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 ♡
  - "Econometric Theory and Methods", R. Davidson and J. G. MacKinnon, Oxford University Press (New York)
  - "Advanced econometrics", Takeshi Amemiya
  - "Econometrics", Fumiohayashi
  - "Probability and Random Processes", G. R. Grimmett, D. R. Stirzaker
- A course is designed to familiarize students with statistical methods employed in analysis of economic and financial data. Emphasis will be places ona 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** – $(\Omega, \mathcal{F}, \mathbb{P})$
  - Ω is a set called sample space
  - $\mathcal{F}$ is a family of events (an event is an element of $\mathcal{F}$).
  - $\mathbb{P}$ is a probability measure on $(\Omega, \mathcal{F}, \mathbb{P})$.
- **Random variable** is a function $x : \Omega \to \mathbb{R}$ with a property that $\{\omega \in \Omega : X(\omega) \leq x\} \in \mathcal{F} \ \forall \ x \in \mathbb{R}$.
- **Distribution function** of a random variable X is the function $F : \mathbb{R} \to [0, 1]$ given by $F(x) = \mathbb{P}(X \leq x)$. Random variable X is **continuous** if its distribution function can be expressed as $F(x) = \int_{-\infty}^{x} f(u)du$, $x \in \mathbb{R}$ for some integrable function $f : \mathbb{R} \to [0, \infty)$.

3. Convergence of random variables

Let $X_1, \ldots, X_n$ be random variables on some probability space $(\Omega, \mathcal{F}, \mathbb{P}$. We say that

- $X_n \xrightarrow{a.s.} X$ **almost surely** if $\{\omega \in \Omega : X_n(\omega) \to X(\omega)$ as $n \to \infty\}$ is an event whose probability is 1.
- $X_n \xrightarrow{r} X$ **in r'th mean**, where $r \geq 1$ if $\mathbb{E}|X_n^r| < \infty$ for all n and $\mathbb{E}(|X_n - X|^r) \to 0$ as $n \to \infty$.
- $X_n \xrightarrow{\mathbb{P}} X$ **in probability** if $\mathbb{P}(|X_n - X| > \varepsilon) \to 0$ as $n \to \infty \ \forall \varepsilon > 0$.
- $X_n \xrightarrow{D} X$ **in distribution** if $\mathbb{P}(X_n \leq X) \to \mathbb{P}(X \leq x)$ as $n \to \infty$ for all points at which $F_n(x) = \mathbb{P}(X \leq x)$ is continuous.

4. Implications

- $X_n \xrightarrow{a.s.} X / X_n \xrightarrow{r} X \Rightarrow X_n \xrightarrow{\mathbb{P}} X \Rightarrow X_n \xrightarrow{D} X$ $(r \geq 1)$.
- If $r > s \geq 1$, then $X_n \xrightarrow{r} X \Rightarrow X_n \xrightarrow{s} X$.
- If $X_n \xrightarrow{D} c$, where c is constant, then $X_n \xrightarrow{\mathbb{P}} c$.
- If $X_n \xrightarrow{D} X$ and $\mathbb{P}(|X_h| \leq k) = 1 \ \forall h$ and some k then $X_n \xrightarrow{r} X \ \forall r \geq 1$.
- If $\sum_n \mathbb{P}(|X_n - X| > \varepsilon) < \infty \ \forall \varepsilon > 0$, then $X_n \xrightarrow{a.s.} X$.
- If $X_n \xrightarrow{\mathbb{P}} X$ then $X_n \xrightarrow{D} X$.
  Converse is false: Let X be Bernoulli variable with parameter 1/2. Let $X_1, \ldots, X_n$ be identical random variables given by $X_n = X \ \forall n$. Then $X_n \xrightarrow{D} X$. Now let $Y = 1 - X$. Clearly $X_n \xrightarrow{D} Y$. We can't converge in any other mode as $|X_n - Y| = 1$ always.

Proof. Suppose $X_n \xrightarrow{\mathbb{P}} X$. Let's write $F_n(x) = \mathbb{P}(X_n \le x)$, $F(x) = \mathbb{P}(X \le x)$.

$$F_n(x) = \mathbb{P}(X_n \le x) = \mathbb{P}(X_n \le x \cap X \le x+\varepsilon) + \mathbb{P}(X_n \le x \cap X > x+\varepsilon) \le$$
$$\le F(x+\varepsilon) + \mathbb{P}(|X_n - X| > \varepsilon)(\xrightarrow{n\to\infty} 0)$$
$$F(x-\varepsilon) = \mathbb{P}(X \le x-\varepsilon) = \mathbb{P}(X \le x-\varepsilon \cap X_n \le x) + \mathbb{P}(X \le x-\varepsilon \cap X_n > x) \le$$
$$\le F_n(x) + \mathbb{P}(|X_n - X| > \varepsilon)(\xrightarrow{n\to\infty} 0)$$

So we obtain $F(x-\varepsilon) \le \liminf F_n(x) \le \limsup F_n(x) \le F(x+\varepsilon)$ $\forall \varepsilon > 0$. If F is continuous at x then $F(x-\varepsilon) \uparrow F(x)$ and $F(x+\varepsilon) \downarrow F(x)$ as $\varepsilon \to 0$. Since ε is arbitrary, $F_n(x) \xrightarrow{D} F(x)$. □

5. Other

- **Markov's inequality**
  If X is any random variable with finite mean then $\mathbb{P}(|X| \ge a) \le \frac{\mathbb{E}|X|}{a}$ for any $a > 0$.

