Cheatsheet for Econ 203A–C Comp 2019 by Jesper Bojeryd – jesperbojeryd@ucla.edu

Distributions

A CDF $F(t)$ is right continuous, i.e., $\lim_{\Delta\downarrow 0} F(t+\Delta) = F(t)$. *CADLAG*.

Transformation of distr.

If g is a 1–1, differentiable function, then $Y = g(X)$ has pdf

$$f_Y(y) = f_X(g^{-1}(y)) \left| \frac{\partial g^{-1}(y)}{\partial y} \right|.$$

In multivariate case, analogously,

$$f_Y(y_1, \dots, y_n) = \\ = f_X(g_1^{-1}(y_1), \dots, g_n^{-1}(y_n))|J|.$$

Student's theorem

Let $X_1, \dots, X_n$ be i.i.d. r.v. $\sim \mathcal{N}(\mu, \sigma^2)$ and $S_n^2 \equiv \frac{1}{n-1}\sum_{i=1}^n (X_i - \bar{X}_n)^2$. Then,

1. $\bar{X}_n = \mathcal{N}(\mu, \sigma^2/n)$;
2. $\bar{X}_n$ and S_n^2 are independent;
3. $(n-1)S_n^2/\sigma^2 \sim \chi^2(n-1)$;
4. $T \equiv \dfrac{\bar{X}_n - \mu}{S_n/\sqrt{n}} \sim t(n-1)$.

Gamma distribution

$\text{supp}(X) = \mathbb{R}^+$

PDF: $f(x) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x}$, CDF: *omit.*

MGF: $M(t) = (1 - t/\beta)^{-\alpha}$ for $t < \beta$

$\mathbb{E}[X] = \alpha/\beta$, $\mathbb{V}\text{ar}[X] = \alpha/\beta^2$

Notes: $\Gamma(1) = 1$, $\Gamma(\alpha) = (\alpha-1)\Gamma(\alpha-1)$, if $1 < \alpha \in \mathbb{N}$, $\Gamma(\alpha) = (\alpha-1)!$, $\Gamma(1/2) = \sqrt{\pi}$.

$\chi^2(r)$ distribution

$\text{supp}(X) = \mathbb{R}^+$, PDF: *omit.*

MGF: $M(t) = (1-2t)^{-r/2}$ for $t < 1/2$

$\mathbb{E}[X] = r$, $\mathbb{V}\text{ar}[X] = 2r$

Notes: For a seq. $\{X_i \sim \chi^2(r_i)\}_{i=1}^n$ of indep. r.v., $\sum_{i=1}^n X_i \sim \chi^2(\sum_{i=1}^n r_i)$.

If $Z \sim \mathcal{N}(0,1)$, then $Z^2 \sim \chi^2(1)$

Poisson distribution

$\text{supp}(X) = \{0, 1, 2, \dots\}$

PMF: $f(x) = \frac{\lambda^x \exp(-\lambda)}{x!}$, CMF: *omit.*

MGF: $M(t) = \exp(\lambda(e^t - 1))$

$\mathbb{E}[X] = \lambda$, $\mathbb{V}\text{ar}[X] = \lambda$

Binomial distribution

$X = k$ successes in a sequence of n i.i.d. draws.

PDF: $f(k) = \binom{n}{k} p^k (1-p)^{n-k}$

CDF: $F(k) = \sum_{i=1}^k f(i)$

MGF: $M(t) = (1 - p + p\exp(t))^n$

$\mathbb{E}[X] = np$, $\mathbb{V}\text{ar}[X] = np(1-p)$

Bernoulli(p) distribution

$$X = \begin{cases} 1, & \text{w/ prob. } p \\ 0, & \text{w/ prob. } 1-p \end{cases}$$

PMF: $f(x) = p^x(1-p)^{1-x}$

MGF: $M(t) = q + pe^t$

$\mathbb{E}[X] = p$, $\mathbb{V}\text{ar}[X] = p - p^2$

Normal distribution $\mathcal{N}(\mu, \sigma^2)$

PDF: $f(x) = \dfrac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\dfrac{(x-\mu)^2}{2\sigma^2}\right)$

CDF: $F(x) = \Phi((x-\mu)/\sigma)$

MGF: $M(t) = \exp(\mu t + \sigma^2 t^2/2)$

Notes: For a seq. normally distr. independent r.v. $\{X_i \overset{\text{indep.}}{\sim} \mathcal{N}(\mu_i, \sigma_i^2)\}_{i=1}^n$, $\sum_{i=1}^n X_i \sim \mathcal{N}(\sum_{i=1}^n \mu_i, \sum_{i=1}^n \sigma_i^2)$.

If $X \sim \mathcal{N}(\mu, \Sigma)$, then the linear comb. $Ax + b \sim \mathcal{N}(A\mu + b, A\Sigma A')$.

For $\mathcal{N}(0,1)$: $\mathbb{E}[X] = 0$, $\mathbb{E}[X^2] = 1$, $\mathbb{E}[X^3] = 0$, $\mathbb{E}[X^4] = 3$, $\mathbb{E}[X^5] = 0$, $\mathbb{E}[X^6] = 105$

Uniform distribution $U[a,b]$

$\text{supp}(X) = [a,b]$

PDF: $f(x) = \frac{1}{b-a}$

CDF: $F(x) = \frac{x-a}{b-a}$

MGF: $M(t) = \begin{cases} \frac{\exp(tb) - \exp(ta)}{t(b-a)} & \text{if } t \neq 0 \\ 1 & \text{if } t = 0 \end{cases}$

$\mathbb{E}[X] = (a+b)/2$, $\mathbb{V}\text{ar}[X] = (b-a)^2/12$

Student's t-distribution $t(\nu)$

Def: $t(\nu) = \dfrac{\mathcal{N}(0,1)}{\sqrt{\chi^2(n)/n}}$, $\nu = n-1$

MGF: $\nexists$, $\mathbb{E}[X] = 0$ if $\nu > 1$, $\mathbb{E}[X] =$ undef. if $\nu \leq 1$, $\mathbb{V}\text{ar}[X] = \nu/(\nu-2)$ if $\nu > 2$, $\mathbb{V}\text{ar}[X] = \infty$ if $\nu \leq 2$

