# Advanced Econometrics: Finite Sample Theory

José Jaén Delgado

**Introduction** - In this paper solutions to most exercises from Chapter 1 of Fumio Hayashi's *Econometrics* are provided

# 1 Questions for Review

## 1.1 The Classical Linear Regression Model

**1.** Let us estimate a Linear Regression Model with OLS. Specifically, let it take the semi-log form:

$$\log(wage_i) = \beta_1 + \beta_2 S_i + \beta_3 tenure_i + \beta_4 exp_i + \varepsilon_i$$

We now show that changes in measurement units make a difference:

$$\begin{aligned}
\log(wage_i * 100) &= \beta_1 + \beta_2 S_i + \beta_3 tenure_i + \beta_4 exp_i + \varepsilon_i \\
\log(wage_i) + \log(100) &= \beta_1 + \beta_2 S_i + \beta_3 tenure_i + \beta_4 exp_i + \varepsilon_i \\
\log(wage_i) &= (\beta_1 - 2) + \beta_2 S_i + \beta_3 tenure_i + \beta_4 exp_i + \varepsilon_i
\end{aligned}$$

**2.** Let us prove $\mathrm{E}[\varepsilon_i \varepsilon_j | \mathbf{X}] = \mathrm{E}[\varepsilon_i | \mathbf{x}_i]\mathrm{E}[\varepsilon_j | \mathbf{x}_j]$ for random sampling when $i \neq j$

$$\begin{aligned}
\mathrm{E}[\varepsilon_i \varepsilon_j | \mathbf{X}] &= \mathrm{E}[\mathrm{E}[\varepsilon_i \varepsilon_j | \mathbf{X}, \varepsilon_j] | \mathbf{X}] \quad \text{(by LIE)} \\
&= \mathrm{E}[\varepsilon_j \mathrm{E}[\varepsilon_i | \mathbf{X}, \varepsilon_j] | \mathbf{X}] \quad \text{(by linearity of conditional expectations)} \\
&= \mathrm{E}[\varepsilon_j | \mathbf{x}_j]\mathrm{E}[\mathrm{E}[\varepsilon_i | \mathbf{X}, \varepsilon_j] | \mathbf{X}] \quad \text{(by random sampling properties)} \\
&= \mathrm{E}[\varepsilon_j | \mathbf{x}_j]\mathrm{E}[\varepsilon_i | \mathbf{x}_i]
\end{aligned}$$

Where we leveraged that $\mathrm{E}[\mathrm{E}[Y|X,Z]|X] = \mathrm{E}[Y|X]$ and $\mathrm{E}[\varepsilon_i|\mathbf{X}] = \mathrm{E}[\varepsilon_i|\mathbf{x}_i]$

**3.** Let us show that A.1 (Linearity) and A.2 (Strict Exogeneity) imply:

$$\mathrm{E}[y_i|\mathbf{X}] = \mathbf{x}_i'\boldsymbol{\beta}$$

We start by the general expression of any Regression Model:

$$\begin{aligned}
y_i &= \mathrm{E}[y_i|\mathbf{X}] + \varepsilon_i \\
y_i &= \mathbf{x}_i'\boldsymbol{\beta} + \varepsilon_i \quad \big(\text{by A.1: } \mathrm{E}[\mathbf{Y}|\mathbf{X}] = \mathbf{X}\boldsymbol{\beta}\big) \\
\mathrm{E}[y_i|\mathbf{X}] &= \mathrm{E}[\mathbf{x}_i'\boldsymbol{\beta} + \varepsilon_i|\mathbf{X}] \quad \big(\text{taking conditional expectations}\big) \\
\mathrm{E}[y_i|\mathbf{X}] &= \mathrm{E}[\mathbf{x}_i'\boldsymbol{\beta}|\mathbf{X}] + \mathrm{E}[\varepsilon_i|\mathbf{X}] \\
\mathrm{E}[y_i|\mathbf{X}] &= \mathbf{x}_i'\boldsymbol{\beta} \quad \big(\text{by A.2: } \mathrm{E}[\boldsymbol{\varepsilon}|\mathbf{X}] = 0\big)
\end{aligned}$$

Conversely, assuming $\mathrm{E}[y_i|\mathbf{X}] = \mathbf{x}_i'\boldsymbol{\beta}$:

$$\begin{aligned}
y_i &= \mathrm{E}[y_i|\mathbf{X}] + \varepsilon_i \\
y_i &= \mathbf{x}_i'\boldsymbol{\beta} + \varepsilon_i \quad \big(\text{A.1 is now proved}\big) \\
y_i - \mathbf{x}_i'\boldsymbol{\beta} &= \varepsilon_i \\
\mathrm{E}[y_i - \mathbf{x}_i'\boldsymbol{\beta}|\mathbf{X}] &= \mathrm{E}[\varepsilon_i|\mathbf{X}] \\
\mathbf{x}_i'\boldsymbol{\beta} - \mathbf{x}_i'\boldsymbol{\beta} &= \mathrm{E}[\varepsilon_i|\mathbf{X}] \\
\mathrm{E}[\varepsilon_i|\mathbf{X}] &= 0 \quad \big(\text{A.2 is now proved}\big)
\end{aligned}$$

**4.** Considering a random sample on Consumption and Disposable Income such that $(CON_i, YD_i) \overset{i.i.d}{\sim} N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, we proceed to prove A.1 (Linearity), A.2 (Strict Exogeneity), A.3 (No Multicollinearity) and A.4 (Spherical Error Variance):

A.3 is pretty straightforward: $\Pr(\text{rank}(\mathbf{X}'\mathbf{X}) = K) = 1$ since we are dealing with a random sample. Note that $\mathbf{X}$ is a $n \,\text{x}\, K$ matrix, so we claim it is of full-column rank.

Since $CON_i$ and $YD_i$ are jointly normally distributed, their Conditional Expectation Function (CEF) is linear and the conditional variance of the target variable is equal to the unconditional one.

$$\mathrm{E}[CON_i|\mathbf{YD}] = \beta_1 + \beta_2\mathbf{YD}$$

$$\mathrm{V}[CON_i|\mathbf{YD}] = \mathrm{V}[CON_i]$$

A.1 is satisfied as CEF is linear due to properties of the normal distribution. Then:

$$\begin{aligned}
\mathrm{E}[CON_i|\mathbf{YD}] &= \mathrm{E}[\beta_1 + \beta_2\mathbf{YD} + \varepsilon_i|\mathbf{YD}] \\
\mathrm{E}[CON_i|YD_i] &= \mathrm{E}[\beta_1 + \beta_2 YD_i + \varepsilon_i|YD_i] \quad \text{(by random sampling)} \\
&= \beta_1 + \beta_2 YD_i + \mathrm{E}[\varepsilon_i|YD_i]
\end{aligned}$$

Note that $\mathrm{E}[\varepsilon_i|YD_i] = 0$ since $\mathrm{E}[CON_i|\mathbf{YD}] = \beta_1 + \beta_2\mathbf{YD}$. A.2 is thus proved.

For A.4 let us derive the expression of $\mathrm{E}[\varepsilon_i 2|\mathbf{YD}]$ where $\mathbf{X}$ represents the matrix of regressors $\mathbf{YD}$, and $y_i$ the dependent variable $CON_i$:

$$\begin{aligned}
\mathrm{E}[(y_i - \mathbf{x}_i'\boldsymbol{\beta})^2|\mathbf{X}] &= \mathrm{E}[y_i^2 + (\mathbf{x}_i'\boldsymbol{\beta})'(\mathbf{x}_i'\boldsymbol{\beta}) - 2y_i\mathbf{x}_i'\boldsymbol{\beta}|\mathbf{X}] \\
&= \mathrm{E}[y_i^2|\mathbf{X}] + (\mathbf{x}_i'\boldsymbol{\beta})'(\mathbf{x}_i'\boldsymbol{\beta}) - 2\mathbf{x}_i'\boldsymbol{\beta}\mathrm{E}[y_i|\mathbf{X}] \\
&= \underbrace{\mathrm{V}[y_i|\mathbf{X}] + \mathrm{E}[y_i|\mathbf{X}]^2}_{\mathrm{E}[y_i|\mathbf{X}]^2} + (\mathbf{x}_i'\boldsymbol{\beta})'(\mathbf{x}_i'\boldsymbol{\beta}) - 2\mathbf{x}_i'\boldsymbol{\beta}\mathrm{E}[y_i|\mathbf{X}] \\
&= \mathrm{V}[y_i|\mathbf{X}] + \mathrm{E}[y_i|\mathbf{X}]^2 + (\mathbf{x}_i'\boldsymbol{\beta})'(\mathbf{x}_i'\boldsymbol{\beta}) - 2(\mathbf{x}_i'\boldsymbol{\beta})'\underbrace{(\mathbf{x}_i'\boldsymbol{\beta})}_{\mathrm{E}[y_i|\mathbf{X}]} \\
&= \mathrm{V}[y_i|\mathbf{X}] \quad (\text{since } \mathrm{E}[y_i|\mathbf{X}]^2 = (\mathbf{x}_i'\boldsymbol{\beta})'(\mathbf{x}_i'\boldsymbol{\beta})) \\
&= \mathrm{V}[y_i] \quad \text{(by normal distribution properties)} \\
&= \sigma^2
\end{aligned}$$

