This document details mathematical proofs of notions that we have covered in class.

1 Proofs of standard statistical results

Proposition 1. *Linear property of the Expectation:* $E(aX+b) = aE(X)+b$ *when* X *is a random variable defined on support* Ω *and* a *and* b *are constants.*

Proof. $E(aX+b) = \int_\Omega (ax+b)f(x)dx = a\int_\Omega xf(x)dx + b = aE(X)+b.$ □

Proposition 2. *Property of the variance:* $V(X) = E[(X-E(X))^2] = E(X^2) - E(X)^2$ *in which* X *is a random variable defined on support* Ω.

Proof.

$$\begin{aligned} V(X) &= E[(X-E(X))^2] \\ &= E(X^2) + E(X)^2 - 2E(X)^2 \\ &= E(X^2) - E(X)^2 \end{aligned}$$

□

Proposition 3. *Scaling property of the variance:* $V(aX) = a^2V(X)$ *in which* X *is a random variable defined on support* Ω *and* a *is a constant.*

Proof.

$$\begin{aligned} V(aX) &= E(a^2X^2) - a^2E(X)^2 \\ &= a^2(E(X^2) - a^2E(X)^2) = a^2V(X). \end{aligned}$$

□

Proposition 4. *Let us consider* T *random variables* $\{Y_t\}_{t=1}^T \sim_{i.i.d.} N(\mu, \sigma^2)$. *Then* $\bar{Y} = \frac{\sum_{t=1}^T Y_t}{T}$ *is an unbiased estimator of* μ *(i.e.* $E(\bar{Y}) = \mu$*).*

Proof.

$$\begin{aligned} E(\bar{Y}) &= E(\sum_{t=1}^T Y_t/T) \\ &= \frac{1}{T}\sum_{t=1}^T E(Y_t) \\ &= \mu \end{aligned}$$

$\square$

Proposition 5. *Let us assume T random variables* $\{Y_t\}_{t=1}^T \sim_{i.i.d.} N(\mu, \sigma^2)$. *Then* $\hat{\sigma}^2 = \frac{\sum_{t=1}^T (Y_t - \bar{Y})^2}{T-1}$ *is an unbiased estimator of* σ^2.

Proof.

$$\begin{aligned}
E(\hat{\sigma}^2) &= (T-1)^{-1} E(\sum_{t=1}^T ((Y_t - \mu) + (\mu - \bar{Y}))^2) \\
&= (T-1)^{-1} E(\sum_{t=1}^T [(y_t - \mu)^2 + (\mu - \bar{Y})^2 + 2(\mu - \bar{Y})(Y_t - \mu)]) \\
&= (T-1)^{-1} (E[\sum_{t=1}^T (y_t - \mu)^2] + E[\sum_{t=1}^T (\mu - \bar{Y})^2] - 2E[(\mu - \bar{Y}) \sum_{t=1}^T (\mu - y_t)]) \\
&= (T-1)^{-1} (T\sigma^2 + TE[(\mu - \bar{Y})^2] - 2TE[(\mu - \bar{Y})(\mu - \bar{Y})]) \\
&= (T-1)^{-1} (T\sigma^2 - E[T(\mu - \bar{Y})^2]) = (T-1)^{-1}(T-1)\sigma^2
\end{aligned}$$

where the last equality holds because $V(\bar{Y}) = E[(\mu - \bar{Y})^2] = \frac{\sigma^2}{T}$. In fact $V(\bar{Y}) = \frac{1}{T^2} \sum_{t=1}^T V(Y_t) = \frac{\sigma^2}{T}$. $\square$

Proposition 6. *Law of iterated expectations:* $E_X(X) = E_Y(E_X(X|Y))$ *in which* X *and* Y *are random variables (with supports* Ω_x *and* Ω_y*, respectively).*

Proof.

$$\begin{aligned}
E_X(X) &= \int_{\Omega_x} x f(x) dx \\
&= \int_{\Omega_x} x \int_{\Omega_y} f(x,y) dy dx \\
&= \int_{\Omega_y} \int_{\Omega_x} x f(x,y) dx dy \\
&= \int_{\Omega_y} [\int_{\Omega_x} x f(x|y) dx] f(y) dy \\
&= E_Y(E_X(X|Y)).
\end{aligned}$$

$\square$

Proposition 7. *Let us assume M and N random variables* $\{X_i\}_{i=1}^m$,$\{Y_j\}_{j=1}^n$. *The covariance of* $Z_1 = \sum_{i=1}^m X_i$ *and* $Z_2 = \sum_{j=1}^n Y_i$ *is given by* $Cov(\sum_{i=1}^m X_i, \sum_{j=1}^n Y_i) = \sum_i \sum_j Cov(X_i, Y_j)$.