  Proof. Let $A = \{|X| \ge a\}$. Then $|X| \ge aI_A$, where $I_A(\omega) = 1$ if $\omega \in A$, 0 otherwise.
  So $\mathbb{E}|X| \ge a\mathbb{P}(|X| \ge a)$. □

- **Skorokhod's representation theorem**
  If $\{X_n\}$ and X with distribution function $\{F_n\}$ and F are such that $X_n \xrightarrow{D} X$, then there exists a probability space $(\Omega', \mathcal{F}', \mathbb{P}')$ and random variables $\{Y_n\}$ and Y', which map Ω' into $\mathbb{R}$ such that
  - $\{Y_n\}$ and Y have distribution functions $\{F_n\}$ and F
  - $Y_n \xrightarrow{a.s.} Y$ as $n \to \infty$.
- **Corollary**
  If $X_n \xrightarrow{D} X$ and $g : \mathbb{R} \to \mathbb{R}$ is continuous, then $g(X_n) \xrightarrow{D} g(X)$.

  Proof. By Skorokhod's there exists a sequence $\{Y_n\}$ distributed identically to $\{X_n\}$ which converges almost surely to Y, which is a copy of X. Since g is continuous $Y_n(\omega) \to Y(\omega)$ implies $g(Y_n(\omega)) \xrightarrow{a.s.} g(Y(\omega))$. It means that $\{\omega : Y_n(\omega) \to Y(\omega)\} \subseteq \{\omega : g(Y_n(\omega)) \to g(Y(\omega))\}$ and $\mathbb{P}\{\omega : Y_n(\omega) \to Y(\omega)\} = 1$ (a.s. convergence), then $g(Y_n) \xrightarrow{a.s.} g(Y) \Rightarrow g(Y_n) \xrightarrow{D} g(Y) \Rightarrow g(X_n) \xrightarrow{D} g(X)$. □

6. Laws of Large Numbers (LLN)
   Let $\{X_n\}$ 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 $\frac{1}{n}\sum_{i=1}^n X_i \xrightarrow{a.s.} \mu$ 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 $VarX_i = \sigma_i^2$.
  If $\sum_{i=1}^n \frac{\sigma_i^2}{i^2} < \infty$ then $\frac{1}{n}\sum_{i=1}^n X_i - \frac{1}{n}\sum_{i=1}^n \mu_i \xrightarrow{a.s} 0$ (or written $\bar{X}_n - \bar{\mu}_n \xrightarrow{a.s} 0$).

7. Central Limit Theorems (CLT)

- **Lindeberg-Levy CLT**
  Let $X_1, X_2, \ldots$ be a sequence of i.id. random variables with finite means μ and finite non-zero vairances σ^2. Let $S_n = \sum_{i=1}^n X_i$.
  Then $\frac{S_n - n\mu}{\sqrt{n\sigma^2}} \xrightarrow{D} X \sim \mathcal{N}(0,1)$ or $\sqrt{n}(\frac{1}{n}\sum_{i=1}^n(\frac{X_i-\mu}{\sigma})) \xrightarrow{D} X \sim \mathcal{N}(0,1)$.
- **Lindeberg-Feller CLT**
  Let $X_1, X_2, \ldots$ be a sequence of independent random variables with $\mathbb{E}X_i = \mu_i$, $VarX_i = \sigma_i^2 < \infty$ and distribution function F_i.
  Then $\sqrt{n}\frac{\frac{1}{n}S_n - \frac{1}{n}\sum_{i=1}^n \mu_i}{\sqrt{\frac{1}{n}\sum_{i=1}^n \sigma_i^2}} \xrightarrow{D} \mathcal{N}(0,1)$ (random variable with normal distribution)
  and $\lim_{n\to\infty} \max_{1\le i\le n} n^{-1}(\frac{\sigma_i^2}{\bar{\sigma}_n^2}) = 0$ (where $\bar{\sigma}_n^2 = \frac{1}{n}\sum_{i=1}^n \sigma_i^2$).
  CLTs above are satisfied if and only if for any $\varepsilon > 0$ $\lim_{n\to\infty} \bar{\sigma}_n^{-2} n^{-1} \sum_{i=1}^n \int_{(x-\mu_i)^2 > \varepsilon n \sigma_n^2} (x-\mu_i)^2 dF_i(x) = 0$ (**Lindeberg condition** – it restricts average contribution form tails of the distribution to the variance).
- **Lyapunov's CLT**
  Let $X_1, X_2, \ldots$ be a sequence of independent random variables with $\mathbb{E}X_i = \mu_i$, $VarX_i = \sigma_i^2$, $\sigma_i^2 \ne 0$ and $\mathbb{E}|X_h - \mu_h|^{2+\delta} < M < \infty$ for some $\delta > 0$ $\forall h$.
  If $\bar{\sigma}_h^2 > \delta > 0$ $\forall h$ sufficiently large, then $\sqrt{n}(\frac{\frac{1}{n}S_n - \frac{1}{n}\sum_{i=1}^n \mu_i}{\sqrt{\frac{1}{n}\sum_{i=1}^n \sigma_i^2}}) \xrightarrow{D} \mathcal{N}(0,1)$.