A.4 is satisfied as $\mathrm{E}[\varepsilon_i^2|\mathbf{X}] = \sigma^2 \ \forall i$

**5.** Let us show that the full-column rank condition, i.e: $\Pr(\mathrm{rank}(X) = K) = 1$ in the Simple Regression Model ($K = 2$) implies that $x_{i2} \neq x_{j2}$ for some pairs $i \neq j$.

$$\mathbf{X} = \begin{bmatrix} \mathbf{X}_1' \\ \mathbf{X}_2' \end{bmatrix} = \begin{bmatrix} x_{11} & x_{12} \\ x_{21} & x_{22} \\ \vdots & \vdots \\ x_{n1} & x_{n2} \end{bmatrix} = \begin{bmatrix} 1 & x_{12} \\ 1 & x_{22} \\ \vdots & \vdots \\ 1 & x_{n2} \end{bmatrix}$$

Note that $x_{i1} = 1 \ \forall i$ since a SRM is composed by two regressors: a constant and a random variable.

The rank of any matrix determines how many columns are linearly independent of each other. Thus, if we want the full-column rank condition to hold, there should be $K$ linearly independent columns. Since $K = 2$, the only way to assure $\text{rank}(\mathbf{X}) = 2$ is that each column is different from each other, i.e: $x_{i2} \neq x_{j2}$

**6.** We prove that A.2: $\mathrm{E}[\boldsymbol{\varepsilon}|\mathbf{X}] = 0$ and A.4: $\mathrm{E}[\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}'|\mathbf{X}] = \sigma^2\mathbf{I}$ and $\mathrm{E}[\varepsilon_i\varepsilon_j|\mathbf{X}] = 0$ imply:

$$\mathrm{V}[\varepsilon_i] = \sigma^2 \ \ \forall i$$
$$\text{cov}(\varepsilon_i, \varepsilon_j) = 0 \quad \forall i \neq j$$

$$\begin{aligned}
\mathrm{E}[\varepsilon_i] &= \mathrm{E}[\mathrm{E}[\varepsilon_i|\mathbf{X}]] \quad \big(\text{by LIE}\big) \\
&= 0 \quad \big(\text{by A.2}\big) \\
\mathrm{E}[x_{jk}\varepsilon] &= \mathrm{E}[\mathrm{E}[x_{jk}\varepsilon_i|x_{jk}]] \quad \big(\text{by LIE}\big) \\
&= \mathrm{E}[x_{jk}\mathrm{E}[\varepsilon_i|x_{jk}]] \quad \big(\text{by linearity of conditional expectations}\big) \\
&= 0 \quad \big(\text{by A.2}\big)
\end{aligned}$$

Thus:

$$\begin{aligned}
\mathrm{V}[\varepsilon_i] &= \mathrm{E}[\varepsilon_i^2] - \mathrm{E}[\varepsilon_i]^2 \\
&= \mathrm{E}[\varepsilon_i^2] \quad \big(\text{following our results}\big) \\
&= \mathrm{E}[\mathrm{E}[\varepsilon_i^2|\mathbf{X}]] \\
&= \sigma^2 \quad \big(\text{by A.4}\big) \\
\text{cov}(\varepsilon_i, \varepsilon_j) &= \mathrm{E}[\varepsilon_i\varepsilon_j] - \mathrm{E}[\varepsilon_i]\mathrm{E}[\varepsilon_j] \\
&= \mathrm{E}[\varepsilon_i\varepsilon_j] \quad \big(\text{following our results}\big) \\
&= \mathrm{E}[\mathrm{E}[\varepsilon_i\varepsilon_j|\mathbf{X}]] \\
&= 0 \quad \big(\text{by A.4 if } i \neq j\big)
\end{aligned}$$

## 1.2 The Algebra of Least Squares

**1.** We prove $\mathbf{X}'\mathbf{X}$ if Positive Definite if $\mathbf{X}$ is of full column rank.

Let $\dim(\mathbf{X}) = n \,\mathrm{x}\, K$ and $\mathbf{Xc} = \mathbf{z}$, we need to show that $\mathbf{c}'\mathbf{X}'\mathbf{X}\mathbf{c} > 0 \;\; \forall \mathbf{c} \neq 0$

$$\begin{aligned} \mathbf{c}'\mathbf{X}'\mathbf{X}\mathbf{c} &= (\mathbf{Xc})'\mathbf{Xc} \\ &= \mathbf{z}'\mathbf{z} \\ &= \sum_{i=1}^{n} z_i^2 \end{aligned}$$

By construction: $\sum_{i=1}^{n} z_i^2 \geq 0$ but note that if $\Pr(\mathrm{rank}(\mathbf{X}) = K) = 1$ then $\sum_{i=1}^{n} z_i^2 > 0$

Thus, if $\mathbf{X}$ is of full-column rank, $\mathbf{X}'\mathbf{X}$ is Positive Definite.

**2.** We verify the following equalities:

$$\mathbf{X}'\mathbf{X}/n = \frac{1}{n}\sum_{i}^{n} \mathbf{x}_i\mathbf{x}_i' \qquad \mathbf{X}'\mathbf{y}/n = \frac{1}{n}\sum_{i=1}^{n} \mathbf{x}_i y_i$$

$$\begin{aligned} \mathbf{X}'\mathbf{X}/n &= \frac{1}{n}\begin{bmatrix} x_{11} & x_{21} & \dots & x_{n1} \\ x_{12} & x_{22} & \dots & x_{n2} \\ \vdots & \vdots & \ddots & \vdots \\ x_{1k} & x_{2k} & \dots & x_{nk} \end{bmatrix} \begin{bmatrix} x_{11} & x_{12} & \dots & x_{1k} \\ x_{21} & x_{22} & \dots & x_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \dots & x_{nk} \end{bmatrix} \\ &= \frac{1}{n}\begin{bmatrix} \sum_{i=1}^{n} x_{i1}^2 & \sum_{i=1}^{n} x_{i1}x_{i2} & \dots & \sum_{i=1}^{n} x_{i1}x_{ik} \\ \sum_{i=1}^{n} x_{i2}x_{i1} & \sum_{i=1}^{n} x_{i2}^2 & \dots & \sum_{i=1}^{n} x_{i2}x_{ik} \\ \vdots & \vdots & \ddots & \vdots \\ \sum_{i=1}^{n} x_{ik}x_{i1} & \sum_{i=1}^{n} x_{ik}x_{i2} & \dots & \sum_{i=1}^{n} x_{ik}^2 \end{bmatrix} \end{aligned}$$

Where we leveraged that $x_i' = x_i$ for scalars. Also, note that:

$$\mathbf{x}_i = \begin{bmatrix} x_{i1} \\ x_{i2} \\ \vdots \\ x_{ik} \end{bmatrix}$$

Thus:

$$\frac{1}{n}\sum_{i=1}^{n}\mathbf{x}_i\mathbf{x}_i' = \frac{1}{n}\begin{bmatrix} \sum_{i=1}^{n} x_{i1}^2 & \sum_{i=1}^{n} x_{i1}x_{i2} & \dots & \sum_{i=1}^{n} x_{i1}x_{ik} \\ \sum_{i=1}^{n} x_{i2}x_{i1} & \sum_{i=1}^{n} x_{i2}^2 & \dots & \sum_{i=1}^{n} x_{i2}x_{ik} \\ \vdots & \vdots & \ddots & \vdots \\ \sum_{i=1}^{n} x_{ik}x_{i1} & \sum_{i=1}^{n} x_{ik}x_{i2} & \dots & \sum_{i=1}^{n} x_{ik}^2 \end{bmatrix}$$

