LECTURE 1, 17/02/2012

1. INTRODUCTION

- Instructor: Piotr Eliasz, language: English
- Wednesday 9-10 (confirm by email – peliasz@mimuw.edu.pl)
- Bibliography:
  - "Estimation and Inference in Econometrics", R. Davidson and J. G. MacKinnon
  - "Advanced econometrics", Takeshi Amemiya
  - "Econometrics", Fumio Hayashi

A course is designed to familiarize students with statistical methods employed in analysis of economic and financial data. Emphasis will be places on a thorough review of statistical techniques employed in small and large sample inference. Specifically, we will start with a review of matrix algebra and probability. Next, we will cover the following concepts: standard linear model in small and large samples, violations of assumptions; maximum likelihood estimation; generalized method of moments.

- Grading: final exam (50%) and take-home exercises and empirical applications (50%)

2. DEFINITIONS

- **Probability space** – (Ω, F, P)
  - Ω is a set called sample space
  - F is a family of events (an event is an element of F).
  - P is a probability measure on (Ω, F, P).

- **Random variable** is a function X : Ω → R with a property that {ω ∈ Ω : X(ω) ≤ x} ∈ F ∀ x ∈ R.

- **Distribution function** of a random variable X is the function F : R → [0, 1] given by F(x) = P(X ≤ x).

Random variable X is **continuous** if its distribution function can be expressed as F(x) = ∫ −∞x f(u)du, x ∈ R for some integrable function f : R → [0, ∞).

3. CONVERGENCE OF RANDOM VARIABLES

Let X₁, ..., Xₙ be random variables on some probability space (Ω, F, P). We say that

- **Xₙ \xrightarrow{a.s.} X almost surely** if {ω ∈ Ω : Xₙ(ω) → X(ω) as n → ∞} is an event whose probability is 1.
- **Xₙ \xrightarrow{L} X in r’th mean**, where r ≥ 1 if E(Xₙ^r) < ∞ for all n and E(|Xₙ - X|^r) → 0 as n → ∞.
- **Xₙ \xrightarrow{D} X in probability** if P(|Xₙ - X| > ε) → 0 as n → ∞ ∀ε > 0.
- **Xₙ \xrightarrow{P} X in distribution** if P(Xₙ ≤ x) → P(X ≤ x) as n → ∞ for all points at which Fₙ(x) = P(X ≤ x) is continuous.

4. IMPLICATIONS

- **Xₙ \xrightarrow{a.s.} X/Xₙ \xrightarrow{L} X ⇒ Xₙ \xrightarrow{P} X ⇒ Xₙ \xrightarrow{D} X (r ≥ 1).**
- If r > s ≥ 1, then Xₙ \xrightarrow{L} X ⇒ Xₙ \xrightarrow{a.s.} X.
- If Xₙ \xrightarrow{D} c, where c is constant, then Xₙ \xrightarrow{P} c.
- If Xₙ \xrightarrow{D} X and P(|Xₙ| ≤ k) = 1 ∀h and some k then Xₙ \xrightarrow{L} X ∀r ≥ 1.
- If \sum P(|Xₙ - X| > ε) < ∞ ∀ε > 0, then Xₙ \xrightarrow{a.s.} X.
- If Xₙ \xrightarrow{P} X then Xₙ \xrightarrow{D} X.

Converse is false: Let X be Bernoulli variable with parameter 1/2. Let X₁, ..., Xₙ be identical random variables given by Xₙ = X ∀n. Then Xₙ \xrightarrow{D} X. Now let Y = 1 − X. Clearly Xₙ \xrightarrow{D} Y. We can’t converge in any other mode as |Xₙ − Y| = 1 always.
Proof. Suppose \( X_n \overset{D}{\rightarrow} X \). Let’s write \( F_n(x) = P(X_n \leq x) \), \( F(x) = P(X \leq x) \).

\[
F_n(x) = P(X_n \leq x) = P(X_n \leq x \cap X \leq x + \varepsilon) + P(X_n \leq x \cap X > x + \varepsilon) \leq F(x + \varepsilon) + P(|X_n - X| > \varepsilon) \rightarrow 0
\]

So we obtain \( F(x - \varepsilon) \leq \liminf F_n(x) \leq \limsup F_n(x) \leq F(x + \varepsilon) \forall \varepsilon > 0 \). If \( F \) is continuous at \( x \) then \( F(x - \varepsilon) \uparrow F(x) \) and \( F(x + \varepsilon) \downarrow F(x) \) as \( \varepsilon \rightarrow 0 \). Since \( \varepsilon \) is arbitrary, \( F_n(x) \overset{D}{\rightarrow} F(x) \).

\( \square \)

5. Other

- Markov’s inequality
  If \( X \) is any random variable with finite mean then \( P(|X| \geq a) \leq \frac{E(X)}{a} \) for any \( a > 0 \).

\( \square \)

- Skorokhod’s representation theorem
  If \( \{X_n\} \) and \( X \) with distribution function \( \{F_n\} \) and \( F \) are such that \( X_n \overset{D}{\rightarrow} X \), then there exists a probability space \( (\Omega', \mathcal{F}', P') \) and random variables \( \{Y_n\} \) and \( Y \), which map \( \Omega' \) into \( \mathbb{R} \) such that
  - \( \{Y_n\} \) and \( Y \) have distribution functions \( \{F_n\} \) and \( F \)
  - \( Y_n \overset{D}{\rightarrow} Y \) as \( n \rightarrow \infty \).

\( \square \)

- Corollary
  If \( X_n \overset{D}{\rightarrow} X \) and \( g: \mathbb{R} \rightarrow \mathbb{R} \) is continuous, then \( g(X_n) \overset{D}{\rightarrow} g(X) \).

\( \square \)

6. Laws of Large Numbers (LLN)

Let \( \{X_i\} \) be a sequence of random variables with partial sums \( S_n = \sum_{i=1}^{n} X_i \).

- Kolmogorov’s LLN
  Let \( X_1, X_2, \ldots \) be independent identically distributed (i.i.d.) random variables.
  Then \( \frac{1}{n} \sum_{i=1}^{n} X_i \overset{a.s.}{\rightarrow} \mu \) if and only if \( \mathbb{E}|X_i| < \infty \) and \( \mathbb{E}X_i = \mu \).

