \qquad \boxed{\begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{pmatrix}}$$

Collect the observations $\{Y_1, ..., Y_T\}$, $\{X_{1,1}, ..., X_{T,1}\}$ and $\{X_{1,2}, ..., X_{T,2}\}$ into vectors

$$\begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_T \end{pmatrix}, \quad \begin{pmatrix} X_{1,1} \\ X_{2,1} \\ \vdots \\ X_{T,1} \end{pmatrix}, \quad \begin{pmatrix} X_{1,2} \\ X_{2,2} \\ \vdots \\ X_{T,2} \end{pmatrix}$$

$$\boxed{\begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_T \end{pmatrix}} \qquad \boxed{Y} = \boxed{X}\,\boxed{\beta} + \boxed{\varepsilon} \qquad \boxed{\begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_T \end{pmatrix}}$$

$$\boxed{\begin{pmatrix} 1 & X_{1,1} & X_{1,2} \\ 1 & X_{2,1} & X_{2,2} \\ \vdots & \vdots & \vdots \\ 1 & X_{T,1} & X_{T,2} \end{pmatrix}} \qquad \boxed{\begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{pmatrix}}$$

Let A be symmetric $n \times n$

$$A^{\dagger} = A \Leftrightarrow A_{kj} = A_{jk}, \quad k, j = 1, ..., n.$$

Let β be a vector $n \times 1$. Consider the quadratic form for the case n=2

$$\begin{aligned} f(\beta) &= \beta^{\dagger} A \beta = \sum_k \beta_k \, (A\,\beta)_k = \sum_k \beta_k \sum_j A_{kj}\,\beta_j = \sum_{kj} A_{kj} \beta_k \, \beta_j \\ &= A_{11}\,\beta_1^2 + 2\,A_{12}\beta_1\,\beta_2 + A_{22}\beta_2^2 \Rightarrow \quad f : \mathbb{R}^2 \to \mathbb{R} \\ \frac{\partial f(\beta)}{\partial \beta_1} &= 2\,A_{11}\,\beta_1 + 2\,A_{12}\,\beta_2, \quad \frac{\partial f(\beta)}{\partial \beta_2} = 2\,A_{21}\,\beta_1 + 2\,A_{22}\,\beta_2. \end{aligned}$$

In summary

$$\begin{pmatrix} \frac{\partial f(\beta)}{\partial \beta_1} \\ \frac{\partial f(\beta)}{\partial \beta_2} \end{pmatrix} = 2 \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix} \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix}$$

In matrix notation (for any n, not only $n = 2$)

$$\frac{\partial \left(\beta^{\dagger} A \beta\right)}{\partial \beta} = 2\,A\,\beta \qquad \text{or also} \qquad \nabla_{\beta}\left(\beta^{\dagger} A \beta\right) = 2\,A\,\beta.$$

The objective function is

$$
\begin{aligned}
\mathcal{O}\left(\beta_0,\beta_1,\beta_2\right) &\doteq \sum_{t=1}^{T}\varepsilon_t^2 = \varepsilon_1^2+\cdots+\varepsilon_T^2 = \left(\varepsilon_1,\cdots,\varepsilon_T\right)\begin{pmatrix}\varepsilon_1\\ \vdots\\ \varepsilon_T\end{pmatrix}\\
&= \left(Y - X\,\beta\right)^\dagger\left(Y - X\,\beta\right) = \left(Y^\dagger - \left(X\,\beta\right)^\dagger\right)\left(Y - X\,\beta\right)\\
&= \left(Y^\dagger - \beta^\dagger\,X^\dagger\right)\left(Y - X\,\beta\right) =\\
&= Y^\dagger\,Y - Y^\dagger\,X\,\beta - \beta^\dagger\,X^\dagger\,Y + \beta^\dagger\,X^\dagger\,X\,\beta =\\
&= Y^\dagger\,Y - \underbrace{\left(\beta^\dagger X^\dagger\,Y\right)^\dagger}_{\text{It's a scalar!}} - \beta^\dagger\,X^\dagger\,Y + \beta^\dagger\,X^\dagger\,X\,\beta =\\
&= Y^\dagger\,Y - 2\,\beta^\dagger\,X^\dagger\,Y + \beta^\dagger\,X^\dagger\,X\,\beta.
\end{aligned}
$$