Consequently, $\mathbf{X}'\mathbf{X}/n = \frac{1}{n}\sum_{i=1}^{n}\mathbf{x}_i\mathbf{x}_i'$

$$\begin{aligned} \mathbf{X}'\mathbf{y}/n &= \frac{1}{n}\begin{bmatrix} x_{11} & x_{21} & \dots & x_{n1} \\ x_{12} & x_{22} & \dots & x_{n2} \\ \vdots & \vdots & \ddots & \vdots \\ x_{1k} & x_{2k} & \dots & x_{nk} \end{bmatrix}\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} \\ &= \frac{1}{n}\begin{bmatrix} \sum_{i=1}^{n} x_{i1}y_i \\ \sum_{i=1}^{n} x_{i2}y_i \\ \vdots \\ \sum_{i=1}^{n} x_{ik}y_i \end{bmatrix} \\ &= \frac{1}{n}\sum_{i=1}^{n}\mathbf{x}_i y_i \end{aligned}$$

**3.** For the Simple Regression Model ($K = 2$, $x_{i1} = 1\ \forall i$) let us prove that:

$$\mathbf{S_{xx}} = \begin{bmatrix} 1 & \bar{x}_2 \\ \bar{x}_2 & \frac{1}{n}\sum_{i=1}^{n} x_{i2}^2 \end{bmatrix} \qquad \mathbf{S_{xy}} = \begin{bmatrix} \bar{y} \\ \frac{1}{n}\sum_{i=1}^{n} x_{i2}y_i \end{bmatrix}$$

$$b_2 = \frac{\frac{1}{n}\sum_{i=1}^{n}(x_{i2}-\bar{x}_2)(y_i-\bar{y})}{\frac{1}{n}\sum_{i=1}^{n}(x_{i2}-\bar{x}_2)^2} \qquad b_1 = \bar{y} - \bar{x}_2 b_2$$

Where

$$\bar{y} := \frac{1}{n}\sum_{i=1}^{n} y_i \qquad \bar{x}_2 := \frac{1}{n}\sum_{i=1}^{n} x_{i2}$$

$$\mathbf{S}_{xx} = \mathbf{X}'\mathbf{X}/n = \frac{1}{n}\begin{bmatrix} x_{11} & x_{21} & \dots & x_{n1} \\ x_{12} & x_{22} & \dots & x_{n2} \end{bmatrix}\begin{bmatrix} x_{11} & x_{12} \\ x_{21} & x_{22} \\ \vdots & \vdots \\ x_{n1} & x_{n2} \end{bmatrix}$$

$$= \frac{1}{n}\begin{bmatrix} \sum_{i=1}^{n} x_{i1}^2 & \sum_{i=1}^{n} x_{i1}x_{i2} \\ \sum_{i=1}^{n} x_{i2}x_{i1} & \sum_{i=1}^{n} x_{i2}^2 \end{bmatrix} = \begin{bmatrix} 1 & \bar{x}_2 \\ \bar{x}_2 & \frac{1}{n}\sum_{i=1}^{n} x_{i2}^2 \end{bmatrix} \quad (\text{since} \quad x_{i1} = 1\ \forall i)$$

$$\mathbf{S}_{xy} = \mathbf{X}'\mathbf{y}/n = \frac{1}{n}\begin{bmatrix} x_{11} & x_{21} & \dots & x_{n1} \\ x_{12} & x_{22} & \dots & x_{n2} \end{bmatrix}\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}$$

$$= \frac{1}{n}\begin{bmatrix} \sum_{i=1}^{n} x_{i1}y_i \\ \sum_{i=1}^{n} x_{i2}y_i \end{bmatrix} = \begin{bmatrix} \bar{y} \\ \frac{1}{n}\sum_{i=1}^{n} x_{i2}y_i \end{bmatrix}$$

By minimizing the objective function $\mathbf{e}'\mathbf{e}$ we obtain $\mathbf{b} = \mathbf{S_{xx}^{-1}S_{xy}}$ and applying brute force matrix inversion:

$$\mathbf{S}_{xx}^{-1} = \frac{1}{\frac{1}{n}\sum_{i=1}^{n} x_{i2}^2 - \bar{x}_2^2} \begin{bmatrix} \frac{1}{n}\sum_{i=1}^{n} x_{i2}^2 & -\bar{x}_2 \\ -\bar{x}_2 & 1 \end{bmatrix}$$

We will leverage the following equality to derive $\mathbf{b}$:

$$\frac{1}{n}\sum_{i=1}^{n} x_{i2}^2 - \bar{x}_2^2 = \frac{1}{n}\sum_{i=1}^{n} (x_{i2} - \bar{x}_2)^2$$

Thus:

$$\begin{aligned} \mathbf{b} &= \frac{1}{\frac{1}{n}\sum_{i=1}^{n} (x_{i2} - \bar{x}_2)^2} \begin{bmatrix} \frac{1}{n}\sum_{i=1}^{n} x_{i2}^2 & -\bar{x}_2 \\ -\bar{x}_2 & 1 \end{bmatrix} \begin{bmatrix} \bar{y} \\ \frac{1}{n}\sum_{i=1}^{n} x_{i2}y_i \end{bmatrix} \\ &= \frac{1}{\frac{1}{n}\sum_{i=1}^{n} (x_{i2} - \bar{x}_2)^2} \begin{bmatrix} \bar{y}\frac{1}{n}\sum_{i=1}^{n} x_{i2}^2 - \bar{x}_2\frac{1}{n}\sum_{i=1}^{n} x_{i2}y_i \\ \frac{1}{n}\sum_{i=1}^{n} x_{i2}y_i - \bar{x}_2\bar{y} \end{bmatrix} \end{aligned}$$

Notice that the estimator vector $\mathbf{b}$ consists of two elements, $b_1$ and $b_2$:

$$b_2 = \frac{\frac{1}{n}\sum_{i=1}^{n} x_{i2}y_i - \bar{x}_2\bar{y}}{\frac{1}{n}\sum_{i=1}^{n}(x_{i2} - \bar{x}_2)^2} = \frac{\frac{1}{n}\sum_{i=1}^{n}(x_{i2} - \bar{x}_2)(y_i - \bar{y})}{\frac{1}{n}\sum_{i=1}^{n}(x_{i2} - \bar{x}_2)^2}$$

Since:

$$\begin{aligned} \frac{1}{n}\sum_{i=1}^{n}(x_{i2} - \bar{x}_2)(y_i - \bar{y}) &= \frac{1}{n}\sum_{i=1}^{n} x_{i2}y_i - \bar{y}\frac{1}{n}\sum_{i=1}^{n} x_{i2} - \bar{x}_2\frac{1}{n}\sum_{i=1}^{n} y_i + \frac{1}{n}\sum_{i=1}^{n}\bar{x}_2\bar{y} \\ &= \frac{1}{n}\sum_{i=1}^{n} x_{i2}y_i - \bar{y}\bar{x}_2 - \bar{x}_2\bar{y} + \bar{x}_2\bar{y} = \frac{1}{n}\sum_{i=1}^{n} x_{i2}y_i - \bar{x}_2\bar{y} \end{aligned}$$

$$
\begin{aligned}
b_1 &= \frac{\bar{y}\frac{1}{n}\sum_{i=1}^{n}\left(x_{i2}^2 - 2^2 \bar{+} \bar{x}_2^2\right) - \bar{x}_2\frac{1}{n}\sum_{i=1}^{n} x_{i2}y_i}{\frac{1}{n}\sum_{i=1}^{n}(x_{i2}-\bar{x}_2)^2} \\
&= \bar{y} - \bar{x}_2\frac{\frac{1}{n}\sum_{i=1}^{n} x_{i2}y_i}{\frac{1}{n}\sum_{i=1}^{n}(x_{i2}-\bar{x}_2)^2} - \bar{y}\bar{x}_2^2 = \bar{y} - \bar{x}_2 \underbrace{\frac{\frac{1}{n}\sum_{i=1}^{n} x_{i2}y_i - \bar{x}_2\bar{y}}{\frac{1}{n}\sum_{i=1}^{n}(x_{i2}-\bar{x}_2)^2}}_{b_2} \\
&= \bar{y} - \bar{x}_2 b_2
\end{aligned}
$$

**4.** Let the projection matrix be denoted as $\mathbf{P} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$ and the annihilator matrix as $\mathbf{M} = (\mathbf{I}_n - \mathbf{P})$, we prove they are both symmetric and idempotent.