TUTORIAL 1, 17/02/2012

1. Exercise 1

   Let X, Y – Bernoulli with parameter 1/2. Consider $X+Y$ and $|X-Y|$.

   Note

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

   $\mathbb{P}(X+Y=0,|X-Y|=0) = \frac{1}{4}$
   $\mathbb{P}(X+Y=0)\mathbb{P}(|X-Y|=0) = \frac{1}{4}\cdot\frac{1}{2} \neq \frac{1}{4}$
   So correlation is 0*, but variables are not independent.*

2. Exercise 2

   Let X and Y have joint probability distrubtion (bivariate normal, where ρ is a constant $-1<\rho<1$)

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

   Now, $f_Y(y) = \int_{-\infty}^{\infty} g(x,y)h(y)dx = h(y)\int_{\infty}^{\infty} g(x,y)dx = h(y)\cdot 1$ *(*1 *– as it's a normal density).*
   By symmetry $f_X(x) = \frac{1}{\sqrt{2\pi}}\exp[-\frac{1}{2}x^2]$ *(*$\mathcal{N}(0,1)$*).*

$$Cov(X,Y) = \int\int xyf(x,y)dxdy = \textit{ (only this, because } \mu = 0\textit{)}$$
$$= \int\int xyg(x,y)h(y)dxdy = \int yh(y)[\int xg(x,y)dx]dy = \int yh(y)\rho y dy =$$
$$= \rho\int y^2h(y)dy = \rho\cdot 1 \textit{ (as variance } = 1\textit{)}$$

   If $\rho = 0$ *then* $f(x,y) = \frac{1}{\sqrt{2\pi}}\exp(-\frac{1}{2}x^2)\frac{1}{\sqrt{2\pi}}\exp(-\frac{1}{2}y^2) = f_X(x)f_Y(y)$.
   So X *and* Y *are independent.*

3. Exercise 3

   - If $VarX = 0 \Rightarrow X$ is a constant.
     Note
     $\mathbb{E}(X^2) = \sum_x x^2\mathbb{P}(X=x) = 0 \Rightarrow \mathbb{P}(X=x) = 0\ \forall x\neq 0 \Rightarrow \mathbb{P}(X=0)=1$
     $VarX = 0 \Rightarrow \mathbb{P}(X-\mathbb{E}X=0) = 1 \Rightarrow X = constant$
   - *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 $b\neq 0$ $\quad \Delta = 4b^2\mathbb{E}(XY)^2 - 4\mathbb{E}(X^2)b^2\mathbb{E}(Y^2) \leq 0$
     $\mathbb{E}(XY)^2 \leq \mathbb{E}(X^2)\mathbb{E}(Y^2)$ – ***Cauchy-Schwartz inequality***
     Note
     $\mathbb{E}(XY)^2 = \mathbb{E}(X^2)\mathbb{E}(Y^2)$ *only if* $\mathbb{P}(aX=bY)=1$ *(*$Z=0$*) for some* $a,b\in\mathbb{R}$ *and* $b\neq 0$.
   - *From Cauchy-Schwartz*
     $\mathbb{E}[(X-\mathbb{E}X)(Y-\mathbb{E}Y)]^2 \leq \mathbb{E}[(X-\mathbb{E}X)^2]\mathbb{E}[(Y-\mathbb{E}Y)^2] = VarXVarY$
     Taking square roots
     $|Cov(X,Y)| = \mathbb{E}[(X-\mathbb{E}X)(Y-\mathbb{E}Y)] \leq \sqrt{VarXVarY} \Rightarrow |\rho(X,Y)| = \frac{|Cov(X,Y)|}{\sqrt{VarXVarY}} \leq 1$

4. ◊ Exercise 4

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

LECTURE 2, 24/02/2012

1. Introduction

- Let us consider a set of data $\{z_t\}_{t=1,\dots,T}$ (not necessarily a time series). Let this data be distributed as $f(z_t, \theta_0)$ (known p.d.f. – probability density function), where θ_0 is the true value of parameter θ.
- 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 $f(z_t, \theta_0) = f_1(y_t|x_t, \theta_0) f_2(x_t, \theta_0) = f_1(y_t|x_t, \theta_{01}) f_2(x_t, \theta_{02})$ and our interest is in θ_0.
- For instance we can take a function $f_1(y_t|x_t = x, \theta_{01}) \sim \mathcal{N}(\mu(x, \theta_{01}), \sigma^2(x, \theta_0))$. For example y – income, x – consumption.
- Our interest often is in $\mathbb{E}(y_t|x_t)$.

2. Linear regression model (standard linear model) makes an assumption that this conditional expectation is linear in x i.e. $\mathbb{E}(y_t|x_t) = \beta x_t$ or in matrix notation $\mathbb{E}(Y_t|X_t) = \beta' X_t$, where $\beta = [\beta_1, \dots, \beta_n]'$, $X_t = [X_{1t}, \dots, X_{kt}]'$.