Considering that $X^\dagger\,X$ is always symmetric the FOC's are

$$0 = \nabla_\beta\mathcal{O}\left(\beta_0,\beta_1,\beta_2\right) = -2\,X^\dagger\,Y + 2\,X^\dagger\,X\,\beta = 0 \Rightarrow X^\dagger\,X\,\beta = X^\dagger\,Y$$

Whence, assuming that $X^\dagger\,X$ is invertible, the OLS estimator $\widehat{\beta}$ is

$$\widehat{\beta} = \left(X^\dagger\,X\right)^{-1}X^\dagger\,Y$$

I) The model $Y_t = \beta_0 + \beta_1 X_{t,1} + \beta_2 X_{t,2} + \varepsilon_t$, with $t = 1, \dots, T$, can be written as

$$\underbrace{\begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_T \end{pmatrix}}_{Y = T\times 1} = \underbrace{\begin{pmatrix} 1 & X_{1,1} & X_{1,2} \\ 1 & X_{2,1} & X_{2,2} \\ \vdots & \vdots & \\ 1 & X_{T,1} & X_{T,2} \end{pmatrix}}_{X = T\times 3} \underbrace{\begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{pmatrix}}_{\beta = 3\times 1} + \underbrace{\begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_T \end{pmatrix}}_{\varepsilon = T\times 1}.$$

II) Assuming that $X^\dagger X$ is invertible, the OLS estimator is $\widehat{\beta} = \left(X^\dagger X\right)^{-1} X^\dagger Y$

$$\widehat{\beta} = \left(\underbrace{\begin{pmatrix} 1 & \cdots & 1 \\ X_{1,1} & \cdots & X_{T,1} \\ X_{1,2} & \cdots & X_{T,2} \end{pmatrix}}_{X^\dagger} \underbrace{\begin{pmatrix} 1 & X_{1,1} & X_{1,2} \\ 1 & X_{2,1} & X_{2,2} \\ \vdots & \vdots & \vdots \\ 1 & X_{T,1} & X_{T,2} \end{pmatrix}}_{X} \right)^{-1} \begin{pmatrix} 1 & \cdots & 1 \\ X_{1,1} & \cdots & X_{T,1} \\ X_{1,2} & \cdots & X_{T,2} \end{pmatrix} \begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_T \end{pmatrix}$$

I) The model $Y_t = \beta_0 + \beta_1 X_{t,1} + \beta_2 X_{t,2} + \varepsilon_t$, with $t = 1, ..., T$, can be written as

$$\underbrace{\begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_T \end{pmatrix}}_{Y = T\times 1} = \underbrace{\begin{pmatrix} 1 & X_{1,1} & X_{2,1} \\ 1 & X_{1,2} & X_{2,2} \\ \vdots & \vdots & \\ 1 & X_{t,1} & X_{t,2} \end{pmatrix}}_{X = T\times 3} \underbrace{\begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{pmatrix}}_{\beta = 3\times 1} + \underbrace{\begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_T \end{pmatrix}}_{\varepsilon = T\times 1}.$$

II) Assuming that $X^\dagger X$ is invertible, the OLS estimator is $\widehat{\beta} = \left(X^\dagger X\right)^{-1} X^\dagger Y$

$$\widehat{\beta} = \begin{pmatrix} T & \sum_t X_{t,1} & \sum_t X_{t,2} \\ \sum_t X_{t,1} & \sum_t X_{t,1}^2 & \sum_t X_{t,1} X_{t,2} \\ \sum_t X_{t,2} & \sum_t X_{t,1} X_{t,2} & \sum_t X_{t,2}^2 \end{pmatrix}^{-1} \begin{pmatrix} 1 & \cdots & 1 \\ X_{1,1} & \cdots & X_{T,1} \\ X_{1,2} & \cdots & X_{T,2} \end{pmatrix} \begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_T \end{pmatrix}$$

t	Fund Excess Return	Market Excess Return	SMB factor
1	-0.05	0.3	1.1
2	0.03	0.023	0.2
3	0.02	-0.2	0.1
4	-0.2	-0.01	0.04

I) Y_t = Excess return of the fund.

II) $X_{t,1}$ = Excess return of the market.