Exponential distr. $\text{Exp}(\lambda)$

$\text{supp}(X) = \mathbb{R}^+$

PDF: $f(x) = \lambda\exp(-\lambda x)$

CDF: $F(x) = 1 - \exp(-\lambda x)$

MGF: $\frac{\lambda}{\lambda - t}$, for $t < \lambda$

$\mathbb{E}[X] = \lambda^{-1}$, $\mathbb{V}\text{ar}[X] = \lambda^{-2}$

F distribution

For $U \sim \chi^2(r_1)$, $V \sim \chi^2(r_2)$, $W \equiv \frac{U/r_1}{V/r_2}$. Then $W \sim F(r_1, r_2)$.

PDF: *omit.*, MGF: $\nexists$

$\mathbb{E}[X] = r_2/(r_1 - 2)$ for $r_2 > 2$,

$\mathbb{V}\text{ar}[X] = \dfrac{2r_2^2(r_1 + r_2 - 2)}{r_1(r_2-2)^2(r_2-4)}$ for $r_2 > 4$

$\mathbb{E}[\cdot]$, $\mathbb{P}[\cdot]$, $M(t)$, etc.

LoUS: $\mathbb{E}[g(X)] = \int_\Omega g(x) \mathrm{d}F_X(x)$

MGF: For a r.v. X, $M(t) \equiv \mathbb{E}[e^{Xt}]$ for $-h < t < h$ where the expectation exists.

MGF generalize to joint r.v.'s X_1, X_2: $M_{X_1,X_2}(t_1, t_2) \equiv \mathbb{E}[e^{X_1 t_1 + X_2 t_2}]$.

Also, $M_{X_1}(t_1) = M_{X_1,X_2}(t_1, 0)$.

For pos. integers m, $\mathbb{E}[X^m] = M^{(m)}(0)$.

A useful thm: $F_Y(s) = F_X(s)$, $\forall s \Leftrightarrow M_Y(t) = M_X(t)$, $\forall t \in (-h, h)$.

DeMorgan's Laws:

$$(C_1 \cup C_2)^C = C_1^C \cap C_2^C,$$
$$(C_1 \cap C_2)^C = C_1^C \cup C_2^C.$$

Boole's inequality:

$$\mathbb{P}\left(\cup_{i=1}^n C_i\right) \leq \sum_{i=1}^n \mathbb{P}\left(C_i\right)$$

Bonferroni's inequality:

$$\mathbb{P}\left(C_1 \cap C_2\right) \geq \mathbb{P}\left(C_1\right) + \mathbb{P}\left(C_2\right) - 1$$

Permutations:

$$\mathbb{P}_k^n = \frac{n!}{(n-k)!}$$

Combinations:

$$C_k^n = \binom{n}{k} = \frac{n!}{k!\,(n-k)!}$$

Inequalities

Markov's inequality: If $\mathbb{E}[|X|] < \infty$ and $a > 0$, then

$$\mathbb{P}(|X| > a) \leq \frac{\mathbb{E}[|X|]}{a}$$

Chebyshev's inequality:

$$\mathbb{P}(|X - \mu_X| > b) \leq \frac{\sigma_X^2}{b^2}$$

Jenses's inequality: For a convex ϕ,

$$\phi(\mathbb{E}[X]) \leq \mathbb{E}[\phi(X)]$$

Indep. & cond.

Def. of indep. events:

$$\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$$

Def. of $\mathbb{E}[\cdot|\cdot]$: For r.v.'s X, Y,

$$\mathbb{E}[Y|X] \equiv \arg\min_{\varphi} \mathbb{E}[(Y - \varphi(X))^2]$$

Law of Total Probability:

$$\mathbb{P}(A) = \sum_{i=1}^n \mathbb{P}(A|C_i)\mathbb{P}(C_i)$$

Law of Total Expectation:

$$\mathbb{E}(X) = \sum_{i=1}^n \mathbb{E}[X|C_i]\mathbb{P}(C_i)$$

Bayes' rule:

$$\mathbb{P}(A|B) = \mathbb{P}(B|A)\frac{\mathbb{P}(B)}{\mathbb{P}(A)}$$

Law of Iterated Expectations:

$$\mathbb{E}[Y] = \mathbb{E}[\mathbb{E}[Y|X]]$$

Take-out-what's-known:

$$\mathbb{E}[h(X)Y|X] = h(X)\mathbb{E}[Y|X]$$

Convergences

Monotone Conv. Thm.

For a seq. meas., non-neg. func. $\{f_n\}_n$ on $(\Omega, \mathcal{F}, \mathbb{P})$, s.t. $f_n < f_{n+1}$ and $f_n \to f$,

$$\lim_{n\to\infty} \mathbb{E}[f_n] = \mathbb{E}[f].$$

Dominated Conv. Thm.