Proof. We start by showing that $Cov(\sum_{i=1}^{m} X_i, Y_1) = \sum_{i=1}^{m} Cov(X_i, Y_1)$.

$$\begin{aligned} Cov(\sum_{i=1}^{m} X_i, Y_1) &= E([\sum_{i=1}^{m} X_i]Y_1) - E([\sum_{i=1}^{m} X_i])E(Y_1) \\ &= \sum_{i=1}^{m} E(X_i Y_1) - E(X_i)E(Y_1), \\ &= \sum_{i=1}^{m} Cov(X_i, Y_1). \end{aligned}$$

Since the covariance is symmetric (i.e. $Cov(X, Y) = Cov(Y, X)$), the converse result also holds true, that is $Cov(X_1, \sum_{j=1}^{n} Y_j) = \sum_{j=1}^{n} Cov(X_1, Y_j)$.

$$\begin{aligned} Cov(\sum_{i=1}^{m} X_i, \sum_{j=1}^{n} Y_i) &= \sum_{i=1}^{m} Cov(X_i, \sum_{j=1}^{n} Y_i), \\ &= \sum_{i=1}^{m} \sum_{j=1}^{n} Cov(X_i, Y_j) \end{aligned}$$

□

Corollary 1. *Let us assume* M *random variables* $\{X_i\}_{i=1}^{m}$. *The sum of these random variables is denoted by* $Z = \sum_{i=1}^{m} X_i$. *Then,* $Var(Z) = \sum_{i=1}^{m} \sum_{j=1}^{m} Cov(X_i, X_j)$. *Note also that* $Z = \sum_{i=1}^{m} X_i = \mathbb{1}'_m X$ *in which* $\mathbb{1}_m = (1, 1, \ldots, 1)' \in \Re^{mx1}$. *So the result implies that* $Var(\mathbb{1}'_m X) = \mathbb{1}'_m Var(X) \mathbb{1}_m$.

Proof. Since $\text{Var}(\sum_{i=1}^{m} X_i) = Cov(\sum_{i=1}^{m} X_i, \sum_{j=1}^{m} X_j)$, we just apply proposition 7 to prove the result. □

Proposition 8. *Let us denote by* $f_X(x)$ *the density function of a random variable* X. *Assuming that* $Z = X - a$ *in which* a *is a constant, then the density function of* Z *is given by* $f(z) = f_X(z + a)$.

Proof. The cumulative density function (cdf) of Z is given by

$$P[Z \leq z] = P[X \leq z + a].$$

The probability density function is the derivative of the cdf. Therefore, we have

$$\begin{aligned} f(z) = \frac{dP[Z \leq z]}{dz} &= \frac{dP[X \leq z + a]}{d(z + a)} \frac{d(z + a)}{dz}, \\ &= f_X(z + a). \end{aligned}$$

□

Proposition 9. *Let us denote by $f_X(x)$ the density function of a random variable X. Assuming that $Z = \frac{X-a}{b}$ in which a and b are constant with $b \neq 0$, then the density function of Z is given by $f(z) = f_X(bz+a)b$.*

Proof. The cumulative density function (cdf) of Z is given by

$$P[Z \leq z] \quad = \quad P[X \leq bz + a].$$

The probability density function is the derivative of the cdf. Therefore, we have

$$\begin{aligned} f(z) = \frac{dP[Z \leq z]}{dz} &= \frac{dP[X \leq bz+a]}{d(bz+a)} \frac{d(bz+a)}{dz}, \\ &= b f_X(bz+a). \end{aligned}$$

□

Corollary 2. *Let us assume that $X \sim N(\mu, \sigma^2)$. Its density function is given by $f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\big(-\frac{(x-\mu)^2}{2\sigma^2}\big)$. Then, the random variable $Z = \frac{X-\mu}{\sigma}$ exhibits a density function given by $f(z) = \frac{1}{\sqrt{2\pi}} \exp\big(-\frac{z^2}{2}\big)$, i.e. $Z \sim N(0,1)$.*

Proof. By applying proposition 9 with $a = \mu$ and $b = \sigma$, we have that $f(z) = \sigma f_X(\sigma z + \mu)$. It leads to the following simplifications:

$$\begin{aligned} f(z) &= \sigma \frac{1}{\sqrt{2\pi\sigma^2}} \exp\big(-\frac{(\sigma z + \mu - \mu)^2}{2\sigma^2}\big), \\ &= \frac{1}{\sqrt{2\pi}} \exp\big(-\frac{z^2}{2}\big). \end{aligned}$$