- Kolmogorov’s LLN
  Let \( X_1, X_2, \ldots \) be independent (but not identical) with \( \mathbb{E}X_i = \mu_i \) and \( \text{Var}X_i = \sigma_i^2 \).
  If \( \sum_{i=1}^{n} \frac{\sigma_i^2}{n} < \infty \) then \( \frac{1}{n} \sum_{i=1}^{n} X_i - \frac{1}{n} \sum_{i=1}^{n} \mu_i \overset{a.s.}{\rightarrow} 0 \) (or written \( \bar{X}_n - \mu \overset{a.s.}{\rightarrow} 0 \)).

7. Central Limit Theorems (CLT)

- Lindeberg-Lévý CLT
  Let \( X_1, X_2, \ldots \) be a sequence of i.i.d. random variables with finite means \( \mu \) and finite non-zero variances \( \sigma^2 \). Let \( S_n = \sum_{i=1}^{n} X_i \).
  Then \( \frac{S_n - n \mu}{\sqrt{n} \sigma} \overset{D}{\rightarrow} \mathcal{N}(0, 1) \) or \( \sqrt{n} \left( \frac{1}{n} \sum_{i=1}^{n} (X_i - \mu) \right) \overset{D}{\rightarrow} \mathcal{N}(0, 1) \).

- Lindeberg-Feller CLT
  Let \( X_1, X_2, \ldots \) be a sequence of independent random variables with \( \mathbb{E}X_i = \mu_i \), \( \text{Var}X_i = \sigma_i^2 \) and distribution function \( F_i \).
  Then \( \sqrt{n} \left( \frac{\sum_{i=1}^{n} X_i - n \mu}{\sqrt{n} \sum_{i=1}^{n} \sigma_i} \right) \overset{D}{\rightarrow} \mathcal{N}(0, 1) \) (random variable with normal distribution)

  and \( \lim_{n \rightarrow \infty} \max_{1 \leq i \leq n} n^{-1} \left( \frac{\sigma_i^2}{n} \right) = 0 \) (where \( \sigma_n^2 = \frac{1}{n} \sum_{i=1}^{n} \sigma_i^2 \)).

CLTs above are satisfied if and only if for any \( \varepsilon > 0 \), \( \lim_{n \rightarrow \infty} \sigma_n^{-2} n^{-1} \sum_{i=1}^{n} \int_{(x-\mu_i)^2 > \varepsilon n \sigma_i^2} (x-\mu_i)^2 dF_i(x) = 0 \).

- Lyapunov’s CLT
  Let \( X_1, X_2, \ldots \) be a sequence of independent random variables with \( \mathbb{E}X_i = \mu_i \), \( \text{Var}X_i = \sigma_i^2 \), \( \sigma_i^2 \neq 0 \) and
  \( \mathbb{E}|X_i - \mu_i|^2 + \delta < M < \infty \) for some \( \delta > 0 \) \( \forall \delta \).

If \( \sigma_n^2 > \delta > 0 \) \( \forall \delta \) sufficiently large, then \( \sqrt{n} \left( \frac{\sum_{i=1}^{n} X_i - n \mu}{\sqrt{n} \sum_{i=1}^{n} \sigma_i} \right) \overset{D}{\rightarrow} \mathcal{N}(0, 1) \).
1. Exercise 1
Let $X, Y \sim \text{Bernoulli with parameter } 1/2$. Consider $X + Y$ and $|X - Y|$. 

Note

\[ \text{Cov}(X + Y, |X - Y|) = \mathbb{E}((X + Y)|X - Y|) - \mathbb{E}(X + Y)\mathbb{E}|X - Y| = \]
\[ = \frac{1}{4} + \frac{1}{4} \quad \text{(only when } X = 0, Y = 1 \text{ or } X = 1, Y = 0) - \left( \frac{1}{4} + \frac{1}{4} + \frac{1}{2} \right) = 0 \]

\[ \mathbb{P}(X + Y = 0, |X - Y| = 0) = \frac{1}{4} \]
\[ \mathbb{P}(X + Y = 0)\mathbb{P}(|X - Y| = 0) = \frac{1}{4} \cdot \frac{1}{2} \neq \frac{1}{4} \]

So correlation is 0, but variables are not independent.

2. Exercise 2
Let $X$ and $Y$ have joint probability distribution (bivariate normal, where $\rho$ is a constant $-1 < \rho < 1$)

\[ f(x, y) = \frac{1}{2\pi \sqrt{1 - \rho^2}} \exp\left[ -\frac{1}{2(1 - \rho^2)}(x^2 - 2\rho xy + y^2) \right] \]
\[ = \frac{1}{\sqrt{2\pi}} \exp\left[ -\frac{1}{2(1 - \rho^2)}(x - \rho y)^2 \right] \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}y^2\right) = g(x, y)h(y) \]

Now, \[ f_Y(y) = \int_{-\infty}^{\infty} g(x, y)h(y)dx = h(y) \int_{-\infty}^{\infty} g(x, y)dx = h(y) \cdot 1 \quad (1 \text{ as it's a normal density).} \]

By symmetry \[ f_X(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}x^2\right) \mathcal{N}(0, 1). \]

\[ \text{Cov}(X, Y) = \iint x y f(x, y)\,dxdy = \text{(only this, because } \mu = 0) \]
\[ = \iint x y g(x, y)h(y)\,dxdy = \int y h(y)\left[ \int x g(x, y)\,dx \right]dy = \int y h(y)\rho ydy = \]
\[ = \rho \int y^2 h(y)dy = \rho \cdot 1 \quad \text{(as variance = 1)} \]

If $\rho = 0$ then $f(x, y) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}x^2\right) \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}y^2\right) = f_X(x)f_Y(y)$. 

So $X$ and $Y$ are independent.

3. Exercise 3

- If $\text{Var}X = 0 \Rightarrow X$ is a constant.