For a seq. meas. func. $\{f_n\}_n$ on $(\Omega, \mathcal{F}, \mathbb{P})$, suppose p-w conv. a.s. to a func. f, and $\exists g > 0$, and

$$|f_n(x)| \leq |g(x)| \; \forall x, n, \text{ and } \mathbb{E}[|g|] < \infty.$$

Then $\mathbb{E}[|f|] < \infty$, and

$$\lim_{n\to\infty} \mathbb{E}[f_n] = \mathbb{E}[f].$$

Central limit theorem:
If X_i is an i.i.d. seq.,
$$\sqrt{n}(\bar{X}_n - \mathbb{E}[X]) \xrightarrow{d} \mathcal{N}(0,\sigma^2)$$

Delta method:
If $\sqrt{n}(X_n-\theta) \xrightarrow{d} \mathcal{N}(0,\sigma^2)$ and $g(\cdot)$ is continuously differentiable at θ, then
$$\sqrt{n}(g(X_n)-g(\theta)) \xrightarrow{d} \mathcal{N}(0,\sigma^2[g'(\theta)]^2).$$

Convergence in $\mathbb{P}$, L^r, d, $a.s.$
$\xrightarrow{a.s.}$: $P(\{w : X_n(w) \to X(w)\}) = 1$
$\xrightarrow{L^r}$: $E[|X_n - X|^r] \to 0$
$\xrightarrow{\mathbb{P}}$: If $\forall\epsilon>0$, we have
$$\lim_{n\to\infty} P[|X_n - X| \ge \epsilon] = 0 \equiv$$
$$\equiv \lim_{n\to\infty} P[(X_n - X) < \epsilon] = 1.$$
$\xrightarrow{d}$: For $\{X_n\}$ and X, let the cdfs be F_{X_n} and F_X. Let $C(F_X)$ denote the set of x where F_X is continuous. $X_n \xrightarrow{d} X$ if
$$\lim_{n\to\infty} F_{X_n}(x) = \lim_{n\to\infty} P(X_n \in (-\infty, x]) =$$
$$= F_X(x), \forall x \in C(F_X).$$

Bounded in prob.: if $\forall\epsilon>0$, $\exists B_\epsilon>0$ s.t. $\forall n \ge N_\epsilon \in \mathbb{Z}$, $\mathbb{P}[|X_n| \le B_\epsilon] \ge 1-\epsilon$, then X_n is bound. in prob.

About $\mathcal{O}_p(1), o_p(1)$, etc.

- If $X_n \xrightarrow{d} X$, then $X_n + o_p(1) \xrightarrow{d} X$
- $(a+o_p(1)) + (b+o_p(1))X_n \xrightarrow{d} a+bX$
- $\mathcal{O}_p(1) + \mathcal{O}_p(1) = \mathcal{O}_p(1)$
- $\mathcal{O}_p(1) \cdot \mathcal{O}_p(1) = \mathcal{O}_p(1)$
- $\mathcal{O}_p(1) \cdot o_p(1) = o_p(1)$
- $o_p(1) + o_p(1) = o_p(1)$
- $o_p(1) \cdot o_p(1) = o_p(1)$
- $(X+o_p(1)) + (Y+o_p(1)) =$
  $= X + Y + o_p(1)$
- $(X+o_p(1)) \cdot (Y+o_p(1)) =$
  $= X \cdot Y + o_p(1)$
- $g(a+o_p(1)) = g(a) + o_p(1)$
- For r.v. X, $X \cdot o_p(1) = o_p(1)$.

$\xrightarrow{d}$ does **not** imply $\xrightarrow{L^s}$.

$$\begin{array}{ccc} \xrightarrow{L^r} & \xrightarrow{\text{if } s<r} & \xrightarrow{L^s} \\ & & \Downarrow \\ \xrightarrow{a.s.} & \Longrightarrow \xrightarrow{\mathbb{P}} \Longrightarrow & \xrightarrow{d} \end{array}$$

$X_n \xrightarrow{d} X$ and $Y_n \xrightarrow{d} Y$ (*marginal* conv.) does **not** imply $(X_n, Y_n) \xrightarrow{d} (X, Y)$ (*joint* conv.).

Cont. Mapp. Thm and Slutsky:
If $X_n \xrightarrow{d} X$ and $g(\cdot)$ is cont., then $g(X_n) \xrightarrow{d} g(X)$. If $A_n \xrightarrow{\mathbb{P}} a$ and $B_n \xrightarrow{\mathbb{P}} b$ (a, b are const.), then $A_nX_n + B_n \xrightarrow{d} aX + b$.

Identification

Def: h^* is identified within $H \Leftrightarrow \forall h \neq h^* \in H$, $F_{Y,X}(\,\cdot\,; h) \neq F_{Y,X}(\,\cdot\,; h^*)$.

Linear regressions

Geometric intuition

$$\mathbb{P}_n \equiv \mathbb{X}_n(\mathbb{X}_n'\mathbb{X}_n)^{-1}\mathbb{X}_n',\ \mathbb{M}_n \equiv I_n - \mathbb{P}_n$$
$$\mathbb{P}_n\mathbb{P}_n = \mathbb{P}_n,\ \mathbb{M}_n\mathbb{M}_n = \mathbb{M}_n,\ \mathbb{P}_n\mathbb{X}_n = \mathbb{X}_n$$
$$\mathbb{P}_n\mathbb{M}_n = 0,\ \mathbb{P}_n\mathbb{M}_n = 0,\ \mathbb{M}_n\mathbb{X}_n = 0$$
Also, $\|a\|^2 = \|\mathbb{P}_n a\|^2 + \|\mathbb{M}_n a\|^2$.

OLS regression

Estimator:
$$\hat{\beta}_n = (\mathbb{X}_n'\mathbb{X}_n)^{-1}\mathbb{X}_n'\mathbb{Y}_n =$$
$$= \left(\frac{1}{n}\sum_{i=1}^n X_iX_i\right)^{-1}\frac{1}{n}\sum_{i=1}^n X_iY_i$$
$$\mathbb{P}_n\mathbb{Y}_n = \mathbb{X}_n\hat{\beta}_n$$

OLS assumptions:

1. $\{Y_i, X_i\}_{i=1}^n$ is i.i.d.;
2. $\mathbb{E}[Y^2] < \infty$;
3. $\mathbb{E}[XX'] < \infty$ and invertible,
4. $\mathbb{E}[\|X\|^2U^2] < \infty$

$\hat{\beta}_n$ is a consistent estimator if OLS 1–3 hold.

Asympt. norm.: If OLS 1, 3, 4 hold,
$$\sqrt{n}(\hat{\beta}_n - \beta_0) \xrightarrow{d}$$
$$\mathcal{N}\big(0, (\mathbb{E}[XX'])^{-1}\mathbb{E}[XX'U^2](\mathbb{E}[XX'])^{-1}\big)$$
Homosk. If $\mathbb{E}[U^2|X] = \sigma^2$ w.p. 1 over X.