- An alternative way to write this is $Y_t = \beta' X_t + \mu_t$ $(\star)$, $t = 1, \dots, T$ where $\mathbb{E}(\mu_t|x_t) = 0$
- **Terminology**:
  Y_t – endogenous, dependent variable
  X_t – exogenous, independent variable, regressors
- **Remark**:
  If X_t is fixed (deterministic), then $\mathbb{E}(Y_t) = \beta' X_t$ and $\mathbb{E}(\mu_t) = 0$ (it'll often be this case).
- **Matrix notation**
  Take $Y = [y_1, \dots, y_T]'_{T\times 1}$, $X = \begin{bmatrix} x_{11} & \dots & x_{1k} \\ & \dots & \\ x_{T1} & \dots & x_{TK} \end{bmatrix}_{T\times K}$ (T observations, K variables),
  $\beta = [\beta_1, \dots, \beta_K]_{K\times 1}$, $\mu = [\mu_1, \dots, \mu_T]_{T\times 1}$
- We can now write $(\star)$ as $Y = X\beta + \mu$.
  We have a data set X, Y and we want to make inferences about parameter β (from observed data (Y, X)). For example we can write an objective function.

3. Objective function Q

- Q is such that $\hat{\beta} = \operatorname{argmin}_\beta Q(\beta; Y, X)$.
- One obvious candidate for Q is a function which squares deviations of Y_t from their mean level $\beta' X_t$. $\hat{\beta} = \operatorname{argmin}_\beta Q(\beta)$ 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 ful column rank, then $\hat{\beta} = (X'X)^{-1} X'Y$.
- **Second Order Conditions** (SOC) for the optimization:
  $\frac{\partial^2 Q}{\partial \beta \partial \beta'} = 2X'X > 0$ (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 $\hat{\mu}_t = y_t - \hat{\beta}' x_t$, $t = 1, \dots, T$ or in matrix notation $\hat{\mu} = Y - X\hat{\beta}$ (#).
- **Remark 1**
  Residuals are orthogonal to X i.e. $X'\hat{\mu} = 0$ (##), where $\hat{\mu} = [\hat{\mu}_1, \dots, \hat{\mu}_T]$.

  Proof. 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 $\mathbb{I}'\hat{\mu} = 0$ (or $\sum_t \hat{\mu}_t = 0$).

4. Statistical properties of OLS
Assumptions

I X is non-stochastic and finite $T \times K$, II $X'X\beta$ is non-singular $\forall T \geq K$,

III $\mathbb{E}(\mu) = 0$,

IV $\mu \sim \mathcal{N}(0, \sigma_0^2 I)$,

V $\lim_{T\to\infty}(\frac{X'X}{T}) = Q$ is positive definite.

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

a) Under I and II $\hat{\beta}$ exists and is unique.

b) Under I to III $\mathbb{E}(\hat{\beta}) = \beta_0$, so $\hat{\beta}$ is unbiased estimator of β_0.

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

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 $\tilde{\beta} = D^\star Y$, where $D^\star$ does not depend on the data Y and let $D = D^\star - (X'X)^{-1}X'$. 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 $\hat{\beta}$ is unbiased, so we must have $DX = 0$. Now

$$\begin{aligned} Var\tilde{\beta} &= \mathbb{E}(\tilde{\beta} - \beta_0)(\tilde{\beta} - \beta_0)' = (D + (X'X)^{-1}X')\mathbb{E}(\mu\mu')(D' + X(X'X)^{-1}) = \\ &= const \cdot \sigma^2 \cdot const = \sigma^2[(DD') + DX(X'X)^{-1} + (X'X)^{-1}X'D' + (X'X)^{-1}] = \\ &= \sigma^2(DD' + (X'X)^{-1}) = Var\hat{\beta} + \sigma^2 DD' > Var\hat{\beta} \end{aligned}$$

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

$$\begin{aligned} Var\hat{\beta} &= \mathbb{E}[(X'X)^{-1}\mu\mu'X(X'X)^{-1}] = (X'X)^{-1}X'\mathbb{E}(\mu\mu')X(X'X)^{-1} = \\ &= \sigma^2(X'X)^{-1}X'X(X'X)^{-1} = \sigma^2(X'X)^{-1}. \end{aligned}$$

□

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

e) Under I to V $\hat{\beta}$ is consistent for β_0.

Proof. We have $\hat{\beta} - \beta_0 = (X'X)^{-1}X'\mu = (\frac{X'X}{T})^{-1}(\frac{X'\mu}{T}) \xrightarrow{T\to\infty} Q^{-1} \cdot \star$.
From remark above mean of $\star$ equals 0. Let us consider the second moment

$$Var(\frac{X'\mu}{T})_{K\times K} = \mathbb{E}[(\frac{1}{T}\sum_t X_t\mu_t)_{K\times 1, 1\times 1}(\frac{1}{T}\sum_t X_t\mu_t)' = \frac{1}{T^2}\sum_t X_t X_t' \mathbb{E}\mu_t^2 = \frac{\sigma^2}{T}\frac{\sum_t X_t X_t'}{T} \xrightarrow{\mathbb{P}} 0$$

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

There exist other estimators (here – quadratic loss; other moments, absolute loss, assymetric losses).