Note that a symmetric matrix $\mathbf{A}$ satisfies $\mathbf{A} = \mathbf{A}'$ and a matrix $\mathbf{C}$ is said to be idempotent if $\mathbf{C} = \mathbf{C}^2$

$$
\begin{aligned}
\mathbf{P}' &= (\mathbf{X}(\mathbf{X}'\mathbf{X}^{-1})\mathbf{X})' \\
&= \underbrace{\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'}_{\mathbf{P}} \quad \big(\text{since} \quad (\mathbf{AB})' = \mathbf{B}'\mathbf{A}'\big) \\
\mathbf{P}^2 &= [\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\underbrace{\mathbf{X}'][\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}}_{\mathbf{I}_k}\mathbf{X}'] \\
&= \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' \\
\mathbf{M}' &= (\mathbf{I}_n - \mathbf{P})' \\
&= \mathbf{I}_n' - \mathbf{P}' = \mathbf{I}_n - \mathbf{P} \quad \big(\text{since } \mathbf{P} \text{ and } \mathbf{I}_n \text{ are symmetric}\big) \\
\mathbf{M}^2 &= (\mathbf{I}_n - \mathbf{P})(\mathbf{I}_n - \mathbf{P}) \\
&= \mathbf{I}_n - \mathbf{P} - \mathbf{P} + \mathbf{P}^2 \\
&= \mathbf{I}_n - \mathbf{P} \quad \big(\text{since } \mathbf{P} \text{ is idempotent and } \mathbf{AI} = \mathbf{A}\big)
\end{aligned}
$$

**5.** Let us prove some matrix algebra properties of fitted values and residuals:

i) $\hat{\mathbf{y}} = \mathbf{P}\mathbf{y}$

$$\begin{aligned}
\hat{\mathbf{y}} &= \mathbf{X}\mathbf{b} \quad \big(\text{by definition of } \ \hat{\mathbf{y}}\big) \\
&= \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} \quad \big(\text{since } \ \mathbf{b} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}\big) \\
&= \mathbf{P}\hat{\mathbf{y}}
\end{aligned}$$

ii) $\mathbf{e} = \mathbf{M}\mathbf{y}$ and $\mathbf{e} = \mathbf{M}\boldsymbol{\varepsilon}$

$$\begin{aligned}
\mathbf{e} &= \mathbf{y} - \hat{\mathbf{y}} \\
&= \mathbf{y} - \mathbf{P}\mathbf{y} \quad \big(\text{as we proved}\big) \\
&= (\mathbf{I}_n - \mathbf{P})\mathbf{y} \\
&= \mathbf{M}\mathbf{y} \\
\mathbf{e} &= \mathbf{y} - \mathbf{X}\mathbf{b} \\
&= \mathbf{y} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\underbrace{(\mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon})}_{\mathbf{y}} \\
&= \mathbf{y} - \mathbf{X}\boldsymbol{\beta} + \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\varepsilon} \\
&= \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} - \mathbf{X}\boldsymbol{\beta} + \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\varepsilon} \\
&= (\mathbf{I}_n - \mathbf{P})\boldsymbol{\varepsilon} \\
&= \mathbf{M}\boldsymbol{\varepsilon}
\end{aligned}$$

iii) SSR $= \boldsymbol{\varepsilon}'\mathbf{M}\boldsymbol{\varepsilon}$

$$\begin{aligned}
SSR &= \mathbf{e}'\mathbf{e} \\
&= (\mathbf{M}\boldsymbol{\varepsilon})'\mathbf{M}\boldsymbol{\varepsilon} \\
&= \boldsymbol{\varepsilon}'\mathbf{M}'\mathbf{M}\boldsymbol{\varepsilon} \\
&= \boldsymbol{\varepsilon}'\mathbf{M}\boldsymbol{\varepsilon} \quad \big(\text{since } \mathbf{M} \text{ is idempotent}\big)
\end{aligned}$$

**6.** $R^2$ is not altered by changes in the unit of measurement of $(\mathbf{y}, \mathbf{X})$:

$$R^2 = \frac{\displaystyle\sum_{i=1}^{n}(\mathbf{k}\hat{y}_i - \mathbf{k}\bar{y})^2}{\displaystyle\sum_{i=1}^{n}(\mathbf{k}y_i - \mathbf{k}\bar{y})^2}$$

$$= \frac{\mathbf{k}}{\mathbf{k}}\frac{\displaystyle\sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2}{\displaystyle\sum_{i=1}^{n}(y_i - \bar{y})^2}$$

In the case of regressors, let us study whether $\mathbf{b}$ is modified or not:

$$\begin{aligned}\mathbf{b} &= (\mathbf{kX'kX})^{-1}\mathbf{kX'y} \\ &= [\mathbf{k}(\mathbf{X'X})]^{-1}\mathbf{kX'y} \\ &= (\mathbf{X'X})^{-1}\mathbf{X'y}\end{aligned}$$

**7.** Given that $R_{uc}^2 = 1 - \dfrac{\mathbf{e'e}}{\mathbf{y'y}}$ and $R^2 = 1 - \dfrac{\sum_{i=1}^{n} e_i^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2}$

We prove the following relationship:

$$1 - R^2 = \left(1 + \frac{n\bar{y}^2}{\displaystyle\sum_{i=1}^{n}(y_i - \bar{y})^2}\right)(1 - R_{uc}^2)$$

$$1 - R^2 = \frac{\displaystyle\sum_{i=1}^{n} e_i^2}{\displaystyle\sum_{i=1}^{n}(y_i - \bar{y})^2}$$

Note that:

$$
\begin{aligned}
\sum_{i=1}^{n}(y_i-\bar{y})^2 &= \sum_{i=1}^{n} y_i^2 + \sum_{i=1}^{n}\bar{y}^2 - 2\sum_{i=1}^{n} y_i\bar{y} \\
&= \sum_{i=1}^{n} y_i^2 + n\bar{y}^2 - 2\bar{y}\sum_{i=1}^{n} y_i \\
&= \sum_{i=1}^{n} y_i^2 + n\bar{y}^2 - 2n\bar{y}^2 \\
&= \sum_{i=1}^{n} y_i^2 - n\bar{y}^2
\end{aligned}
$$

Then:

$$
\begin{aligned}
1-R^2 &= \frac{\displaystyle\sum_{i=1}^{n} e_i^2}{\displaystyle\sum_{i=1}^{n} y_i^2 - n\bar{y}^2} = \frac{\mathbf{e}'\mathbf{e}}{\mathbf{y}'\mathbf{y}-n\bar{y}^2} \\
&= \frac{(1-R_{uc}^2)\mathbf{y}'\mathbf{y}}{\mathbf{y}'\mathbf{y}-n\bar{y}^2} \quad (\text{since } \mathbf{e}'\mathbf{e} = (1-R_{uc}^2)\mathbf{y}'\mathbf{y}) \\
&= \frac{\mathbf{y}'\mathbf{y}}{\mathbf{y}'\mathbf{y}-n\bar{y}^2}(1-R_{uc}^2) \\
&= \frac{\displaystyle\sum_{i=1}^{n} y_i^2}{\displaystyle\sum_{i=1}^{n} y_i^2 - n\bar{y}^2}(1-R_{uc}^2) \\
&= \frac{\sum_{i=1}^{n} y_i^2 - n\bar{y}^2 + n\bar{y}^2}{\sum_{i=1}^{n} y_i^2 - n\bar{y}^2}(1-R_{uc}^2)
\end{aligned}
$$

Notice that we applied the add-and-subtract strategy and since:

$$\sum_{i=1}^{n}(y_i - \bar{y})^2 = \sum_{i=1}^{n} y_i^2 - n\bar{y}^2$$

We finally obtain:

$$1 - R^2 = \left( 1 + \frac{n\bar{y}^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \right)(1 - R^2_{uc})$$

**8.** Let us show that $R^2_{uc} = \dfrac{\mathbf{y}'\mathbf{P}\mathbf{y}}{\mathbf{y}'\mathbf{y}}$

$$\begin{aligned}\hat{\mathbf{y}} &= \mathbf{X}\mathbf{b} \\ &= \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} \\ &= \mathbf{P}\mathbf{y}\end{aligned}$$