III) $X_{t,2}$ = SmallMinusBig Excess return = historic excess returns of small caps over big caps in the fund.

$$X = \begin{pmatrix} 1 & 0.300 & 1.10 \\ 1 & 0.023 & 0.20 \\ 1 & -0.200 & 0.10 \\ 1 & -0.010 & 0.04 \end{pmatrix}$$

```
>> X=[1, 0.3,1.1;1,0.023,0.2;1,-0.2,0.1;1,-0.01,0.04]

X =

    1.0000    0.3000    1.1000
    1.0000    0.0230    0.2000
    1.0000   -0.2000    0.1000
    1.0000   -0.0100    0.0400
```

```
>> X'*X

ans =

    4.0000    0.1130    1.4400
    0.1130    0.1306    0.3142
    1.4400    0.3142    1.2616
```

```
>> det(X'*X)

ans =

    0.0796
```

```
>> inv(X'*X)

ans =

    0.8302    3.8935   -1.9173
    3.8935   37.3509  -13.7462
   -1.9173  -13.7462    6.4046
```

```
>> Y=[-0.05;0.03;0.02;-0.2]

Y =

   -0.0500
    0.0300
    0.0200
   -0.2000
```

```
>> beta = inv(X'*X)*X'*Y

beta =

   -0.1241
   -0.6318
    0.2554
```

$$X = \begin{pmatrix} 1 & 0.300 & 1.10 \\ 1 & 0.023 & 0.20 \\ 1 & -0.200 & 0.10 \\ 1 & -0.010 & 0.04 \end{pmatrix}$$

```
>> X=[1, 0.3,1.1;1,0.023,0.2;1,-0.2,0.1;1,-0.01,0.04]

X =

    1.0000    0.3000    1.1000
    1.0000    0.0230    0.2000
    1.0000   -0.2000    0.1000
    1.0000   -0.0100    0.0400
```

```
>> X'*X
```

```
>> inv(X'*X)
```

The estimated model is thus

$$\widehat{Y}_t = -0.1241 - 0.6316\, X_{1,t} + 0.2554\, X_{2,t}$$

```
>> Y=[-0.05;0.03;0.02;-0.2]

Y =

   -0.0500
    0.0300
    0.0200
   -0.2000
```

```
>> beta = inv(X'*X)*X'*Y

beta =

   -0.1241
   -0.6318
    0.2554
```

t	Fund Excess Return	Market Excess Return	Recession?
1	-0.1	0.3	1
2	0.3	0.023	0
3	0.2	-0.2	0
4	-0.18	-0.01	1
5	0.5	0.2	0
6	-0.3	0.01	1

I) Y_t = Excess return of the fund.

II) $X_{t,1}$ = Excess return of the market.

III) $X_{t,2}$ = A dummy variable for recessions.

Classroom exercise: Which is the discrepancy, in expected Fund Excess Return, between recessions and normal periods?

```
>> X=[1, 0.3,1;1,0.023,0;1,-0.2,0;1,-0.01,1;1,0.2,0;1,0.01,1]

X =

    1.0000    0.3000    1.0000
    1.0000    0.0230         0
    1.0000   -0.2000         0
    1.0000   -0.0100    1.0000
    1.0000    0.2000         0
    1.0000    0.0100    1.0000
```

```
>> X'*X

ans =

    6.0000    0.3230    3.0000
    0.3230    0.1707    0.3000
    3.0000    0.3000    3.0000
```

```
>> inv(X'*X)

ans =

    0.3338   -0.0545   -0.3283
   -0.0545    7.1148   -0.6569
   -0.3283   -0.6569    0.7273
```

```
Y =

   -0.0100
    0.3000
    0.2000
   -0.1800
    0.5000
   -0.3000
```

```
>> beta = inv(X'*X)*X'*Y

beta =

    0.3277
    0.7402
   -0.5650
```

Estimated model:

$$\widehat{Y}_t = 0.3277 + 0.7402\, X_{1,t} - 0.5650\, X_{2,t}$$

```
>> Yhat = 0.3277+0.7402*X(:,2)-0.5650*X(:,3)

Yhat =

   -0.0152
    0.3447
    0.1797
   -0.2447
    0.4757
   -0.2299
```