Note
\[ \mathbb{E}(X^2) = \sum x^2 \mathbb{P}(X = x) = 0 \Rightarrow \mathbb{P}(X = x) = 0 \forall x \neq 0 \Rightarrow \mathbb{P}(X = 0) = 1 \]
\[ \text{Var}X = 0 \Rightarrow \mathbb{P}(X - EX = 0) = 1 \Rightarrow X = \text{constant} \]

- Take $\mathbb{E}(X^2) > 0$, $\mathbb{E}(Y^2) > 0$. For $a, b \in \mathbb{R}$, let $Z = aX - bY$

\[ 0 \leq \mathbb{E}(Z^2) = a^2\mathbb{E}(X^2) - 2ab\mathbb{E}XY + b^2\mathbb{E}(Y^2) \]

Consider its as a quadratic \(- if b \neq 0 \quad \Delta = 4b^2\mathbb{E}(XY)^2 - 4\mathbb{E}(X^2)b^2\mathbb{E}(Y^2) \leq 0\)

\[ \mathbb{E}(XY)^2 \leq \mathbb{E}(X^2)\mathbb{E}(Y^2) - \text{Cauchy-Schwartz inequality} \]

Note
\[ \mathbb{E}(XY)^2 = \mathbb{E}(X^2)\mathbb{E}(Y^2) \text{ only if } \mathbb{P}(aX = bY) = 1 \text{ (Z = 0) for some } a, b \in \mathbb{R} \text{ and } b \neq 0. \]

- From Cauchy-Schwartz
\[ \mathbb{E}[(X - EX)(Y - EY)]^2 \leq \mathbb{E}[(X - EX)^2]\mathbb{E}[(Y - EY)^2] = \text{Var}X\text{Var}Y \]

Taking square roots
\[ |\text{Cov}(X, Y)| = \mathbb{E}[(X - EX)(Y - EY)] \leq \sqrt{\text{Var}X \text{Var}Y} \Rightarrow |\rho(X, Y)| = \frac{|\text{Cov}(X, Y)|}{\sqrt{\text{Var}X \text{Var}Y}} \leq 1 \]

4. Exercise 4
Let $X_1, X_2, \ldots$ be a sequence of random variables with $\mathbb{E}X_t = 0$ and $\text{Var}(X_t) = \sigma_t^2 < c < \infty$. Let $\text{corr}(X_s, X_t) = \rho_{st}$. Show that if $\rho_{st} \to 0$ as $|s - t| \to \infty$ then $\bar{X}_n \xrightarrow{\text{ms}} 0$ (convergence in mean-square, $r = 2$).
LECTURE 2, 24/02/2012

1. INTRODUCTION

• Let us consider a set of data \{z_t\}_{t=1}^T (not necessarily a time series). Let this data be distributed as \(f(z_t, \theta_0)\) (known p.d.f. – probability density function), where \(\theta_0\) is the true value of parameter \(\theta\).

• In this course we will be concerned mostly with the case when \(z_t = \{y_t, x_t\}\), \(f(z_t, \theta_0)\) can be decomposed in the following way \(f(z_t, \theta_0) = f_1(y_t|x_t, \theta_0)f_2(x_t, \theta_0)\) and our interest is in \(\theta_0\).

• For instance we can take a function \(f_1(y_t|x_t = x, \theta_0) \sim N(\mu(x, \theta_0), \sigma^2(x, \theta_0))\).

• Our interest often is in \(E(y_t|x_t)\).

2. LINEAR REGRESSION MODEL (standard linear model)

makes an assumption that this conditional expectation is linear in \(x\) i.e.
\[E(y_t|x_t) = \beta x_t\]
or in matrix notation
\[E(Y_t|X_t) = \beta' X_t,\]
where \(\beta = [\beta_1, \ldots, \beta_n]'\), \(X_t = [X_{t1}, \ldots, X_{tk}]'\).

• An alternative way to write this is \(Y_t = \beta' X_t + \mu_t\) (\#), \(t = 1, \ldots, T\) where \(E(\mu_t|x_t) = 0\)

• Terminology:
  \(Y_t\) – endogenous, dependent variable
  \(X_t\) – exogenous, independent variable, regressors

• Remark:
  If \(X_t\) is fixed (deterministic), then \(E(Y_t) = \beta' X_t\) and \(E(\mu_t) = 0\) (it’ll often be this case).

• Matrix notation

  Take \(Y = [y_1, \ldots, y_T]'_{T \times 1}\), \(X = \begin{bmatrix} x_{t1} & \ldots & x_{tk} \\ \vdots & \ddots & \vdots \\ x_{T1} & \ldots & x_{TK} \end{bmatrix}_{T \times K}\) (\(T\) observations, \(K\) variables),
\[\beta = [\beta_1, \ldots, \beta_K]'_{K \times 1}, \mu = [\mu_1, \ldots, \mu_T]'_{T \times 1}\]

• We can now write (\#) as \(Y = X\beta + \mu\).

  We have a data set \(X, Y\) and we want to make inference about parameter \(\beta\) (from observed data \((Y, X)\)). For example we can write an objective function.

3. OBJECTIVE FUNCTION \(Q\)

• \(Q\) is such that \(\hat{\beta} = \text{argmin}_\beta Q(\beta; Y, X)\).

• One obvious candidate for \(Q\) is a function which squares deviations of \(Y_t\) from their mean level \(\beta' X_t\).

  \(\hat{\beta} = \text{argmin}_\beta Q(\beta)\) is an Ordinary Least Squares estimator (OLS).

\[Q(\beta) = \sum_{t=1}^T (y_t - \beta' x_t)^2 = (y - X\beta)'(y - X\beta)\] – matrix notation

• First Order Conditions (FOC) for the optimization:
\[\frac{\partial Q}{\partial \beta} |_{\hat{\beta}} = 0 \Rightarrow -2X'Y + 2X'X\hat{\beta} = 0.
\]
If \(X'X\) is of full column rank, then \(\hat{\beta} = (X'X)^{-1}X'Y\).

• Second Order Conditions (SOC) for the optimization:
\[\frac{\partial^2 Q}{\partial \beta \partial \beta'} |_{\hat{\beta}} = 2X'X > 0\] (ok, if \(X'X\) is positive definite).

• If FOC and SOC are satisfied, then \(\hat{\beta}\) will minimize \(Q(\beta)\).

• Least Square Residuals

  are defined by \(\hat{\mu}_t = y_t - \hat{\beta}' x_t, t = 1, \ldots, T\) or in matrix notation \(\hat{\mu} = Y - X\hat{\beta}\) (\#).

• Remark 1

  Residuals are orthogonal to \(X\) i.e. \(X'\hat{\mu} = 0\) (\#\#), where \(\hat{\mu} = [\hat{\mu}_1, \ldots, \hat{\mu}_T]\).

  \[\text{Proof.} \text{ If we substitute (\#) to (\#\#) } X'\hat{\mu} = X'Y - X'X\hat{\beta} = X'Y - X'X(X'X)^{-1}X'Y = 0. \]

• Remark 2

  If there is a constant among regressors, then \(X'\hat{\mu} = 0\) (or \(\sum_t \hat{\mu}_t = 0\)).

4. STATISTICAL PROPERTIES OF OLS

Assumptions

1. \(X\) is non-stochastic and finite \(T \times K\),
II $X'X\beta$ is non-singular $\forall T \geq K$,
III $E(\mu) = 0$,
IV $\mu \sim N(0, \sigma_0^2 I)$,
V $\lim_{T \to \infty} \frac{X'X}{T} = Q$ is positive definite.