Since $\mathbf{P}$ is symmetric and idempotent:

$$\mathbf{y}'\mathbf{P}\mathbf{y} = \mathbf{y}'\mathbf{P}'\mathbf{P}\mathbf{y} = \hat{\mathbf{y}}'\hat{\mathbf{y}}$$

Then:

$$R^2_{uc} = \frac{\hat{\mathbf{y}}'\hat{\mathbf{y}}}{\mathbf{y}'\mathbf{y}} = \frac{\mathbf{y}'\mathbf{P}\mathbf{y}}{\mathbf{y}'\mathbf{y}}$$

**9.** Proof that regression coefficients and related statistics can be calculated from sample averages $\mathbf{S_{xx}}, \mathbf{S_{xy}}, \bar{y}, \mathbf{y}'\mathbf{y}/n$. Thus, they only need to be computed once.

$$\begin{aligned}\mathbf{b} &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} = (n^{-1}\mathbf{X}'\mathbf{X})^{-1}n^{-1}\mathbf{X}'\mathbf{y} \\ &= \mathbf{S}_{\mathbf{xx}}^{-1}\mathbf{S_{xy}}\end{aligned}$$

$$
\begin{aligned}
SSR &= (\mathbf{y} - \mathbf{X}\mathbf{b})'(\mathbf{y} - \mathbf{X}\mathbf{b}) \\
&= \mathbf{y}'\mathbf{y} - \mathbf{y}'\mathbf{X}\mathbf{b} - (\mathbf{X}\mathbf{b})'\mathbf{y} + (\mathbf{X}\mathbf{b})'(\mathbf{X}\mathbf{b}) \\
&= \mathbf{y}'\mathbf{y}/n - 2\mathbf{b}\mathbf{X}'\mathbf{y}/n + \mathbf{b}'\mathbf{X}'\mathbf{X}/n\mathbf{b} \\
&= \mathbf{y}'\mathbf{y}/n - 2\mathbf{b}\mathbf{S}_{\mathbf{xy}} + \mathbf{b}'\mathbf{S}_{\mathbf{xx}}\mathbf{b}
\end{aligned}
$$

$$
\begin{aligned}
S^2 &= \frac{SSR}{n-k} \\
&= \frac{\mathbf{y}'\mathbf{y}/n - 2\mathbf{b}\mathbf{S}_{\mathbf{xy}} + \mathbf{b}'\mathbf{S}_{\mathbf{xx}}\mathbf{b}}{n-k}
\end{aligned}
$$

$$
\begin{aligned}
R^2 &= 1 - \frac{SSR}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \\
&= 1 - \frac{\mathbf{y}'\mathbf{y}/n - 2\mathbf{b}\mathbf{S}_{\mathbf{xy}} + \mathbf{b}'\mathbf{S}_{\mathbf{xx}}\mathbf{b}}{\mathbf{y}'\mathbf{y}/n - \bar{y}^2}
\end{aligned}
$$

## 1.3 Finite Sample Properties of OLS

**1.** We prove Gauss-Markov Theorem. Additonally, we prove the adequacy of $S^2$ as an estimator for $\sigma^2$. Our Linear Regression Model satisfies the following assumptions:

- A.1 (Linearity):

$$\mathrm{E}[\mathbf{Y}|\mathbf{X}] = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$$

- A.2 (Strict Exogeneity):

$$\mathrm{E}[\boldsymbol{\varepsilon}|\mathbf{X}] = 0$$

- A.3 (Rank Condition):

$$\Pr(\mathrm{rank}(\mathbf{X}) = K) = 1$$

- A.4 (Spherical Error Variance):

$$\mathrm{E}[\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}'|\mathbf{X}] = \sigma^2\mathbf{I} \qquad \mathrm{E}[\varepsilon_i\varepsilon_j|\mathbf{X}] = 0 \quad \text{if } i \neq j$$

i) The OLS estimator $\mathbf{b}$ is unbiased:

$$
\begin{aligned}
\mathrm{E}[\mathbf{b}|\mathbf{X}] &= \mathrm{E}[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}|\mathbf{X}] \quad \text{(by minimizing SSR)} \\
&= \mathrm{E}[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'(\mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon})] \quad \text{(by A.1)} \\
&= \mathrm{E}[\boldsymbol{\beta} + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\varepsilon}|\mathbf{X}] \\
&= \boldsymbol{\beta} + \mathrm{E}[\mathbf{A}\boldsymbol{\varepsilon}|\mathbf{X}] \quad (\mathbf{A} := (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}') \\
&= \boldsymbol{\beta} + \mathbf{A}\mathrm{E}[\boldsymbol{\varepsilon}|\mathbf{X}] \quad \text{(by linearity of conditional expectations)} \\
&= \boldsymbol{\beta} \quad \text{(by A.2)}
\end{aligned}
$$

Note that $\mathbf{X}'\mathbf{X}$ is invertible by A.3, so all previous results hold. Now let us focus on the variance of $\mathbf{b}$:

$$
\begin{aligned}
\mathrm{V}[\mathbf{b}|\mathbf{X}] &= \mathrm{V}[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}|\mathbf{X}] \\
&= \mathrm{V}[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'(\mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon})] \quad \text{(by A.1)} \\
&= \mathrm{V}[\boldsymbol{\beta} + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\varepsilon}|\mathbf{X}] \\
&= \mathrm{V}[\mathbf{A}\boldsymbol{\varepsilon}|\mathbf{X}] \quad (\text{as } \boldsymbol{\beta} \text{ is a fixed parameter}) \\
&= \mathbf{A}\mathrm{V}[\boldsymbol{\varepsilon}|\mathbf{X}]\mathbf{A}' \quad \text{(by properties of the variance)} \\
&= \mathbf{A}\sigma^2\mathbf{I}\mathbf{A}' \quad \text{(by A.4)} \\
&= \sigma^2[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}] \\
&= \sigma^2(\mathbf{X}'\mathbf{X})^{-1}
\end{aligned}
$$

Where we claimed $\boldsymbol{\beta}$ is a fixed parameter as we are taking a Frequentist Approach (may Bayesians forgive me).

ii) The OLS estimator $\mathbf{b}$ is efficient.

Let $\hat{\boldsymbol{\beta}}$ be an unbiased estimator for $\boldsymbol{\beta}$ and linear in $\mathbf{y}$ such that:

$$\hat{\boldsymbol{\beta}} = \mathbf{C}\mathbf{y}$$

Where $\mathbf{C}$ is a function of $\mathbf{X}$. Let $\mathbf{D} = \mathbf{C}$ - $\mathbf{A}$, where $\mathbf{A} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$
Then $\hat{\boldsymbol{\beta}} = (\mathbf{D} + \mathbf{A})\mathbf{y}$

$$
\begin{aligned}
\mathrm{E}[\hat{\boldsymbol{\beta}}|\mathbf{X}] &= \mathrm{E}[(\mathbf{D}+\mathbf{A})\mathbf{y}|\mathbf{X}] \quad \text{(by definition of } \mathbf{C}\text{)} \\
&= \mathrm{E}[\mathbf{D}\mathbf{y}+\mathbf{A}\mathbf{y}|\mathbf{X}] \\
&= \mathrm{E}[\mathbf{D}(\mathbf{X}\boldsymbol{\beta}+\boldsymbol{\varepsilon})+(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'(\mathbf{X}\boldsymbol{\beta}+\boldsymbol{\varepsilon})|\mathbf{X}] \\
&= \mathrm{E}[\mathbf{D}\mathbf{X}\boldsymbol{\beta}+\mathbf{D}\boldsymbol{\varepsilon}+\boldsymbol{\beta}+\mathbf{A}\boldsymbol{\varepsilon}|\mathbf{X}] \\
&= \mathbf{D}\mathbf{X}\boldsymbol{\beta}+\boldsymbol{\beta}+\mathbf{D}\mathrm{E}[\boldsymbol{\varepsilon}|\mathbf{X}]+\mathbf{A}\mathrm{E}[\boldsymbol{\varepsilon}|\mathbf{X}] \\
&= \mathbf{D}\mathbf{X}\boldsymbol{\beta}+\boldsymbol{\beta}
\end{aligned}
$$