Estimated model:

$$\widehat{Y}_t = 0.3277 + 0.7402\, X_{1,t} - 0.5650\, X_{2,t}$$

	$\widehat{Y}_1$	$\widehat{Y}_2$	$\widehat{Y}_3$	$\widehat{Y}_4$	$\widehat{Y}_5$	$\widehat{Y}_6$	Mean
$X_{2,t} = 1$							
$X_{2,t} = 0$							

Estimated model:

$$\widehat{Y}_t = 0.3277 + 0.7402\, X_{1,t} - 0.5650\, X_{2,t}$$

	$\widehat{Y}_1$	$\widehat{Y}_2$	$\widehat{Y}_3$	$\widehat{Y}_4$	$\widehat{Y}_5$	$\widehat{Y}_6$	Mean
$X_{2,t} = 1$	-0.0152	-0.2203	-0.38	-0.24	-0.09	-0.22	
$X_{2,t} = 0$							

```
>> Yhat_dummy1 = 0.3277+0.7402*X(:,2)-0.5650

Yhat_dummy1 =

   -0.0152
   -0.2203
   -0.3853
   -0.2447
   -0.0893
   -0.2299
```

Estimated model:

$$\widehat{Y}_t = 0.3277 + 0.7402\,X_{1,t} - 0.5650\,X_{2,t}$$

	$\widehat{Y}_1$	$\widehat{Y}_2$	$\widehat{Y}_3$	$\widehat{Y}_4$	$\widehat{Y}_5$	$\widehat{Y}_6$	Mean
$X_{2,t} = 1$	-0.0152	-0.2203	-0.38	-0.24	-0.09	-0.22	-0.1975
$X_{2,t} = 0$							

```
>> mean(Yhat_dummy1)

ans =

   -0.1975
```

Estimated model:

$$\widehat{Y}_t = 0.3277 + 0.7402\, X_{1,t} - 0.5650\, X_{2,t}$$

	$\widehat{Y}_1$	$\widehat{Y}_2$	$\widehat{Y}_3$	$\widehat{Y}_4$	$\widehat{Y}_5$	$\widehat{Y}_6$	Mean
$X_{2,t} = 1$	-0.0152	-0.2203	-0.38	-0.24	-0.09	-0.22	-0.1975
$X_{2,t} = 0$	0.5498	0.3347	0.1797	0.3203	0.4757	0.3351	

```
>> Yhat_dummy0 = 0.3277+0.7402*X(:,2)

Yhat_dummy0 =

    0.5498
    0.3447
    0.1797
    0.3203
    0.4757
    0.3351
```

Estimated model:

$$\widehat{Y}_t = 0.3277 + 0.7402\, X_{1,t} - 0.5650\, X_{2,t}$$

	$\widehat{Y}_1$	$\widehat{Y}_2$	$\widehat{Y}_3$	$\widehat{Y}_4$	$\widehat{Y}_5$	$\widehat{Y}_6$	Mean
$X_{2,t} = 1$	-0.0152	-0.2203	-0.38	-0.24	-0.09	-0.22	-0.1975
$X_{2,t} = 0$	0.5498	0.3347	0.1797	0.3203	0.4757	0.3351	0.3675

```
>> mean(Yhat_dummy0)

ans =

    0.3675
```

Estimated model:

$$\widehat{Y}_t = 0.3277 + 0.7402\, X_{1,t} - 0.5650\, X_{2,t}$$

	$\widehat{Y}_1$	$\widehat{Y}_2$	$\widehat{Y}_3$	$\widehat{Y}_4$	$\widehat{Y}_5$	$\widehat{Y}_6$	Mean
$X_{2,t} = 1$	-0.0152	-0.2203	-0.38	-0.24	-0.09	-0.22	-0.1975
$X_{2,t} = 0$	0.5498	0.3347	0.1797	0.3203	0.4757	0.3351	0.3675

$$\mathbb{E}\left[\widehat{Y}_t \,\middle|\, X_{2,t} = 1\right] - \mathbb{E}\left[\widehat{Y}_t \,\middle|\, X_{2,t} = 0\right] =$$