Under these assumptions, we have the following:
(existence and uniqueness, unbiasedness, BLUE, normal distribution, consistent)

a) Under I and II

b) Under I to IV $\hat{\beta} = \beta_0$, so $\hat{\beta}$ is unbiased estimator of $\beta_0$.

Proof. $E(\hat{\beta}) = E[(X'X)^{-1}X'Y] = E[(X'X)^{-1}(X'X)\beta_0 + (X'X)^{-1}X'\mu] = \beta_0 + (X'X)^{-1}X'E\mu = \beta_0$ as $E\mu = 0$.

c) Under I to III $\hat{\beta}$ is the Best Linear Unbiased Estimator (BLUE) in a sense that the covariance matrix of any other linear unbiased estimator exceeds that of $\hat{\beta}$ by a positive definite matrix (Gauss-Markov theorem).

Proof. Consider another linear estimator $\tilde{\beta} = D'Y$, where $D'$ does not depend on the data $Y$ and let $D = D* - (X'X)^{-1}X'$. With this we have:

$$\tilde{\beta} = [D + (X'X)^{-1}X']Y = [D + (X'X)^{-1}X'](X\beta_0 + \mu) = (DX + I)\beta_0 + (D + (X'X)^{-1}X)\mu$$

As $X$ is fixed, the expected value of the second part equals 0. So far $\tilde{\beta}$ is unbiased, so we must have $DX = 0$. Now

$$Var(\tilde{\beta}) = E(\tilde{\beta} - \beta_0)(\tilde{\beta} - \beta_0)' = (D + (X'X)^{-1}X')E(\mu\mu')(D' + X(X'X)^{-1}) =$$

$$= const \cdot \sigma^2 \cdot const = \sigma^2[(DD') + DX(X'X)^{-1} + (X'X)^{-1}X'D' + (X'X)^{-1}] =$$

$$= \sigma^2(DD' + (X'X)^{-1}) = Var(\tilde{\beta}) + \sigma^2DD' > Var(\hat{\beta})$$

So $\tilde{\beta}$ is a worse estimator than $\hat{\beta}$.
Recall

$$\hat{\beta} = (X'X)^{-1}X'Y = (X'X)^{-1}X'X\beta_0 + (X'X)^{-1}X'\mu = \beta_0 + (X'X)^{-1}X'\mu.$$

So

$$Var(\hat{\beta}) = E[(X'X)^{-1}\mu\mu'(X'X)^{-1}] = (X'X)^{-1}X'E(\mu\mu')(X'X)^{-1} =$$

$$= \sigma^2(X'X)^{-1}X'(X'X)^{-1} = \sigma^2(X'X)^{-1}.$$

d) Under I to IV $\hat{\beta} \sim N(\beta_0, \sigma^2(X'X)^{-1})$.

e) Under I to V $\hat{\beta}$ is consistent for $\beta_0$.

Proof. We have $\hat{\beta} - \beta_0 = (X'X)^{-1}X'\mu = (X'X)^{-1}(X'X')^{-1} \xrightarrow{T \to \infty} Q^{-1}$. Let us consider the second moment

$$Var(\frac{X'\mu}{T})_{K \times K} = E[(\frac{1}{T} \sum_{t} X_t \mu_t)_{K \times 1,1 \times 1} (\frac{1}{T} \sum_{t} X_t \mu_t)' = \frac{1}{T^2} \sum_{t} X_t X_t' E\mu_t = \frac{\sigma^2}{T} \sum_{t} X_t X_t' \xrightarrow{T \to \infty} 0$$

By Markov’s $\frac{X'\mu}{T} \xrightarrow{p} 0 \Rightarrow \hat{\beta} \xrightarrow{p} \beta_0$ (by Slutsky theorem).

There exist other estimators (here - quadratic loss; other moments, absolute loss, assymetric losses).
TUTORIAL 2, 24/02/2012

1. ◊ Exercise 4 (Tutorial 1)

\[ \text{corr}(X_s, X_t) = \rho_{st} = \frac{\text{Cov}(X_s, X_t)}{\sigma_s \sigma_t} = \frac{E(X_s X_t)}{\sigma_s \sigma_t} \]

So \( E(X_s X_t) = \rho_{st} \sigma_s \sigma_t \).

We have to show that \( \rho_{st} \xrightarrow{|s-t| \to \infty} 0 \Rightarrow X_n \xrightarrow{2} 0 \) (i.e. \( E(X_n^2) \xrightarrow{n \to \infty} 0 \)),

\[ E(X_n) = \frac{1}{n}\sum_{i=1}^{n} X_i + 2 \sum_{i=1, i \neq j}^{n} X_i X_j = \frac{1}{n^2}(\sum_{i=1}^{n} E X_i^2 + 2 \sum_{i=1, i \neq j}^{n} \rho_{ij} \sigma_i \sigma_j) \leq \frac{1}{n^2}(nc + 2c \sum \rho_{ij}) < \frac{1}{n^2}(nc + 2c(nN + n(n-1)/2)\varepsilon) < \frac{1}{n^2}(nc + 2c(2n) + c^2 \varepsilon) = \frac{c + 2cN}{n} + c \varepsilon \to 0 \]

Because \( \rho_{st} \xrightarrow{|s-t| \to \infty} 0 \Leftrightarrow \forall \varepsilon > 0 \exists N |\rho_{st}| < \varepsilon \) if \( |s-t| > N \).

2. Exercise 1

Let \( X_1, X_2, \ldots \) be a sequence of i.i.d. (independent and identically distributed) random variables with \( E X_t = \mu, \text{Var} X_t = \sigma^2 < \infty \). Show that Lindeberg condition is satisfied

\[ \lim_{n \to \infty} \sigma_n^{-2} n^{-1} \sum_{i=1}^{n} f(x - \mu_i)^2 dF(x) = 0. \]

Note that Lyapunov condition is stronger, it is better to show the Lindeberg condition if possible. Notation: \( \sigma = \sqrt{\sigma_n^2} \).

\[ \sum_{i=1}^{n} E(X_i - \mu_i)^2 I_{\{|X_i - \mu_i| > \varepsilon \sigma_n \sigma^2 \}} = \frac{n}{\sigma^2} E(X_i - \mu_i)^2 I_{\{|X_i - \mu_i| > \varepsilon \sigma_n \sigma^2 \}} \xrightarrow{n \to \infty} 0 \]

We can use the theorem about monotone convergence, because

\[ E|X_1 - \mu|^2 I_{\{|X_i - \mu_i| > \varepsilon \sigma_n \sigma^2 \}} \leq E|X_1 - \mu|^2 = \sigma^2 \]

3. Exercise 2 (Proof of Lyapunov’s CLT)

Let \( X_1, X_2, \ldots \) be a sequence of independent random variables with \( E X_t = \mu_t, \text{Var} X_t = \sigma_t^2 < \infty \) and \( E|X_t - \mu_t|^{2+\delta} \leq M < \infty \) for some \( \delta > 0 \) \( \forall t \) and \( \exists \delta' > 0 \) such that \( \sigma_n^2 > \delta' \) for all \( n \) sufficiently large.