OLS w/ intercept
$$(\alpha, \beta) = \arg\min_{a,b}\sum_{i=1}^n(Y_i - a - Z_i'b)^2$$
$$\hat{\beta}_n = \frac{\sum_{i=1}^n(Y_i - \bar{Y}_n)(Z_i - \bar{Z}_n)}{\sum_{i=1}^n(Z_i - \bar{Z}_n)^2}.$$

Conditional mean is linear
Assume $\mathbb{E}[Y|X] = X'\gamma_0$, i.e., linear. By LoIE, $\mathbb{E}[(Y - X'\gamma_0)X] = \mathbb{E}[YX] - \mathbb{E}[YX] = 0$, which implies that $\mathbb{E}\left[(Y - X'\gamma_0)X\right] = \mathbb{E}\left[(Y - X'\beta_0)X\right]$ or that $\beta_0 = \gamma_0$, whenever $\mathbb{E}[XX']$ is full rank.
Thus, the OLS estimand β_0 corresponds to the cond. exp. param. in $E[Y|X] = X'\beta_0$.

Best linear predictor
If $Y \in \mathbb{R}$, $X \in \mathbb{R}$, $\mathbb{E}[Y^2] < \infty$, and $\mathbb{E}[XX']$ is full rank, then
$$\beta_0 \equiv \arg\min_{b\in\mathbb{R}^d} E\left[(Y - X'b)^2\right] =$$
$$= \arg\min_{b\in\mathbb{R}^d} E\left[(E[Y|X] - X'b)^2\right].$$

Meas. of fit: $R^2 \equiv 1 - \frac{RSS}{TSS}$ where,
$$RSS \equiv \frac{1}{n}\sum_{i=1}^n((Y_i - \bar{Y}_n) - (X_i - \bar{X}_n)'\hat{\beta}_n)^2$$
$$TSS \equiv \frac{1}{n}\sum_{i=1}^n(Y_i - \bar{Y}_n)^2$$

The Wald test: Suppose for the single linear restriction we choose a $r \in \mathbb{R}^d$ and $b \in \mathbb{R}$ and we are interested in testing: $H_0 : r'\beta_0 = b$ versus $H_1 : r'\beta_0 \neq b$. A special case would be setting r to a vector of zeros with one one and setting $b = 0$, which would test a single regressor is equal to zero.
The Wald test is set up so we reject for large values:
$$\phi_n \equiv \mathbb{1}\left(\frac{1}{\sqrt{r'\hat{\Sigma}_n r}}\left|\sqrt{n}(r'\hat{\beta}_n - b)\right| > c_{1-\alpha/2}\right)$$
Thm Let the OLS-2 hold, $\sigma^2 \equiv r'\Sigma_0 r$, $Z \sim \mathcal{N}(0,1)$, and
$$\Sigma_0 \equiv \mathbb{E}\left[XX'\right]^{-1}\mathbb{E}\left[XX'U^2\right]\mathbb{E}\left[XX'\right]^{-1}.$$
If $r'\beta_0 = b$, then it follows:
$$\left|\sqrt{n}\left\{r'\hat{\beta}_n - b\right\}\right| \xrightarrow{d} |\sigma_0 Z| \sim \left|\mathcal{N}\left(0, r'\Sigma_0 r\right)\right|$$

Thm If OLS-2 holds, $\hat{\Sigma}_n \xrightarrow{p} \Sigma_0$, $\sigma_0 > 0$, and $r'\beta_0 = b$, then (where $c_{1-\alpha}$ is the $1-\alpha$ quantile of the standard normal random variable):
$$\lim_{n\to\infty}\mathbb{P}\left(\frac{\sqrt{n}}{\sqrt{r'\hat{\Sigma}_n r}}\left|r'\hat{\beta}_n - b\right| > c_{1-\alpha/2}\right) = \alpha.$$

IV regression

IVs are used when X is correlated with ϵ, in which case OLS results will be biased. Such correlation may occur

1. when changes in Y change the value of at least one of the covariates ("reverse" causation);
2. when there are omitted variables that affect both Y and X; or
3. when X is subject to non-random measurement error.

X is *endogenous* if any of above cases apply. If an instrument is available, consistent estimates may be obtained. An instrument is a variable that does not itself belong in the explanatory equation but is correlated with the endogenous explanatory variables, conditional on the value of other covariates.

Estimand: β_0 solves
$$\mathbb{E}[(Y - X'\beta_0)Z] = 0$$

Estimator:
$$\hat{\beta}_n = \arg\min_{b\in\mathbb{R}^{d_x}}\left\|\frac{1}{n}\sum_{i=1}^n(Y_i - X_i'b)Z_i\right\|^2$$
$$= (\mathbb{X}_n'\mathbb{Z}_n\hat{\Omega}_n\mathbb{Z}_n'\mathbb{X}_n')^{-1}\mathbb{X}_n'\mathbb{Z}_n\hat{\Omega}_n\mathbb{Z}_n'\mathbb{Y}_n$$

Consistency: If $\{Y_i \in \mathbb{R}, X_i \in \mathbb{R}^{d_x}, Z_i \in \mathbb{R}^{d_z}\}_{i=1}^n$ is i.i.d. and holds for the moment condition for some β_0; $\hat{\Omega}_n \to \Omega$ for some Ω, rank$(\mathbb{E}[XZ']) = d_x$; and $\mathbb{E}[\|XZ'\|] < \infty$, then $\hat{\beta}_n$ is a consistent estimator of β_0.

Asympt. norm.: Under same assumptions as for consistency, and $\mathbb{E}[ZZ'U^2] < \infty$, then the limit distr. of $\sqrt{n}(\hat{\beta}_n - \beta_0)$ is
$$\mathcal{N}(0, K\mathbb{E}[XZ']\Omega\mathbb{E}[ZZ'U^2]\Omega\mathbb{E}[ZX']K)$$
where $K = (\mathbb{E}[XZ']\Omega\mathbb{E}[ZX'])^{-1}$.