Estimated model:

$$\widehat{Y}_t = 0.3277 + 0.7402\, X_{1,t} - 0.5650\, X_{2,t}$$

	$\widehat{Y_1}$	$\widehat{Y_2}$	$\widehat{Y_3}$	$\widehat{Y_4}$	$\widehat{Y_5}$	$\widehat{Y_6}$	Mean
$X_{2,t} = 1$	-0.0152	-0.2203	-0.38	-0.24	-0.09	-0.22	-0.1975
$X_{2,t} = 0$	0.5498	0.3347	0.1797	0.3203	0.4757	0.3351	0.3675

$$\mathbb{E}\left[\widehat{Y}_t \,\middle|\, X_{2,t} = 1\right] - \mathbb{E}\left[\widehat{Y}_t \,\middle|\, X_{2,t} = 0\right] = -0.1975 - 0.3675 =$$

Estimated model:

$$\widehat{Y}_t = 0.3277 + 0.7402\, X_{1,t} - 0.5650\, X_{2,t}$$

	$\widehat{Y}_1$	$\widehat{Y}_2$	$\widehat{Y}_3$	$\widehat{Y}_4$	$\widehat{Y}_5$	$\widehat{Y}_6$	Mean
$X_{2,t} = 1$	-0.0152	-0.2203	-0.38	-0.24	-0.09	-0.22	-0.1975
$X_{2,t} = 0$	0.5498	0.3347	0.1797	0.3203	0.4757	0.3351	0.3675

$$\mathbb{E}\left[\widehat{Y}_t \,\middle|\, X_{2,t} = 1\right] - \mathbb{E}\left[\widehat{Y}_t \,\middle|\, X_{2,t} = 0\right] = -0.1975 - 0.3675 =$$

Estimated model:

$$\widehat{Y}_t = 0.3277 + 0.7402\, X_{1,t} - 0.5650\, X_{2,t}$$

	$\widehat{Y}_1$	$\widehat{Y}_2$	$\widehat{Y}_3$	$\widehat{Y}_4$	$\widehat{Y}_5$	$\widehat{Y}_6$	Mean
$X_{2,t} = 1$	-0.0152	-0.2203	-0.38	-0.24	-0.09	-0.22	-0.1975
$X_{2,t} = 0$	0.5498	0.3347	0.1797	0.3203	0.4757	0.3351	0.3675

$$\mathbb{E}\left[\widehat{Y}_t \,\middle|\, X_{2,t} = 1\right] - \mathbb{E}\left[\widehat{Y}_t \,\middle|\, X_{2,t} = 0\right] = -0.1975 - 0.3675 = -0.5650$$

Estimated model:

$$\widehat{Y}_t = 0.3277 + 0.7402\, X_{1,t} \underline{-\ 0.5650}\, X_{2,t}$$

	$\widehat{Y}_1$	$\widehat{Y}_2$	$\widehat{Y}_3$	$\widehat{Y}_4$	$\widehat{Y}_5$	$\widehat{Y}_6$	Mean
$X_{2,t} = 1$	-0.0152	-0.2203	-0.38	-0.24	-0.09	-0.22	-0.1975
$X_{2,t} = 0$	0.5498	0.3347	0.1797	0.3203	0.4757	0.3351	0.3675

$$\mathbb{E}\left[\widehat{Y}_t \,\middle|\, X_{2,t} = 1\right] - \mathbb{E}\left[\widehat{Y}_t \,\middle|\, X_{2,t} = 0\right] = -0.1975 - 0.3675 = \underline{-0.5650 = \widehat{\beta}_2}$$

I) The model $Y_t = \beta_0 + \beta_1 X_{t,1} + \beta_2 X_{t,2} + \cdots + \beta_K X_{t,K} + \varepsilon_t$, with $t = 1, \ldots, T$, can be written as

$$\underbrace{\begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_T \end{pmatrix}}_{Y = T \times 1} = \underbrace{\begin{pmatrix} 1 & X_{1,1} & \cdots & X_{1,K} \\ 1 & X_{2,1} & \cdots & X_{2,K} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & X_{T,1} & \cdots & X_{T,K} \end{pmatrix}}_{X = T \times (K+1)} \underbrace{\begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_K \end{pmatrix}}_{\beta = (K+1) \times 1} + \underbrace{\begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_T \end{pmatrix}}_{\varepsilon = T \times 1}.$$