TUTORIAL 2, 24/02/2012

1. $\Diamond$ Exercise 4 (Tutorial 1)

$$corr(X_s, X_t) = \rho_{st} = \frac{Cov(X_s, X_t)}{\sqrt{\sigma_s^2 \sigma_t^2}} = \frac{\mathbb{E}(X_t X_s)}{\sigma_t \sigma_s}$$

So $\mathbb{E}(X_t X_s) = \rho_{st}\sigma_s\sigma_t$.
We have to show that $\rho_{st} \xrightarrow{|s-t|\to\infty} 0 \Rightarrow \bar{X}_n \xrightarrow{2} 0$ *(i.e.* $\mathbb{E}(\bar{X}_n)^2 \xrightarrow{n\to\infty} 0$*),*

$$\mathbb{E}(\bar{X}_n)^2 = \frac{1}{n^2}\mathbb{E}(X_1^2 + \ldots + X_n^2) + 2\sum_{i,j=1, i\neq j}^{n} X_i X_j =$$
$$= \frac{1}{n^2}(\sum \mathbb{E}X_i^2 + 2\sum_{i,j=1,i\neq j}^{n} \rho_{ij}\sigma_i\sigma_j) \leq \frac{1}{n^2}(nc + 2c\sum \rho_{ij}) <$$
$$< \frac{1}{n^2}(nc + 2c(nN + \frac{n(n-1)}{2}\varepsilon)) < \frac{1}{n^2}(n(c+2cN) + cn^2\varepsilon) = \frac{c+2cN}{n} + c\varepsilon \to 0$$

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 $\mathbb{E}X_t = \mu$, $VarX_t = \sigma^2 < \infty$. Show that Lindeberg condition is satisfied

$$\lim_{n\to\infty} \bar{\sigma}_n^{-2} n^{-1} \sum_{i=1}^{n} \int_{(x-\mu_i)^2 > \varepsilon n \sigma_n^2} (x-\mu_i)^2 dF_i(x) = 0.$$

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

$$\frac{\sum_{i=1}^{n} \mathbb{E}(X_i-\mu)^2 \mathbb{I}_{\{|X_i-\mu|>\varepsilon n\sigma\}}}{n\sigma^2} = \frac{n\mathbb{E}(X_1-\mu)^2\mathbb{I}_{\{|X_1-\mu|>\varepsilon n\sigma\}}}{n\sigma^2} \xrightarrow{n\to\infty} \frac{\mathbb{E}0}{\sigma^2} = 0$$

We can use the theorem about monotone convergence, because

$$\mathbb{E}|X_1-\mu|^2\mathbb{I}_{\{|X_1-\mu|>\varepsilon n\sigma\}} \leq \mathbb{E}|X_1-\mu|^2 = \sigma^2$$

3. Exercise 2 (proof of Lyapunov's CLT)
Let $X_1, X_2, \ldots$ be a sequence of independent random variables with $\mathbb{E}X_t = \mu_t$, $VarX_t = \sigma_t^2 < \infty$ and $\mathbb{E}|X_t - \mu_t|^{2+\delta} < M < \infty$ for some $\delta > 0$ $\forall t$ and $\exists\delta' > 0$ such that $\bar{\sigma}_n^2 > \delta'$ for all n sufficiently large.
Then $\sqrt{n}\frac{\bar{X}_n - \bar{\mu}_n}{\bar{\sigma}_n} \xrightarrow{D} \mathcal{N}(0,1)$.
Hint: Try Lindeberg condition $\bar{\sigma}_n^{-2} n^{-1} \sum_{i=1}^{n} \mathbb{E}|X_i - \mu_i|^2 \mathbb{I}_{\{|X_i-\mu_i|>\varepsilon n\bar{\sigma}^2\}}$.

$$\mathbb{E}|X_i-\mu_i|^2\mathbb{I}_{\{|X_i-\mu_i|>\varepsilon n\bar{\sigma}^2\}} \leq (\mathbb{E}|X_t-\mu_t|^{2+\delta})^{\frac{1}{1+\frac{\delta}{2}}} (\mathbb{E}\mathbb{I}_{\{|X_t-\mu_t|^2>\varepsilon n\bar{\sigma}^2\}})^{\frac{\frac{\delta}{2}}{1+\frac{\delta}{2}}} \leq$$
$$\leq M \cdot (\mathbb{E}\mathbb{I}_{\{|X_t-\mu_t|^2>\varepsilon n\bar{\sigma}^2\}})^{\frac{\frac{\delta}{2}}{1+\frac{\delta}{2}}} \xrightarrow{n\to\infty} 0$$

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

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

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) $y'Ay \sim \chi_p^2$,

b) $y'A_1y$ and $y'A_2y$ are independent if and only if $A_1A_2 = 0$.

a) Since A is symmetric and idempotent we can orthogonalize this matrix i.e. $A = S\Lambda S'$ (S is the matrix of eigenvectors), where $\Lambda = \begin{bmatrix} 1 & & & & & \\ & 1 & & & & \\ & & 1 & & & \\ & & & \dots & & \\ & & & & 0 & \\ & & & & & 0 \end{bmatrix}$

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 Λ.
Now $y'Ay = y'S\Lambda S'y = t'\Lambda t$, 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 $t'\Lambda t = \sum_{i=1}^{n} t_i^2 \lambda_i = \sum_{i=1}^{p} t_i^2 \sim \chi_p^2$.

b) We need $A_1 y$ and $A_2 y$ to be independent (transformation).

$$Cov(A_1 y, A_2 y) = \mathbb{E}(A_1 y y' A_2) = A_1 \mathbb{E}(yy') A_2 = A_1 A_2.$$

y is $\mathcal{N}(0, I)$, so $Cov(A_1 y, A_2 y) = 0$ is sufficient for independence. Therefore $A_1 A_2 = 0$.