2SLS and 3SLS/Choice of Ω:

Choosing $\hat{\Omega}_n = \left(\frac{1}{n}\sum_{i=1}^{n}\sum Z_i Z_i'\right)^{-1}$ is **2SLS**, equiv. with the following algorithm:

1. Regress $\mathbb{X}_n$ on $\mathbb{Z}_n$ and fit $\hat{\mathbb{X}}_n$.
2. Regress $\mathbb{Y}_n$ on $\hat{\mathbb{X}}_n$ to estimate β_0.

We can see the equivalence from plugging in the 2SLS $\hat{\Omega}_n$ into the $\hat{\beta}_n$ equation:

$$\hat{\beta}_n = [\mathbb{X}_n'\mathbb{Z}_n(\mathbb{Z}_n'\mathbb{Z}_n)^{-1}\mathbb{Z}_n'\mathbb{X}_n]^{-1}$$
$$\mathbb{X}_n'\mathbb{Z}_n(\mathbb{Z}_n'\mathbb{Z}_n)^{-1}\mathbb{Z}_n'\mathbb{Y}_n =$$
$$= [(\mathbb{P}_n^Z\mathbb{X}_n)'(\mathbb{P}_n^Z\mathbb{X}_n)]^{-1}(\mathbb{P}_n^Z\mathbb{X}_n)'\mathbb{Y}_n =$$
$$= (\hat{X}X_n'\hat{X}X_n)^{-1}\hat{X}X_n'\mathbb{Y}_n.$$

Remember: $\mathbb{P}_n^Z \equiv \mathbb{Z}_n(\mathbb{Z}_n'\mathbb{Z}_n)^{-1}\mathbb{Z}_n'$.

3SLS corresponds to the following:

1. Obtain $\tilde{\beta}_n$ that is consistent for β_0.
2. Create $\tilde{U}_i = (Y_i - X_i'\tilde{\beta}_n)$ and set $\hat{\Omega}_n = (\frac{1}{n}\sum_{i=1}^{n}\sum Z_i Z_i'\tilde{U}_i^2)$.
3. Solve $\hat{\beta}_n =$
$= [\mathbb{X}_n'\mathbb{Z}_n\hat{\Omega}_n\mathbb{Z}_n'\mathbb{X}_n]^{-1}\mathbb{X}_n'\mathbb{Z}_n\hat{\Omega}_n\mathbb{Z}_n'\mathbb{Y}_n$ with $\hat{\Omega}_n$.

LATE – Local aver. treat. effect

$Y \in \mathbb{R}$, $D \in \{0,1\}$, and $Z \in \{0,1\}$ such that we observe: $Y = Y(0) + D(Y(1) - Y(0))$, $D(1)$ if $Z = 1$, and $D(0)$ if $Z = 0$; i.e., Z affects the treatment decision. Similarly we observe: $D = D(0) + Z(D(1) - D(0))$.

We make two LATE assumptions:
LATE-1

(a) $(Y(1), Y(0), D(1), D(0)) \perp Z$;
(b) $\mathbb{P}(D(1) \neq D(0)) > 0$.

LATE-2:

(a) monotonicity: $D(1) \geq D(0)$ a.s. – no defiers.

We can obtain the LATE estimator using 2SLS to estimate the β that we assume solves the moment restrictions:

$$\mathbb{E}[(Y - \beta_{00} - D\beta_{01})\begin{bmatrix}1\\Z\end{bmatrix}] = 0$$

If we solved this out, we would find that:

$$\beta_{01} = \frac{\mathbb{C}\text{ov}[Y,Z]}{\mathbb{C}\text{ov}[D,Z]}.$$

By Law of total expectations,

$$\beta_{01} = \frac{\mathbb{E}[Y|Z=1] - \mathbb{E}[Y|Z=0]}{\mathbb{E}[D|Z=1] - \mathbb{E}[D|Z=0]},$$

or using LATE-1 that we have:

$$\beta_{01} = \frac{\mathbb{E}[(Y(1) - Y(0))(D(1) - D(0))]}{\mathbb{E}[D(1) - D(0)]}$$

Under LATE-2 we then have:

$$\beta_{01} = E[Y(1) - Y(0)|D(1) - D(0) = 1].$$

This β_{01} is the TE on the compliers/LATE.

Panel data

Clustered data

Thm Assuming

1. $\{Y_i, X_i\}_{i=1}^n$ is i.i.d.;
2. $\mathbb{E}\left[\sum_{t=1}^{T} X_{it}U_{it}\right] = 0$;
3. $\mathbb{E}[X_i'X_i]$ is finite and invertible; and,
4. $\sum_{t=1}^{T}\mathbb{E}\left[\|X_{it}\|^2 U_{it}^2\right] < \infty$;

then:

$$\sqrt{n}(\hat{\beta}_n - \beta_0) \xrightarrow{d}$$
$$\mathcal{N}(0, (S\mathbb{E}[(\sum_{t=1}^{T} X_{it}U_{it})(\sum_{t=1}^{T} X_{it}U_{it})']S),$$

where $S = \mathbb{E}[X_i'X_i])^{-1}$.

We can then estimate the asympt. variance of $\hat{\beta}_n$ with the sample analogue:

$$\hat{S}\frac{1}{n}\sum_{i=1}^{n}\left((\sum_{t=1}^{T} X_{it}\hat{U}_{it})(\sum_{t=1}^{T} X_{it}\hat{U}_{it})'\right)\hat{S}$$

where $\hat{U}_{it} \equiv Y_{it} - X_{it}'\hat{\beta}_n$ and $\hat{S} = \left(\frac{1}{n}\sum_{i=1}^{n} X_i'X_i\right)^{-1}$.