Then \( \sqrt{n} \sum_{t} \frac{X_t - \mu_t}{\sigma_n} \xrightarrow{D} \mathcal{N}(0, 1) \).

Hint: Try Lindeberg condition \( \sigma_n^{-2} n^{-1} \sum_{i=1}^{n} E|X_i - \mu_i|^2 I_{\{|X_i - \mu_i| > \varepsilon \sigma_n \sigma^2 \}} \).

\[ E|X_i - \mu_i|^2 I_{\{|X_i - \mu_i| > \varepsilon \sigma_n \sigma^2 \}} \leq (E|X_i - \mu_i|^{2+\delta})^{\frac{1}{2+\delta}} (E|X_i - \mu_i|^{2+\delta})^{\frac{\delta}{2+\delta}} \leq M \cdot (E|X_i - \mu_i|^2 I_{\{|X_i - \mu_i| > \varepsilon \sigma_n \sigma^2 \}})^{\frac{\delta}{2+\delta}} \xrightarrow{n \to \infty} 0 \]

So Lindeberg condition is satisfied. We have used facts that:

- Schwartz inequality: \( E|xy| \leq (E|x|^p)^{\frac{1}{p}} (E|y|^q)^{\frac{1}{q}}, \frac{1}{p} + \frac{1}{q} = 1 \)
- \( \mathbb{P}(|X_i - \mu_i|^2 > \varepsilon \sigma_n \sigma^2) \leq \mathbb{P}(|X_i - \mu_i|^2 > \varepsilon \delta' n) \to 0 \)

4. Exercise 3

Let \( y_{n \times 1} \sim \mathcal{N}(0, I) \) and \( A \) be a symmetric, idempotent matrix of order \( n \) and rank \( p \). Show that

a) \( y' A y \sim \chi^2_p \)
b) \( y' A_1 y \) and \( y' A_2 y \) are independent if and only if \( A_1 A_2 = 0 \).
a) Since $A$ is symmetric and idempotent we can orthogonalize this matrix i.e. $A = SAS'$ $(S$ is the matrix of eigenvectors), where $\Lambda = \begin{bmatrix} 1 & 1 \\ & & \ldots & 0 \\ & & & 0 \end{bmatrix}$.

Eigenvalues equals either 0 or 1 since $A$ is idempotent i.e. $A'A = A$.

There are $n - p$ zeros and $p$ ones in matrix $\Lambda$.

Now $y'Ag = y'SAS'y = t'At$, where $t = S'y$.

Since $y \sim \mathcal{N}(0, I)$, we have that $t \sim \mathcal{N}(0, S'S) = \mathcal{N}(0, I)$ and therefore $t'At = \sum_{i=1}^{n} t_i^2 \lambda_i = \sum_{i=1}^{p} t_i^2 \sim \chi_p^2$.

b) We need $A_1y$ and $A_2y$ to be independent ($\text{transformation}$).

$$\text{Cov}(A_1y, A_2y) = E(A_1yy' A_2) = A_1 E(yy') A_2 = A_1 A_2.$$ $y$ is $\mathcal{N}(0, I)$, so $\text{Cov}(A_1y, A_2y) = 0$ is sufficient for independence. Therefore $A_1 A_2 = 0$.

\[ \text{Lecture 3, 02/03/2012} \]

1. Linear regression $(y_t = \beta' x_t + \mu_t)$ – continuation

- **Assumptions:**
  
  I $X$ is fixed (deterministic)
  
  II $\text{rank}(X) = k$
  
  III $E\mu = 0$, $E\mu^2 = \sigma^2$, $E\mu \mu^\prime = 0$ ($\forall t \neq \tau$)
  
  IV $\lim_{T \to \infty} \frac{X'X}{T} = Q$ is positive definite

- **Recall** $\hat{\beta} = (X'X)^{-1}X'Y$ and so $X_T^\prime K \hat{\beta}_{K \times 1} = X(X'X)^{-1}X'Y = P_X Y$, where $P_X = X(X'X)^{-1}X'$ – projects onto space spanned by $X$. $M_X X = X - X(X'X)^{-1}X'X = 0$

- **Consider** $M_X = I - P_X = I - X(X'X)^{-1}X'$. It projects onto space orthogonal to $X$ (annihilates $X$).

- **Note** that $P_X X = X(X'X)^{-1}X'X = X$ and $P_X A X = AX(X'A'AX)^{-1}(X'A'AX) = AX$

**Geometry of OLS** (linear regression – orthogonal projection!)

$P_X$ and $M_X$ are symmetric and idempotent that is $M_X M_X = M_X$ and $P_X P_X = P_X$, $P_X + M_X = I$.

In our case $Y = \beta X + \mu$, $\hat{\beta} = (X'X)^{-1}X'Y$,$\hat{\beta} = \beta X + \nu = \beta X + \beta' \nu = \beta (X + \nu)$. It projects onto space orthogonal to $X$ (annihilates $X$).

- **Last week** we showed that $\hat{\beta} \xrightarrow{p} \beta$. Now we show that the same holds for variance of the residuals $\mu$.

  Recall that $\sigma^2 = E\mu^2$ and $\hat{\sigma}^2 = \frac{1}{T} \sum \mu^2$ ($\mu_t$ is not observable!).

  Note that $\hat{\beta}$ is close to $\beta$, so we can expect that $\hat{\mu}_t = y_t - \hat{\beta} x_t$ to be close to $\mu_t$. Thus we can consider $\frac{1}{T} \sum \hat{\mu}_t$ as an estimator for $\sigma^2$.

**Proposition**

$\hat{\sigma}^2 \rightarrow \sigma^2$, where $\hat{\sigma}^2 = \frac{1}{T} \sum \hat{\mu}_t$.