LECTURE 3, 02/03/2012

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

- Assumptions:
  - I X is fixed (deterministic)
  - II $rank(X) = k$
  - III $\mathbb{E}\mu_t = 0$, $\mathbb{E}\mu_t^2 = \sigma^2$, $\mathbb{E}\mu_t \mu_\tau = 0$ ($\forall t \neq \tau$)
  - IV $\lim_{T\to\infty} \frac{X'X}{T} = Q$ is positive definite
- Recall $\hat{\beta} = (X'X)^{-1}X'Y$ and so $X_{T\times K}\hat{\beta}_{K\times 1} = X(X'X)^{-1}X'Y = P_X Y$, 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 (anihilates X).
- Note that $P_X X = X(X'X)^{-1}X'X = X$ and $P_{AX} AX = AX(X'A'AX)^{-1}(X'A'AX) = AX$
- **Geomtry of OLS** (linear regression – orthogonal projection!)
  P_X and M_X are symmetric and idemponent that is $M_X M_X = M_X$ and $P_X P_X = P_X$, $P_X + M_X = I$.
  In our case $Y = \beta X + \mu$, $\hat{\beta} = (X'X)^{-1}X'Y$,
  $\hat{Y} = \hat{\beta} X = P_X Y$ – fitted values, projection on X,
  $\hat{\mu} = y - \hat{y} = (I - P_X)Y = M_X Y$ – residuals, projection on Y.
- Last week we showed that $\hat{\beta} \xrightarrow{\mathbb{P}} \beta$. Now we show that the same holds for varince of the residuals μ.
  Recall that $\sigma^2 = \mathbb{E}\mu_t^2$ and $\hat{\sigma}^2 = \frac{1}{T}\sum \mu_t^2$ (μ_t is not observable!).
  Note that $\hat{\beta}$ is close to β, so we can expect that $\hat{\mu}_t = y_t - \hat{\beta}' x_t$ to be close to μ_t. Thus we can consider $\frac{1}{T}\sum \hat{\mu}_t$ as an estimator for σ^2.
- **Proposition**
  $\hat{\sigma}^2 \to \sigma^2$, where $\hat{\sigma}^2 = \frac{1}{T}\sum_t \hat{\mu}_t$.

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

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

  $$\begin{aligned} tr M_X &= tr(I_{T\times T} - X_{T\times \text{ (by) } K}(X'X)^{-1}X') = \\ &= tr I_{t\times T} - tr(X(X'X)^{-1}X') = T - tr((X'X)^{-1}X'X)_{K\times K} = T - K. \end{aligned}$$

  So we get that $\mathbb{E}(\hat{\mu}'\hat{\mu}) = (T-K)\sigma^2$.
  This says that $\mathbb{E}(\frac{1}{T-K}\sum \mu_t^2) = \sigma^2 \Rightarrow \hat{\sigma}^2 = \frac{1}{T-K}\sum \hat{\mu_t^2}$.
  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 $X_n \xrightarrow{D} X$ (in distrbution) $\Leftrightarrow \lambda' X_n \xrightarrow{D} \lambda' X$ $\forall \lambda \neq 0$.
  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 $A_n \xrightarrow{\mathbb{P}} A$ 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:
  a) $(X'X)^{1/2}(\hat{\beta} - \beta) \sim \mathcal{N}(0, \sigma^2 I)$
  b) $(T-k)\frac{\hat{\sigma}^2}{\sigma^2} \sim \chi^2_{T-k}$ – properly scaled estimator of σ^2 has the χ^2_{T-k} distribution.

  Moreover $\hat{\beta}$ and $\hat{\sigma}^2$ are independent (where $\hat{\sigma}^2 = \frac{1}{T-k}\Sigma\hat{\mu}_t^2$).

  Proof. a)

$$(X'X)^{1/2}(\hat{\beta} - \beta) = (X'X)^{1/2}X'\mu = (X'X)^{-1/2}\sum_t x_t\mu_t \sim$$
$$\sim \mathcal{N}(0, \sigma^2(X'X)^{1/2}X'X(X'X)^{-1/2}) = \mathcal{N}(0, \sigma^2 I_k)$$

  b) Note $(T-k)\frac{\hat{\sigma}^2}{\sigma^2} = \frac{1}{\sigma^2}\mu' M\mu$.
  Since $\sigma^{-1}\mu \sim \mathcal{N}(0, I)$, then $\frac{\mu' M\mu}{\sigma^2} \sim \chi^2_{rank(M)=T-k}$.
  From Tut. 2 Ex. 3 recall that $X'AX$ and βY are independent only if $AB = 0$. Note that $M_X X(X'X)^{-1} = 0$, because $M_X X = 0$ (M_X anihilates X).

  □

- **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 $\frac{X'X}{T} \to Q$.
  Therefore $(\frac{X'X}{T})^{1/2}T^{1/2}(\hat{\beta} - \beta) \sim^a \mathcal{N}(0, \sigma^2 I)$.
  Using Cramer this says that $T^{1/2}(\hat{\beta} - \beta) \sim^a \mathcal{N}(0, \sigma^2 Q^{-1})$ (asymptotic distribution).