Split the middle term into a "standard term" and a 2nd corr.-within-cluster term:

$$\frac{1}{n}\sum_{i=1}^{n}\left((\sum_{t=1}^{T} X_{it}\hat{U}_{it})(\sum_{t=1}^{T} X_{it}\hat{U}_{it})'\right) =$$
$$= \frac{1}{n}\sum_{i=1}^{n}\sum_{t=1}^{T} X_{it}X_{it}'\hat{U}_{it}^2 +$$
$$+ \frac{2}{n}\sum_{i=1}^{n}\sum_{t=1}^{T-1}\sum_{t'>t} X_{it}X_{it'}'\hat{U}_{it}\hat{U}_{it'}.$$

Random Effects

Assume U_{it} contains an individual error A_i and an individual-time error V_{it}, s.t.:

$$Y_{it} = X_{it}'\beta_0 + \underbrace{A_i + V_{it}}_{U_{it}}$$

In an RE model, we assume that X_{it} is exogenous, i.e. uncorrelated with (A_i, V_{it}) as well as that A_i and V_{it} are i.i.d.. In FE we are concerned with corr. b/w A_i and X_{it}.

RE-1:

(i) $\{Y_i, X_i\}_{i=1}^n$ is i.i.d. and satisfies the RE model;
(ii) $\mathbb{E}[A_i|X_i] = 0$ and $\mathbb{E}[V_i|X_i, A_i] = 0$;
(iii) $\mathbb{E}[A_i^2|X_i] = \sigma_A^2$ and $\mathbb{E}[V_iV_i'|X_i, A_i] = \sigma_V^2 I_T$

RE-1 (ii) implies $E[(Y_{it} - X_{it}'\beta_0)X_{it}] = 0, \forall 1 \leq t \leq T \wedge 1 \leq \tilde{t} \leq T \implies E[X_i'\Omega(Y_i - X_i\beta_0)] = 0$

RE-2:

(i) $\hat{\Omega}_n \xrightarrow{p} \Omega \wedge E[X_i'\Omega X_i]$ is full rank;
(ii) $E\left[\|X_i\|^2\right] < \infty$.

Given RE-2 (i),

$$\hat{\beta}_n^{\text{re}} = (\frac{1}{n}\sum_{i=1}^{n} X_i'\hat{\Omega}_n X_i)^{-1}(\frac{1}{n}\sum_{i=1}^{n} X_i'\hat{\Omega}_n Y_i).$$

Asympt. norm.: Let $\Sigma \equiv \mathbb{E}[U_iU_i'|X_i]$ and RE-1 and RE-2 hold. Then:

$$\sqrt{n}\left(\hat{\beta}_n^{\text{re}} - \beta_0\right) \xrightarrow{d}$$
$$\mathcal{N}\left(0, S\mathbb{E}\left[X_i'\Omega\Sigma\Omega X_i\right]S\right)$$

where $S \equiv (\mathbb{E}[X_i'\Omega X_i])^{-1}$

From the asympt. var. for the FE estimator, we can see that, as in IV, there is an efficient $\Omega = \Sigma^{-1} \equiv (E[U_iU_i'|X_i])^{-1}$.

We have the following procedure for RE:

1. Obtain a $\tilde{\beta}_n$ that is consistent for β_0 – for instance by solving the sample analogue of the moment conditions with $\hat{\Omega}_n = I_T$.
2. Employing $\tilde{\beta}_n$ create residuals $\tilde{U}_{it} = \left(Y_{it} - X_{it}'\tilde{\beta}_n\right)$ and motivated by the structure of $E[U_iU_i'|X_i]$ let:
$$\hat{\alpha}_A^2 \equiv \frac{1}{nT(T-1)/2} \cdot \sum_{i=1}^{n}\sum_{t=1}^{T-1}\sum_{\tilde{t}=t+1}^{T}\tilde{U}_{it}\tilde{U}_{i\tilde{t}}$$
$$\hat{\alpha}_B^2 \equiv \frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\left(\tilde{U}_{it}\right)^2 - \hat{\sigma}_A^2$$
3. Employing $\hat{\sigma}_A^2$ and $\hat{\sigma}_B^2$, compute $\hat{\beta}_n^{\text{re}}$ by solving the $\hat{\beta}_n^{\text{re}}$ estimation problem with $\hat{\Omega}_n$ set to equal $\hat{\sigma}_A^2 + \hat{\sigma}_B^2$ on the diagonal and $\hat{\sigma}_A^2$ o/w.

Fixed Effects

In FE model, we maintain the same model but change our assumption on the residual U_{it}:

$$Y_{it} = X_{it}'\beta_0 + \underbrace{A_i + V_{it}}_{U_{it}}$$

where U_{it} contains an individual specific A_i, and an individual and time specific V_{it}, and:

FE-1

(i) $\{Y_i, X_i\}_{i=1}^n$ is i.i.d. from FE model;
(ii) $\mathbb{E}[V_i|X_i, A_i] = 0$

We no longer require that $\mathbb{E}[A_i|X_i] = 0$.

FE as Demeaning We have: $\dot{Y}_{it} \equiv Y_{it} - \overline{Y} = \dot{X}_{it}'\beta_0 + \dot{V}_{it}$ and

$$\mathbb{E}[\dot{V}_{it}\dot{X}_{it}] = \mathbb{E}[(V_{it} - \overline{V}_i)(X_{it} - \dot{X}_i)] =$$
$$= \mathbb{E}[\mathbb{E}[(V_{it} - \overline{V}_i)|X_i](X_{it} - \dot{X}_i)] = 0.$$

Define FE regressor as:

$$\hat{\beta}_n^{\text{fe}} \equiv \arg\min_{b\in\mathbb{R}^d}\frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\left(\dot{Y}_{it} - \dot{X}_{it}'b\right)^2.$$

Thus this becomes the usual OLS problem.