**Proof**. Note that $\hat{\mu} = M_X \hat{\mu}$ and

$$E(\mu' \hat{\mu} = E(\mu' M_X \mu) = E(\mu' M \mu) = E(tr(\mu' M \mu)) = E(tr(M \mu \mu')) = \sigma^2 tr(M_X),$$

$$\text{tr}M_X = \text{tr}(I_{T \times K} - X_{T \times K} (by) K (X'X)^{-1}X') = \text{tr}I_{T \times K} - \text{tr}(X(X'X)^{-1}X') = T - \text{tr}((X'X)^{-1}X'X)_{K \times K} = T - K.$$ So we get that $E(\hat{\mu}^2) = (T - K) \sigma^2$.

This says that $E(\frac{1}{T-K} \sum \mu^2) = \sigma^2 \Rightarrow \hat{\sigma}^2 = \frac{1}{T-K} \sum \hat{\mu}^2$.

So $\hat{\sigma}^2$ is unbiased (this is also a consistent estimator – we will show it later).

- **Cramer-Wald device**

  Let $\{X_n\}$ be a sequence of $k \times 1$ random variables. Then $X_n \overset{D}{\rightarrow} X$ (in distribution) $\Leftrightarrow X'X_n \overset{D}{\rightarrow} X'X$ $\forall \lambda \neq 0$.

  **Comment:** We get a scalar problem, which is much more convenient to solve than a vector problem.
• Cramer
Let \( \{X_n\} \) be a sequence of \( k \times 1 \) random variables and assume that \( X_n = A_n Z_n \). Suppose in addition that \( A_n \xrightarrow{p} A \) which is positive definite and \( Z_n \xrightarrow{D} \mathcal{N}(\mu, \Sigma) \). Then \( A_n Z_n \xrightarrow{D} \mathcal{N}(A \mu, A \Sigma A') \).

• Proposition
If we add an assumption
\[ V \mu_t \sim \mathcal{N}(0, \sigma^2) \]
we will have:
\begin{enumerate}[(a)]  
• \( (X'X)^{1/2}(\hat{\beta} - \beta) \sim \mathcal{N}(0, \sigma^2 I) \)
• \( (T - k)\frac{\hat{\sigma}^2}{\sigma^2} \sim \chi^2_{T-k} \) proper scaled estimator of \( \sigma^2 \) has the \( \chi^2_{T-k} \) distribution.
\end{enumerate}
Moreover \( \hat{\beta} \) and \( \hat{\sigma}^2 \) are independent (where \( \hat{\sigma}^2 = \frac{1}{T-k} \Sigma \hat{\mu}^2 \)).

Proof.  
\begin{enumerate}[(a)]  
• Note \( (T - k)\frac{\hat{\sigma}^2}{\sigma^2} = \frac{\hat{\sigma}^2}{\sigma^2} \Sigma \).  
• From Tut. 2 Ex. 3 recall that \( X'AX \) and \( \beta Y \) are independent only if \( AB = 0 \). Note that \( M_X X (X'X)^{-1} = 0 \), because \( M_X X = 0 \) (\( M_X \) annihilates \( X \)). \hfill \Box
\end{enumerate}

• Corollary
Under our assumptions the asymptotic distribution of \( T^{1/2}(\hat{\beta} - \beta) \sim \mathcal{N}(0, \sigma^2 Q^{-1}) \).

Proof. That comes \( \hat{\beta} - \beta \sim \mathcal{N}(0, \sigma^2 (X'X)^{-1}) \) and \( X'X \rightarrow Q \).
Therefore \( (X'X)^{1/2} T^{1/2}(\hat{\beta} - \beta) \sim_a \mathcal{N}(0, \sigma^2 I) \).
Using Cramer this says that \( T^{1/2}(\hat{\beta} - \beta) \sim_a \mathcal{N}(0, \sigma^2 Q^{-1}) \) (asymptotic distribution). \hfill \Box

2. Relaxing assumption I

• Now \( X \) are random variables, not fixed numbers. Let us consider stochastic regressors (\( X \) are random regressors), so replace I with I:\`

I' Random variables \( x_t \) are i.i.d. with \( \mathbb{E}(x_t x'_t) = \Sigma_x \) positive definite.

• Proposition
Under I', II and III, \( \hat{\beta} \) is consistent.

Proof. \( \hat{\beta} - \beta = (X'X)^{-1}X'\mu = (X'X)^{-1} (X'\mu) \).
By LLN \( \frac{X'X}{n} \xrightarrow{p} \Sigma_X \) (\( *, \)).
Since \( \Sigma_X \) is positive definite, then by Slutsky theorem \( (X'X)^{-1} \xrightarrow{p} \Sigma_X^{-1} \).
Now consider \( \frac{1}{T} \sum x_t \mu_t = \frac{1}{T} \sum z_t. \) With \( x_t \) iid. and \( \mu_t \) iid., \( z_t \) is also iid.
Therefore \( \mathbb{E}(x_t \mu_t) = \mathbb{E} x_t \mathbb{E} \mu_t = 0. \) So employing LLN \( \frac{1}{T} \sum z_t \xrightarrow{p} 0. \) We only have to show consistency of \( \hat{\beta} - \beta \).
In effect \( \hat{\beta} - \beta = (X'X)^{-1} (X'\mu) \xrightarrow{p} \Sigma_X^{-1} \cdot 0 = 0. \)
So \( \hat{\beta} \) is consistent. \hfill \Box

• Proposition
Under I', III and \( \mu_t \) iid., \( \hat{\sigma}^2 \xrightarrow{p} \sigma^2. \)

Proof. \( \hat{\sigma}^2 = \frac{\hat{\beta}^2}{T-k} = \frac{\hat{\beta}^2}{T-k} \frac{\hat{\mu}' \hat{\mu}}{\hat{\mu}' \hat{\mu}} = \frac{\hat{\beta}^2}{T-k} \frac{\hat{\mu}' \hat{\mu} - \hat{\mu}' \hat{\mu}}{\hat{\mu}' \hat{\mu} - \hat{\mu}' \hat{\mu}} \xrightarrow{p} 1[\sigma^2 - 0 \cdot \Sigma_X^{-1} \cdot 0] = \sigma^2. \)
Therefore \( \hat{\sigma}^2 \) is consistent. \hfill \Box

• Remarks: \( \hat{\sigma}^2 \) is an estimator of the volatility. Estimators \( \hat{\sigma}^2 \) and \( \hat{\beta} \) are consistent!