□

2 Linear regression

In this section, we consider a multiple linear regression given by

$$\begin{aligned} y_t &= \sum_{i=1}^{K} \beta_i x_{t,i} + \epsilon_t, \\ &= x_t' \beta + \epsilon_t, \end{aligned}$$

in which $\beta = (\beta_1, \ldots, \beta_K)'$ and $x_t = (x_{t,1}, \ldots, x_{t,K})'$. The sum of squared residuals (SSR), i.e. $SSR(\hat{\beta}) = \sum_{t=1}^{T} \hat{\epsilon}_t^2 = \sum_{t=1}^{T} (y_t - x_t'\hat{\beta})^2$, is the standard criterion used to derive the estimators of β. Note that we can write the linear regression in a matrix expression as follows

$$y \quad = \quad X\beta + \epsilon,$$

where $X = \begin{pmatrix} x'_1 \\ x'_2 \\ ... \\ x'_T \end{pmatrix} = \begin{pmatrix} x_{1,1} & \dots & x_{1,K} \\ x_{2,1} & \dots & x_{2,K} \\ ... & & \\ x_{T,1} & \dots & x_{T,K} \end{pmatrix}$, $y = \begin{pmatrix} y_1 \\ y_2 \\ ... \\ y_T \end{pmatrix}$ and $\epsilon = \begin{pmatrix} \epsilon_1 \\ \epsilon_2 \\ ... \\ \epsilon_T \end{pmatrix}$.

Proposition 10. *The OLS estimator is given by* $\hat{\beta} = (X'X)^{-1}X'y$.

Proof. We first remind that $\frac{d(a'\beta)}{d\beta} = a$ and that $\frac{d(\beta' A\beta)}{d\beta} = 2A\beta$ in which A is a matrix and a is a vector. We want to minimize the sum of squared residuals given by $\text{Argmin}_\beta \epsilon'\epsilon = \sum_{t=1}^{T} \epsilon_t^2$. It leads to

$$\begin{aligned} \epsilon'\epsilon &= (y - X\beta)'(y - X\beta), \\ &= y'y + \beta'(X'X)\beta - 2y'X\beta \\ \frac{d\epsilon'\epsilon}{d\beta} &= 2(X'X)\beta - 2X'y \\ \hat{\beta} &= (X'X)^{-1}X'y. \end{aligned}$$

The solution minimizes the SSR function since $\frac{d^2e'e}{d\beta^2} = (X'X)$ is a definite positive matrix. Note that the proof implies that $X'y - (X'X)\hat{\beta} = X'(y - X\hat{\beta}) = X'e = 0$. For instance, if the regression exhibits a constant (i.e. the first explanatory variable is fixed $x_{t,1} = 1$ for all t), we have that $\sum_{t=1}^{T} e_t = 0$. So, the error term average is equal to zero. □

Proposition 11. *If the no-multicollinearity assumption and the strict exogeneity assumption hold, then the OLS estimator is unbiased.*

Proof. Note that the OLS estimators can be equivalently expressed as

$$\begin{aligned} \hat{\beta} &= (X'X)^{-1}X'y, \\ &= (X'X)^{-1}X'(X\beta + \epsilon), \\ &= \beta + (X'X)^{-1}X'\epsilon. \end{aligned}$$

The strict exogeneity assumption implies that $E(\epsilon|X) = 0$. It means that any random variable such as $Z = f(X)\epsilon$, in which $f(X)$ is a function of the explanatory variable, has an expectation equal to zero because $E(Z|X) = E(f(X)\epsilon|X) = f(X)E(\epsilon|X) = 0$. For the OLS estimator, we apply this property with $f(X) = (X'X)^{-1}X'$ and we have that

$$\begin{aligned} E(\hat{\beta}|X) &= \beta + E((X'X)^{-1}X'\epsilon|X), \\ &= \beta. \end{aligned}$$

□

Proposition 12. *If the no-multicollinearity assumption and the strict exogeneity assumption hold, then the OLS estimator is consistent if a law of large number applies for $\{x_t x_t'\}$ as well as for $\{x_t \epsilon_t\}$ (i.e. $\frac{1}{T}\sum_{t=1}^{T} x_t \epsilon_t \to_p E(x_t \epsilon_t) < \infty$ and $\frac{1}{T}\sum_{t=1}^{T} x_t x_t' \to_p E(x_t x_t') < \infty$).*

Proof. Note that the OLS estimators can be equivalently expressed as

$$\begin{aligned}
\hat{\beta} &= (X'X)^{-1}X'y, \\
&= (X'X)^{-1}X'(X\beta + \epsilon), \\
&= \beta + (X'X)^{-1}X'\epsilon, \\
&= \beta + (\frac{1}{T}\sum_{t=1}^{T} x_t x_t')^{-1}(\frac{1}{T}\sum_{t=1}^{T} x_t \epsilon_t),
\end{aligned}$$