FE-2

(i) $\sum_{t=1}^{T}\mathbb{E}\left[\dot{X}_{it}\dot{X}_{it}'\right]$ is full rank;
(ii) $\mathbb{E}\left[(\sum_{t=1}^{T}\dot{X}_{it}\dot{V}_{it})(\sum_{t=1}^{T}\dot{X}_{it}\dot{V}_{it})'\right] < \infty$

Asympt. norm.: Given FE-1 and FE-2, then:

$$\sqrt{n}\left(\hat{\beta}_n^{\text{fe}} - \beta_0\right) \xrightarrow{d}$$
$$\mathcal{N}\left(0, B\mathbb{E}\left[(\sum_{t=1}^{T}\dot{X}_{it}\dot{V}_{it})(\sum_{t=1}^{T}\dot{X}_{it}\dot{V}_{it})'\right]B\right)$$

where $B \equiv (\sum_{t=1}^{T} \mathbb{E}\left[\dot{X}_{it}\dot{X}'_{it}\right])^{-1}$. We can estimate this variance with sample analogues.

Hypothesis testing

Basic concepts: A **test** φ is a procedure to choose between two hypotheses H_0 and H_1. $\{H_0, H_1\}$ is a partition.

$$H_0 : \theta \in \Theta_0 \subset \Theta \text{ v.s. } H_1 : \theta \in \Theta_1 \subset \Theta.$$

$$\text{or,} \quad H_0 : \mathbb{P} \in \mathbf{P}_0 \text{ v.s. } H_1 : \mathbb{P} \in \mathbf{P}_1$$

The **critical region** C_φ characterizes φ. Given data X, φ rejects H_0 if $X \in C_\varphi$.

Type-1 error is when H_0 is rejected but true. The prob. of type-1 error is

$$P_\theta(\text{test } \varphi \text{ rejects } H_0) = P_\theta(C_\varphi),\ \theta \in \Theta_0$$

Type-2 error is when H_0 is accepted and false.

$$P_\theta(\text{test } \varphi \text{ rejects } H_1) = P_\theta(C'_\varphi),\ \theta \in \Theta_1.$$

Size of a test is the highest prob. of type-1 error over all $\theta \in \Theta_0$. That is,

$$\text{Size of test } \varphi = \sup_{\theta \in \Theta_0} P_\theta(C_\varphi) = \sup_{\mathbf{P}_0} \mathbb{E}_{\mathbb{P}_0}[\phi_n]$$

Test φ has sign. lvl α if the size of φ is $\le \alpha$. A test has several sign. lvl., i.e., $\alpha \in$ [size of $\varphi, 1$].

The power of a test is the highest prob. of φ rejecting H_0 when H_1 is true.

$$\text{Power of } \varphi = P_\theta(C_\varphi),\ \text{for any } \theta \in \Theta_1.$$

A p-value is "the probability under the null hypothesis of observing a more extreme outcome than the data X".

$$\text{p-value} \equiv \inf\{\alpha \in [0,1] : X \in C_\varphi^\alpha\}.$$

UMP Test and N-P Lemma

The power function of φ is

$$K_\varphi(\theta) \equiv P_\theta(\varphi \text{ rejects } H_0).$$

Uniformly Most Powerful: A test φ is UMP with sign. lvl α if

$$\sup_{\theta \in \Theta_0} K_\varphi(\theta) \le \alpha, \text{ and}$$

$$K_\varphi(\theta) \ge K_{\varphi_*}(\theta) \text{ for any } \theta \in \Theta_1,$$

where $\varphi_* \neq \varphi$ has sign. lvl α.

A UMP test exists when the null and alternative hypothesis are simple.

N-P Lemma: Let $X = (X_1, \dots, X_n)$ have pdf $f(x;\theta)$ and define the (simple) hypotheses

$$H_0 : \theta = \theta_0 \text{ v.s. } H_1 : \theta = \theta_1.$$

Consider a test φ with critical region

$$C_\varphi \equiv \{x \in \Omega_{X,n} : \frac{f(x;\theta_1)}{f(x;\theta_0)} > k_\alpha\},$$

where k_α is chosen so that the size of φ is α. Then:

- φ is a UMP test with sign. lvl α;
- any UMP test with sign. lvl α must be a size-α test;
- if $f(x;\theta_1) \neq k_\alpha f(x;\theta_0)$ a.s., then all level-α UMP tests are identical a.s.

Different tests

Likelihood ratio test: A LRT has critical region

$$C_\varphi = \{x = X : \lambda(x) \equiv \frac{\max_{\theta \in \Theta} f(x;\theta)}{\max_{\theta \in \Theta_0} f(x;\theta)} > k_\alpha\},$$

where the significance level of the test is

$$\alpha = \max_{\theta \in \Theta_0} P_\theta(\lambda(x) > k_\alpha).$$

A thm: If certain ML regularity conditions hold, then, under Θ_0,

$$LRT_n = 2\ln\left(\frac{\max_{\theta \in \Theta} f(x;\theta)}{\max_{\theta \in \Theta_0} f(x;\theta)}\right) \xrightarrow{d}$$

$$\xrightarrow{d} \chi^2(r),\ \text{as } n \to \infty,$$

where $\Theta_0 = \{\theta \in \Theta : h(\theta) = 0_r\}$, and r is # restr. imposed on the parameters by Θ_0.

Wald test: Take a function $h : \mathbb{R}^k \to \mathbb{R}^r$, where k is # parameters and r # restr. Then,

$$H_0 : h(\theta) = 0_r,\ \text{v.s. } H_1 : h(\theta) \neq 0_r.$$

Suppose $\hat{\theta}_n$ satisfies

$$\sqrt{n}(\hat{\theta}_n - \theta) \xrightarrow{d} \mathcal{N}(0, V(\theta)) \text{ as } n \to \infty.$$

Also, suppose there is a consistent estimator $\hat{V}(\hat{\theta}_n) \xrightarrow{p} V(\theta)$. Then,

$$W_n \equiv \sqrt{n}(h(\hat{\theta}_n))' \left[H(\hat{\theta}_n)\hat{V}_n(\hat{\theta}_n)H'(\hat{\theta}_n)\right]^{-1} \sqrt{n}h(\hat{\theta}_n) \xrightarrow{d} \chi^2(r).$$

H denotes Jacobian. The Wald test has

$$C_\varphi = \{X_n \in \Omega_{X,n} : W_n(X_n) > k_\alpha\},$$

where k_α is the α quantile of $\chi^2(r)$.