As $\hat{\boldsymbol{\beta}}$ is unbiased by construction: $\mathrm{E}[\hat{\boldsymbol{\beta}}|\mathbf{X}]=\boldsymbol{\beta}$

$$
\begin{aligned}
\mathrm{E}[\hat{\boldsymbol{\beta}}|\mathbf{X}] &= \mathbf{D}\mathbf{X}\boldsymbol{\beta}+\boldsymbol{\beta} \\
\boldsymbol{\beta} &= \mathbf{D}\mathbf{X}\boldsymbol{\beta}+\boldsymbol{\beta} \\
\mathbf{D}\mathbf{X}\boldsymbol{\beta} &= 0 \implies \mathbf{D}\mathbf{X}=0
\end{aligned}
$$

Deriving the sampling error:

$$
\begin{aligned}
\hat{\boldsymbol{\beta}} &= (\mathbf{D}+\mathbf{A})\mathbf{y} \\
&= \mathbf{D}\mathbf{X}\boldsymbol{\beta}+\mathbf{D}\boldsymbol{\varepsilon}+\mathbf{A}\mathbf{y} \\
&= \mathbf{D}\boldsymbol{\varepsilon}+\mathbf{A}(\mathbf{X}\boldsymbol{\beta}+\boldsymbol{\varepsilon}) \quad \text{(since } \mathbf{D}\mathbf{X}=0\text{)} \\
&= \mathbf{D}\boldsymbol{\varepsilon}+\boldsymbol{\beta}+\mathbf{A}\boldsymbol{\varepsilon} \quad \text{(since } \mathbf{A}\mathbf{y}=\boldsymbol{\beta}+\mathbf{A}\boldsymbol{\varepsilon}\text{)} \\
\hat{\boldsymbol{\beta}}-\boldsymbol{\beta} &= (\mathbf{D}+\mathbf{A})\boldsymbol{\varepsilon}
\end{aligned}
$$

Then:

$$
\begin{aligned}
\mathrm{V}[\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}|\mathbf{X}] &= \mathrm{V}[(\mathbf{D}+\mathbf{A})\mathbf{y}|\mathbf{X}] \\
\mathrm{V}[\hat{\boldsymbol{\beta}}|\mathbf{X}] &= \mathrm{V}[(\mathbf{D}+\mathbf{A})\mathbf{y}|\mathbf{X}] \quad \text{(since } \boldsymbol{\beta} \text{ is a parameter)} \\
&= (\mathbf{D}+\mathbf{A})\mathrm{V}[\boldsymbol{\varepsilon}|\mathbf{X}](\mathbf{D}+\mathbf{A})' \\
&= \sigma^2[(\mathbf{D}+\mathbf{A})(\mathbf{D}+\mathbf{A})'] \quad \text{(by A.4)} \\
&= \sigma^2[\mathbf{D}\mathbf{D}'+\mathbf{A}\mathbf{A}'] \\
&= \sigma^2\mathbf{D}\mathbf{D}'+\sigma^2(\mathbf{X}'\mathbf{X})^{-1}
\end{aligned}
$$

The step that was skipped is reproduced below:

$$\begin{aligned}\mathrm{V}[\hat{\boldsymbol{\beta}}|\mathbf{X}] &= \sigma^2[(\mathbf{D}+\mathbf{A})(\mathbf{D}+\mathbf{A})'] \\ &= \sigma^2[\mathbf{D}\mathbf{D}' + \mathbf{D}\mathbf{A}' + \mathbf{A}\mathbf{D}' + \mathbf{A}\mathbf{A}']\end{aligned}$$

Note that

$$\begin{aligned}\mathbf{D}\mathbf{A}' + \mathbf{A}\mathbf{D}' &= 2\mathbf{D}\mathbf{A}' \quad \big(\text{since } (\mathbf{A}\mathbf{B})' = \mathbf{B}'\mathbf{A}'\big) \\ &= 2\mathbf{D}\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \\ &= 0 \quad \big(\text{since } \mathbf{D}\mathbf{X} = 0\big)\end{aligned}$$

Clearly:

$$\sigma^2[\mathbf{D}\mathbf{D}' + \mathbf{D}\mathbf{A}' + \mathbf{A}\mathbf{D}' + \mathbf{A}\mathbf{A}'] = \sigma^2[\mathbf{D}\mathbf{D}' + \mathbf{A}\mathbf{A}']$$

Consequently:

$$\mathrm{V}[\hat{\boldsymbol{\beta}}] = \sigma^2\mathbf{D}\mathbf{D}' + \underbrace{\sigma^2(\mathbf{X}'\mathbf{X})^{-1}}_{\mathrm{V}[\mathbf{b}|\mathbf{X}]}$$

Since **DD**' is Positive Semidefinite by construction, it entails $\mathbf{DD} \geq 0$. Thus:

$$\begin{aligned}\sigma^2\mathbf{D}\mathbf{D}' + \sigma^2(\mathbf{X}'\mathbf{X})^{-1} &\geq \sigma^2(\mathbf{X}'\mathbf{X})^{-1} \\ \mathrm{V}[\hat{\boldsymbol{\beta}}] &\geq \mathrm{V}[\mathbf{b}|\mathbf{X}]\end{aligned}$$

Indeed, the OLS estimator **b** exhibits the lowest variance within the class of linear unbiased estimators. We have just proved Gauss-Markov Theorem.

iii) $S^2$ is an unbiased estimator for $\sigma^2$.

Notice how $\sigma^2$ has been present throughout the entirety of the Gauss-Markov Theorem proof. It is actually of little practicality to deal with a parameter in our expressions, as we want to estimate a Linear Regression Model. Since it is not observable, we have to estimate $\sigma^2$ as well, for which we propose $S^2$, defined as

$$SSR/n - k = \mathbf{e}'\mathbf{e}/n - k$$

We proved $\mathbf{e}'\mathbf{e} = \boldsymbol{\varepsilon}'\mathbf{M}\boldsymbol{\varepsilon}$ earlier, so:

$$\begin{aligned}
\mathrm{E}[S^2|\mathbf{X}] &= \mathrm{E}\left[\frac{\boldsymbol{\varepsilon}'\mathbf{M}\boldsymbol{\varepsilon}}{n-k}|\mathbf{X}\right] \\
&= \frac{1}{n-k}\mathrm{E}\left[\sum_{i=1}^{n}\sum_{j=1}^{n} m_{ij}\varepsilon_i\varepsilon_j \big| \mathbf{X}\right] \\
&= \frac{1}{n-k}\sum_{i=1}^{n}\sum_{j=1}^{n} m_{ij}\mathrm{E}\left[\varepsilon_i\varepsilon_j \big| \mathbf{X}\right] \\
&= \frac{1}{n-k}\sum_{i=1}^{n}\sum_{j=1}^{n} m_{ii}\sigma^2 \quad \big(\text{by A.4}\big) \\
&= \frac{\sigma^2}{n-k}\mathrm{trace}(\mathbf{M})
\end{aligned}$$

Only for $i \neq j$ does $\mathrm{E}[\varepsilon_i\varepsilon_j] = \sigma^2$ then $m_{ij} \neq 0$ in the same case. Consequently it becomes $m_{ii}$ (or $m_{jj}$) which includes only the diagonal elements of the matrix $\mathbf{M}$. Since we are summing the diagonal of a matrix, the trace operator kicks in.

$$\begin{aligned}
\mathrm{trace}(\mathbf{M}) &= \mathrm{trace}(\mathbf{I}_n - \mathbf{P}) \quad \big(\text{by definition of } \mathbf{M}\big) \\
&= n - \mathrm{trace}(\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}') \quad \big(\text{by definition of } \mathbf{P}\big) \\
&= n - \mathrm{trace}((\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X}) \quad \big(\mathrm{trace}(\mathbf{AB}) = \mathrm{trace}(\mathbf{BA})\big) \\
&= n - k \quad \big(\text{since } (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X} = \mathbf{I}_k\big)
\end{aligned}$$

Thus:

$$\begin{aligned}
\mathrm{E}[S^2|\mathbf{X}] &= \frac{\sigma^2}{n-k}\mathrm{trace}(\mathbf{M}) \\
&= \frac{\sigma^2}{n-k}\, n-k \\
&= \sigma^2
\end{aligned}$$

We have just proved that the estimator $S^2$ is unbiased. Note that dividing $SSR$ by the degrees of freedom $n-k$ was imperative to achieve unbiasedness.

