Advanced Econometrics: Large Sample Theory

José Jaén Delgado

Introduction - In this paper solutions to most exercises from Chapter 2 of Fumio Hayashi's *Econometrics* are provided

1 Questions for Review

1.1 Limit Theorems for Sequences of Random Variables

1. Does $\lim_{n\to\infty} z_n = \alpha \implies \text{plim}_{n\to\infty} z_n = \alpha$?

A sequence of random variables $\{z_n\}$ converges in probability to α if $\forall \varepsilon > 0:$

$$\lim_{n\to\infty} \Pr(|z_n - \alpha| \geq \varepsilon) = 0$$

It is easy to conclude from the first fact that:

$$\lim_{n\to\infty} z_n = \alpha \implies \lim_{n\to\infty} |z_n - \alpha| = 0$$

Consequently, for a sufficiently large sample size:

$$\Pr(|z_n - \alpha| \geq \varepsilon) = 0 \implies \underset{n\to\infty}{\text{plim}}\, z_n = \alpha$$

2. Verify that convergence in mean square for random vectors is equivalent to:

$$\lim_{n\to\infty} \mathrm{E}[(\boldsymbol{z}_n - \boldsymbol{z})'(\boldsymbol{z}_n - \boldsymbol{z})] = 0$$

The definition of convergence in the context of sequences of random scalars can be extended to a sequence of random vectors by requiring element-by-element convergence. Note, however, that this is not the case for convergence in distribution, whose connection can be established by applying the Multivariate Convergence in Distribution Theorem.

Thus, the proof boils down to showing that:

$$\lim_{n\to\infty} \mathrm{E}[(\boldsymbol{z}_n - \boldsymbol{z})'(\boldsymbol{z}_n - \boldsymbol{z})] = \lim_{n\to\infty} \mathrm{E}\Big[\sum_{j=1}^{k}(z_{nj} - z_j)^2\Big]$$

Expanding the LHS term:

$$\begin{aligned}\mathrm{E}[(\boldsymbol{z}_n - \boldsymbol{z})'(\boldsymbol{z}_n - \boldsymbol{z})] &= \mathrm{E}[\boldsymbol{z}_n'\boldsymbol{z}_n - \boldsymbol{z}_n'\boldsymbol{z} - \boldsymbol{z}'\boldsymbol{z}_n + \boldsymbol{z}'\boldsymbol{z}] \\ &= \mathrm{E}[\boldsymbol{z}_n'\boldsymbol{z}_n - 2(\boldsymbol{z}_n'\boldsymbol{z}) + \boldsymbol{z}'\boldsymbol{z}] \text{ (since } \boldsymbol{z}_n'\,\boldsymbol{z} = \boldsymbol{z}'\boldsymbol{z}_n)\end{aligned}$$

Note that $\boldsymbol{z}_n$ and $\boldsymbol{z}$ are column vectors:

$$\boldsymbol{z}_n = \begin{bmatrix} z_{n1} \\ z_{n2} \\ \vdots \\ z_{nk} \end{bmatrix}_{(K\times 1)} \qquad \boldsymbol{z} = \begin{bmatrix} z_1 \\ z_2 \\ \vdots \\ z_k \end{bmatrix}_{(K\times 1)}$$

Consequently, bearing in mind that $z_{nj}' = z_{nj}\ \forall j$ (since it is a scalar):

$$\boldsymbol{z}_n'\boldsymbol{z}_n = \begin{bmatrix} z_{n1}' & z_{n2}' & \dots & z_{nk}' \end{bmatrix} \begin{bmatrix} z_{n1} \\ z_{n1} \\ \vdots \\ z_{nk} \end{bmatrix} = z_{n1}^2 + z_{n2}^2 + \cdots + z_{nk}^2 = \sum_{j=1}^{k} z_{nj}^2$$

$$\boldsymbol{z}'\boldsymbol{z} = \sum_{j=1}^{k} z_j^2 \qquad -2(\boldsymbol{z}_n'\boldsymbol{z}) = -2\sum_{j=1}^{k} z_{nj}z_j$$

Combining all the results: $(\boldsymbol{z}_n - \boldsymbol{z})'(\boldsymbol{z}_n - \boldsymbol{z}) = \sum_{j=1}^{k} z_{nj}^2 + \sum_{j=1}^{k} z_j^2 - 2\sum_{j=1}^{k} z_{nj}z_j$

Thus:

$$\lim_{n\to\infty} \mathrm{E}[(\boldsymbol{z}_n - \boldsymbol{z})'(\boldsymbol{z}_n - \boldsymbol{z})] = \lim_{n\to\infty} \mathrm{E}\Big[\sum_{j=1}^{k}(z_{nj} - z_j)^2\Big]$$

And so we can conclude that $\boldsymbol{z}_n \xrightarrow{m.s} \boldsymbol{z}$ if $z_{nk} \xrightarrow{m.s} z_k \ \forall k$

3. Let us prove $\boldsymbol{A}_n\boldsymbol{x}_n \xrightarrow{d} \boldsymbol{A}\boldsymbol{x}$ from Slutzky´s Theorem:

Considering $\boldsymbol{x}_n \xrightarrow{d} \boldsymbol{x}$ and $\boldsymbol{A}_n \xrightarrow{p} \boldsymbol{A}$ since convergence in probability implies convergence in distribution: $\boldsymbol{x}_n + \boldsymbol{A}_n \xrightarrow{d} \boldsymbol{x} + \boldsymbol{A}$

Nevertheless, convergence in distribution can imply convergence in probability if the limiting random variable is a constant. For instance, let $\boldsymbol{y}_n \xrightarrow{p} 0 \implies \boldsymbol{y}_n'\boldsymbol{x}_n \xrightarrow{p} 0$

Note that $\boldsymbol{A}_n\boldsymbol{x}_n$ can be expressed as $(\boldsymbol{A}_n - \boldsymbol{A})\boldsymbol{x}_n + \boldsymbol{A}\boldsymbol{x}_n$

Then $\boldsymbol{A}_n - \boldsymbol{A} \xrightarrow{p} 0$ since $\boldsymbol{A}_n \xrightarrow{p} \boldsymbol{A}$ and leveraging this fact: $(\boldsymbol{A}_n - \boldsymbol{A})\boldsymbol{x}_n \xrightarrow{p} 0$

Consequently, $(\boldsymbol{A}_n - \boldsymbol{A})\boldsymbol{x}_n + \boldsymbol{A}\boldsymbol{x}_n \xrightarrow{d} \boldsymbol{A}\boldsymbol{x}$ (since $\mathbf{A}$ already represents the limiting random variable to be attained and $\boldsymbol{x}_n \xrightarrow{d} \boldsymbol{x}$)

4. Let us show that $\hat{\theta} \xrightarrow{p} \theta$ supposing $\sqrt{n}(\hat{\theta} - \theta) \xrightarrow{d} N(0, \sigma^2)$

From a theoretical point of view, we are given the very definition of a $\sqrt{n}$-consistent or *CAN* (Consistent and Asymptotically Normal) estimator $\hat{\theta}$ so clearly: $\hat{\theta} \xrightarrow{p} \theta$

Analytically, let us multiply and divide by $\sqrt{n}$: $\frac{1}{\sqrt{n}}\sqrt{n}(\hat{\theta} - \theta)$

Clearly, $\lim_{n\to\infty} \frac{1}{\sqrt{n}} = 0 \implies \text{plim}_{n\to\infty} \frac{1}{\sqrt{n}} = 0$ (as we showed in the first exercise).

Applying Slutzky´s Theorem and recalling that $\boldsymbol{z}_n'\sqrt{n}(\hat{\theta} - \theta) \xrightarrow{p} 0$ if $\boldsymbol{z}_n \xrightarrow{p} 0$:

$$\underset{n\to\infty}{\text{plim}} \frac{1}{\sqrt{n}}\sqrt{n}(\hat{\theta} - \theta) = 0$$

Since the limiting cumulative density function of $\sqrt{n}(\hat{\theta} - \theta)$ exists and is finite.

Thus, it is easy to conclude that $\hat{\theta} \xrightarrow{p} \theta$

P.D: From now on we will refer to existing and finite moments as simply ‘finite’. Therefore, if $\mathrm{E}[\phi(x)] < \infty$ then the first moment of $\phi(x)$ is said to exist and be finite.

5. Let $\{z_i\}$ be i.i.d with $\mathrm{E}[z_i] = \mu$ and $\mathrm{V}[z_i] = \sigma^2 < \infty$

Our goal is to prove:

$$\sqrt{n}\left(\frac{1}{\bar{z_n}} - \frac{1}{\mu}\right) \xrightarrow{d} N\left(0, \frac{\sigma^2}{\mu^4}\right)$$

Firstly, by Lindeberg-Levy CLT (Central Limit Theorem):

$$\sqrt{n}(\bar{z_n} - \mu) \xrightarrow{d} N(0, \sigma^2)$$

This holds since:

$$\mathrm{E}[\bar{z_n}] = \mathrm{E}\left[\frac{1}{n}\sum_{i=1}^{n} z_i\right] = \mathrm{E}\left[\frac{1}{n} n\mu\right] = \mu$$

$$\mathrm{V}[\bar{z_n}] = \mathrm{V}\left[\frac{1}{n}\sum_{i=1}^{n} z_i\right] = \frac{1}{n^2}\mathrm{V}\big[n\sigma^2\big] = \frac{\sigma^2}{n} < \infty$$

$$\left(\sum_{i=1}^{n}\sum_{i\neq j}^{n}\mathrm{cov}(z_i, z_j) = 0 \text{ since } \{z_i\} \text{ is i.i.d } \forall i\right)$$

Notice that for any existing continuous transformation $\mathrm{a}(x) = x^{-1}$ applying the Delta Method we retrieve:

$$\sqrt{n}\left(\mathrm{a}(\bar{z_n}) - \mathrm{a}(\mu)\right) \xrightarrow{d} \mathrm{A}(\mu)F(z)$$

Thus:

$$\sqrt{n}\left(\frac{1}{\bar{z_n}} - \frac{1}{\mu}\right) \xrightarrow{d} N\left(0, \frac{\sigma^2}{\mu^4}\right)$$

Since:

- $\mathrm{A}(\mu) = \dfrac{\partial \mathrm{a}(\mu)}{\partial \mu'} = \dfrac{-1}{\mu^2}$
- $\mathrm{A}(\mu)N(0, \sigma^2) = N(0, \mathrm{A}(\mu)\sigma^2\mathrm{A}(\mu)') = N\left(0, \dfrac{\sigma^2}{\mu^4}\right)$
- Note that μ is a scalar so $\mu' = \mu \implies -\mu^{-2}\sigma^2(-\mu^{-2})' = \sigma^2\mu^{-4}$
- We assume $\mu \neq 0$ in order to apply the Continuous Mapping Theorem (CMT) underlying the Delta Method

1.2 Fundamental Concepts in Time-Series Analysis

1. Let us prove that $\mathbf{\Gamma}_j = \mathbf{\Gamma}'_{-j}$ under covariance stationarity:

Firstly:

$$\mathbf{\Gamma}_j \coloneqq \text{cov}(\boldsymbol{x}_i, \boldsymbol{x}_{i-j}) \qquad \mathbf{\Gamma}_{-j} \coloneqq \text{cov}(\boldsymbol{x}_{i+j}, \boldsymbol{x}_i)$$

Since $\{\boldsymbol{x}_i\}$ is weakly stationary, it follows that $\text{cov}(\boldsymbol{x}_i, \boldsymbol{x}_{i-j}) = \text{cov}(\boldsymbol{x}_{i+j}, \boldsymbol{x}_i)$

3. Let $\{x_i\}$ be a martingale with respect to $\{\boldsymbol{z}_i\}$:

$$\text{E}[x_i|\boldsymbol{z}_{i-1}, \boldsymbol{z}_{i-2}, \ldots, \boldsymbol{z}_1] = x_{i-1}$$

Let the information set $\mathbf{Z}$ contain $\{\boldsymbol{z}_{i-1}, \boldsymbol{z}_{i-2} \ldots, \boldsymbol{z}_1\}$ and $\mathbf{\Omega}$ contain $\{\boldsymbol{z}_i, \boldsymbol{z}_{i-1} \ldots, \boldsymbol{z}_1\}$
Clearly: $\mathbf{Z} \subset \mathbf{\Omega}$

$$\begin{aligned}
\text{E}[x_{i+j}|z_{i-1}, z_{i-2}, \ldots z_1] &= \text{E}\left[\text{E}\left[x_{i+j}|\boldsymbol{z}_i, \boldsymbol{z}_{i-1}, \ldots, \boldsymbol{z}_1\right]|\boldsymbol{z}_{i-1}, \boldsymbol{z}_{i-2}, \ldots, \boldsymbol{z}_1\right] \\
&= \text{E}[x_i|\boldsymbol{z}_{i-1}, \boldsymbol{z}_{i-2}, \ldots \boldsymbol{z}_1] \text{ (by LIE)} \\
&= x_{i-1} \text{(by definition)}
\end{aligned}$$

Where we applied the Law of Iterated Expectations (LIE) leveraging $\mathbf{Z} \subset \mathbf{\Omega}$

Now let us focus on $\text{E}[x_{i+j+1} - x_{i+j}|z_{i-1}, z_{i-2}, \ldots z_1]$, which can be rewritten as:

$$\begin{aligned}
&\text{E}[\text{E}[x_{i+j+1} - x_{i+j}|\boldsymbol{z}_{i+j}, \boldsymbol{z}_{i+j-1}, \ldots, \boldsymbol{z}_1]|\boldsymbol{z}_{i-1}, \boldsymbol{z}_{i-2}, \ldots, \boldsymbol{z}_1] \\
&= \text{E}[x_{i+j} - x_i|\boldsymbol{z}_{i-1}, \boldsymbol{z}_{i-2}, \ldots \boldsymbol{z}_1] \text{ (by LIE)} \\
&= 0 \text{ (following our results)}
\end{aligned}$$