The strict exogeneity assumption implies that $E(\epsilon|X) = 0$. It means that any random variable such as $Z = f(X)\epsilon$, in which $f(X)$ is a function of the explanatory variable, has an expectation equal to zero because $E(Z|X) = E(f(X)\epsilon|X) = f(X)E(\epsilon|X) = 0$. In addition, we have assumed that $\frac{1}{T}\sum_{t=1}^{T} x_t \epsilon_t \to_p E(x_t \epsilon_t)$ and $\frac{1}{T}\sum_{t=1}^{T} x_t x_t' \to_p E(x_t x_t')$. Since convergence in probability is preserved with continuous transformation and because $X'X$ is invertible (by assumption of no-multicollinearity), we have that

$$\begin{aligned}
\hat{\beta} &\to_p \beta + E(x_t x_t')^{-1} E(x_t \epsilon_t), \\
&= \beta \quad \text{since } E(x_t \epsilon_t) = 0.
\end{aligned}$$

□

Proposition 13. *If assumptions 1 to 6 of the finite sample framework hold (so from linearity to Normally distributed error terms), the OLS estimator is normally distributed.*

Proof. First note that $\hat{\beta} = (X'X)^{-1}X'y = \beta + (X'X)^{-1}X'\epsilon$. Therefore, we have that

$$\begin{aligned}
\text{Var}(\hat{\beta} - \beta|X) &= (X'X)^{-1}X'\text{Var}(\epsilon|X)X(X'X)^{-1}, \\
&= \Sigma^2 (X'X)^{-1}.
\end{aligned}$$

Since $\hat{\beta} - \beta = X(X'X)^{-1}X'\epsilon$ is a linear combination of random variables that are jointly normally distributed (i.e. $\epsilon \sim N(0, \sigma^2 I_T)$), $\hat{\beta} - \beta|X$ is also normally distributed (to proof that a linear combination of normally distributed random variables remains normally distributed, we can use the characteristic function). In addition, we know the conditional expectation and variance of the Normal distribution. So, $\hat{\beta} - \beta | X \sim N(0, \Sigma^2 (X'X)^{-1})$. $\square$

2.1 Asymptotic framework

Proposition 14. *Cauchy-Schwartz:* $|E(XY)| \leq \sqrt{E(X^2)}\sqrt{E(Y^2)}$

Proof. We define the following optimization program: $c^* = \text{Arg min.}_{c^*} E([X - cY]^2) = c^2 E(Y^2) - 2cE(XY) + E(X^2)$. Note that the objective function is convex with respect to c (if a minimum exists, it is the global minimum). The global minimum is given by

$$\begin{aligned} \frac{d}{dc} &= 2cE(Y^2) - 2E(XY), \\ c^* &= \frac{E(XY)}{E(Y^2)}. \end{aligned}$$

Note that $E([X - cY]^2) \geq 0 \; \forall c \in \Re$. Of course, this is also true for the global minimizer c^*. Evaluating the objective function at the minimum leads to

$$\begin{aligned} E([X - c^* Y]^2) &= (\frac{E(XY)}{E(Y^2)})^2 E(Y^2) - 2\frac{E(XY)}{E(Y^2)} E(XY) + E(X^2), \\ &= E(X^2) - \frac{E(XY)^2}{E(Y^2)} \quad (\geq 0), \\ E(X^2) - \frac{E(XY)^2}{E(Y^2)} &\geq 0, \\ E(XY)^2 &\leq E(X^2)E(Y^2), \\ |E(XY)| &\leq \sqrt{E(X^2)E(Y^2)}. \end{aligned}$$

If $|E(XY)| = \sqrt{E(X^2)E(Y^2)}$, we have that $E([X - c^* Y]^2) = 0$ implying that $X - c^* Y = 0$ and $X = c^* Y = \frac{E(XY)}{E(Y^2)} Y$. $\square$

Proposition 15. ***Markov's inequality.*** *Let us assume a non-negative random variable X with finite expectation, i.e. $E(X) < \infty$. Then, $\forall a > 0$: $P[X \geq a] \leq \frac{E(X)}{a}$.*

Proof. We consider the following inequality: $\mathbb{1}_{\{x \geq a\}} \leq \frac{x}{a}$. If we multiply both side by the density function $f(x)$ (note that $f(x) \geq 0 \;\; \forall x$) and then we integrate the function with respect to x, we get

$$\begin{aligned} f(x)\mathbb{1}_{\{x\geq a\}} &\leq f(x)\frac{x}{a}, \\ \int_{\Re^+} f(x)\mathbb{1}_{\{x\geq a\}}dx &\leq \frac{1}{a}\int_{\Re^+} xf(x)dx, \\ P[X\geq a] &\leq E(\frac{X}{a}). \end{aligned}$$

□

Proposition 16. ***Chebishev's inquality.*** *Let us assume a random variable X with finite expectation and variance, i.e. $\mu \equiv E(X) < \infty$ and $\sigma^2 \equiv E\big((X-\mu)^2\big) < \infty$. Then, $\forall a > 0$: $P[|X-\mu| \geq a] \leq \frac{\sigma^2}{a^2}$*