**2.** We proved Gauss-Markov Theorem which establishes that the OLS estimator $\mathbf{b}$ is unbiased and efficient in the class of linear unbiased estimators. This does not entail that any other linear estimator is excluded from unbiasedness. Let us show an example for the following Linear Regression model satisfying the previously stated assumptions:

$$CON_i = \beta_1 + \beta_2 YD_i + \varepsilon_i$$

Let our synthetic estimator be denoted as:

$$\hat{\beta}_2 = \tfrac{CON_2 - CON_1}{YD_2 - YD_1}$$

We proceed to show it is unbiased:

$$\mathrm{E}[\hat{\beta}_2|\mathbf{YD}] = \mathrm{E}\left[\frac{\beta_1 + \beta_2 YD_2 + \varepsilon_2 - \beta_1 - \beta_2 YD_1 - \varepsilon_1}{YD_2 - YD_1}|\mathbf{YD}\right]$$

$$= \frac{1}{YD_2 - YD_1}\mathrm{E}\left[\beta_2 YD_2 - \beta_2 YD_1 + \varepsilon_2 - \varepsilon_1|\,\mathbf{YD}\right]$$

$$= \frac{1}{YD_2 - YD_1}\mathrm{E}[\beta_2(YD_2 - YD_1)|\mathbf{YD}] + \mathrm{E}[\varepsilon_2 - \varepsilon_1|\mathbf{YD}]$$

$$= \beta_2\left(\frac{YD_2 - YD_1}{YD_2 - YD_1}\right) + \frac{1}{YD_2 - YD_1}\left(\mathrm{E}[\varepsilon_2|\mathbf{YD}] - \mathrm{E}[\varepsilon_1|\mathbf{YD}]\right)$$

$$= \beta_2 \quad \big(\text{by A.2: Strict Exogeneity Assumption}\big)$$

**3.** Likewise, there exist linear but not necessarily unbiased estimators that have smaller variance than the OLS estimator $\mathbf{b}$. We present an example here:

Let $\hat{\beta} = c$, where $c$ represents any real number (and thus a constant). Unless $\beta = c$, our estimator $\hat{\beta}$ is badly biased. However, note that $\mathrm{V}[\hat{\beta}] = 0$ by construction, as $c$ lacks any sort of stochasticity.

Clearly: $\mathrm{V}[\mathbf{b}] > \mathrm{V}[\hat{\beta}]$

The purpose of questions **2)** and **3)** is to show that unbiasedness and low variance are not sufficient properties by themselves to determine whether an estimator is good or not, as well as to clarify what Gauss-Markov Theorem really means.

**4.** We prove the following equality for the unconditional variance:

$$\mathrm{V}[\hat{\boldsymbol{\beta}}] = \mathrm{E}\left[\mathrm{V}[\hat{\boldsymbol{\beta}}|\mathbf{X}]\right] + \mathrm{V}\left[\mathrm{E}[\hat{\boldsymbol{\beta}}|\mathbf{X}]\right]$$

Using the add-and-subtract strategy and defining:

$$\mathbf{d} := \hat{\boldsymbol{\beta}} - \mathrm{E}[\hat{\boldsymbol{\beta}}|\mathbf{X}] \qquad \mathbf{a} := \hat{\boldsymbol{\beta}} - \mathrm{E}[\hat{\boldsymbol{\beta}}] \qquad \mathbf{c} := \mathrm{E}[\hat{\boldsymbol{\beta}}|\mathbf{X}] - \mathrm{E}[\hat{\boldsymbol{\beta}}]$$

Then

$$\mathbf{d} = \mathbf{a}-\mathbf{c} \qquad \mathrm{E}[\mathbf{dd}'] = \mathrm{E}[(\mathbf{a}-\mathbf{c})(\mathbf{a}-\mathbf{c})'] = \mathrm{E}[\mathbf{aa}'] - \mathrm{E}[\mathbf{ca}'] - \mathrm{E}[\mathbf{ac}'] + \mathrm{E}[\mathbf{cc}']$$

By definition:

$$\mathrm{V}[\mathbf{x}] = \sum_{i=1}^{n} (x_i - \mathrm{E}[\mathbf{x}])^2 = (\mathbf{x} - \mathrm{E}[\mathbf{x}])(\mathbf{x} - \mathrm{E}[\mathbf{x}])'$$

Following the previous expression:

$$\mathrm{E}[\mathbf{aa}'] = \mathrm{E}\left[\left(\hat{\boldsymbol{\beta}} - \mathrm{E}[\hat{\boldsymbol{\beta}}]\right)\left(\hat{\boldsymbol{\beta}} - \mathrm{E}[\hat{\boldsymbol{\beta}}]\right)'\right] = \mathrm{V}[\hat{\boldsymbol{\beta}}]$$

$$\begin{aligned}
\mathrm{E}[\mathbf{cc}'] &= \mathrm{E}\left[\left(\mathrm{E}[\hat{\boldsymbol{\beta}}|\mathbf{X}] - \mathrm{E}[\hat{\boldsymbol{\beta}}]\right)\left(\mathrm{E}[\hat{\boldsymbol{\beta}}|\mathbf{X}] - \mathrm{E}[\hat{\boldsymbol{\beta}}]\right)'\right] \\
&= \mathrm{V}\left[\mathrm{E}[\hat{\boldsymbol{\beta}}|\mathbf{X}]\right] \quad \left(\text{since } \mathrm{E}\left[\mathrm{E}[\hat{\boldsymbol{\beta}}|\mathbf{X}]\right] = \mathrm{E}[\hat{\boldsymbol{\beta}}]\right)
\end{aligned}$$

$$\begin{aligned}
\mathrm{E}[\mathbf{ca}'] &= \mathrm{E}[\mathrm{E}[\mathbf{ca}'|\mathbf{X}]] \quad (\text{by LIE}) \\
&= \mathrm{E}\left[\mathrm{E}\left[\left(\mathrm{E}[\hat{\boldsymbol{\beta}}|\mathbf{X}] - \mathrm{E}[\hat{\boldsymbol{\beta}}]\right)\left(\hat{\boldsymbol{\beta}} - \mathrm{E}[\hat{\boldsymbol{\beta}}]\right)' | \mathbf{X}\right]\right] \\
&= \mathrm{E}\left[\left(\mathrm{E}[\hat{\boldsymbol{\beta}}|\mathbf{X}] - \mathrm{E}[\hat{\boldsymbol{\beta}}]\right)\left(\mathrm{E}[\hat{\boldsymbol{\beta}}|\mathbf{X}] - \mathrm{E}[\hat{\boldsymbol{\beta}}]\right)'\right] \\
&= \mathrm{E}\left[\mathrm{V}\left[\mathrm{E}[\hat{\boldsymbol{\beta}}|\mathbf{X}]\right]\right] = \mathrm{V}\left[\mathrm{E}[\hat{\boldsymbol{\beta}}|\mathbf{X}]\right]
\end{aligned}$$

This last equality holds since

$$\begin{aligned} \mathrm{E}\left[\mathrm{E}[\hat{\boldsymbol{\beta}}|\mathbf{X}]|\mathbf{X}\right] &= \mathrm{E}[\hat{\boldsymbol{\beta}}|\mathbf{X}] \\ \mathrm{E}\left[\mathrm{E}[\hat{\boldsymbol{\beta}}|\mathbf{X}]\right] &= \mathrm{E}[\hat{\boldsymbol{\beta}}] \end{aligned}$$

Note that $\mathrm{E}[\mathbf{ca}'] = \mathrm{E}[\mathbf{ac}']$. Thus:

$$\begin{aligned} \mathrm{E}[\mathbf{dd}'] &= \mathrm{E}\left[\mathrm{E}[\mathbf{dd}'|\mathbf{X}]\right] \\ &= \mathrm{E}\left[\mathrm{E}\left[\left(\hat{\boldsymbol{\beta}} - \mathrm{E}[\hat{\boldsymbol{\beta}}|\mathbf{X}]\right)\left(\hat{\boldsymbol{\beta}} - \mathrm{E}[\hat{\boldsymbol{\beta}}|\mathbf{X}]\right)'|\mathbf{X}\right]\right] \\ &= \mathrm{E}\left[\mathrm{V}[\hat{\boldsymbol{\beta}}|\mathbf{X}]\right] \end{aligned}$$