4. Let $\{\varepsilon_i\} \overset{i.i.d}{\sim} WN(0, \sigma^2_\varepsilon)$ and $\{x_i\}$ be a sequence of real numbers changing over time:

i) $\{x_i\varepsilon_i\}$ is not independent and identically distributed as x_{i+j} depends on x_{i+j-1}

ii) $\text{cov}(x_i\varepsilon_i,\, x_{i-j}\varepsilon_{i-j}) = x_i x_{i-j}\text{cov}(\varepsilon_i, \varepsilon_{i-j}) = 0$ as $\{\varepsilon_i\}$ is i.i.d and $\{x_i\}$ is composed of constants

iii) $\{x_i\varepsilon_i\}$ is an m.d.s (martingale difference sequence):

For any $\{x_i\}$ an m.d.s is defined as:

$$\text{E}[x_i|x_{i-1}, x_{i-2}, \ldots, x_1] = 0$$

Thus, in our case and considering that $\mathrm{E}[Y|\mathcal{G}] = \mathrm{E}[\mathrm{E}[Y|\mathcal{F}]|\mathcal{G}]$ for information sets such that $\mathcal{F} \supset \mathcal{G}$ where $\mathcal{F} := \{x_i, x_{i-1}, \ldots, x_1, \varepsilon_{i-1}, \varepsilon_{i-2}, \ldots, \varepsilon_1\}$ and $\mathcal{G} := \{x_{i-1}\varepsilon_{i-1}, \ldots, x_1\varepsilon_1\}$

$$\begin{aligned}
\mathrm{E}[x_i\varepsilon_i|\mathcal{G}] &= \mathrm{E}[\mathrm{E}[x_i\varepsilon_i|x_i, x_{i-1}, \ldots, x_1, \varepsilon_{i-1}, \varepsilon_{i-2}, \ldots, \varepsilon_1]|x_{i-1}\varepsilon_{i-1}, \ldots, x_1\varepsilon_1] \\
&= \mathrm{E}[x_i\mathrm{E}[\varepsilon_i|x_i, x_{i-1}, \ldots, x_1, \varepsilon_{i-1}, \varepsilon_{i-2}, \ldots, \varepsilon_1]|x_{i-1}\varepsilon_{i-1}, \ldots, x_1\varepsilon_1] \\
&= 0 \ \big(\text{by linearity of conditional expectations and since } \{\varepsilon_i\} \text{ is i.i.d}\big)
\end{aligned}$$

iv) $\{x_i\varepsilon_i\}$ is clearly not stationary as its central second moment varies across i:

$$\mathrm{V}[x_i\varepsilon_i] = x_i^2\mathrm{V}[\varepsilon_i] = x_i^2\sigma^2$$

Where we have been assuming that the sequence of real numbers stemming from $\{x_i\}$ is not entirely stochastic once $x_i^{(t)}$ has been set.

5. Let a Random Walk $\{\boldsymbol{z}_i\}$ with $\{\boldsymbol{g}_i\} \overset{i.i.d}{\sim} WN(0, \boldsymbol{\sigma}_g^2)$ be expressed as:

$$\boldsymbol{z}_i = \boldsymbol{z}_{i-1} + \boldsymbol{g}_i$$

Note that

$$\begin{gathered}
\boldsymbol{z}_1 = \boldsymbol{z}_0 + \boldsymbol{g}_1 \\
\boldsymbol{z}_2 = \boldsymbol{z}_1 + \boldsymbol{g}_2 = \boldsymbol{z}_2 = \boldsymbol{z}_0 + \boldsymbol{g}_1 + \boldsymbol{g}_2 \\
\boldsymbol{z}_3 = \boldsymbol{z}_2 + \boldsymbol{g}_3 = \boldsymbol{z}_3 = \boldsymbol{z}_0 + \boldsymbol{g}_1 + \boldsymbol{g}_2 + \boldsymbol{g}_3 \\
\vdots \\
\boldsymbol{z}_i = \boldsymbol{z}_0 + \sum_{j=1}^{i} \boldsymbol{g}_j
\end{gathered}$$

Let us further assume that $\mathrm{E}[\boldsymbol{z}_0] = \omega$ and $\mathrm{V}[\boldsymbol{z}_0] = \omega^2$ where said central moments do not vary across i:

$$\begin{aligned}
\mathrm{V}[\boldsymbol{z}_i] = \mathrm{V}[\boldsymbol{z}_0] + \mathrm{V}\left[\sum_{j=1}^{i} \boldsymbol{g}_j\right] &= \omega^2 + \mathrm{cov}(z_0, g_j) + 2\sum_{j=1}^{n}\sum_{j\neq s}^{n}\mathrm{cov}(g_j, g_s) \\
&= \omega^2 + i\sigma_g^2 \ \big(\text{since } \{g_i\} \text{ is an independent white noise stochastic process}\big)
\end{aligned}$$

6. Let $\{\boldsymbol{z}_i\}$ be a martingale and $\boldsymbol{g}_i \coloneqq \boldsymbol{z}_i - \boldsymbol{z}_{i-1}$

Note that we can construct the sequence $\{\boldsymbol{g}_i\}$ from $\{\boldsymbol{z}_i\}$. Thus, by LIE:

$$\begin{aligned}
\mathrm{E}[\boldsymbol{z}_i - \boldsymbol{z}_{i-1}|\boldsymbol{g}_{i-1}, \boldsymbol{g}_{i-2}, \ldots, \boldsymbol{g}_1] &= \mathrm{E}[\mathrm{E}[\boldsymbol{z}_i - \boldsymbol{z}_{i-1}|\boldsymbol{z}_{i-1}, \boldsymbol{z}_{i-2}, \ldots, \boldsymbol{z}_1]|\boldsymbol{g}_{i-1}, \boldsymbol{g}_{i-2}, \ldots, \boldsymbol{g}_1] \\
&= \mathrm{E}[\mathrm{E}[\boldsymbol{z}_i|\boldsymbol{z}_{i-1}, \boldsymbol{z}_{i-2}, \ldots, \boldsymbol{z}_1] - \boldsymbol{z}_{i-1}|\boldsymbol{g}_{i-1}, \boldsymbol{g}_{i-2}, \ldots, \boldsymbol{g}_1] \\
&= \mathrm{E}[\boldsymbol{z}_{i-1} - \boldsymbol{z}_{i-1}|\boldsymbol{g}_{i-1}, \boldsymbol{g}_{i-2}, \ldots, \boldsymbol{g}_1] \\
&= 0 \;\; \big(\text{since } \mathrm{E}[\boldsymbol{z}_i|\boldsymbol{z}_{i-1}, \boldsymbol{z}_{i-2}, \ldots, \boldsymbol{z}_1] = \boldsymbol{z}_{i-1}\big)
\end{aligned}$$

7. Let $g_i \coloneqq \varepsilon_i\varepsilon_{i-1}$ where $\{\varepsilon_i\} \overset{i.i.d}{\sim} WN(0, \sigma_\varepsilon^2)\ \forall i$

It is easy to see that $\{g_i\}$ is not i.i.d since for the index vector $(i, j) = (2,\, 1)$:

$$\begin{aligned}
\mathrm{cov}\big(\varepsilon_i\varepsilon_{i-1}, \varepsilon_{i-j}\varepsilon_{i-j-1}\big) &= \mathrm{cov}\big(\varepsilon_2\varepsilon_1, \varepsilon_1\varepsilon_0\big) \\
&= \mathrm{cov}\big(\varepsilon_1, \varepsilon_1\big) \\
&= \sigma_\varepsilon^2
\end{aligned}$$

However, it can be shown to be an m.d.s noting that $\{g_{i-1}, \ldots, g_1\} \subset \{\varepsilon_{i-1}, \ldots, \varepsilon_1\}$:

$$\begin{aligned}
\mathrm{E}[\varepsilon_i\varepsilon_{i-1}|g_{i-1}, g_{i-2}, \ldots, g_1] &= \mathrm{E}[\mathrm{E}[\varepsilon_i\varepsilon_{i-1}|\varepsilon_{i-1}, \varepsilon_{i-2}, \ldots, \varepsilon_1]|g_{i-1}, g_{i-2}, \ldots, g_1] \\
&= \mathrm{E}[\mathrm{E}[\varepsilon_i|\varepsilon_{i-1}, \varepsilon_{i-2}, \ldots, \varepsilon_1]\varepsilon_{i-1}|g_{i-1}, g_{i-2}, \ldots, g_1] \\
&= 0 \;\big(\text{as } \{\varepsilon_i\} \text{ is i.i.d}\big)
\end{aligned}$$

8. Let $\{y_i\}$ be a stochastic process with $\mathrm{E}[y_i|y_{i-1}, y_{i-2}, \ldots, y_1] < \infty$ and define:

$$r_{i1} \coloneqq \mathrm{E}[y_i|y_{i-1}, y_{i-2}, \ldots, y_1] \text{ - } \mathrm{E}[y_i|y_{i-2}, y_{i-3}, \ldots, y_1]$$

We can prove $\{r_{i1}\}$ $(i \geq 2)$ is and m.d.s with respect to $\{y_i\}$

Firstly, let us present the relevant information set notation:

- Let $\{y_{i-1}, y_{i-2}, \ldots, y_1\}$ be denoted as $\mathcal{F}$
- Let $\{y_{i-2}, y_{i-3}, \ldots, y_1\}$ be denoted as $\mathcal{G}$
- Clearly: $F \supset G$

Thus

$$\begin{aligned}
\mathrm{E}[r_{i1}|y_{i-1}, y_{i-2}, \dots y_1] &= \mathrm{E}[\mathrm{E}[y_i|y_{i-1}, y_{i-2}, \dots y_1] - \mathrm{E}[y_i|y_{i-2}, y_{i-3}, \dots y_1]|\mathcal{F}] \\
&= \mathrm{E}[y_i|\mathcal{F}] - \mathrm{E}[\mathrm{E}[\mathrm{E}[y_i|\mathcal{F}]|\mathcal{G}]|\mathcal{F}] \text{ (applying LIE on the RHS)} \\
&= \mathrm{E}[y_i|\mathcal{F}] - \mathrm{E}[y_i|\mathcal{F}] \text{ (by LIE and linearity of conditional expectations)} \\
&= 0
\end{aligned}$$

9. Proof that the Ergodic Stationary Martingale Differences CLT (Billingsley) is more general than Lindeberg-Levy CLT:

Let us reproduce the assumptions of each CLT:

Ergodic Stationary Martingale Differences CLT

- The ergodic stationary stochastic process $\{\boldsymbol{g}_i\}$ is a vector m.d.s
- $\mathrm{E}[\boldsymbol{g}_i \boldsymbol{g}_i'] = \boldsymbol{\Sigma} < \infty$
- $\sqrt{n}(\bar{\boldsymbol{g}} - \mathrm{E}[\boldsymbol{g}_i]) \xrightarrow{d} N(\mathbf{0}, \boldsymbol{\Sigma})$

Lindeberg-Levy CLT

- The stochastic process $\{\boldsymbol{z}_i\}$ is i.i.d
- $\mathrm{E}[\boldsymbol{z}_i] = \boldsymbol{\mu}$ and $\mathrm{V}[\boldsymbol{z}_i] = \boldsymbol{\Sigma} < \infty$
- $\sqrt{n}(\bar{\boldsymbol{z}}_n - \boldsymbol{\mu}) \xrightarrow{d} N(\mathbf{0}, \boldsymbol{\Sigma})$

Let $\{\boldsymbol{x}_i\}$ be an i.i.d vector sequence with $\mathrm{E}[\boldsymbol{x}_i] = \boldsymbol{\mu}$ and $\mathrm{V}[\boldsymbol{x}_i] = \boldsymbol{\Sigma} < \infty$

Then the stochastic process $\{\bar{\boldsymbol{x}}_n - \boldsymbol{\mu}\}$ satisfies the following conditions:

$$\begin{aligned}
E[\bar{\boldsymbol{x}}_n - \boldsymbol{\mu}] &= E\left[\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{x}_i - \boldsymbol{\mu}\right] \\
&= \frac{1}{n}\sum_{i=1}^{n}\boldsymbol{\mu} - \boldsymbol{\mu} = 0 \ (\text{since } \mathrm{E}[\boldsymbol{x}_i] = \boldsymbol{\mu}\ \forall i)
\end{aligned}$$

$$= \frac{1}{n^2} \mathrm{V} \left[\sum_{i=1}^{n} \boldsymbol{x}_i + 2 \sum_{i=1}^{n} \sum_{i \neq j}^{n} \mathrm{Cov}(\boldsymbol{x}_i, \boldsymbol{x}_j) \right]$$
$$= \mathrm{n}^{-1} \boldsymbol{\Sigma} \ \big(\text{as } \mathrm{Cov}(\boldsymbol{x}_i, \boldsymbol{x}_j) = 0 \text{ for } \forall i \neq j \text{ since } \boldsymbol{x}_i \text{ is i.i.d}\big)$$

Consequently, said stochastic process is covariance stationary as its central tendency moments do not vary across i.

Since $\{\boldsymbol{x}_i\}$ is i.i.d with $\mathrm{E}[\boldsymbol{x}_i] = \boldsymbol{\mu}$ then $\{\boldsymbol{g}_i\}$ is an m.d.s where $\boldsymbol{g}_i \coloneqq \bar{\boldsymbol{x}}_n - \boldsymbol{\mu}$ as $\mathrm{E}[\bar{\boldsymbol{x}}_i - \boldsymbol{\mu} | \boldsymbol{g}_{i-1}, \boldsymbol{g}_{i-2}, \ldots, \boldsymbol{g}_1] = 0$

Combining all these facts it is easy to see that $\sqrt{n}(\bar{\boldsymbol{x}}_n - \boldsymbol{\mu}) \xrightarrow{d} N(\mathbf{0}, \boldsymbol{\Sigma})$

Thus we have proved that an i.i.d sequence with finite central tendency moments is an ergodic stationary m.d.s and so Lindeberg-Levy CLT is stronger than Billingsley CLT.