  □

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. $\hat{\beta} - \beta = (X'X)^{-1}X'\mu = (\frac{X'X}{T})^{-1}(\frac{X'\mu}{T})$.
  By LLN $\frac{X'X}{T} \xrightarrow{\mathbb{P}} \Sigma_X$ ($\star$).
  Since Σ_X is positive definite, then by Slutsky theorem $(\frac{X'X}{T})^{-1} \xrightarrow{\mathbb{P}} \Sigma_X^{-1}$.
  Now consider $\frac{1}{T}\sum x_t\mu_t = \frac{1}{T}\sum z_t$. With x_t iid. and μ_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 $\frac{1}{T}\sum z_t \xrightarrow{\mathbb{P}} 0$. We only have to show consistency of $\hat{\beta} - \beta$.
  In effect $\hat{\beta} - \beta = (\frac{X'X}{T})^{-1}(\frac{X'\mu}{T}) \xrightarrow{\mathbb{P}} \Sigma_X^{-1} \cdot 0 = 0$.
  So $\hat{\beta}$ is consistent. □

- **Proposition**
  Under I', III and μ_t iid., $\hat{\sigma}^2 \xrightarrow{\mathbb{P}} \sigma^2$.

  Proof. $\hat{\sigma}^2 = \frac{\hat{\mu}'\hat{\mu}}{T-k} = \frac{\mu' M_X\mu}{T-k} = \frac{T}{T-K}[\frac{1}{T}\mu'\mu - \frac{X'\mu}{T}(\frac{X'X}{T})^{-1}\frac{X'\mu}{T}] \xrightarrow{\mathbb{P}} 1[\sigma^2 - 0 \cdot \Sigma_X^{-1}$ (from $\star$) $\cdot\, 0] = \sigma^2$.
  Therefore $\hat{\sigma}^2$ is consistent. □

- 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 μ_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 $\mathbb{E}\mu_t^4<\infty$ then

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

Proof. a) We have that $T^{1/2}(\hat\beta-\beta)=(\frac{X'X}{T})^{-1}\frac{X'\mu}{T^{1/2}}$. We know already that $(\frac{X'X}{T})^{-1}\to\Sigma_X^{-1}$.
What remains to prove is $(\frac{X'\mu}{T^{1/2}})\to\mathcal{N}(0,\sigma^2\Sigma_X)$.

$$\frac{X'\mu}{T^{1/2}}\equiv T^{-1/2}\sum x_t\mu_t=T^{-1/2}\sum z_t\to\mathcal{N}(0,\sigma^2\sum_X)$$

where z_t is a vector of random variables.
By Cramer-Wald we need to show that $\lambda' z_t$ converges $\forall\lambda\neq 0$.
μ_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-:evy CLT $\frac{1}{T^{1/2}}\sum\lambda' z_t\to\mathcal{N}(0,\sigma^2\lambda'\Sigma_X\lambda)$.
Since λ is arbitrary, $T^{-1/2}\sum z_t\to\mathcal{N}(0,\sigma^2\Sigma_X)$.

b) $\hat\sigma^2=\frac{1}{T-k}\sum \hat{mu}_t^2$. Consider $\tilde\sigma^2=\frac{1}{T}\sum\hat\mu_t^2$ (asymptotically it is the same).
Recall that $M_X=I-X(X'X)^{-1}X'$.

$$T^{1/2}(\tilde\sigma^2-\sigma^2)=T^{-1/2}\hat\mu'\hat\mu-T^{1/2}\sigma^2=\frac{1}{T^{1/2}}(\mu' M\mu-T\sigma^2)=$$
$$=T^{-1/2}\sum(\mu_t^2-\sigma^2)-T^{1/2}(\frac{X'X}{T})(\frac{X'X}{T})^{-1}(\frac{X'\mu}{T})\to\mathcal{N}(0,\mathbb{E}\mu_t^2-\sigma^4)$$

because $(\frac{X'X}{T})(\frac{X'X}{T})^{-1}(\frac{X'\mu}{T})\to 0$.
Next $T^{1/2}(\hat\sigma^2-\sigma^2)=(T-k)\frac{1}{T}T^{1/2}(\tilde\sigma^2-\sigma^2)+\frac{K}{T^{1/2}}\sigma^2\to 1\cdot T^{1/2}(\tilde\sigma^2-\sigma^2)+0$.
By Cramer $T^{1/2}(\hat\sigma^2-\sigma^2)\mathcal{N}(0,\mathbb{E}\mu^4-\sigma^4)$.