Proof. We consider the positive random variable $Z^2 = (X-\mu)^2 \leq 0$. Since Z^2 is a positive random variable, we can apply the Markov's inequality. It leads to

$$\begin{aligned} P[Z^2 \geq a^2] &\leq E(\frac{Z^2}{a^2}), \\ P[|Z| \geq a] &\leq \frac{E((X-\mu)^2)}{a^2}, \\ &\leq \frac{\sigma^2}{a^2}. \end{aligned}$$

□

Proposition 17. *Mean square convergence implies convergence in probability: $Z_t \to_{m.s.} \alpha$ then $Z_t \to_p \alpha$.*

Proof. We consider the random variable $Q_t = Z_t - \alpha$. By applying the Chebyshev's inquality with $a = \epsilon > 0$, we have that

$$\begin{aligned} P[|Q_t| \geq \epsilon] &\leq \frac{E((Z_t-\alpha)^2)}{\epsilon^2}, \\ lim_{t\to\infty} P[|Q_t| \geq \epsilon] &\leq lim_{t\to\infty}\frac{E((Z_t-\alpha)^2)}{\epsilon^2}, \\ lim_{t\to\infty} P[|Z_t-\alpha| \geq \epsilon] &\leq 0. \end{aligned}$$

□

Proposition 18. ***Delta method.*** *Let us assume a series of $K-$dimensional r.v. $\{Z_t\}$, and $\sqrt{t}(Z_t - \beta) \to_d Z$ and $Z_t \to_p \beta \in \Re^K$. Let us assume a continuous function $a(\cdot) : \Re^K \to \Re^r$ with continuous first derivatives given by $A(\beta) \in \Re^{r,K}$ then $\sqrt{t}(a(Z_t) - a(\beta)) \to_d A(\beta)Z$.*

Proof. Using the mean value theorem, it exists $y_t \in [Z_t, \beta]$ such that $A(y_t) = \frac{a(\beta)-a(Z_t)}{\beta - Z_t}$. It implies that $A(y_t)(\beta - Z_t) = a(\beta) - a(Z_t)$. Therefore, we have

$$A(y_t)\sqrt{t}(\beta - Z_t) \quad = \quad \sqrt{t}(a(\beta) - a(Z_t)).$$

Since $Z_t \rightarrow_p \beta$ and $y_t \in [Z_t, \beta]$, we conclude that $y_t \rightarrow_p \beta$. In addition, since $A(\cdot)$ is a continuous function by assumption, the convergence in probability is preserved and so, $A(y_t) \rightarrow_p A(\beta)$. As $\sqrt{t}(\beta - Z_t) \rightarrow_d Z$ (by assumption), we have that $\sqrt{t}(a(\beta) - a(Z_t)) = A(y_t)\sqrt{t}(\beta - Z_t) \rightarrow_d A(\beta)Z$. □

Proposition 19. *Weak white noise is not a martingale difference sequence.*

Proof. Let us assume that $\{\epsilon_t\}$ is a weak white noise. We know that $E(\epsilon_t) = 0$ and $\text{Var}(\epsilon_t) = \sigma^2 < \infty$. Using the expected law of iteration, we have that $E(\epsilon_t) = E(E(\epsilon_t|\epsilon_{t-1}, ...)) = 0$. However, this does not imply that $E(\epsilon_t|\epsilon_{t-1}, ...) = 0$ (the condition for being a m.d.s.). In fact, it only implies that the random variable $E(\epsilon_t|\epsilon_{t-1}, ...)$ (which can be a function of $\epsilon_{t-1}, ...$) exhibits an expectation of zero. A simple example of w.w.n. that is not an m.d.s. is the following process: $\epsilon_t = \epsilon_{t-1}$ with $\epsilon_0 \sim N(0, 1)$.
On the contrary, if $\{\epsilon_t\}$ is a m.d.s., then $E(\epsilon_t) = E(E(\epsilon_t|\epsilon_{t-1}, ...)) = 0$ and so $E(\epsilon_t) = 0$. An m.d.s. is therefore a w.w.n. □

Proposition 20. *Why is an m.d.s. necessarily uncorrelated ?*

Proof. Let us assume that $\{\epsilon_t\}$ is an m.d.s. We have that

$$\begin{aligned}\gamma_j &= E(\epsilon_t \epsilon_{t-j}),\\ &= E(E(\epsilon_t|\epsilon_{t-1}, ...)\epsilon_{t-j}) = 0 \qquad \forall j > 0.\end{aligned}$$

□

Proposition 21. *Why is an m.d.s. more general than an i.i.d. process ?*

Proof. Let us assume that $\{\epsilon_t\}$ is an m.d.s. We do not assume anything about higher moments such as $\text{Var}(\epsilon_t|\epsilon_{t-1}, ...) = E(\epsilon_t^2|\epsilon_{t-1}, ...) =?$. In addition, $\{\epsilon_t\}$ is not necessarily stationary. So, the random variables composing the m.d.s. can be distributed differently over time. □