1.3 Large Sample Distribution of the OLS Estimator

1. Proof that supposing $\mathrm{E}[y_i|\boldsymbol{x}_i] = \boldsymbol{x}_i'\boldsymbol{\beta}$ implies that $\boldsymbol{x}_i$ is orthogonal to ε_i:

Note that by definition of a regression model: $\varepsilon_i = y_i - \mathrm{E}[y_i|\boldsymbol{x}_i]$, so:

$$\mathrm{E}[\varepsilon_i|\boldsymbol{x}_i] = \mathrm{E}[y_i|\boldsymbol{x}_i] - \mathrm{E}[\boldsymbol{x}_i'\boldsymbol{\beta}|\boldsymbol{x}_i] \text{ (taking conditional expectations on both sides)}$$
$$\mathrm{E}[\varepsilon_i|\boldsymbol{x}_i] = \boldsymbol{x}_i'\boldsymbol{\beta} - \boldsymbol{x}_i'\boldsymbol{\beta} \text{ (by assumption and linearity of conditional expectations)}$$
$$\mathrm{E}[\varepsilon_i|\boldsymbol{x}_i] = 0$$

Thus

$$\begin{aligned}
\mathrm{E}[\boldsymbol{x}_i\varepsilon_i] &= \mathrm{E}[\mathrm{E}[\boldsymbol{x}_i\varepsilon_i|\boldsymbol{x}_i]] \text{ (by LIE)} \\
&= \mathrm{E}[\boldsymbol{x}_i\mathrm{E}[\varepsilon_i|\boldsymbol{x}_i]] \text{ (by linearity of conditional expectations)} \\
&= 0 \ \big(\text{as } \mathrm{E}[\varepsilon_i|\boldsymbol{x}_i] = 0\big)
\end{aligned}$$

2. Only when one of the regressors include a constant (namely, $x_{i1} = 1\,\forall i$) can we assure that $\mathrm{E}[\varepsilon_i^2] < \infty$. So linearity, ergodic stationarity or even orthogonality of regressors to the error term are not sufficient conditions.

Let $\boldsymbol{g}_i \coloneqq \boldsymbol{x}_i \varepsilon_i$ it is easy to see

$$\begin{aligned}
\mathrm{E}[\boldsymbol{g}_i \boldsymbol{g}_i'] &= \mathrm{E}[\boldsymbol{x}_i \varepsilon_i (\boldsymbol{x}_i \varepsilon_i)'] \\
&= \mathrm{E}[\boldsymbol{x}_i \varepsilon_i \varepsilon_i' \boldsymbol{x}_i'] \ \big(\text{since } (\boldsymbol{AB})' = \boldsymbol{B}'\boldsymbol{A}'\big) \\
&= \mathrm{E}[\varepsilon_i^2 \boldsymbol{x}_i \boldsymbol{x}_i'] \ \big(\text{since } \varepsilon_i \text{ is a scalar}\big)
\end{aligned}$$

Then (noting that $x_{i1} = 1\ \forall i$):

$$\mathrm{E}[\varepsilon_i^2 \boldsymbol{x}_i \boldsymbol{x}_i'] = \begin{bmatrix}
\mathrm{E}[\varepsilon_i^2] & \mathrm{E}[\varepsilon_i^2 \boldsymbol{x}_{i2}'] & \cdots & \mathrm{E}[\varepsilon_i^2 \boldsymbol{x}_{ik}'] \\
\mathrm{E}[\varepsilon_i^2 \boldsymbol{x}_{i2}'] & \mathrm{E}[\varepsilon_i^2 \boldsymbol{x}_{i2}^2] & \cdots & \mathrm{E}[\varepsilon_i^2 \boldsymbol{x}_{i2} \boldsymbol{x}_{ik}'] \\
\vdots & \vdots & \ddots & \vdots \\
\mathrm{E}[\varepsilon_i^2 \boldsymbol{x}_{ik}'] & \mathrm{E}[\varepsilon_i^2 \boldsymbol{x}_{ik} \boldsymbol{x}_{i2}'] & \cdots & \mathrm{E}[\varepsilon_i^2 \boldsymbol{x}_{ik}^{2\,\prime}]
\end{bmatrix}$$

Assuming that $\mathrm{E}[\boldsymbol{g}_i \boldsymbol{g}_i'] < \infty$ then $\mathrm{E}[\varepsilon_i^2] < \infty$

Moreover, if $\{y_i, \boldsymbol{x}_i\}$ is jointly ergodic stationary, so is $\{\varepsilon_i\}$, as it is a function of the former ($\varepsilon_i = y_i - \boldsymbol{x}_i'\beta$).

Thus, the second central tendency moment $\mathrm{E}[\varepsilon_i^2] < \infty$ and is constant across i.

3. Let $f(\boldsymbol{x}_i) \coloneqq \mathrm{E}[\varepsilon_i^2 | \boldsymbol{x}_i]$, it can be proved that $\boldsymbol{S}$ ($\coloneqq \mathrm{E}[\varepsilon_i^2 \boldsymbol{x}_i \boldsymbol{x}_i']$) can be written as

$$\boldsymbol{S} = \mathrm{E}[f(\boldsymbol{x}_i)\boldsymbol{x}_i \boldsymbol{x}_i']$$

The proof is as follows:

$$\begin{aligned}
\boldsymbol{S} &= \mathrm{E}[f(\boldsymbol{x}_i)\boldsymbol{x}_i \boldsymbol{x}_i'] \ \big(\text{by definition}\big) \\
&= \mathrm{E}[\mathrm{E}[\varepsilon_i^2 \boldsymbol{x}_i \boldsymbol{x}_i' | \boldsymbol{x}_i]] \ \big(\text{by LIE}\big) \\
&= \mathrm{E}[\mathrm{E}[\varepsilon_i^2\ | \boldsymbol{x}_i] \boldsymbol{x}_i \boldsymbol{x}_i'] \ \big(\text{since } \boldsymbol{x}_i \boldsymbol{x}_i' \text{ is a function of } \boldsymbol{x}_i\big) \\
&= E[f(\boldsymbol{x}_i)\boldsymbol{x}_i \boldsymbol{x}_i']
\end{aligned}$$

4. Given the following key assumptions (A.1: Linearity, A.2: Ergodic Stationarity, A.3: Orthogonality Condition, A.4: Rank Condition, A.5: $\{\boldsymbol{g}_i\}$ is an m.d.s) the error variance can be consistently estimated, namely:

$$S^2 := \frac{1}{n-k}\sum_{i=1}^{n}\hat{\varepsilon}_i^2 \xrightarrow{p} \mathrm{E}[\varepsilon_i^2]$$

Notice that by multiplying and dividing by the sample size we can rewrite the expression above as:

$$S^2 = \frac{n}{n-k}\left(\frac{1}{n}\sum_{i=1}^{n}\hat{\varepsilon}_i^2\right)$$

Where applying properties of the limiting behavior of variables it is easy to see:

$$\lim_{n\to\infty}\frac{n}{n-k} = 1 \implies \operatorname*{plim}_{n\to\infty}\frac{n}{n-k} = 1$$

For the other part of the term note that A.1 (linearity) ensures that residuals can be expressed as:

$$\begin{aligned}
\frac{1}{n}\sum_{i=1}^{n}\hat{\varepsilon}_i^2 &= \frac{1}{n}\sum_{i=1}^{n}(y_i - \boldsymbol{x}_i'\mathbf{b})^2 \\
&= \frac{1}{n}\sum_{i=1}^{n}(\boldsymbol{x}_i'\boldsymbol{\beta} + \varepsilon_i - \boldsymbol{x}_i'\mathbf{b})^2 \quad (\text{since by A.1 } y_i = \boldsymbol{x}_i'\boldsymbol{\beta} + \varepsilon_i) \\
&= \frac{1}{n}\sum_{i=1}^{n}(\varepsilon_i - \boldsymbol{x}_i'(\mathbf{b} - \boldsymbol{\beta}))^2 \\
&= \frac{1}{n}\sum_{i=1}^{n}\varepsilon_i^2 - 2\frac{1}{n}\sum_{i=1}^{n}\varepsilon_i\boldsymbol{x}_i'(\mathbf{b} - \boldsymbol{\beta}) + (\mathbf{b} - \boldsymbol{\beta})'\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{x}_i\boldsymbol{x}_i'(\mathbf{b} - \boldsymbol{\beta})
\end{aligned}$$

Where $\mathbf{b}$ is the K-dimensional vector containing OLS estimators for each population parameter in our linear regression model. Note that A.1 - A.4 along with Ergodic LLN ensure that $\mathbf{b} \xrightarrow{p} \boldsymbol{\beta}$, so clearly $\mathbf{b}$ - $\boldsymbol{\beta} \xrightarrow{p} 0$

- Since $\mathbf{b}$ - $\boldsymbol{\beta} \xrightarrow{p} 0$, by Continuous Mapping Theorem (CMT): $(\mathbf{b}$ - $\boldsymbol{\beta})' \xrightarrow{p} 0$

  By A.4 (Rank condition) the KxK full column rank matrix of moments $\mathrm{E}[\boldsymbol{x}_i\boldsymbol{x}_i']$ exists and is finite.

Also, $\{\boldsymbol{x}_i\boldsymbol{x}_i'\}$ is ergodic stationary as it is a function of $\{\boldsymbol{x}_i\}$, which in turn is an ergodic stationary stochastic process by A.2 (Ergodic Stationary).

Thus, by Ergodic LLN:

$$\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{x}_i\boldsymbol{x}_i' \xrightarrow{p} \mathrm{E}[\boldsymbol{x}_i\boldsymbol{x}_i'] < \infty$$

Consequently, applying Slutzky's Theorem (ST), the last term vanishes, as all limits exist and are finite:

$$(\mathbf{b}-\boldsymbol{\beta})'\ \frac{1}{n}\sum_{i=1}^{n}\boldsymbol{x}_i\boldsymbol{x}_i'(\mathbf{b}-\boldsymbol{\beta}) \xrightarrow{p} 0$$

- Let $\boldsymbol{g}_i \coloneqq \boldsymbol{x}_i\varepsilon_i$ since according to A.5 it is an m.d.s, it follows that $\mathrm{E}[\boldsymbol{g}_i] = 0$ by LIE.

  Also, $\{\boldsymbol{g}_i\}$ is ergodic stationary as it is a function of $\{\varepsilon_i, \boldsymbol{x}_i\}$ which is jointly ergodic stationary by A.2 (ε_i satisfies this condition as it is a linear function of y_i by A.1).

  Thus, by Ergodic LLN:

$$\frac{1}{n}\sum_{i=1}^{n}\varepsilon_i\boldsymbol{x}_i \xrightarrow{p} 0$$

  Consequently, by ST, the middle term also vanishes, as all limits exist and are finite:

$$\frac{1}{n}\sum_{i=1}^{n}\varepsilon_i\boldsymbol{x}_i(\mathbf{b}-\boldsymbol{\beta}) \xrightarrow{p} 0$$

- Lastly, either introducing an additional assumption that $\mathrm{E}[\varepsilon_i^2] < \infty$ or leveraging the fact that an intercept is introduced ($x_{i1} = 1$, $\forall i$) as virtually all econometric applications, then the second moment of the error term exists and is finite.

  Also, $\{\varepsilon_i\}$ is ergodic stationary as it is a function of $\{y_i\}$ by A.1, which in turn is ergodic stationary by A.2.

  Thus, by Ergodic LLN:

$$\frac{1}{n}\sum_{i=1}^{n}\hat{\varepsilon}_i^2 \xrightarrow{p} \mathrm{E}[\varepsilon_i^2]$$

So clearly, by ST:

$$\frac{n}{n-k}\left(\frac{1}{n}\sum_{i=1}^{n}\hat{\varepsilon}_i^2\right) \xrightarrow{p} \mathrm{E}[\varepsilon_i^2]$$

And so we conclude that $S^2 \xrightarrow{p} \mathrm{E}[\varepsilon_i^2]$

5. We can go further than the previous exercise and prove that S^2 is consistent with any $\hat{\boldsymbol{\beta}} \xrightarrow{p} \boldsymbol{\beta}$ and relaxing the orthogonality condition. In other words, said assumption now becomes A.3': $\mathrm{E}[\boldsymbol{x}_i\varepsilon_i] \neq 0$ but still $< \infty$

$$\begin{aligned}
\frac{1}{n}\sum_{i=1}^{n}\hat{\varepsilon}_i^2 &= \frac{1}{n}\sum_{i=1}^{n}(y_i - \boldsymbol{x}_i'\hat{\boldsymbol{\beta}})^2 \\
&= \frac{1}{n}\sum_{i=1}^{n}(\boldsymbol{x}_i'\boldsymbol{\beta} + \varepsilon_i - \boldsymbol{x}_i'\hat{\boldsymbol{\beta}})^2 \quad \text{(since by A.1 } y_i = \boldsymbol{x}_i'\boldsymbol{\beta} + \varepsilon_i\text{)} \\
&= \frac{1}{n}\sum_{i=1}^{n}(\varepsilon_i - \boldsymbol{x}_i'(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}))^2 \\
&= \frac{1}{n}\sum_{i=1}^{n}\varepsilon_i^2 - 2\frac{1}{n}\sum_{i=1}^{n}\varepsilon_i\boldsymbol{x}_i'(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) + (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})'\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{x}_i\boldsymbol{x}_i'(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})
\end{aligned}$$

- Since $\hat{\boldsymbol{\beta}}$ - $\boldsymbol{\beta} \xrightarrow{p} 0$, by Continuous Mapping Theorem (CMT): $(\hat{\boldsymbol{\beta}}$ - $\boldsymbol{\beta})' \xrightarrow{p} 0$

  By A.4 (Rank condition) the KxK full column rank matrix of moments $\mathrm{E}[\boldsymbol{x}_i\boldsymbol{x}_i']$ exists and is finite.