□

3. Remarks

a) **To convergence results**

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

b) **About consistency of $\hat\beta$**
If $\frac{X'\mu}{T}\xrightarrow{a.s.}0$ and $\frac{X'X}{T}-M_T\xrightarrow{a.s}0$,
where M_n 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 $\frac{1}{T}X'X\to\Sigma_X$ fixed and positive definite.
Note that $det(\frac{X'X}{T})-det(M_T)\xrightarrow{a.s.}0$,
because $<_T$ is bounded and determinant is continuous.
Since $\{M_T\}$ is uniformly positive definite, $\exists\delta>0$ such that determinant $det(\frac{X'X}{T})>\delta$ for sufficiently alrge T.
Then $(\frac{X'X}{T})^{-1}$ exists alsmo 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

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

□

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) $Y_nX_n\xrightarrow{D}cX$,

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 β in the linear regression: $R\beta = r$, where R is $m \times k$ and r is $m \times 1$ and $m < k$, $rank(R) = m$.
- 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: $\frac{g}{\partial \beta} = -2X'Y + 2X'X\beta + 2R\lambda' = 0$, $\frac{\partial g}{\partial \lambda} = R\beta - r = 0$.
It is easier to solve it in the matrix form

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

$$\begin{bmatrix} \hat{\beta} \\ \lambda \end{bmatrix} = \begin{bmatrix} X'X & R' \\ R & 0 \end{bmatrix}^{-1} \begin{bmatrix} X'Y \\ r \end{bmatrix} =$$
$$= \begin{bmatrix} (X'X)^{-1}(I - R'(R(X'X)^{-1}R')^{-1}R(X'X)^{-1} & (X'X)^{-1}R(R(X'X)^{-1}R')^{-1} \\ -(R(X'X)^{-1}R')^{-1}R(X'X)^{-1} & -(R(X'X)^{-1}R')^{-1} \end{bmatrix} \begin{bmatrix} X'Y \\ r \end{bmatrix}$$

This gives a constrained least sqaures estimator

$$\hat{\beta} = (X'X)^{-1}(I - R'(R(X'X)^{-1}R')^{-1}R(X'X)^{-1})X'Y + (X'X)^{-1}R'(R(X'X)^{-1}R')^{-1}r =$$
$$= \beta^{OLS} - (X'X)^{-1}R'(R(X'X)^{-1}R')^{-1}(R\beta^{OLS} - r)$$

If we have constraints in linear regression, we get $\hat{\beta} = \beta^{OLS} -$ correction term.

$$\hat{\beta} = \beta^{OLS} \text{ if } \beta^{OLS} \text{ satisfies our constraint } R\beta^{OLS} - r = 0.$$

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) $\mu_t \sim \mathcal{N}(0, \sigma^2 I)$

  where β solves $R\beta = r$ with $rank(R) = m$

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

Remarks:
$\hat{\beta}$ 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 $\frac{q_1/n_1}{1_2/n_2} \sim F_{n_1,n_2}$ $(\star)$.

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! $Q \cdot Q = Q$), $rank(Q) = m$.
So we have that $\mu'Q\mu \sim \chi^2_m$ and $(T-k)\frac{\hat{\sigma}^2}{\sigma^2} = \frac{\mu' M_X \mu}{\sigma^2} \sim \chi^2_{T-k}$. So from $(\star)$, we have $(\odot)$. In addition $M_X Q = 0 \Rightarrow q_1$ and q_2 are independent. ☐

TUTORIAL 3, 02/03/2012

1. Exercise 1
Establish consistency of β in the following linear model $y_t = \alpha + \beta t + \mu t$, $t = 1, \ldots, T$, where $\mathbb{E}\mu_t = 0$, $\mathbb{E}\mu_t^2 = \sigma^2$ and $\mathbb{E}(\mu_t \mu_s) = 0 \ \forall t \neq s$. 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 $\mathbb{E}\hat{\beta} = \beta$
$Var\hat{\beta} = \sigma^2(X'X)^{-1} = \frac{\sigma^2}{\sum_t (t-\bar{t})^2}$, *where* $\bar{t} = \frac{1}{T}\sum_{t=1}^T t$.
To show consistecy it is sufficient to show that $Var\hat{\beta} \to 0$ *(it follows from Markov's inequality that checking the second moment is enough for consistency of estimator).*
$$\frac{\sigma^2}{\sum_t (t-\bar{t})^2} = \frac{\sigma^2}{\sum t^2 - T\bar{t}^2} = \frac{\sigma^2}{\frac{1}{6}T(T+1)(2T+1) - \frac{1}{4}T(T+1)^2} = \frac{12\sigma^2}{T(T^2-1)} \to 0$$
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*
$$Var(T^{3/2}(\hat{\beta} - \beta)) = T^3 Var(\hat{\beta} - \beta) = \frac{12\sigma^2}{1 - T^2} \to 12\sigma^2$$
So $T^{3/2}(\hat{\beta} - \beta) \xrightarrow{a} \mathcal{N}(0, 12\sigma^2)$.

2. ♢ Exercise 2
Consider a regression model $y_t = \beta_1 x_t^2 + \beta_2 x_t^2 + \mu_t$, where x_t^1 and x_t^2 are centered (zero mean). Let ρ be a simple correlation of x_t^1 and x_t^2
$$\hat{\rho} = \frac{Cor(x_t^1, x_t^2)}{\sqrt{Varx_t^1 Varx_t^2}}.$$
Show that $corr(\hat{\beta}_1, \hat{\beta}_2) = -\hat{\rho}$. 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 β_1 and β_2.
Hint:
Matrix form: $Y = \beta_1 X_1 + \beta_2 X_2 + \mu$
Show that $\hat{\beta}_1 = (X_1' M_2 X_1)^{-1} X_1' M_2 Y$ and similarly $\hat{\beta}_2 = (X_2' M_1 X_2)^{-1} X_2' M_1 Y$, 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)