II) Assuming that $X^\dagger X$ is invertible, the OLS estimator is $\widehat{\beta} = \left(X^\dagger X\right)^{-1} X^\dagger Y$

$$\widehat{\beta} = \left(\underbrace{\begin{pmatrix} 1 & \cdots & 1 \\ X_{1,1} & \cdots & X_{T,1} \\ \vdots & \ddots & \vdots \\ X_{1,K} & \cdots & X_{T,K} \end{pmatrix}}_{X^\dagger} \underbrace{\begin{pmatrix} 1 & X_{1,1} & \cdots & X_{1,K} \\ 1 & X_{2,1} & \cdots & X_{2,K} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & X_{T,1} & \cdots & X_{T,K} \end{pmatrix}}_{X} \right)^{-1} \begin{pmatrix} 1 & \cdots & 1 \\ X_{1,1} & \cdots & X_{T,1} \\ \vdots & \ddots & \vdots \\ X_{1,K} & \cdots & X_{T,K} \end{pmatrix} \begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_T \end{pmatrix}$$

I) The model $Y_t = \beta_0 + \beta_1 X_{t,1} + \beta_2 X_{t,2} + \cdots + \beta_K X_{t,K} + \varepsilon_t$, with $t = 1, ..., T$, can be written as

$$\underbrace{\begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_T \end{pmatrix}}_{Y=T\times 1} = \underbrace{\begin{pmatrix} 1 & X_{1,1} & \cdots & X_{1,K} \\ 1 & X_{2,1} & \cdots & X_{2,K} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & X_{T,1} & \cdots & X_{T,K} \end{pmatrix}}_{X=T\times(K+1)} \underbrace{\begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_K \end{pmatrix}}_{\beta=(K+1)\times 1} + \underbrace{\begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_T \end{pmatrix}}_{\varepsilon=T\times 1}.$$

II) Assuming that $X^\dagger X$ is invertible, the OLS estimator is $\widehat{\beta} = \left(X^\dagger X\right)^{-1} X^\dagger Y$

$$\widehat{\beta} = \begin{pmatrix} T & \sum_t X_{t,1} & \sum_t X_{t,2} & \cdots & \sum_t X_{t,K} \\ \sum_t X_{t,1} & \sum_t X_{t,1}^2 & \sum_t X_{t,1} X_{t,2} & \cdots & \sum_t X_{t,1} X_{t,K} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \sum_t X_{t,K} & \sum_t X_{t,K} X_{t,1} & \sum_t X_{t,K} X_{t,2} & \cdots & \sum_t X_{t,K}^2 \end{pmatrix}^{-1} \begin{pmatrix} 1 & \cdots & 1 \\ X_{1,1} & \cdots & X_{T,1} \\ \vdots & \ddots & \vdots \\ X_{1,K} & \cdots & X_{T,K} \end{pmatrix} \begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_T \end{pmatrix}$$

$$Y_t = \beta_0 + \beta_1 X_{t,1} + \cdots + \beta_K X_{t,K} + \varepsilon_t$$

Bias of the estimator (not the relative bias!):

$$\mathbb{E}\left[\widehat{\beta} - \beta\right] = \mathbb{E}\left[\begin{pmatrix} \widehat{\beta}_0 \\ \vdots \\ \widehat{\beta}_K \end{pmatrix} - \begin{pmatrix} \beta_0 \\ \vdots \\ \beta_K \end{pmatrix}\right] = \begin{pmatrix} \mathbb{E}\left[\widehat{\beta}_0 - \beta_0\right] \\ \vdots \\ \mathbb{E}\left[\widehat{\beta}_K - \beta_K\right] \end{pmatrix}.$$

The X-conditional variance-covariance matrix of the estimator:

$$\Sigma_X^2 \doteq \mathbb{E}\left[\left(\widehat{\beta} - \beta\right)\left(\widehat{\beta} - \beta\right)^{\dagger} \middle| X\right] = \mathbb{E}\left[\begin{pmatrix} \widehat{\beta}_0 - \beta_0 \\ \vdots \\ \widehat{\beta}_K - \beta_K \end{pmatrix}\left(\widehat{\beta}_0 - \beta_0, \cdots, \widehat{\beta}_K - \beta_K\right) \middle| X\right]$$

To clarify, consider the case $K = 2$.