• Theorem
Under I', II, III with \( \mu_t \) iid., we have
a) \( T^{1/2}(\hat{\beta} - \beta) \xrightarrow{D} \mathcal{N}(0, \sigma^2 \Sigma_X^{-1}) \)

If in addition we assume that \( \mathbb{E}\mu_4^i < \infty \) then

b) \( T^{1/2}(\hat{\sigma}^2 - \sigma^2) \xrightarrow{D} \mathcal{N}(0, \mathbb{E}\mu_4^i - \sigma^4) \).

**Proof.** a) We have that \( T^{1/2}(\hat{\beta} - \beta) = (\frac{X'X}{T})^{-1} \frac{X'\mu}{\sqrt{T}} \). We know already that \( (\frac{X'X}{T})^{-1} \rightarrow \Sigma_X^{-1} \).

What remains to prove is \( \frac{X'\mu}{\sqrt{T}} \rightarrow \mathcal{N}(0, \sigma^2 \Sigma_X) \).

\[
\frac{X'\mu}{\sqrt{T}} \equiv T^{-1/2} \sum x_i \mu_i = T^{-1/2} \sum z_t \rightarrow \mathcal{N}(0, \sigma^2 \Sigma_X)
\]

where \( z_t \) is a vector of random variables.

By Cramer-Wald we need to show that \( \lambda'z_t \) converges \( \forall \lambda \neq 0 \).

\( \mu_t \) and \( x_t \) are iid., so \( z_t \) are iid. and \( \lambda'z_t \) are iid. We also know that \( \mathbb{E}(\lambda'z_t) = 0 \).

By the Lindeberg-Levy CLT, \( T^{-1/2} \sum z_t \rightarrow \mathcal{N}(0, \sigma^2 \Sigma_X) \).

Since \( \lambda \) is arbitrary, \( T^{-1/2} \sum z_t \rightarrow \mathcal{N}(0, \sigma^2 \Sigma_X) \).

b) \( \hat{\sigma}^2 = \frac{1}{T-k} \sum m_i \hat{\sigma}^2 \).

Consider \( \tilde{\sigma}^2 = \frac{1}{T} \sum \tilde{\mu}^2_i \) (asymptotically it is the same).

Recall that \( M_X = I - X(X'X)^{-1}X' \).

\[
T^{1/2}(\hat{\sigma}^2 - \sigma^2) = T^{1/2}\tilde{\mu}' \tilde{\mu} - T^{1/2}\sigma^2 = \frac{1}{T^{1/2}}(\mu'M\mu - T\sigma^2) = T^{-1/2} \sum (\mu^2 - \sigma^2) = T^{-1/2}(X'X)(\frac{X'X}{T})^{-1}(\frac{X'\mu}{T}) \rightarrow \mathcal{N}(0, \mathbb{E}\mu_4^i - \sigma^4)
\]

because \( (\frac{X'X}{T})^{-1}(\frac{X'\mu}{T}) \rightarrow 0 \).

Next \( T^{1/2}(\hat{\sigma}^2 - \sigma^2) = (T-k) \frac{1}{T} T^{1/2}(\tilde{\sigma}^2 - \sigma^2) + \frac{k}{T} \sigma^2 \rightarrow 1 \cdot T^{1/2}(\tilde{\sigma}^2 - \sigma^2) + 0 \).

By Cramer \( T^{1/2}(\hat{\sigma}^2 - \sigma^2) \xrightarrow{D} \mathcal{N}(0, \mathbb{E}\mu_4^i - \sigma^4) \).

\( \square \)

3. **Remarks**

a) **To convergence results**

- We do not need normality of the errors \( \mu_t \).
- If \( \mathbb{E}\mu_4^i > 0 \) for some \( \delta > 0 \) is finite, then we can drop iid. assumption and use Lyapunov instead (Lindeberg-Levy CLT).
- If \( \mu_t \) are normally distributed \( T^{1/2}(\hat{\sigma}^2 - \sigma^2) \xrightarrow{D} \mathcal{N}(0, 2\sigma^4) \), because \( \mathbb{E}\mu_4^i = 3\sigma^4 \).

b) **About consistency of \( \hat{\beta} \)**

If \( \frac{X'X}{T} \xrightarrow{a.s.} 0 \) and \( \frac{X'X}{T} - M_T \xrightarrow{a.s.} 0 \), where \( M_T \) is bounded and uniformly positive definite (matrix), then \( \hat{\beta} \) exists almost surely for all \( T \) sufficiently large and \( \hat{\beta} \xrightarrow{a.s.} \beta \).

**Proof.** Recall that earlier we had \( \frac{1}{T} X'X \rightarrow \Sigma_X \) fixed and positive definite.

Note that \( \text{det}(\frac{X'X}{T}) - \text{det}(M_T) \xrightarrow{a.s.} 0 \),

because \( <T \) is bounded and determinant is continuous.

Since \( \{M_T\} \) is uniformly positive definite, \( \exists \delta > 0 \) such that determinant \( \text{det}(\frac{X'X}{T}) > \delta \) for sufficiently large \( T \).

Then \( (\frac{X'X}{T})^{-1} \) exists almost surely and \( \hat{\beta} = (\frac{X'X}{T})^{-1} \frac{X'Y}{T} \).

So \( \hat{\beta} - \beta = (\frac{X'X}{T})^{-1} \frac{X'\mu}{T} \) and

\[
\hat{\beta} - (\beta + M_T^{-1} \cdot 0) \xrightarrow{a.s.} 0 \Rightarrow \hat{\beta} \xrightarrow{a.s.} \beta.
\]

\( \square \)

c) **Slutsky theorem**

Let \( \{X_n\} \), \( \{Y_n\} \) be sequences of scalar/vector/matrix random elements. If \( X_n \) converges in distribution to a random element \( X \), and \( Y_n \) converges in probability to a constant \( c \), then

a) \( X_n + Y_n \xrightarrow{D} X + c \),

b) \( Y_n X_n \xrightarrow{D} cX \),

c) \( Y_n^{-1} X_n \xrightarrow{D} c^{-1}X \) provided that \( c \) is invertible.
4. Restricted Least Squares

- Suppose there are \( m \) linearly independent constraints on parameters of \( \beta \) in the linear regression: \( R\beta = r \), where \( R \) is \( m \times k \) and \( r \) is \( m \times 1 \) and \( m < k \), \( \text{rank}(R) = m \).
- How do we estimate a regression with constraints? We add a Lagrangian to the loss function.

\[
g(\beta, \lambda) = (Y - X\beta)'(Y - X\beta) + 2\lambda'(R\beta - r)
\]

FOC: \( \frac{\partial g}{\partial \beta} = -2X'Y + 2X'X\beta + 2Rx = 0 \), \( \frac{\partial g}{\partial \lambda} = R\beta - r = 0 \).