Proposition 22. *Let $\{y_1, ..., y_T\}$ be a stationary and ergodic stochastic process. Then,*

$$\hat{\rho}_j = \frac{\hat{\gamma}_j}{\hat{\gamma}_0} = \frac{T^{-1}\sum_{t=j+1}^{T}(y_t - \bar{y}_T)(y_{t-j} - \bar{y}_T)}{T^{-1}\sum_{t=1}^{T}(y_t - \bar{y}_T)^2} \to_p \rho_j = \frac{E((y_t - \mu)(y_{t-j} - \mu))}{E((y_t - \mu)^2)}$$

Proof. Focusing first on the numerator, we have that

$$\begin{aligned}
\text{Num of } \hat{\rho}_j \quad &= \quad \frac{1}{T}\sum_{t=j+1}^{T}(y_t y_{t-j} + \bar{y}_T^2 - y_{t-j}\bar{y}_T - y_t\bar{y}_T), \\
\text{First term:} \quad &= \quad \frac{1}{T}\sum_{t=j+1}^{T} y_t y_{t-j}, \\
&= \quad \frac{T-j}{T}\frac{1}{T-j}\sum_{t=j+1}^{T} y_t y_{t-j}, \\
&\to_p \quad 1E(y_t y_{t-j}), \\
\text{Second term:} \quad &= \quad \frac{1}{T}\sum_{t=j+1}^{T} \bar{y}_T^2, \\
&= \quad \bar{y}_T^2\frac{T-j}{T} \quad (\to_p E(y_t)^2), \\
\text{Third and fourth terms:} \quad &= \quad \frac{1}{T}\sum_{t=j+1}^{T} y_{t-j}\bar{y}_T, \\
&= \quad \bar{y}_T\frac{T-j}{T}\frac{1}{T-j}\sum_{t=j+1}^{T} y_{t-j} \quad (\to_p E(y_t)^2).
\end{aligned}$$

We conclude that the numerator converges to

$$\begin{aligned}
\frac{1}{T}\sum_{t=j+1}^{T}(y_t - \bar{y}_T)(y_{t-j} - \bar{y}_T) \quad &\to_p \quad E(y_t y_{t-j}) + E(y_t)^2 - 2E(y_t)^2, \\
&= \quad E(y_t y_{t-j}) - E(y_t)^2 \quad (= Cov(y_t, y_{t-j})).
\end{aligned}$$

The denominator can be decomposed as follows

$$\begin{aligned}
\text{Deno of } \hat{\rho}_j \quad &= \quad \frac{1}{T}\sum_{t=1}^{T}(y_t^2 + \bar{y}_T^2 - 2y_t\bar{y}_T), \\
&\to_p \quad E(y_t^2) - E(y_t)^2 \quad (= \text{Var}(y_t)).
\end{aligned}$$

We have a continuous function of two quantities that converge in probability. Since continuous transformation preserves the convergence in probability, we conclude that

$$\frac{\hat{\gamma}_j}{\hat{\gamma}_0} \to_p \frac{\gamma_j}{\gamma_0} \quad (= Corr(y_t, y_{t-j})).$$

□

Proposition 23. *Let a stationary and ergodic m.d.s. $\{\epsilon_t\}$ and a process $y_t = \mu + \epsilon_t$. then $\sqrt{T}\hat{\gamma} \to_d N(0, \sigma^4 I_d)$.*

Proof. Note first that $e_t = y_t - \bar{y}_T = \mu + \epsilon_t - \bar{y}_T$. We have that $\hat{\gamma}_j = \frac{1}{T}\sum_{t=j+1}^{T}(y_t - \bar{y}_T)(y_{t-j} - \bar{y}_T) = \frac{1}{T}\sum_{t=j+1}^{T} e_t e_{t-j}$. Expanding the terms give the following mathematical simplifications,

$$\begin{aligned}\frac{1}{T}\sum_{t=j+1}^{T} e_t e_{t-j} &= \frac{1}{T}\sum_{t=j+1}^{T}\epsilon_t\epsilon_{t-j} + (\mu - \bar{y}_T)\frac{1}{T}\sum_{t=j+1}^{T}\epsilon_t + (\mu - \bar{y}_T)\frac{1}{T}\sum_{t=j+1}^{T}\epsilon_{t-j} + (\mu - \bar{y}_T)^2,\\ &\to_p E(\epsilon_t\epsilon_{t-j}) + 0E(\epsilon_t) + 0E(\epsilon_t) + 0,\\ &\to_p E(\epsilon_t\epsilon_{t-j}).\end{aligned}$$

If we multiply by $\sqrt{T}$, we have that

$$\frac{1}{\sqrt{T}}\sum_{t=j+1}^{T} e_t e_{t-j} = \frac{1}{\sqrt{T}}\sum_{t=j+1}^{T}\epsilon_t\epsilon_{t-j} + (\mu - \bar{y}_T)\frac{1}{\sqrt{T}}\sum_{t=j+1}^{T}\epsilon_t + (\mu - \bar{y}_T)\frac{1}{\sqrt{T}}\sum_{t=j+1}^{T}\epsilon_{t-j} + \sqrt{T}(\mu - \bar{y}_T)^2.$$