  Also, $\{\boldsymbol{x}_i\boldsymbol{x}_i'\}$ is ergodic stationary as it is a function of $\{\boldsymbol{x}_i\}$, which in turn is an ergodic stationary stochastic process by A.2 (Ergodic Stationary).

  Thus, by Ergodic LLN:

  $$\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{x}_i\boldsymbol{x}_i' \xrightarrow{p} \mathrm{E}[\boldsymbol{x}_i\boldsymbol{x}_i'] < \infty$$

  Consequently, applying Slutzky's Theorem (ST), the last term vanishes, as all limits exist and are finite:

$$(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})' \ \frac{1}{n}\sum_{i=1}^{n}\boldsymbol{x}_i\boldsymbol{x}_i'(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \xrightarrow{p} 0$$

- Let $\boldsymbol{g}_i := \boldsymbol{x}_i\varepsilon_i$ and recall that according to A.3' $\mathrm{E}[\boldsymbol{g}_i] < \infty$

  Also, $\{\boldsymbol{g}_i\}$ is ergodic stationary as it is a function of $\{\varepsilon_i, \boldsymbol{x}_i\}$ which is jointly ergodic stationary by A.2 (ε_i satisfies this condition as it is a linear function of y_i by A.1).

  Thus, by Ergodic LLN:

  $$\frac{1}{n}\sum_{i=1}^{n}\varepsilon_i\boldsymbol{x}_i \xrightarrow{p} \mathrm{E}[\boldsymbol{x}_i\varepsilon_i] < \infty$$

  Consequently, by ST, the middle term also vanishes, as all limits exist and are finite:

  $$\frac{1}{n}\sum_{i=1}^{n}\varepsilon_i\boldsymbol{x}_i(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \xrightarrow{p} 0$$

- Lastly, either introducing an additional assumption that $\mathrm{E}[\varepsilon_i^2] < \infty$ or leveraging the fact that an intercept is introduced ($x_{i1} = 1$, $\forall i$) as virtually all econometric applications, then the second moment of the error term exists and is finite.

  Also, $\{\varepsilon_i\}$ is ergodic stationary as it is a function of $\{y_i\}$ by A.1, which in turn is ergodic stationary by A.2.

  Thus, by Ergodic LLN:

  $$\frac{1}{n}\sum_{i=1}^{n}\hat{\varepsilon}_i^2 \xrightarrow{p} \mathrm{E}[\varepsilon_i^2]$$

So clearly, by ST:

$$\frac{n}{n-k}\left(\frac{1}{n}\sum_{i=1}^{n}\hat{\varepsilon}_i^2\right) \xrightarrow{p} \mathrm{E}[\varepsilon_i^2]$$

And so we conclude that $S^2 \xrightarrow{p} \mathrm{E}[\varepsilon_i^2]$ for consistent estimators different from OLS.

1.4 Hypothesis Testing

1. Proof that the Heteroskedasticity-consistent standard error or White's standard error vanishes as the sample size increases:

Firstly, let us present the notation for said standard error:

$$SE^*(b_k) \coloneqq \sqrt{\frac{1}{n}(\mathbf{S}_{\mathbf{xx}}^{-1}\hat{\mathbf{S}}\mathbf{S}_{\mathbf{xx}}^{-1})_{kk}}$$

Then, note the following facts:

- $\mathbf{S}_{\mathbf{xx}} \coloneqq n^{-1}\mathbf{X}'\mathbf{X}$
- $\lim\limits_{n\to\infty} \dfrac{1}{n} = 0 \implies \operatorname*{plim}\limits_{n\to\infty} \dfrac{1}{n} = 0$
- $\hat{\mathbf{S}} \xrightarrow{p} \mathbf{S}$ where $\mathbf{S} \coloneqq \mathrm{E}[\varepsilon^2 \boldsymbol{x}_i \boldsymbol{x}_i']$ and $\hat{\mathbf{S}} \coloneqq \dfrac{1}{n}\sum\limits_{i=1}^{n} \hat{\varepsilon}_i^2 \boldsymbol{x}_i \boldsymbol{x}_i'$
- The process $\{\boldsymbol{x}_i \boldsymbol{x}_i'\}$ is a function of $\{\boldsymbol{x}_i\}$ which is ergodic stationary by A.2 (so the former also satisfies this condition).

  Furthermore, by A.4 $\mathrm{E}[\boldsymbol{x}_i \boldsymbol{x}_i'] = \boldsymbol{\Sigma}_{\boldsymbol{xx}}$ which is a KxK full column matrix that exists and is finite. Thus, it is also nonsingular and Positive Definite, entailing it can be inverted. Clearly, as the sample size increases $\mathbf{S}_{\mathbf{xx}}$ is also Positive Definite and nonsingular, hence invertible.

  By Ergodic LLN and Continuous Mapping Theorem:

$$\left(\frac{1}{n}\sum_{i=1}^{n} \boldsymbol{x}_i \boldsymbol{x}_i'\right)^{-1} \xrightarrow{p} \boldsymbol{\Sigma}_{\boldsymbol{xx}}^{-\mathbf{1}} < \infty$$

Then, by ST:

$$\left(\frac{1}{n}\sum_{i=1}^{n} \boldsymbol{x}_i \boldsymbol{x}_i'\right)^{-1} \frac{1}{n}\sum_{i=1}^{n} \hat{\varepsilon}_i^2 \boldsymbol{x}_i \boldsymbol{x}_i' \left(\frac{1}{n}\sum_{i=1}^{n} \boldsymbol{x}_i \boldsymbol{x}_i'\right)^{-1} \xrightarrow{p} \boldsymbol{\Sigma}_{\boldsymbol{xx}}^{-\mathbf{1}} \mathbf{S} \boldsymbol{\Sigma}_{\boldsymbol{xx}}^{-\mathbf{1}} < \infty$$

Since this is a finite limit, adding the remaining part pointed out by the second bullet point and applying Continuous Mapping Theorem to the square root expression results in:

$$\sqrt{\tfrac{1}{n}(\mathbf{S}_{\mathbf{xx}}^{-1}\hat{\mathbf{S}}\mathbf{S}_{\mathbf{xx}}^{-1})_{kk}} \xrightarrow{p} 0$$

2 . Let $K = 1$ and let b be the consistent OLS estimate of β.

Furthermore, let $\text{SE}(b) = \sqrt{\widehat{\text{Avar}(b)}/n}$ and suppose $\lambda = -\ln\beta$ so a natural estimator implied by OLS is $\hat{\lambda} = -\ln b$. Let us get $\text{SE}(\hat{\lambda})$

Note that $\hat{\lambda} = \text{a}(b)$ where $\text{a}(b)$ = -$\ln b$

Clearly, if $b > 0$ then by Slutzky´s Theorem $\text{a}(b) \xrightarrow{p} \text{a}(\beta)$

Since b is consistent and supposing the rest of relevant assumptions hold, by CLT: $\sqrt{n}(b-\beta) \xrightarrow{d} N(0, Avar(b))$ so applying Delta Method:

$$\sqrt{n}(\text{a}(b) - \text{a}(\beta)) \xrightarrow{d} N(0, \text{A}(\beta)Avar(b)\text{A}(\beta)')$$

Where $\text{A}(\beta) \coloneqq \dfrac{\partial \text{a}(\beta)}{\partial \beta'} = -\dfrac{1}{\beta}$

Thus, $\widehat{\text{Avar}(\hat{\lambda})} = \dfrac{1}{b^2}\widehat{\text{Avar}(b)}$

Consequently, $\text{SE}(\hat{\lambda}) = \dfrac{1}{b}\sqrt{\widehat{\text{Avar}(b)}}$

3. The asymptotic distribution of W (Wald Statistic) remains unaffected by different choices of $\mathbf{R}$ and $\mathbf{r}$ as there is no unique way to write the linear hypothesis $\mathbf{R}\boldsymbol{\beta} = \mathbf{r}$. For any $\#\mathbf{r}x\#\mathbf{r}$ nonsingular matrix $\mathbf{F}$, the same set of restrictions can be represented as $\tilde{\mathbf{R}}\boldsymbol{\beta} = \tilde{\mathbf{r}}$ where $\tilde{\mathbf{R}} \coloneqq \mathbf{FR}$ and $\tilde{\mathbf{r}} \coloneqq \mathbf{Fr}$

Note that $\tilde{\mathbf{R}}\mathbf{b} - \tilde{\mathbf{r}} = \tilde{\mathbf{F}}(\mathbf{Rb} - \mathbf{r})$ and since the asymptotic distribution of W can be expressed as $\sqrt{n}(\mathbf{Rb} - \mathbf{r})'[\mathbf{R}\widehat{\text{Avar}(b)}\mathbf{R}']^{-1}(\mathbf{Rb} - \mathbf{r}) \xrightarrow{d} \chi^2(\#\mathbf{r})$, by ST:

$$\sqrt{n}(\tilde{\mathbf{R}}\mathbf{b} - \tilde{\mathbf{r}})' \left[\tilde{\mathbf{R}}\widehat{\text{Avar}(b)}\tilde{\mathbf{R}}'\right]^{-1} (\tilde{\mathbf{R}}\mathbf{b} - \tilde{\mathbf{r}}) \xrightarrow{d} \chi^2(\#\mathbf{r})$$

Since:

$$\begin{aligned}
\dim(\tilde{\mathbf{R}}\widehat{\text{Avar}(b)}\tilde{\mathbf{R}}') &= (\#\mathbf{r}\text{x}\#\mathbf{r})(\#\mathbf{r}\text{x}\mathbf{K})(\mathbf{K}\text{x}\mathbf{K})(\mathbf{K}\text{x}\#\mathbf{r})(\#\mathbf{r}\text{x}\#\mathbf{r}) \\
&= (\#\mathbf{r}\text{x}\mathbf{K})(\mathbf{K}\text{x}\#\mathbf{r}) \\
&= (\#\mathbf{r}\text{x}\#\mathbf{r})
\end{aligned}$$

Where $\dim(\cdot)$ is a linear operator that retrieves the dimension of any data matrix or vector. It is based on the function from Programming Language **R**. Think of numpy.shape() in the case of **Python** or size() in **MATLAB**.

1.5 Estimating $\mathrm{E}[\varepsilon_i^2 \boldsymbol{x}_i \boldsymbol{x}_i']$ Consistently

1. Leveraging the OLS estimator properties, sample moments need to be computed just once to obtain relevant statistics. Let us show that it is also the case for $\hat{\mathbf{S}}$ (which is the estimator of $\mathbf{S}$ defined ***here***).

Let $\hat{\mathbf{S}} \coloneqq \dfrac{1}{n}\displaystyle\sum_{i=1}^{n} \hat{\varepsilon}_i{}^2 \boldsymbol{x}_i \boldsymbol{x}_i'$ and $\mathrm{SE}^*(\mathbf{b}_k) \coloneqq \sqrt{\frac{1}{n}(\mathbf{S}_{\mathbf{xx}}^{-1}\hat{\mathbf{S}}\mathbf{S}_{\mathbf{xx}}^{-1})_{kk}}$

$$\mathbf{S}_{\mathbf{xx}}^{-1} = \left(\frac{1}{n}\mathbf{X}'\mathbf{X}\right)^{-1} = \left(\frac{1}{n}\sum_{i=1}^{n} \boldsymbol{x}_i \boldsymbol{x}_i'\right)^{-1}$$

$$\begin{aligned}
\hat{\varepsilon}_i^2 &= (y_i - \boldsymbol{x}_i'\mathbf{b})^2 \\
&= y_i^2 + (\boldsymbol{x}_i'\mathbf{b})'(\boldsymbol{x}_i'\mathbf{b}) - 2y_i\boldsymbol{x}_i'\mathbf{b}
\end{aligned}$$

$$\frac{1}{n}\sum_{i=1}^{n} \hat{\varepsilon}_i^2 = \frac{1}{n}\boldsymbol{y}'\boldsymbol{y} + \left(\mathbf{S}_{\mathbf{xx}}^{-1}\mathbf{S}_{\mathbf{xy}}\right)' \frac{1}{n}\sum_{i=1}^{n} \boldsymbol{x}_i \boldsymbol{x}_i' \left(\mathbf{S}_{\mathbf{xx}}^{-1}\mathbf{S}_{\mathbf{xy}}\right) - 2\left(\mathbf{S}_{\mathbf{xx}}^{-1}\mathbf{S}_{\mathbf{xy}}\right)\sum_{i=1}^{n} \boldsymbol{x}_i y_i$$

So computing sample moments is only required once.