Finally:

$$\begin{aligned} \mathrm{E}\left[\mathrm{V}[\hat{\boldsymbol{\beta}}|\mathbf{X}]\right] &= \mathrm{V}[\hat{\boldsymbol{\beta}}] - \mathrm{V}\left[\mathrm{E}[\hat{\boldsymbol{\beta}}|\mathbf{X}]\right] \\ \mathrm{V}[\hat{\boldsymbol{\beta}}] &= \mathrm{E}\left[\mathrm{V}[\hat{\boldsymbol{\beta}}|\mathbf{X}]\right] + \mathrm{V}\left[\mathrm{E}[\hat{\boldsymbol{\beta}}|\mathbf{X}]\right] \end{aligned}$$

Leveraging this derivation we show that $\mathrm{V}[\hat{\boldsymbol{\beta}}] \geq \mathrm{V}[\mathbf{b}]$

Note that

$$\mathrm{V}\left[\mathrm{E}[\hat{\boldsymbol{\beta}}|\mathbf{X}]\right] = \mathrm{V}\left[\mathrm{E}[\mathbf{b}|\mathbf{X}]\right] = \mathrm{V}[\boldsymbol{\beta}] = 0$$

So

$$\begin{aligned} \mathrm{V}[\hat{\boldsymbol{\beta}}] &= \mathrm{E}\left[\mathrm{V}[\hat{\boldsymbol{\beta}}|\mathbf{X}]\right] = \sigma^2((\mathbf{X}'\mathbf{X})^{-1} + \mathbf{DD}') \\ \mathrm{V}[\mathbf{b}] &= \mathrm{E}\left[\mathrm{V}[\mathbf{b}|\mathbf{X}]\right] = \sigma^2(\mathbf{X}'\mathbf{X})^{-1} \end{aligned}$$

Clearly:

$$\mathrm{E}\left[\mathrm{V}[\hat{\boldsymbol{\beta}}|\mathbf{X}]\right] \geq \mathrm{E}\left[\mathrm{V}[\mathbf{b}|\mathbf{X}]\right] \implies \mathrm{V}[\hat{\boldsymbol{\beta}}] \geq \mathrm{V}[\mathbf{b}]$$

**5.** It was previously noted that dividing by the degrees of freedom actually contributed to obtaining an unbiased estimator for $\sigma^2$. Think of this procedure as an operation similar to Bessel's correction. We now prove that it would not be necessary if we had data on $\boldsymbol{\varepsilon}$, i.e, the error term were observable.

Let our theoretical estimator be denoted as:

$$\Psi := \frac{\boldsymbol{\varepsilon}'\boldsymbol{\varepsilon}}{n}$$

Then

$$\begin{aligned}
\mathrm{E}[\Psi|\mathbf{X}] &= \mathrm{E}\left[\frac{\boldsymbol{\varepsilon}'\boldsymbol{\varepsilon}}{n}|\mathbf{X}\right] = \frac{1}{n}\mathrm{E}[\boldsymbol{\varepsilon}'\boldsymbol{\varepsilon}|\mathbf{X}] \\
&= \frac{1}{n}\mathrm{E}\left[\sum_{i=1}^{n}\varepsilon_i^2|\mathbf{X}\right] \\
&= \frac{1}{n}\sum_{i=1}^{n}\mathrm{E}[\varepsilon_i^2|\mathbf{X}] \\
&= \frac{1}{n}\sum_{i=1}^{n}\sigma^2 \quad \big(\text{by A.4}\big) \\
&= \frac{n}{n}\sigma^2 = \sigma^2
\end{aligned}$$

**7.** Let us prove that the $i$-th element of the matrix $\mathbf{P}$ is bounded: $p_i \in [0,1]$

$$p_i := \mathbf{x}_i'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_i$$

Since trace($\mathbf{P}$) $= K$ then $\sum_{i=1}^{n} p_i = K$ so it must be that $p_i \geq 0\ \forall i$ as the diagonal of $\mathbf{P}$ is nonnegative ($\mathbf{P}$ is Positive Semidefinite).

Let $\mathbf{e}_i$ be a vector containing the $i$-th residual so $\mathbf{e}_{-i} = 0$. Then $p_i = \mathbf{e}_i'\mathbf{P}\mathbf{e}_i$

$$\begin{aligned}
p_i &= \mathbf{e}_i'\mathbf{P}\mathbf{e}_i \\
&= \mathbf{e}_i'(\mathbf{I}_n - \mathbf{M})\mathbf{e}_i \quad (\text{since } \mathbf{M} = \mathbf{I}_n - \mathbf{P}) \\
&= \mathbf{e}_i'\mathbf{e}_i - \mathbf{e}_i'\mathbf{M}\mathbf{e}_i \\
&= 1 - \mathbf{e}_i'\mathbf{M}\mathbf{e}_i \\
p_i &\leq 1 \quad (\text{since } \mathbf{M} \text{ is Positive Semidefinite})
\end{aligned}$$

**6.** Under the previously stated assumptions we proceed to prove that the OLS estimator $\mathbf{b}$ and the residual vector $\mathbf{e}$ are conditionally uncorrelated.

$$\mathrm{Cov}(\mathbf{b},\mathbf{e}|\mathbf{X}) = \mathrm{E}\left[(\mathbf{b}-\mathrm{E}[\mathbf{b}|\mathbf{X}])\,(\mathbf{e}-\mathrm{E}[\mathbf{e}|\mathbf{X}])'\,|\mathbf{X}\right]$$

$$\begin{aligned}
\mathbf{b}-\mathrm{E}[\mathbf{b}|\mathbf{X}] &= \mathbf{b}-\boldsymbol{\beta} \quad \text{(since } \mathbf{b} \text{ is unbiased)}\\
&= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}-\boldsymbol{\beta}\\
&= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'(\mathbf{X}\boldsymbol{\beta}+\boldsymbol{\varepsilon})-\beta \quad \text{(by A.1)}\\
&= \boldsymbol{\beta}+(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\varepsilon}-\boldsymbol{\beta}\\
&= \mathbf{A}\boldsymbol{\varepsilon} \quad (\mathbf{A} \coloneqq (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}')
\end{aligned}$$

$$\begin{aligned}
\mathbf{e}-\mathrm{E}[\mathbf{e}|\mathbf{X}] &= \mathbf{b}-\mathrm{E}[\mathbf{y}-\mathbf{X}\mathbf{b}|\mathbf{X}]\\
&= \mathbf{e}-\mathrm{E}[\mathbf{y}|\mathbf{X}]+\mathrm{E}[\mathbf{X}\mathbf{b}|\mathbf{X}]\\
&= \mathbf{e}-\mathrm{E}[\mathbf{X},\boldsymbol{\beta}+\varepsilon|\mathbf{X}]+\mathbf{X}\mathrm{E}[\mathbf{b}|\mathbf{X}]\\
&= \mathbf{e}-\mathbf{X}\boldsymbol{\beta}+\mathrm{E}[\varepsilon|\mathbf{X}]+\mathbf{X}\boldsymbol{\beta}\\
&= \mathbf{e} \quad \text{(by A.2)}
\end{aligned}$$

Then

$$\begin{aligned}
\mathrm{Cov}(\mathbf{b},\mathbf{e}|\mathbf{X}) &= \mathrm{E}[\mathbf{A}\varepsilon\mathbf{e}'|\mathbf{X}]\\
&= \mathrm{E}[\mathbf{A}\varepsilon\boldsymbol{(\mathbf{M}\varepsilon)'}|\mathbf{X}]\\
&= \mathrm{E}[\mathbf{A}\varepsilon\varepsilon'\mathrm{M}|\mathbf{X}] \quad \text{(since } \mathbf{M} \text{ is symmetric)}\\
&= \mathbf{A}\mathrm{E}[\varepsilon\varepsilon'|\mathbf{X}]\mathbf{M}\\
&= \sigma^2\mathbf{A}\mathbf{M}\\
&= \sigma^2\mathbf{M}\mathbf{A}' \quad \text{(since } \mathbf{M} \text{ is symmetric)}\\
&= 0
\end{aligned}$$

This last equality holds since:

$$\begin{aligned}
\mathbf{M}\mathbf{A}' &= \mathbf{M}\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\\
&= (\mathbf{I}_n-\mathbf{P})\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\\
&= \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}-\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\underbrace{\mathbf{X}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}}_{\mathbf{I}_k}\\
&= 0
\end{aligned}$$