It is easier to solve it in the matrix form

\[
\begin{bmatrix} X'X & R' \\ R & 0 \end{bmatrix} \begin{bmatrix} \hat{\beta} \\ \lambda \end{bmatrix} = \begin{bmatrix} X'Y \\ r \end{bmatrix}
\]

This gives a constrained least squares estimator

\[
\hat{\beta} = (X'X)^{-1}(I - R'(R(X'X)^{-1}R')^{-1}R(X'X)^{-1})X'Y + (X'X)^{-1}R'(R(X'X)^{-1}R')^{-1}r = \beta_{\text{OLS}} - (X'X)^{-1}R'(R(X'X)^{-1}R')^{-1}(R\hat{\beta}_{\text{OLS}} - r)
\]

If we have constraints in linear regression, we get \( \hat{\beta} = \beta_{\text{OLS}} - \text{correction term} \).

\[
\hat{\beta} = \beta_{\text{OLS}} \text{ if } \beta_{\text{OLS}} \text{ satisfies our constraint } R\hat{\beta}_{\text{OLS}} - r = 0.
\]

There exists also a geometric interpretation – restricted projection!

- Theorem

If we have a constrained regression with:

a) non-stochastic \( X \)
b) \( X'X \) non singular
c) \( \mu \sim \mathcal{N}(0, \sigma^2 I) \)

where \( \beta \) solves \( R\beta = r \) with \( \text{rank}(R) = m \)

\[
\frac{R\hat{\beta} - r)'(R(X'X)^{-1}R')^{-1}(R\hat{\beta} - r)/m}{T\sigma^2/(T-k)} \sim F_{m, T-k} (\bigcirc).
\]

Remarks:
\( \beta \) is OLS (not constrained).

**F statistic** – ratio of chi-squared variables i.e.

If \( q_1 \sim \chi^2_{n_1} \) and \( q_2 \sim \chi^2_{n_2} \), then \( \frac{q_1/n_1}{q_2/n_2} \sim F_{n_1, n_2} (\bigstar) \).

**Proof.** Note that

\[
[R(\hat{\beta} - \beta)]'[R(X'X)^{-1}R']^{-1}[R(\hat{\beta} - \beta)] = \mu'X(X'X)^{-1}R'[R(X'X)^{-1}R']^{-1}R(X'X)^{-1}X'\mu = \mu'Q\mu
\]

where

\[
Q = X(X'X)^{-1}R'[R(X'X)^{-1}R']^{-1}R(X'X)^{-1}X'
\]

is idempotent (show it! \( Q \cdot Q = Q \), \( \text{rank}(Q) = m \).

So we have that \( \mu'Q\mu \sim \chi^2_m \) and \( (T-k)/\sigma^2 \sim \frac{\mu'X\mu}{\sigma^2} \sim (T-k). \) So from (\bigstar), we have (\bigcirc). In addition \( M_X^2 = 0 \Rightarrow q_1 \) and \( q_2 \) are independent. \( \square \)
1. Exercise 1

Establish consistency of $\beta$ in the following linear model
\[ y_t = \alpha + \beta t + \mu_t, \quad t = 1, \ldots, T, \]
where $E \mu_t = 0$, $E \mu_t^2 = \sigma^2$ and $E(\mu_t \mu_s) = 0 \forall t \neq s$. Obtain a limiting distribution of $T \frac{3}{2} (\hat{\beta} - \beta)$.

Comment:
Because of the specific regressors (time!), we can have the power of 3/2 – we have much faster way of convergence and consistency during the lecture we had 1/2).

We know that
\[ \hat{\beta} = \beta \]
\[ \text{Var} \hat{\beta} = \sigma^2 (X'X)^{-1} = \frac{\sigma^2}{\sum (t - \bar{t})^2}, \]
where $t = \frac{1}{T} \sum_{t=1}^{T} t$.

To show consistency it is sufficient to show that $\text{Var} \hat{\beta} \to 0$ (it follows from Markov’s inequality that checking the second moment is enough for consistency of estimator).

\[
\frac{\sigma^2}{\sum (t - \bar{t})^2} = \frac{\sigma^2}{\sum t^2 - T\bar{t}^2} = \frac{\sigma^2}{\frac{1}{2}T(T+1)(2T+1) - \frac{1}{4}T(T+1)^2} = \frac{12\sigma^2}{T(T^2 - 1)} \to 0
\]

Note that $\sum t = \frac{T(T+1)}{2}$ and $\sum t^2 = \frac{T(T+1)(2T+1)}{6}$. So now we have that

\[ \text{Var}(T^{3/2}(\hat{\beta} - \beta)) = T^{3/2} \text{Var}(\hat{\beta} - \beta) = \frac{12\sigma^2}{1 - \frac{1}{2}} \to 12\sigma^2 \]

So $T^{3/2}(\hat{\beta} - \beta) \xrightarrow{d} N(0, 12\sigma^2)$.

2. ♦ Exercise 2

Consider a regression model
\[ y_t = \beta_1 x_{1t}^2 + \beta_2 x_{2t}^2 + \mu_t, \]
where $x_{1t}$ and $x_{2t}$ are centered (zero mean). Let $\rho$ be a simple correlation of $x_{1t}$ and $x_{2t}$

\[ \hat{\rho} = \frac{\text{Cor}(x_{1t}^2, x_{2t}^2)}{\sqrt{\text{Var}x_{1t}^2} \text{Var}x_{2t}^2}. \]

Show that $\text{corr}(\hat{\beta}_1, \hat{\beta}_2) = -\hat{\rho}$. Think what happens if $\hat{\rho} \to 1$ (then regressors become closer to each other – they are almost the same variables).

Note that $\hat{\beta}_1$ and $\hat{\beta}_2$ are OLS estimators of $\beta_1$ and $\beta_2$.

Hint:
Matrix form: $Y = \beta_1 X_1 + \beta_2 X_2 + \mu$

Show that $\tilde{\beta}_1 = (X_1' M_2 X_1)^{-1} X_1' M_2 Y$ and similarly $\tilde{\beta}_2 = (X_2' M_1 X_2)^{-1} X_2' M_1 Y$, where $M_1 = I - X_1(X_1' X_1)^{-1} X_1'$ and $M_2 = I - X_2(X_2' X_2)^{-1} X_2'$ (projects off the space of $X_1$ and $X_2$ respectively)