2. Let us compare the variance of the GLS estimator and the data matrix representation of the asymptotic variance of the OLS estimator. Expressions are as follows:

$$\mathrm{V}[\hat{\boldsymbol{\beta}}_{\mathrm{GLS}}|\mathbf{X}] = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'(\sigma^2\mathbf{V})\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}$$
$$\widehat{\mathrm{Avar}(\mathbf{b})} = n \cdot (\mathbf{X}'\mathbf{X})^{-1}(\mathbf{X}'\mathbf{B}\mathbf{X})(\mathbf{X}'\mathbf{X})^{-1}$$

$$\mathbf{B} \coloneqq \begin{bmatrix} \hat{\varepsilon}_1^2 & & \\ & \ddots & \\ & & \hat{\varepsilon}_n^2 \end{bmatrix}$$

- The sample size is multiplying $\widehat{\mathrm{Avar}(\mathbf{b})}$
- We ditch out $\mathbf{V}$ for a feasible version
- Clearly data matrix $\mathbf{B}$ is not present in the first expression

1.6 Implications of Conditional Homoskedasticity

1. Without assuming Conditional Homoskedasticity (CH): $\mathrm{E}[\varepsilon_i^2|\mathbf{X}] \neq \sigma^2 \ \forall i$, $\mathrm{Avar}(\mathbf{b}) = \mathbf{\Sigma}_{\mathbf{xx}}^{-1}\mathbf{S}\mathbf{\Sigma}_{\mathbf{xx}}^{-1}$ so in order to show how restricting CH is, let us prove that the asymptotic variance estimator under CH $\widehat{\mathrm{Avar}(\mathbf{b})} = S^2\mathbf{S}_{\mathbf{xx}}^{-1}$ does not converge to the expression given by $\mathrm{Avar}(\mathbf{b})$ above, where $\mathbf{S} = \mathrm{E}[\varepsilon_i^2\boldsymbol{x}_i\boldsymbol{x}_i']$ and $S^2 = \dfrac{1}{n-k}\displaystyle\sum_{i=1}^{n}\hat{\varepsilon}_i^2$

Note that by definition of $\mathbf{S}$:

$$\begin{aligned}
\mathbf{S} &= \mathrm{E}[\mathrm{E}[\varepsilon_i^2\boldsymbol{x}_i\boldsymbol{x}_i'|\mathbf{X}]] \ \ \text{(by LIE)} \\
&= \mathrm{E}[\mathrm{E}[\varepsilon_i^2|\mathbf{X}]\boldsymbol{x}_i\boldsymbol{x}_i'] \ \ \left(\text{as we condition on } \mathbf{X}\right) \\
&= \mathrm{E}[\sigma_i^2\boldsymbol{x}_i\boldsymbol{x}_i'] \\
&= \sigma_i^2\mathbf{\Sigma}_{\mathbf{xx}}
\end{aligned}$$

Thus:

$$\begin{aligned}
\mathrm{Avar}(\mathbf{b}) &= \mathbf{\Sigma}_{\mathbf{xx}}^{-1}\left(\sigma_i^2\mathbf{\Sigma}_{\mathbf{xx}}\right)\mathbf{\Sigma}_{\mathbf{xx}}^{-1} \\
&= \sigma_i^2\mathbf{\Sigma}_{\mathbf{xx}}^{-1}
\end{aligned}$$

Since $S^2\mathbf{S}_{\mathbf{xx}}^{-1} \xrightarrow{p} \sigma^2\mathbf{\Sigma}_{\mathbf{xx}}^{-1}$ by Slutzky's Theorem but σ_i^2 varies across i, clearly this entails: $\widehat{\mathrm{Avar}(\mathbf{b})} \overset{p}{\not\to} \mathrm{Avar}(\mathbf{b})$

Also, since $\mathbf{b} \xrightarrow{p} \boldsymbol{\beta}$ and $\{\boldsymbol{x}_i, y_i\}$ is ergodic stationary, by Lindeberg-Levy CLT:

$$\sqrt{n}(\mathbf{b}-\boldsymbol{\beta}) \xrightarrow{d} N(0, \mathrm{Avar}(\mathbf{b}))$$

Note that the t-ratio is defined as:

$$t_k = \frac{\sqrt{n}(\mathbf{b}_k - \boldsymbol{\beta}_k)}{\sqrt{S^2(\mathbf{S}_{\mathbf{xx}}^{-1})_{kk}}}$$

$t_k \overset{d}{\not\to} N(0,1)$ by Slutzky's Theorem since $\widehat{\mathrm{Avar}(\mathbf{b})} \overset{p}{\not\to} \mathrm{Avar}(\mathbf{b})$ as we showed.

2. If CH assumption holds then $S^2\mathbf{S}_{\mathbf{xx}}^{-1}$ is less computationally expensive than its Heteroskedasticity-consistent counterpart $\mathbf{S}_{\mathbf{xx}}^{-1}\hat{\mathbf{S}}\mathbf{S}_{\mathbf{xx}}^{-1}$. Moreover, the number of parameters to be estimated is reduced under the truth of CH: from two $(\sigma^2, \mathbf{\Sigma}_{\mathbf{xx}}^{-1})$ to three.

3. Since $F(\#\mathbf{r}, n-K) \xrightarrow{d} \chi^2(\#\mathbf{r})$ then at the limit (n - k $\to \infty$) which translates to identical critical values at a given α significance level: $\Pr(F \leq F_\alpha) = \Pr(\chi^2 \leq \chi^2_\alpha)$ for $\#\mathbf{r}$ Degrees of Freedom.

4. Let us prove that without CH the F statistic is not asymptitically distributed as $\chi^2(\#\mathbf{r})$

Firstly, let us show a useful identiy:

$$\frac{SSR_r - SSR_u}{S^2}$$

$$SSR_r = (\mathbf{Y} - \mathbf{X}\mathbf{b} + \mathbf{X}(\mathbf{b} - \hat{\boldsymbol{\beta}}))'(\mathbf{Y} - \mathbf{X}\mathbf{b} + \mathbf{X}(\mathbf{b} - \hat{\boldsymbol{\beta}}))$$

$$\hat{\boldsymbol{\beta}} = \mathbf{b} - (\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}'[\mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}']^{-1}(\mathbf{R}\mathbf{b} - \mathbf{r})$$

$$SSR_u = (\mathbf{Y} - \mathbf{X}\mathbf{b})'(\mathbf{Y} - \mathbf{X}\mathbf{b})$$

So:

$$SSR_r - SSR_u = (\mathbf{R}\mathbf{b} - \mathbf{r})'[\mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}']^{-1}(\mathbf{R}\mathbf{b} - \mathbf{r})$$

$$\frac{SSR_r - SSR_u}{S^2} = (\mathbf{R}\mathbf{b} - \mathbf{r})'[\mathbf{R}S^2(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}']^{-1}(\mathbf{R}\mathbf{b} - \mathbf{r})$$

Note how this last equation if the very definition of the F statistic!

Notice that $\sqrt{n}(\mathbf{R}\mathbf{b} - \mathbf{r}) \xrightarrow{d} N(0, \mathbf{R}\text{Avar}(\mathbf{b})\mathbf{R}')$

Where Avar(b) $= \boldsymbol{\Sigma}_{\mathbf{xx}}^{-1}\mathbf{S}\boldsymbol{\Sigma}_{\mathbf{xx}}^{-1}$ (if $\text{E}[\varepsilon_i^2|\mathbf{X}] = \sigma_i^2$)

So clearly $\sqrt{n}(\mathbf{R}\mathbf{b} - \mathbf{r})[\mathbf{R}S^2(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}']^{-1/2} \overset{d}{\nrightarrow} N(0,1)$

Since $[\mathbf{R}S^2(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}'] \xrightarrow{p} \mathbf{R}\sigma^2\boldsymbol{\Sigma}_{\mathbf{xx}}^{-1}\mathbf{R}'$ which does not equal $\mathbf{R}\boldsymbol{\Sigma}_{\mathbf{xx}}^{-1}\mathbf{S}\boldsymbol{\Sigma}_{\mathbf{xx}}^{-1}\mathbf{R}'$

It is imperative to be aware of the meaning of certain notation used throughout this exercise:

- SSR_r is the Sum of Squared Residuals from the restricted regression under the truth of the null hypothesis
- SSR_u is the Sum of Squared Residuals from the unrestricted regression ($\mathbf{e}$'$\mathbf{e}$)
- $\hat{\boldsymbol{\beta}}$ is the OLS estimator from the restricted regression

5. Under CH, let us prove that $nR^2 \xrightarrow{d} \chi^2(K-1)$

The algebraic relationship between the F-ratio for the null hypothesis and R^2 is:

$$\begin{aligned} F &= \frac{R^2/(K-1)}{(1-R^2)/(n-K)} \\ &= \frac{(n-K)R^2}{(K-1)(1-R^2)} \end{aligned}$$

So rearrenging terms:

$$\begin{aligned} \frac{(K-1)F}{n-K} &= \frac{R^2}{1-R^2} \\ \frac{n-K}{(K-1)F} &= \frac{1-R^2}{R^2} \\ \frac{n-K}{(K-1)F} + 1 &= \frac{1}{R^2} \\ \frac{(n-K)+(K-1)F}{(K-1)F} &= \frac{1}{R^2} \\ R^2 &= \frac{(K-1)F}{n-K+(K-1)F} \end{aligned}$$

So multiplying by n in both sides (equivalently dividing by n on the right-hand side):

$$nR^2 = \frac{(K-1)F}{\frac{1}{n}(n-K)+\frac{1}{n}(K-1)F}$$

Note that $\lim_{n\to\infty}\frac{1}{n} = 0 \implies \operatorname*{plim}_{n\to\infty}\frac{1}{n} = 0$

Under CH: $(K-1) \xrightarrow{d} \chi^2(K-1)$

It is easy to see that $\lim_{n\to\infty}\frac{n-K}{n} = 1 \implies \operatorname*{plim}_{n\to\infty}\frac{n-K}{n} = 1$ and $\frac{1}{n}(K-1)F \xrightarrow{p} 0$ so the denominator becomes 1 by Slutzky's Theorem.

Thus, $nR^2 \xrightarrow{d} \chi^2(K-1)$

1.7 Testing Conditional Homoskedasticity

1. White's Test for Conditional Heteroskedasticity (under a set of assumptions) is expressed as:

$$nR^2 \xrightarrow{d} \chi^2(m)$$

Where R^2 is computed from the following auxiliary regression:

$$\hat{\varepsilon}_i{}^2 = \alpha_0 + \boldsymbol{\psi}_i'\boldsymbol{\delta} + u_i$$

Where $\boldsymbol{\psi}_i$ is a vector collecting unique and nonconstant elements of the square data matrix $\boldsymbol{x}_i\boldsymbol{x}_i'$ and α_0 is a constant.

Now suppose $\boldsymbol{x}_i = (1, q_i, q_i^2, p_i)'$, let us get the dimension of $\boldsymbol{\psi}_i$:

$$\boldsymbol{x}_i\boldsymbol{x}_i' = \begin{bmatrix} 1 \\ q_i \\ q_i^2 \\ p_i \end{bmatrix} \begin{bmatrix} 1 & q_i & q_i^2 & p_i \end{bmatrix} = \begin{bmatrix} 1 & q_i & q_i^2 & p_i \\ q_i & q_i^2 & q_i^3 & q_i p_i \\ q_i^2 & q_i^3 & q_i^4 & q_i^2 p_i \\ p_i & p_i q_i & p_i q_i^2 & p_i^2 \end{bmatrix}$$

So by the provided definition of $\boldsymbol{\psi}_i$:

$$\boldsymbol{\psi}_i = (q_i, q_i^2, p_i, q_i^3, q_i p_i, q_i^4, q_i^2 p_i, p_i^2)'$$

Then length($\boldsymbol{\psi}_i$) = 8, where length($\cdot$) is a linear operator that retrieves the number of elements of a vector. It is based on the function from Programming Language **R**. Think of len($\cdot$) in the case of **Python**.

2 Analytical Exercises

1. Let us show that convergence in probability and convergence in moments need not coincide with the following example:

$$z_n = \begin{cases} 0 & \text{with probability} \quad (n-1)/n \\ n^2 & \text{with probability} \quad 1/n \end{cases}$$

So the random variable z_n is likely to be negligible with little to no chance of exploding. Firstly, let us consider the expected value of said random variable:

$$\begin{aligned} \mathrm{E}[z_n] &= \sum_{i=1}^{n} z_i \Pr(z_n = z_i) \\ &= 0 * \frac{n-1}{n} + n^2 * \frac{1}{n} \\ &= n \end{aligned}$$

Thus:

$$\lim_{n\to\infty} \mathrm{E}[z_n] = \lim_{n\to\infty} n = +\infty$$

On the other hand, by defition of convergence in probability:

$$\begin{aligned} \operatorname*{plim}_{n\to\infty} z_n &= \lim_{n\to\infty} \Pr\left(|z_n| > \varepsilon\right) \quad \text{(where } \forall \varepsilon > 0) \\ &= \lim_{n\to\infty} \frac{1}{n} \text{ (as } 1/n \text{ is the only probability associated with } z_n \text{ being } > 0) \end{aligned}$$

Thus:

$$\lim_{n\to\infty} \frac{1}{n} = 0 \implies \operatorname*{plim}_{n\to\infty} z_n = 0$$

Consequently:

$$\lim_{n\to\infty} z_n \neq \operatorname*{plim}_{n\to\infty} z_n$$

2. Let us prove Chebychev's Weak LLN for the mean of some random variable z_n:

- $\lim_{n\to\infty} \mathrm{E}[\bar{z}_n] = \mu \quad \text{and} \quad \lim_{n\to\infty} \mathrm{V}[\bar{z}_n] = 0$
- $\{z_n\}$ is not necessarily i.i.d

Note that by the definition of convergence in mean square:

$$\lim_{n\to\infty} \mathrm{E}\left[(\bar{z}_n - \mu)^2\right] = 0$$

Applying the add-and-subtract strategy:

$$\lim_{n\to\infty} \mathrm{E}\left[(\bar{z}_n - \mathrm{E}[\bar{z}_n] + \mathrm{E}[\bar{z}_n] - \mu)^2\right] = \lim_{n\to\infty} \mathrm{E}\left[((\bar{z}_n - \mathrm{E}[\bar{z}_n]) + (\mathrm{E}[\bar{z}_n] - \mu))^2\right]$$
$$\hookrightarrow \quad \lim_{n\to\infty} \mathrm{E}\left[(\bar{z}_n - \mathrm{E}[\bar{z}_n])^2 + 2(\bar{z}_n - \mathrm{E}[\bar{z}_n])(\mathrm{E}[\bar{z}_n] - \mu) + (\mathrm{E}[\bar{z}_n] - \mu)^2\right]$$

By Slutzky's Theorem the limiting behavior of a product is the product of the limits (if they exist and are finite).