$$\Sigma_X^2 = \begin{pmatrix} \mathbb{E}\left[\left(\widehat{\beta}_0 - \beta_0\right)^2\right] & \mathbb{E}\left[\left(\widehat{\beta}_0 - \beta_0\right)\left(\widehat{\beta}_1 - \beta_1\right)\right] & \mathbb{E}\left[\left(\widehat{\beta}_0 - \beta_0\right)\left(\widehat{\beta}_2 - \beta_2\right)\right] \\ \mathbb{E}\left[\left(\widehat{\beta}_0 - \beta_0\right)\left(\widehat{\beta}_1 - \beta_1\right)\right] & \mathbb{E}\left[\left(\widehat{\beta}_1 - \beta_1\right)^2\right] & \mathbb{E}\left[\left(\widehat{\beta}_1 - \beta_1\right)\left(\widehat{\beta}_2 - \beta_2\right)\right] \\ \mathbb{E}\left[\left(\widehat{\beta}_0 - \beta_0\right)\left(\widehat{\beta}_2 - \beta_2\right)\right] & \mathbb{E}\left[\left(\widehat{\beta}_1 - \beta_1\right)\left(\widehat{\beta}_2 - \beta_2\right)\right] & \mathbb{E}\left[\left(\widehat{\beta}_2 - \beta_2\right)^2\right] \end{pmatrix}$$

Theorem

Given the regression model

$$Y_t = \beta_0 + \beta_1 X_{t,1} + \beta_2 X_{t,2} + \cdots + \beta_K X_{t,K} + \varepsilon_t,$$

assume that

- *I)* $X^\dagger X$ *is invertible.*
- *II)* $\mathbb{E}\left[\varepsilon_t \mid X\right] = 0$ *(which implies* $\mathbb{E}\left[\varepsilon\right] = 0$*).*

Therefore

$$\mathbb{E}\left[\widehat{\beta}\right] = \beta.$$

$$
\begin{aligned}
\widehat{\beta} &= \left(X^{\dagger} X\right)^{-1} X^{\dagger} Y \\
&= \left(X^{\dagger} X\right)^{-1} X^{\dagger} \left(X \beta + \varepsilon\right) \\
&= \left(X^{\dagger} X\right)^{-1} X^{\dagger} X \beta + \left(X^{\dagger} X\right)^{-1} X^{\dagger} \varepsilon \\
&= \beta + \left(X^{\dagger} X\right)^{-1} X^{\dagger} \varepsilon.
\end{aligned}
$$

Whence

$$
\begin{aligned}
\mathbb{E}\left[\widehat{\beta}\right] &= \beta + \mathbb{E}\left[\left(X^{\dagger} X\right)^{-1} X^{\dagger} \varepsilon\right] \\
&= \beta + \mathbb{E}\left[\mathbb{E}\left[\left(X^{\dagger} X\right)^{-1} X^{\dagger} \varepsilon \,\middle|\, X\right]\right] \\
&= \beta + \mathbb{E}\left[\left(X^{\dagger} X\right)^{-1} X^{\dagger} \mathbb{E}\left[\varepsilon \mid X\right]\right] = \beta.
\end{aligned}
$$

Remark. As a by-product we have proved that:

$$
\widehat{\beta} - \beta = \left(X^{\dagger} X\right)^{-1} X^{\dagger} \varepsilon,
$$

take note of this expression! It will be useful later ...

Theorem

Given the regression model

$$Y_t = \beta_0 + \beta_1 X_{t,1} + \beta_2 X_{t,2} + \cdots + \beta_K X_{t,K} + \varepsilon_t,$$

assume that

I) $X^\dagger X$ *is invertible.*

II) $\mathbb{E}\left[\varepsilon_t \mid X\right] = 0.$

III) $\mathbb{E}\left[\varepsilon_t^2 \mid X\right] = \sigma_\varepsilon^2$, $\mathbb{E}\left[\varepsilon_t \, \varepsilon_s \mid X\right] = \mathbb{E}\left[\varepsilon_t \mid X\right] \mathbb{E}\left[\varepsilon_s \mid X\right] = 0$ *for all* $t \neq s$.