Since $\mu - \bar{y}_T = -\frac{1}{T}\sum_{t=1}^{T}\epsilon_t \to_p 0$, the terms $(\mu - \bar{y}_T)\frac{1}{\sqrt{T}}\sum_{t=j+1}^{T}\epsilon_t \to_p 0N(0,\sigma^2) \equiv 0$ and $(\mu - \bar{y}_T)\frac{1}{\sqrt{T}}\sum_{t=j+1}^{T}\epsilon_{t-j} \to_p 0$. In addition, the term $\sqrt{T}(\mu - \bar{y}_T)^2 = \frac{1}{\sqrt{T}}(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\epsilon_t)^2 \to_p 0$. The final term can be decomposed as follows

$$\begin{aligned}\frac{1}{\sqrt{T}}\sum_{t=j+1}^{T}\epsilon_t\epsilon_{t-j} &= \frac{\sqrt{T-j}}{\sqrt{T}}\frac{1}{\sqrt{T-j}}\sum_{t=j+1}^{T}\epsilon_t\epsilon_{t-j},\\ &\to_d N(E(\epsilon_t\epsilon_{t-j}), \mathrm{Var}(\epsilon_t\epsilon_{t-j})).\end{aligned}$$

where the last result comes from the fact that $\{\epsilon_t\epsilon_{t-j}\}$ is an m.d.s. To prove it, note that $E(\epsilon_t\epsilon_{t-j}) = 0$ and $\mathrm{Var}(\epsilon_t\epsilon_{t-j}) = E(\epsilon_t^2\epsilon_{t-j}^2) = \sigma^4$ following the assumption about $E(\epsilon_t^2|\epsilon_{t-1}, ..., \epsilon_1) = \sigma^2$.
Moreover, $Cov(\epsilon_t\epsilon_{t-i}, \epsilon_t\epsilon_{t-j}) = E(\epsilon_t^2\epsilon_{t-i}\epsilon_{t-j}) = \sigma^2 E(\epsilon_{t-i}\epsilon_{t-j}) = 0$. Using the CLT on m.d.s. (we skip the fact that we should also show that $\lambda'\sqrt{T}\gamma \to_d \lambda' Z$ where $Z \sim N(E(\gamma), V(\gamma))$ $\forall\lambda \neq 0$), we find that

$$\sqrt{T}\begin{pmatrix}\hat{\gamma}_1\\ \hat{\gamma}_2\\ \dots\\ \hat{\gamma}_d\end{pmatrix} = \begin{pmatrix}\frac{1}{\sqrt{T}}\sum_{t=j+1}^{T} e_t e_{t-1}\\ \frac{1}{\sqrt{T}}\sum_{t=j+1}^{T} e_t e_{t-2}\\ \dots\\ \frac{1}{\sqrt{T}}\sum_{t=j+1}^{T} e_t e_{t-d}\end{pmatrix} \to_d N(\begin{pmatrix}0\\ 0\\ \dots\\ 0\end{pmatrix}, \begin{pmatrix}\sigma^4 & 0 & \dots & 0\\ 0 & \sigma^4 & \dots & 0\\ & & \dots & \\ 0 & 0 & \dots & \sigma^4\end{pmatrix}).$$

□

Proposition 24. *If assumptions 1 to 4 hold, the OLS estimator is consistent.*

Proof. As we have shown above, the OLS estimator can be written as follows $\hat{\beta} = (\frac{1}{T}\sum_{t=1}^{T} x_t x_t')^{-1}(\frac{1}{T}\sum_{t=1}^{T} x_t y$ Using the fact that $y_t = x_t'\beta + \epsilon_t$, we find that

$$\begin{aligned} \hat{\beta} &= \beta + (\frac{1}{T}\sum_{t=1}^{T} x_t x_t')^{-1}(\frac{1}{T}\sum_{t=1}^{T} x_t \epsilon_t), \\ &= \beta + S_{XX}^{-1}\bar{g}_T, \end{aligned}$$

where $\bar{g}_T = \frac{1}{T}\sum_{t=1}^{T} x_t \epsilon_t$. Since $\{x_t\}$ is a stationary and ergodic process, we conclude that $S_{XX} \to_p E(x_t x_t')$. Since convergence in probability is preserved by continuous transformation, it leads to $S_{XX}^{-1} \to_p E(x_t x_t')^{-1}$. As $\{x_t, \epsilon_t\}$ is jointly stationary and ergodic (because ϵ_t is function of y_t and x_t that are jointly stationary by assumption), we conclude that $\bar{g}_T \to_p E(x_t \epsilon_t) = 0$ by weak exogeneity. □