$$\lim_{n\to\infty} \mathrm{E}\left[\mathrm{E}[\bar{z}_n] - \mu\right] = \lim_{n\to\infty} \mathrm{E}[\bar{z}_n] - \mu = 0 \qquad \left(\text{since } \lim_{n\to\infty} \mathrm{E}[\bar{z}_n] = \mu\right)$$
$$\lim_{n\to\infty} \mathrm{E}\left[\bar{z}_n - \mathrm{E}[\bar{z}_n]\right] = \lim_{n\to\infty} \mathrm{E}[\bar{z}_n] - \mathrm{E}[\bar{z}_n] = 0$$

So the cross-product in the middle side of the equation vanishes by Slutzky's Theorem. Focusing on the LHS:

$$\lim_{n\to\infty} \mathrm{E}\left[(\bar{z}_n - \mathrm{E}[\bar{z}_n])^2\right] = \lim_{n\to\infty} \mathrm{V}[\bar{z}_n] = 0$$

$$\lim_{n\to\infty} \mathrm{E}\left[\mathrm{E}[(\bar{z}_n] - \mu)^2\right] = \lim_{n\to\infty} \mathrm{E}[\bar{z}_n]^2 - 2\mathrm{E}[\bar{z}_n]\mu + \mu^2$$

Which cancels out as $\lim_{n\to\infty} \mathrm{E}[\bar{z}_n] = \mu$ and said value is a constant (so its limit is itself). Thus:

$$\lim_{n\to\infty} \mathrm{E}[\bar{z}_n]^2 - 2\mathrm{E}[\bar{z}_n]\mu + \mu^2 = \mu^2 - 2\mu^2 + \mu^2 = 0$$

And so we proved $\bar{z}_n \xrightarrow{m.s} \mu$ as $\lim_{n\to\infty} \mathrm{E}[(\bar{z}_n - \mu)^2] = 0$

3. Let us prove the consistency and asymptotic normality of the OLS estimator under the following assumptions:

- A.1) Linearity: $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$
- A.2) Random Sampling: $\{y_i, \boldsymbol{x}_i\}$ is a random sample
- A.3) Orthogonality Condition: $\mathrm{E}[\boldsymbol{x}_i \varepsilon_i] = 0$
- A.4) Rank Condition: $\mathrm{E}[\boldsymbol{x}_i \boldsymbol{x}_i'] = \boldsymbol{\Sigma}_{\mathbf{xx}} < \infty$ which is a KxK nonsingular matrix of full column rank
- A.5) $\mathbf{S} := \mathrm{E}[\boldsymbol{g}_i \boldsymbol{g}_i'] < \infty$ where $\{\boldsymbol{g}_i\}$ is an m.d.s and is defined as $\boldsymbol{g}_i = \boldsymbol{x}_i \varepsilon_i$

Let us start from the sampling error:

$$
\begin{aligned}
\mathbf{b} - \boldsymbol{\beta} &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y} - \boldsymbol{\beta} \ \big(\text{by minimizing } \mathbf{e}'\mathbf{e}\big) \\
&= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'(\mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}) - \boldsymbol{\beta} \ \big(\text{by A.1}\big) \\
&= \boldsymbol{\beta} + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\varepsilon} - \boldsymbol{\beta} \\
&= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\varepsilon} \\
&= (n^{-1}\mathbf{X}'\mathbf{X})^{-1} n^{-1}\mathbf{X}'\boldsymbol{\varepsilon} \ \big(\text{multiplying and dividing by } n\big)
\end{aligned}
$$

Note that $\mathbf{S}_{\mathbf{xx}} := n^{-1}(\mathbf{X}'\mathbf{X})$ which can be expressed as $\mathbf{S}_{\mathbf{xx}} = \dfrac{1}{n}\displaystyle\sum_{i=1}^{n} \boldsymbol{x}_i \boldsymbol{x}_i'$

By A.2 $\{y_i, \boldsymbol{x}_i\}$ is i.i.d and thus so is $\{\boldsymbol{x}_i \boldsymbol{x}_i'\}$ as it is a function of $\{\boldsymbol{x}_i\}$
By A.4 $\mathrm{E}[\boldsymbol{x}_i \boldsymbol{x}_i'] = \boldsymbol{\Sigma}_{\mathbf{xx}} < \infty$
Thus, by Kolmogorov's (Strong) LLN:

$$
\frac{1}{n}\sum_{i=1}^{n} \boldsymbol{x}_i \boldsymbol{x}_i' \xrightarrow{a.s} \mathrm{E}[\boldsymbol{x}_i \boldsymbol{x}_i'] \implies \frac{1}{n}\sum_{i=1}^{n} \boldsymbol{x}_i \boldsymbol{x}_i' \xrightarrow{p} \boldsymbol{\Sigma}_{\mathbf{xx}}
$$

According to Continuous Mapping Theorem (CMT): $\mathrm{a}(z_n) \xrightarrow{p} \mathrm{a}(z)$ if $z_n \xrightarrow{p} z$ (provided $\mathrm{a}(\cdot)$ is a plausible continuous transformation).

Since $\boldsymbol{\Sigma}_{\mathbf{xx}}$ is nonsingular and thus, invertible, by A.4 we can apply CMT:

$$
\mathbf{S}_{\mathbf{xx}}^{-1} \xrightarrow{p} \boldsymbol{\Sigma}_{\mathbf{xx}}^{-1}
$$

$n^{-1}\mathbf{X}'\mathbf{X}$ is also nonsingular by A.2 (columns would be linearly dependent by pure coincidence). Anyway, $n^{-1}\mathbf{S}_{\mathbf{xx}}$ is invertible as the sample size increases since $\boldsymbol{\Sigma}_{\mathbf{xx}}$ is.

$\{\varepsilon_i\}$ is i.i.d since by A.1: $\varepsilon_i = y_i - \boldsymbol{x}_i' \boldsymbol{\beta}$, where $\{y_i, \boldsymbol{x}_i\}$ is i.i.d following A.2.

Also, $\mathrm{E}[\boldsymbol{x}_i \varepsilon_i] = 0$ by A.3 and $n^{-1}\mathbf{X}'\boldsymbol{\varepsilon} = \dfrac{1}{n}\displaystyle\sum_{i=1}^{n} \boldsymbol{x}_i \varepsilon_i$.

Thus, by Kolmogorov's (Strong) LLN:

$$\frac{1}{n}\sum_{i=1}^{n} \boldsymbol{x}_i \varepsilon \xrightarrow{p} \mathrm{E}[\boldsymbol{x}_i \varepsilon_i] \ \ (= 0)$$

By Slutzky's Theorem, the limiting behavior of the product of some elements is the product of the limits of said components (provided they exist and are finite). Thus:

$$\mathbf{S}_{xx}^{-1} \frac{1}{n}\sum_{i=1}^{n} \boldsymbol{x}_i \varepsilon_i \xrightarrow{p} 0$$

This proves $\mathbf{b} \xrightarrow{p} \boldsymbol{\beta}$ and so the OLS estimator is consistent.

Let us get back to the sampling error, but this time we multiply by $\sqrt{n}$:

$$\begin{aligned}\sqrt{n}(\mathbf{b} - \boldsymbol{\beta}) &= \sqrt{n}\mathbf{S}_{\mathbf{xx}}^{-1}\, \bar{\boldsymbol{g}} \ \left(\text{since } \boldsymbol{g}_i \coloneqq \boldsymbol{x}_i \varepsilon_i \ \right) \\ &= \mathbf{S}_{\mathbf{xx}}^{-1} \sqrt{n}\, \bar{\boldsymbol{g}}\end{aligned}$$

Note that $\mathrm{E}[\boldsymbol{g}_i] = \mathrm{E}[\bar{\boldsymbol{g}}]$ as $\{\boldsymbol{x}_i \varepsilon_i\}$ is i.i.d and $\mathrm{E}[\boldsymbol{g}_i] = 0$ by LIE following A.3.

By Kolmogorov's (Strong) LLN: $\bar{\boldsymbol{g}} \xrightarrow{p} \vec{0}$

By A.5: $\mathbf{S} = \mathrm{E}[\boldsymbol{g}_i \boldsymbol{g}_i']$ which is $\mathrm{Avar}(\bar{\boldsymbol{g}})$ since $\mathrm{E}[\bar{\boldsymbol{g}}] = 0$

Thus, by Lindeberg-Levy's CLT:

$$\sqrt{n}\, \bar{\boldsymbol{g}} \xrightarrow{d} N(0, \mathbf{S})$$

Applying Slutzky's Theorem:

$$\mathbf{S}_{\mathbf{xx}}^{-1} \sqrt{n}\, \bar{\boldsymbol{g}} \xrightarrow{d} N(0, \boldsymbol{\Sigma}_{xx}^{-1} \mathbf{S} \boldsymbol{\Sigma}_{xx}^{-1})$$

Note that $\boldsymbol{\Sigma}_{\mathbf{xx}}^{-1\prime} = \boldsymbol{\Sigma}_{\mathbf{xx}}^{-1}$ since it is symmetric.

Consequently:

$$\sqrt{n}(\mathbf{b} - \boldsymbol{\beta}) \xrightarrow{d} N(0, \boldsymbol{\Sigma}_{xx}^{-1} \mathbf{S} \boldsymbol{\Sigma}_{xx}^{-1})$$

We have proved the OLS estimator $\mathbf{b}$ is consistent and asymptotically normal.

4. For simplicity, let us assume $K = 1$, so x_i is a scalar. Furthermore, suppose there exists an estimator for $\mathrm{E}[\boldsymbol{g}_i\boldsymbol{g}_i']$ such that $\hat{\boldsymbol{S}} \xrightarrow{p} \mathbf{S}$ where:

$$\hat{\boldsymbol{S}} := \frac{1}{n}\sum_{i=1}^{n}\hat{\varepsilon}_i^2\boldsymbol{x}_i\boldsymbol{x}_i'$$

Let us keep the assumptions from the previous exercise, relaxing A.2 to $\{y_i, \boldsymbol{x}_i\}$ being an ergodic stationary process (so indeed a random sample is a special case of ergodic stationarity).

$$\begin{aligned}
&\varepsilon_i = y_i - x_i\beta \\
&\hat{\varepsilon}_i = y_i - x_i\hat{\beta} \ \big(\text{as the OLS estimator is consistent}\big) \\
&\hat{\varepsilon}_i^2 = (y_i - x_i\beta)^2 = (x_i\beta + \varepsilon_i - x_i\hat{\beta})^2 = (\varepsilon_i - x_i(\hat{\beta} - \beta))^2 \\
&\quad = \varepsilon_i^2 - 2x_i\varepsilon_i(\hat{\beta} - \beta) + x_i^2(\hat{\beta} - \beta)^2
\end{aligned}$$

Thus:

$$\hat{\boldsymbol{S}} = \frac{1}{n}\sum_{i=1}^{n}\Big(\varepsilon_i^2x_i^2 - 2x_i^3\varepsilon_i(\hat{\beta} - \beta) + x_i^4(\hat{\beta} - \beta)^2\Big)$$

We introduce A.6: $\mathrm{E}[x_i^4] < \infty$, which can be interpreted as large outliers being unlikely.

$\{x_i^4\}$ is ergodic stationary by A.2 as it is a function of $\{x_i\}$

By Ergodic LLN:

$$\frac{1}{n}\sum_{i=1}^{n}x_i^4 \xrightarrow{a.s} \mathrm{E}[x_i^4] \implies \frac{1}{n}\sum_{i=1}^{n}x_i^4 \xrightarrow{p} \mathrm{E}[x_i^4]$$

Thus, by Slutzky's Theorem:

$$\frac{1}{n}\sum_{i=1}^{n}x_i^4(\hat{\beta} - \beta) \xrightarrow{p} 0$$

So the last term of the expression for $\hat{\boldsymbol{S}}$ above vanishes as the sample size increase. Let us now focus on the middle term, applying *Cauchy-Schwartz Inequality*:

$$\mathrm{E}[|f \cdot h|] \leq \sqrt{\mathrm{E}[f^2] \cdot \mathrm{E}[h^2]}$$

This can be proved leveraging trigonometric properties:

$$\cos(\theta) = \frac{\mathbf{x}'\mathbf{y}}{||\mathbf{x}||\,||\mathbf{y}||}$$

Note that $|\cos(\theta)| \leq \ 1$ so:

$$\frac{|\mathbf{x}'\mathbf{y}|}{||\mathbf{x}||\,||\mathbf{y}||} \leq 1$$

$$|\mathbf{x}'\mathbf{y}| \leq ||\mathbf{x}||\,||\mathbf{y}||$$

Particularizing for our case:

$$\mathrm{E}[|x_i^3 \cdot \varepsilon_i|] \leq \sqrt{\mathrm{E}[x_i^2\varepsilon_i^2] \cdot \mathrm{E}[x_i^4]}$$

$\mathrm{E}[x_i^2\varepsilon_i^2] < \infty$ by A.5 and $\mathrm{E}[x_i^4] < \infty$ by A.6, thus, $\mathrm{E}[x_i^3\varepsilon_i]$ is bounded by some finite number, entailing that it exists and is finite.

Since $\{x_i^3\varepsilon_i\}$ is a function of $\{x_i\varepsilon_i\}$ and $\{x_i, \varepsilon_i\}$ is ergodic stationary by A.2, $\{x_i^3\varepsilon_i\}$ is an ergodic stationary stochastic process.

Thus, by Ergodic LLN:

$$\frac{1}{n}\sum_{i=1}^{n} x_i^3\varepsilon_i \xrightarrow{p} \mathrm{E}[x_i^3\varepsilon_i] < \infty$$

By Slutzky's Theorem:

$$\frac{1}{n}\sum_{i=1}^{n} x_i^3\varepsilon_i(\hat{\beta} - \beta) \xrightarrow{p} 0 \quad (\text{since} \quad \hat{\beta} \xrightarrow{p} \beta)$$

Consequently, both, the middle and last term of $\hat{\boldsymbol{S}}$ vanish. Let us analyze the left hand-side one.