Therefore the X*-conditional variance covariance matrix of the estimator* $\widehat{\beta}$ *is*

$$\Sigma_X^2 = \mathbb{E}\left[\left(\widehat{\beta} - \beta\right) \left(\widehat{\beta} - \beta\right)^\dagger \,\middle|\, X\right] = \sigma_\varepsilon^2 \left(X^\dagger X\right)^{-1}.$$

First of all note that

$$\mathbb{E}\left[\varepsilon\,\varepsilon^{\dagger}\right] = \mathbb{E}\left[\begin{pmatrix}\varepsilon_1\\ \vdots \\ \varepsilon_T\end{pmatrix}(\varepsilon_1,\cdots,\varepsilon_T)\right]$$

$$= \begin{pmatrix}\mathbb{E}\left[\varepsilon_1^2\right] & \mathbb{E}\left[\varepsilon_1\,\varepsilon_2\right] & \cdots & \mathbb{E}\left[\varepsilon_1\,\varepsilon_T\right]\\ \mathbb{E}\left[\varepsilon_1\,\varepsilon_2\right] & \mathbb{E}\left[\varepsilon_2^2\right] & \cdots & \mathbb{E}\left[\varepsilon_2\,\varepsilon_T\right]\\ \vdots & \vdots & \ddots & \vdots\\ \mathbb{E}\left[\varepsilon_1\,\varepsilon_T\right] & \mathbb{E}\left[\varepsilon_2\,\varepsilon_T\right] & \cdots & \mathbb{E}\left[\varepsilon_T^2\right]\end{pmatrix}$$

$$= \begin{pmatrix}\sigma_\varepsilon^2 & 0 & \cdots & 0\\ 0 & \sigma_\varepsilon^2 & \cdots & 0\\ \vdots & \vdots & \ddots & \vdots\\ 0 & 0 & \cdots & \sigma_\varepsilon^2\end{pmatrix} = \sigma_\varepsilon^2 \begin{pmatrix}1 & 0 & \cdots & 0\\ 0 & 1 & \cdots & 0\\ \vdots & \vdots & \ddots & \vdots\\ 0 & 0 & \cdots & 1\end{pmatrix} = \sigma_\varepsilon^2\,\mathbb{1}_T.$$

$$
\begin{aligned}
\Sigma_X^2 &= \mathbb{E}\left[\left(\widehat{\beta}-\beta\right)\left(\widehat{\beta}-\beta\right)^{\dagger} \,\middle|\, X\right] = \\
&= \mathbb{E}\left[\left(X^{\dagger}X\right)^{-1} X^{\dagger}\varepsilon \left(\left(X^{\dagger}X\right)^{-1} X^{\dagger}\varepsilon\right)^{\dagger} \,\middle|\, X\right] = \\
\left(\left(A\,B\right)^{\dagger} = B^{\dagger}A^{\dagger}\right) \quad &= \mathbb{E}\left[\left(X^{\dagger}X\right)^{-1} X^{\dagger}\varepsilon\,\varepsilon^{\dagger} X \left(\left(X^{\dagger}X\right)^{-1}\right)^{\dagger} \,\middle|\, X\right] \\
\left(A^{\dagger} = A \Rightarrow \left(A^{-1}\right)^{\dagger} = A^{-1}\right) \quad &= \mathbb{E}\left[\left(X^{\dagger}X\right)^{-1} X^{\dagger}\varepsilon\,\varepsilon^{\dagger} X \left(X^{\dagger}X\right)^{-1} \,\middle|\, X\right] \\
&= \left(X^{\dagger}X\right)^{-1} X^{\dagger}\,\mathbb{E}\left[\varepsilon\,\varepsilon^{\dagger} \,\middle|\, X\right] X \left(X^{\dagger}X\right)^{-1} \\
&= \left(X^{\dagger}X\right)^{-1} X^{\dagger}\,\mathbb{E}\left[\varepsilon\,\varepsilon^{\dagger}\right] X \left(X^{\dagger}X\right)^{-1} \\
&= \sigma_{\varepsilon}^2 \left(X^{\dagger}X\right)^{-1} X^{\dagger}\,\mathbb{1}_T\, X \left(X^{\dagger}X\right)^{-1} \\
&= \sigma_{\varepsilon}^2 \left(X^{\dagger}X\right)^{-1} X^{\dagger} X \left(X^{\dagger}X\right)^{-1} = \sigma_{\varepsilon}^2 \left(X^{\dagger}X\right)^{-1}.
\end{aligned}
$$