Proposition 25. *If assumptions 1 to 5 hold, the OLS estimator is consistent and asymtotically normal (CAN), that is*

$$\sqrt{T}(b - \beta) \overset{d}{\to} N(0, \Sigma_x^{-1} S \Sigma_x^{-1})$$

Proof. As shown in the previous proof, we know that

$$\begin{aligned} \hat{\beta} - \beta &= S_{XX}^{-1}\bar{g}_T, \\ \sqrt{T}(\hat{\beta} - \beta) &= S_{XX}^{-1}\sqrt{T}\bar{g}_T, \end{aligned}$$

where $\bar{g}_T = \frac{1}{T}\sum_{t=1}^{T} x_t \epsilon_t$. Since $\{x_t\}$ we conclude that $S_{XX} \to_p E(x_t x_t')$. As convergence in probability is preserved by continuous transformation, it also leads to $S_{XX}^{-1} \to_p E(x_t x_t')^{-1}$. Assumption 5 implies that $\{x_t \epsilon_t\}$ is a m.d.s. that is stationary and ergodic. Using the CLT related to m.d.s., we find that $\sqrt{T}\bar{g}_T \to_d N(0, \underbrace{E(g_t g_t')}_{S})$. Combining these two convergence results leads to

$$\sqrt{T}(\hat{\beta} - \beta) \to_d N(0, \Sigma_X^{-1} S \Sigma_X^{-1}),$$

where $\Sigma_X = E(x_t x_t')$.

Note that Σ_X is a symmetric matrix because $(x_t x_t')' = x_t x_t'$. □

Proposition 26. *A consistent estimator of S is given by:* $\frac{1}{T}\sum_{t=1}^{T} x_t x_t' e_t^2 \to_p E(x_t x_t' \epsilon_t^2) = S$.

Proof. We know that

$$\begin{aligned} e_t &= y_t - x_t' b \\ &= x_t'(\beta - b) + \epsilon_t. \end{aligned}$$

We develop the proof for $K = 1$. Taking the square of the error term, we get

$$\begin{aligned} e_t^2 &= (\beta - b)^2 x_t^2 + \epsilon_t^2 + 2(\beta - b) x_t \epsilon_t, \\ e_t^2 x_t^2 &= (\beta - b)^2 x_t^4 + \epsilon_t^2 x_t^2 + 2(\beta - b) x_t^3 \epsilon_t, \\ \frac{1}{T}\sum_{t=1}^{T} e_t^2 x_t^2 &= (\beta - b)^2 \frac{1}{T}\sum_{t=1}^{T} x_t^4 + \frac{1}{T}\sum_{t=1}^{T} (\epsilon_t x_t)^2 + 2(\beta - b)\frac{1}{T}\sum_{t=1}^{T} x_t^3 \epsilon_t. \end{aligned}$$

Because b is a consistent estimator, we have that $(\beta - b) \to_p 0$. In addition, $\frac{1}{T}\sum_{t=1}^{T} x_t^4 \to_p E(x_t^4) < \infty$ by assumption. The assumption also implies that $\sum_{t=1}^{T} x_t^3 \epsilon_t \to_p E(x_t^3 \epsilon_t) < \infty$ due to the Cauchy-Schwartz inequality (i.e. $E(x_t^3 \epsilon_t) \leq \sqrt{E(x_t^4) E(x_t^2 \epsilon_t^2)}$ where $E(x_t^2 \epsilon_t^2) = S < \infty$ since $\{x_t \epsilon_t\}$ is an m.d.s.)
We conclude that $\frac{1}{T}\sum_{t=1}^{T} e_t^2 x_t^2 \to_p E(x_t^2 \epsilon_t^2)$.

□

Proposition 27. *A consistent estimator for the variance of the error term is given by* $\frac{1}{T}\sum_{t=1}^{T} e_t^2 \to_p E(\epsilon_t^2) = \sigma^2$.

Proof. The residual can be written as

$$\begin{aligned} e_t &= x_t'(\beta - b) + \epsilon_t, \\ e_t^2 &= \epsilon_t^2 + 2(\beta - b)' x_t \epsilon_t + (\beta - b)'(x_t x_t')(\beta - b), \\ \frac{1}{T}\sum_{t=1}^{T} e_t^2 &= \underbrace{\frac{1}{T}\sum_{t=1}^{T} \epsilon_t^2}_{\to_p \sigma^2} + 2(\beta - b)'[\underbrace{\frac{1}{T}\sum_{t=1}^{T} x_t \epsilon_t}_{\to_p E(x_t \epsilon_t) = 0}] + (\beta - b)'[\underbrace{\frac{1}{T}\sum_{t=1}^{T} (x_t x_t')}_{\to_p \Sigma_X}]\underbrace{(\beta - b)}_{\to_p 0}. \end{aligned}$$

□