Since $\{\varepsilon_i^2 x_i^2\}$ is ergodic stationary by A.2 and $\mathrm{E}[\varepsilon_i^2 x_i^2] < \infty$ by A.5.

By Ergodic LLN:

$$\frac{1}{n}\sum_{i=1}^{n} \varepsilon_i^2 x_i^2 \xrightarrow{p} \mathrm{E}[\varepsilon_i^2 x_i^2]$$

Thus:

$$\hat{\boldsymbol{S}} \xrightarrow{p} \boldsymbol{S}$$

5. Proof that the change in *SSR* divided by σ^2 is asymptotically χ^2 under Conditional Homoskedasticity, namely A.7: $\mathrm{E}[\varepsilon_i^2|\boldsymbol{x}_i] = \sigma^2 > 0 \ \forall i$

From Finite Sample Theory:

$$F = \frac{SSR_r - SSR_u/\#r}{SSR_u/n-k}$$

Let $\tilde{\boldsymbol{\beta}}$ denote $\hat{\boldsymbol{\beta}}$ from the restricted regression

$$\begin{aligned}
SSR_u &= (\mathbf{Y} - \mathbf{X}\tilde{\boldsymbol{\beta}})'(\mathbf{Y} - \mathbf{X}\tilde{\boldsymbol{\beta}}) \\
&= (\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}} + \mathbf{X}\hat{\boldsymbol{\beta}} - \mathbf{X}\tilde{\boldsymbol{\beta}})'(\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}} + \mathbf{X}\hat{\boldsymbol{\beta}} - \mathbf{X}\tilde{\boldsymbol{\beta}}) \\
&= [(\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}}) + \mathbf{X}(\hat{\boldsymbol{\beta}} - \tilde{\boldsymbol{\beta}})]'[(\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}}) + \mathbf{X}(\hat{\boldsymbol{\beta}} - \tilde{\boldsymbol{\beta}})] \\
&= (\mathbf{e} + \mathbf{X}(\hat{\boldsymbol{\beta}} - \tilde{\boldsymbol{\beta}}))'(\mathbf{e} + \mathbf{X}(\hat{\boldsymbol{\beta}} - \tilde{\boldsymbol{\beta}})) \\
&= \mathbf{e}'\mathbf{e} + \mathbf{e}'\mathbf{X}(\hat{\boldsymbol{\beta}} - \tilde{\boldsymbol{\beta}}) + (\hat{\boldsymbol{\beta}} - \tilde{\boldsymbol{\beta}})'\mathbf{X}'\mathbf{e} + (\hat{\boldsymbol{\beta}} - \tilde{\boldsymbol{\beta}})'\mathbf{X}'\mathbf{X}(\hat{\boldsymbol{\beta}} - \tilde{\boldsymbol{\beta}}) \\
&= \mathbf{e}'\mathbf{e} + (\hat{\boldsymbol{\beta}} - \tilde{\boldsymbol{\beta}})'\mathbf{X}'\mathbf{X}(\hat{\boldsymbol{\beta}} - \tilde{\boldsymbol{\beta}})
\end{aligned}$$

This last equality holds since if a Newtonian or closed form solution is feasible, by Normal Equations $\mathbf{X}'\mathbf{e} = 0$ and $\mathbf{e}'X(\hat{\boldsymbol{\beta}} - \tilde{\boldsymbol{\beta}}) = (\hat{\boldsymbol{\beta}} - \tilde{\boldsymbol{\beta}})'\mathbf{X}'\mathbf{e}$

Thus:

$$\begin{aligned}
SSR_r - SSR_u &= \mathbf{e}'\mathbf{e} + (\hat{\boldsymbol{\beta}} - \tilde{\boldsymbol{\beta}})'\mathbf{X}'\mathbf{X}(\hat{\boldsymbol{\beta}} - \tilde{\boldsymbol{\beta}}) - \mathbf{e}'\mathbf{e} \\
&= (\hat{\boldsymbol{\beta}} - \tilde{\boldsymbol{\beta}})'\mathbf{X}'\mathbf{X}(\hat{\boldsymbol{\beta}} - \tilde{\boldsymbol{\beta}}) \\
&= (\mathbf{R}\hat{\boldsymbol{\beta}} - \mathbf{r})'\left[\mathbf{R}(\mathbf{X}'\mathbf{X}^{-1}\mathbf{R}')\right]^{-1}(\mathbf{R}\hat{\boldsymbol{\beta}} - \mathbf{r})
\end{aligned}$$

Plugging this expression to the F-statistic shown above:

$$F = \frac{(\mathbf{R}\hat{\boldsymbol{\beta}} - \mathbf{r}'\left[\mathbf{R}(\mathbf{X}'\mathbf{X}^{-1}\mathbf{R}')\right]^{-1}(\mathbf{R}\hat{\boldsymbol{\beta}} - \mathbf{r})/\#r}{\mathbf{e}'\mathbf{e}/n-k}$$

Recall that $S^2 \coloneqq \mathbf{e}'\mathbf{e}/n-k$ so:

$$F = \frac{(\mathbf{R}\hat{\boldsymbol{\beta}} - \mathbf{r}'\left[\mathbf{R}(\mathbf{X}'\mathbf{X}^{-1}\mathbf{R}')\right]^{-1}(\mathbf{R}\hat{\boldsymbol{\beta}} - \mathbf{r})/\#r}{S^2}$$

Leveraging that we already derived the expression for the sampling error:

$$\begin{aligned}
\mathbf{R}\hat{\boldsymbol{\beta}} - \mathbf{r} &= \mathbf{R}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \text{ (under } H_0\text{)} \\
&= \mathbf{R}((\mathbf{X}'\mathbf{X}^{-1})\mathbf{X}'\boldsymbol{\varepsilon}) \text{ (sampling error)} \\
&= \mathbf{R}(n(\mathbf{X}'\mathbf{X})^{-1}n^{-1}\mathbf{X}'\boldsymbol{\varepsilon}) \\
&= \mathbf{R}(\mathbf{S}_{\mathbf{xx}}^{-1}\bar{\boldsymbol{g}})
\end{aligned}$$

Consequently:

$$\begin{aligned}
SSR_r - SSR_u &= \hat{\boldsymbol{g}}'\mathbf{S}_{\mathbf{xx}}^{-1}\mathbf{R}'\left[\mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}'\right]^{-1}\mathbf{R}\mathbf{S}_{\mathbf{xx}}^{-1}\hat{\boldsymbol{g}} \\
&= (\sqrt{n}\hat{\boldsymbol{g}})'\mathbf{S}_{\mathbf{xx}}^{-1}\mathbf{R}'\left[\mathbf{R}\mathbf{S}_{\mathbf{xx}}^{-1}\mathbf{R}'\right]^{-1}\mathbf{S}_{\mathbf{xx}}^{-1}(\sqrt{n}\hat{\boldsymbol{g}})
\end{aligned}$$

Thus:

$$\begin{aligned}
\frac{SSR_r - SSR_u}{S^2} &= (\sqrt{n}\hat{\boldsymbol{g}})'\mathbf{S}_{\mathbf{xx}}^{-1}\mathbf{R}'\left[\mathbf{R}S^2\mathbf{S}_{\mathbf{xx}}^{-1}\mathbf{R}'\right]^{-1}\mathbf{S}_{\mathbf{xx}}^{-1}(\sqrt{n}\hat{\boldsymbol{g}}) \\
&= (\sqrt{n}\hat{\boldsymbol{g}})'\mathbf{S}_{\mathbf{xx}}^{-1}\mathbf{R}'\left[\mathbf{R}\hat{\boldsymbol{S}}\mathbf{R}'\right]^{-1}\mathbf{S}_{\mathbf{xx}}^{-1}(\sqrt{n}\hat{\boldsymbol{g}})
\end{aligned}$$

Since $\hat{\boldsymbol{S}} = \widehat{\text{Avar}(\hat{\boldsymbol{\beta}})}$ under A.7, then: $F \xrightarrow{d} \chi^2(\#\mathbf{r})$

8. Let us prove the expression for the least squares projection of y on $\mathbf{x}$ when one of the regressors is a constant, i.e: $\widehat{\text{E}^*}[y|\mathbf{x}] = \widehat{\text{E}^*}[y|1, \mathbf{x}] = \mu + \boldsymbol{\gamma}'\tilde{\mathbf{x}}$

Let $\tilde{\mathbf{x}}$ be the vector of stochastic regressors:

$$\mathbf{x} = \begin{bmatrix} 1 \\ \tilde{\mathbf{x}} \end{bmatrix}$$

From the expected values of Normal Equations we get $\text{E}[\mathbf{x}\mathbf{x}']\boldsymbol{\beta} = \text{E}[\mathbf{x}y]$. Consider an optimal estimator such that the forecast error y - $\mathbf{x}'\boldsymbol{\beta}^*$ is orthogonal to $\mathbf{x}$. Assuming $\text{E}[\mathbf{x}\mathbf{x}']$ is nonsingular, then:

$$\boldsymbol{\beta}^* = (\text{E}[\mathbf{x}\mathbf{x}'])^{-1}\text{E}[\mathbf{x}y]$$

$$\mathbf{x}\mathbf{x}' = \begin{bmatrix} 1 \\ \tilde{\mathbf{x}} \end{bmatrix} \begin{bmatrix} 1 & \tilde{\mathbf{x}}' \end{bmatrix} = \begin{bmatrix} 1 & \tilde{\mathbf{x}}' \\ \tilde{\mathbf{x}} & \tilde{\mathbf{x}}\tilde{\mathbf{x}}' \end{bmatrix}$$

$$\mathbf{x}y = \begin{bmatrix} 1 \\ \tilde{\mathbf{x}} \end{bmatrix} y = \begin{bmatrix} y \\ \tilde{\mathbf{x}}y \end{bmatrix}$$

$$\boldsymbol{\beta} = \begin{bmatrix} \mu \\ \boldsymbol{\gamma} \end{bmatrix}$$

Then, $\mathrm{E}[\mathbf{x}\mathbf{x}']\boldsymbol{\beta} = \mathrm{E}[\mathbf{x}y]$ can be rewritten as:

$$\mathrm{E}\begin{bmatrix} \mu + \boldsymbol{\gamma}\tilde{\mathbf{x}} \\ \tilde{\mathbf{x}}\mu + \tilde{\mathbf{x}}\tilde{\mathbf{x}}'\boldsymbol{\gamma} \end{bmatrix} = \mathrm{E}\begin{bmatrix} y \\ \tilde{\mathbf{x}}y \end{bmatrix}$$

Breaking down the matrix into two equations:

$$\mathrm{E}[\mu] + \mathrm{E}[\boldsymbol{\gamma}\tilde{\mathbf{x}}'] = \mathrm{E}[y] \rightarrow \mu + \boldsymbol{\gamma}\mathrm{E}[\tilde{\mathbf{x}}'] = \mathrm{E}[y]$$
$$\mathrm{E}[\tilde{\mathbf{x}}\mu] + \mathrm{E}[\tilde{\mathbf{x}}\tilde{\mathbf{x}}'\boldsymbol{\gamma}] = \mathrm{E}[\tilde{\mathbf{x}}y] \rightarrow \mu\mathrm{E}[\tilde{\mathbf{x}}] + \boldsymbol{\gamma}\mathrm{E}[\tilde{\mathbf{x}}\tilde{\mathbf{x}}'] = \mathrm{E}[\tilde{\mathbf{x}}y]$$

From the first equation: $\mu = \mathrm{E}[y]$ - $\boldsymbol{\gamma}\mathrm{E}[\tilde{\mathbf{x}}']$

Plugging it into the second one:

$$\begin{aligned} (\mathrm{E}[y] - \boldsymbol{\gamma}\mathrm{E}[\tilde{\mathbf{x}}'])\mathrm{E}[\tilde{\mathbf{x}}] + \boldsymbol{\gamma}\mathrm{E}[\tilde{\mathbf{x}}\tilde{\mathbf{x}}'] &= \mathrm{E}[\tilde{\mathbf{x}}y] \\ \mathrm{E}[\tilde{\mathbf{x}}]\mathrm{E}[y] - \boldsymbol{\gamma}\mathrm{E}[\tilde{\mathbf{x}}]\mathrm{E}[\tilde{\mathbf{x}}'] + \boldsymbol{\gamma}\mathrm{E}[\tilde{\mathbf{x}}\tilde{\mathbf{x}}'] &= \mathrm{E}[\tilde{\mathbf{x}}y] \\ \boldsymbol{\gamma}(\mathrm{E}[\tilde{\mathbf{x}}\tilde{\mathbf{x}}'] - \mathrm{E}[\tilde{\mathbf{x}}]\mathrm{E}[\tilde{\mathbf{x}}']) &= \mathrm{E}[\tilde{\mathbf{x}}y] - \mathrm{E}[\tilde{\mathbf{x}}]\mathrm{E}[y] \\ \boldsymbol{\gamma} = \mathrm{V}[\tilde{\mathbf{x}}]^{-1}\mathrm{Cov}(\tilde{\mathbf{x}}, y) \end{aligned}$$

Finally:

$$\begin{aligned} \mu &= \mathrm{E}[y] - \mathrm{V}[\tilde{\mathbf{x}}]^{-1}\mathrm{Cov}(\tilde{\mathbf{x}}, y)\mathrm{E}[V[\tilde{\mathbf{x}}]^{-1}\mathrm{Cov}(\tilde{\mathbf{x}}, y)'] \\ &= \mathrm{E}[y] - \boldsymbol{\gamma}'\mathrm{E}[\tilde{\mathbf{x}}] \end{aligned}$$

9. Let $\{\varepsilon_t\}$ be an ergodic stationary m.d.s with $\mathrm{E}[\varepsilon_t^2|\varepsilon_{t-1}, \varepsilon_{t-2}, \dots, \varepsilon_1] = \sigma^2 < \infty$ Furthermore, let $g_t \coloneqq \varepsilon_t \varepsilon_{t-1}$ and

$$\hat{\gamma}_j = \frac{1}{n} \sum_{t=j+1}^{n} \varepsilon_t \varepsilon_{t-j} \qquad j \in \{0, 1\}, \;\; \hat{\rho}_1 = \frac{\hat{\gamma}_1}{\hat{\gamma}_0}$$

i) Let us show that $\{g_t\}_{t=2}^{+\infty}$ is an m.d.s, i.e: $\mathrm{E}[g_t|g_{t-1}, \dots, g_1] = 0$

$$\begin{aligned}
\mathrm{E}[g_t|g_{t-1}, \dots, g_1] &= \mathrm{E}[\varepsilon_t \varepsilon_{t-1}|g_{t-1}, \dots, g_1] \\
&= \mathrm{E}[\mathrm{E}[\varepsilon_t \varepsilon_{t-1}|\varepsilon_{t-1}, \dots, \varepsilon_1]|g_{t-1}, \dots, g_1] \;\; \text{(by LIE)} \\
&= \mathrm{E}[\varepsilon_{t-1} \mathrm{E}[\varepsilon_t|\varepsilon_{t-1}, \dots, \varepsilon_1]|g_{t-1}, \dots, g_1] \\
&= \mathrm{E}[\varepsilon_{t-1}|g_{t-1}, \dots, g_1] \mathrm{E}[0|g_{t-1}, \dots, g_1] \;\; \text{(since } \{\varepsilon_t\} \text{ is an m.d.s)} \\
&= 0
\end{aligned}$$

ii) Let us show that $\mathrm{E}[g_t^2] = \sigma^4$

$$\begin{aligned}
\mathrm{E}[g_t^2] &= \mathrm{E}[\varepsilon_t \varepsilon_{t-1} \varepsilon_t \varepsilon_{t-1}] \\
&= \mathrm{E}[\varepsilon_t^2 \varepsilon_{t-1}^2] \\
&= \mathrm{E}[\mathrm{E}[\varepsilon_t^2 \varepsilon_{t-1}^2|\varepsilon_{t-1}, \dots, \varepsilon_1]] \\
&= \mathrm{E}[\varepsilon_{t-1}^2 \mathrm{E}[\varepsilon_t^2|\varepsilon_{t-1}, \dots, \varepsilon_1]] \\
&= \mathrm{E}[\varepsilon_{t-1}^2]\sigma^2 \\
&= \sigma^4
\end{aligned}$$

Note that $\mathrm{E}[\varepsilon_{t-1}^2] = \mathrm{E}[\mathrm{E}[\varepsilon_{t-1}^2|\varepsilon_{t-2}, \dots, \varepsilon_1]] = \sigma^2$ by LIE, so our result holds.

iii) Let us show that $\sqrt{n}\hat{\gamma}_1 \xrightarrow{d} N(0, \sigma^4)$

Note that $\hat{\gamma}_1 = \dfrac{1}{n} \displaystyle\sum_{t=2}^{n} \varepsilon_t \varepsilon_{t-1}$

$$\begin{aligned}
\mathrm{E}[\hat{\gamma}_1] &= \frac{1}{n} \mathrm{E}\left[\sum_{t=2}^{n} g_t\right] \\
&= \frac{1}{n} \mathrm{E}\left[\mathrm{E}\left[\sum_{t=2}^{n} g_t|g_{t-1}, \dots, g_1\right]\right] = 0
\end{aligned}$$

$$
\begin{aligned}
\mathrm{V}[\hat{\gamma}_1] &= \frac{1}{n}\mathrm{V}\left[\sum_{t=2}^{n} \varepsilon_t \varepsilon_{t-1}\right] \\
&= \frac{1}{n}\mathrm{V}\left[\sum_{t=2}^{n} g_t\right] \\
&= \frac{1}{n} n\sigma^4 = \sigma^4
\end{aligned}
$$

Since $\{\varepsilon_t\}$ is ergodic stationary so if $\{\varepsilon_t\varepsilon_{t-1}\}$ as it is a function of the former. Also, $\mathrm{E}[\varepsilon_t\varepsilon_{t-1}] = 0$ as $\mathrm{E}[g_t] = 0$ (exists and is finite). Consequently, by Ergodic CLT:

$$
\sqrt{n}\hat{\gamma}_1 \xrightarrow{d} N(0, \sigma^4)
$$

iv) Let us show that $\sqrt{n}\hat{\rho}_1 \xrightarrow{d} N(0, 1)$

Note that $\hat{\rho}_1 = \dfrac{\hat{\gamma}_1}{\hat{\gamma}_0}$ and:

$$
\begin{aligned}
\hat{\gamma}_0 &= \frac{1}{n}\sum_{t=1}^{n} \varepsilon_t^2 \\
\mathrm{E}[\hat{\gamma}_0] &= \mathrm{E}\left[\frac{1}{n}\sum_{t=1}^{n} \varepsilon_t^2\right] \\
&= \frac{1}{n}\mathrm{E}\left[\sum_{t=1}^{n} \varepsilon_t^2\right] \\
&= \sigma^2
\end{aligned}
$$

Since $\{\varepsilon_t\}$ is ergodic stationary so is $\{\varepsilon_t^2\}$. By Ergodic LLN:

$$
\hat{\gamma}_0 \xrightarrow{p} \sigma^2
$$

Note that $\sqrt{n}\hat{\gamma}_1 \xrightarrow{d} N(0, \sigma^4)$. By Slutzky's Theorem:

$$
\sqrt{n}\frac{\hat{\gamma}_1}{\hat{\gamma}_0} \xrightarrow{d} N(0, 1)
$$

Since $\sigma^4/(\sigma^2)^2 = 1$

10. Let $\{\varepsilon_t\} \overset{i.i.d}{\sim} \mathrm{WN}(0,\, \sigma^2_\varepsilon)$ and consider a stochastic process generated by:

$$Y_t = \varepsilon_t + \theta_1\varepsilon_{t-1} + \theta_2\varepsilon_{t-2}$$

Note that $\{Y_t\}$ follows a MA(2) process (Second Order Moving Average).
Let us show the asymptotics of its sample mean and other properties.

i) We claim Y_t is covariance stationary:

$$\begin{aligned}
\mathrm{E}[Y_t] &= \mathrm{E}\left[\sum_{j=0}^{2}\theta_j\varepsilon_{t-j}\right] \\
&= \theta_1\mathrm{E}[\varepsilon_{t-1}] + \theta_2\mathrm{E}[\varepsilon_{t-2}] \\
&= 0
\end{aligned}$$

Since $\mathrm{E}[\varepsilon_t] = 0\ \forall t$ and $\theta_0 = 1$

$$\begin{aligned}
\mathrm{V}[Y_t] &= \mathrm{V}\left[\sum_{j=0}^{2}\theta_j\varepsilon_{t-j}\right] \\
&= \mathrm{V}[\varepsilon_t] + \theta_1^2\mathrm{V}[\varepsilon_{t-1}] + \theta_2^2\mathrm{V}[\varepsilon_{t-2}] \\
&= \sigma^2_\varepsilon(1 + \theta_1^2 + \theta_2^2)
\end{aligned}$$

Since $\mathrm{V}[\varepsilon_t] = \sigma^2_\varepsilon\ \forall t$ and $\mathrm{cov}(\varepsilon_i, \varepsilon_j) = 0$ if $i \neq j$ as ε_t is an independent White Noise process.

$$\begin{aligned}
\mathrm{cov}(Y_t, Y_{t-1}) &= \mathrm{cov}(\varepsilon_t + \theta_1\varepsilon_{t-1} + \theta_2\varepsilon_{t-2},\ \varepsilon_{t-1} + \theta_1\varepsilon_{t-2} + \theta_2\varepsilon_{t-3}) \\
&= \theta_1\sigma^2_\varepsilon + \theta_1\theta_2\sigma^2_\varepsilon \\
&= \theta_1\sigma^2\varepsilon(1 + \theta_2) \\
\mathrm{cov}(Y_t, Y_{t-2}) &= \mathrm{cov}(\varepsilon_t + \theta_1\varepsilon_{t-1} + \theta_2\varepsilon_{t-2},\ \varepsilon_{t-2} + \theta_1\varepsilon_{t-3} + \theta_2\varepsilon_{t-4}) \\
&= \theta_2\sigma^2_\varepsilon \\
\mathrm{cov}(Y_t, Y_{t-3}) &= 0
\end{aligned}$$

Let $\gamma_j \coloneqq \text{cov}(Y_t, Y_{t-j})$, synthesizing what we showed:

$$\gamma_j = \begin{cases} \sigma_\varepsilon^2(1+\theta_1^2+\theta_2^2) & \text{if} \quad j=0 \\ \theta_1\sigma^2\varepsilon(1+\theta_2) & \text{if} \quad j=1 \\ \theta_2\sigma_\varepsilon^2 & \text{if} \quad j=2 \\ 0 & \text{if} \quad j>2 \end{cases}$$

Since neither $\text{E}[Y_t]$, nor γ_j depends on t, $\{Y_t\}$ is a covariance stationary stochastic process.

ii) Let

$$r_{tj} \coloneqq \text{E}[Y_t|Y_{t-j}, Y_{t-j-1}, \ldots, Y_0, Y_{-1}] - \text{E}[Y_t|Y_{t-j-1}, Y_{t-j-2}, \ldots, Y_0, Y_{-1}]$$

Note that $\text{E}[Y_t|Y_{t-j}, \ldots, Y_{-1}] = \text{E}[Y_t|\varepsilon_{t-j}, \ldots, \varepsilon_{-1}]$ as there is a one-to-one mapping relationship between $\{\varepsilon_t\}$ and $\{Y_t\}$. Thus:

$$\begin{aligned} \text{E}[Y_t|Y_{t-j}, \ldots, Y_{-1}] &= \text{E}[Y_t|\varepsilon_{t-j}, \ldots, \varepsilon_{-1}] \\ \text{E}[Y_t|Y_{t-j}, \ldots, Y_{-1}] &= \text{E}[\varepsilon_t + \theta_1\varepsilon_{t-1} + \theta_2\varepsilon_{t-2}|\varepsilon_{t-j}, \ldots, \varepsilon_{-1}] \end{aligned}$$

$$\begin{aligned} r_{t0} = \text{E}[\varepsilon_t &+ \theta_1\varepsilon_{t-1} + \theta_2\varepsilon_{t-2}|\varepsilon_t, \varepsilon_{t-1}, \varepsilon_{t-2}, \ldots, \varepsilon_{-1}] \\ &- \text{E}[\varepsilon_t + \theta_1\varepsilon_{t-1} + \theta_2\varepsilon_{t-2}|\varepsilon_{t-1}, \varepsilon_{t-2}, \varepsilon_{t-3}, \ldots, \varepsilon_{-1}] \\ &= \varepsilon_t + \theta_1\varepsilon_{t-1} + \theta_2\varepsilon_{t-2} - \text{E}[\varepsilon_t|\varepsilon_{t-1}, \varepsilon_{t-2}, \varepsilon_{t-3}, \ldots, \varepsilon_{-1}] - \theta_1\varepsilon_{t-1} - \theta_2\varepsilon_{t-2} \\ &= \varepsilon_t \end{aligned}$$

Since $\{\varepsilon_t\} \overset{i.i.d}{\sim}$ WN(0, σ_ε^2) and so $\text{E}[\varepsilon_t|\varepsilon_{t-1}, \varepsilon_{t-2}, \ldots, \varepsilon_{-1}] = 0$
Clearly:

$$\text{E}[Y_t|\varepsilon_{t-1}, \ldots, \varepsilon_{-1}] = \begin{cases} \varepsilon_t + \theta_1\varepsilon_{t-1} + \theta_2\varepsilon_{t-2} & \text{if} \quad j=0 \\ \theta_1\varepsilon_{t-1} + \theta_2\varepsilon_{t-2} & \text{if} \quad j=1 \\ \theta_2\varepsilon_{t-2} & \text{if} \quad j=2 \\ 0 & \text{if} \quad j>2 \end{cases}$$

So:

$$r_{tj} = \begin{cases} \varepsilon_t & \text{if} \quad j = 0 \\ \theta_1 \varepsilon_{t-1} & \text{if} \quad j = 1 \\ \theta_2 \varepsilon_{t-2} & \text{if} \quad j = 2 \\ 0 & \text{if} \quad j > 2 \end{cases}$$

iii) Let

$$\bar{Y}_T = \frac{1}{n} \sum_{j=1}^{T} Y_j$$

Then its long-run variance is given by $\mathrm{V}[\sqrt{n}\bar{Y}_T]$, which we now derive:

$$\begin{aligned}
\mathrm{V}[\sqrt{n}\bar{Y}_T] &= \mathrm{V}\left[\frac{1}{\sqrt{n}} \sum_{j=1}^{T} Y_j\right] \\
&= \frac{1}{n} \sum_{j=1}^{T} \mathrm{cov}(Y_j, Y_1, \ldots, Y_T) \\
&= \frac{1}{n} \left[\sum_{i=0}^{T-1} \gamma_i + \sum_{j=1}^{T-2} \gamma_j + \cdots + \sum_{k=1}^{T} \gamma_{T-k}\right] \\
&= \frac{1}{n} \left[T\gamma_0 + 2(T-1)\gamma_1 + \cdots + 2(T-j)\gamma_j + \cdots + 2\gamma_{T-1}\right] \\
&= \gamma_0 + 2\sum_{j=1}^{T-1} \left(1 - \frac{j}{n}\gamma_j\right) \\
&= \gamma_0 + 2\left[\left(1 - \frac{1}{T}\right)\gamma_1 + \left(1 - \frac{2}{T}\right)\gamma_2\right]
\end{aligned}$$