Time series

Strict stationarity: X_t is strict. stat. if

$$(X_{t_1+h}, \dots, X_{t_n+h}) \stackrel{d}{=} (X_{t_1}, \dots, X_{t_n})$$

for any h, t_1, and n.

Autocov. func.: If X_t has $\mathbb{E}[X_t^2] < \infty$, $\forall t, s$, then

$$K_X(t,s) \equiv \mathbb{E}\Big[(X_t - \mathbb{E}[X_t])(X_s - \mathbb{E}[X_s])\Big].$$

Also,

$$\Gamma_X(h) \equiv K_X(h, 0),\ \forall h \in \mathbb{Z}.$$

Also, $\Gamma_X(-h) = \Gamma_X(h)$.

Weakly stationarity: A process X_t is weak. stat. if

(a) $\mathbb{E}[X_t^2] < \infty$ for any t;
(b) $\mathbb{E}[X_t] = c \in \mathbb{R}$ for any t;
(c) $K_X(t,s) = K_X(t+h, s+h) = \Gamma_X(t-s)$ for any $t, s, h \in \mathbb{Z}$.

A strict. stat. X_t with $\mathbb{V}\text{ar}[X_t] < \infty$ is also weakly stationary.

A weak. stat. Gaussian process is also strict. stat.

Auto-corr. func: $\rho_X(h)$ of X_t is

$$\rho_X(h) \equiv \frac{\Gamma_X(h)}{\Gamma_X(0)},\quad \forall h \in \mathbb{Z}.$$

IID Noise: X_t is IID noise if obs. are i.i.d., $\mathbb{E}[X_t] = 0$, $\mathbb{E}[X_t^2] = \sigma^2 < \infty$, and if

$$K_X(t,s) = \begin{cases} \sigma^2, & \text{if } s = t, \\ 0, & \text{if } s \neq t. \end{cases}$$

IID noise is stationary.

White noise: A seq. X_t is WN if the autocorr. is zero, $\mathbb{E}[X_t] = 0$, and $\mathbb{E}[X_t^2] = \sigma^2 < \infty$.

IID noise is white noise.

A LLN: If $\{X_t\}_{t \in \mathbb{Z}}$ is weak. stat., and $\Gamma_X(h) \to 0$ as $h \to 0$, then, as $n \to \infty$,

$$\mathbb{V}\text{ar}[\bar{X}_n] = \mathbb{E}\left[(\bar{X}_n - \mathbb{E}[X_t])^2\right] \to 0.$$

Also, if $\sum_{h=-\infty}^{\infty} |\Gamma_X(h)| < \infty$, then

$$n\mathbb{V}\text{ar}[\bar{X}_n] \to \sum_{h=-\infty}^{\infty} \Gamma_X(h) \text{ as } n \to \infty.$$

ARMA processes

Definition: A rand. seq. $\{X_t\}_{t \in \mathbb{Z}}$ is an ARMA(p,q) if it is stationary and

$$X_t + \sum_{k=1}^{p} \phi_k X_{t-k} = u_t + \sum_{i=1}^{q} \theta_i u_{t-i},$$

where $u_j \sim WN(0, \sigma^2)$, and $\phi_k, \theta_i \in \mathbb{R}$.

An alternative representation is

$$\Phi(L)X_t = \theta(L)u_t.$$

Causal repr. of ARMA is an abs. sum. seq. $\{\varphi_k\}_{k=0}^{\infty}$ s.t.

$$X_t = \sum_{k=0}^{\infty} \varphi_k u_{t-k} = \varphi(L)u_t,\ \forall t.$$

Theorem 8: An ARMA process is causal iff the AR part $\phi(L)$ has no roots $|x| < 1$. And $\theta(L)$ has no common roots with $\phi(L)$.

Find causal repr. of an ARMA(p,q): For $\phi(L)X_t = \theta(L)\epsilon_t$:

1. Find roots λ_i of the characteristic polynomial, i.e., $\phi(z) = 0$.
2. Then define caus. repr. $\psi(L)$ by
$$X_t = \psi(L)\epsilon_t \ \Rightarrow$$
$$\Rightarrow\ \theta(L) = \phi(L)\psi(L)$$
3. By matching of coefficients, identify ψ_i from the coefficient of L^i.
4. Use that $\psi_i = \sum_{j=1}^{p} c_j \lambda_j^{-i}$.

Auto-cov. function of ARMA:

Y-W formulas: An ARMA(p,q) $X_t = \varphi(L)u_t$ has $\Gamma_X(h)$ given by:

$$\text{If } h \le q: \quad \Gamma_X(h) + \sum_{k=1}^{p} \phi_k \Gamma_X(h-k) = \sigma^2 \sum_{k=h}^{q} \theta_k \varphi_{k-h}.$$

$$\text{If } h > q: \Gamma_X(h) + \sum_{k=1}^{p} \phi_k \Gamma_X(h-k) = 0.$$

When $u_t \sim WN$:

$$\Gamma_X(h) = \Gamma_u(0) \sum_{k=-\infty}^{\infty} \varphi_k \varphi_{k+h},\ \forall h.$$

Theorem 6: If $\{u_t\}_{t\in\mathbb{Z}}$ is weak. stat., $\mathbb{E}[u_t] = \mu_u$, $\{\varphi_k\}_{k=0}^{\infty}$ is abs. sum., then X_t, defined by

$$X_t \equiv \sum_{k=-\infty}^{\infty} \varphi_k u_{t-k} = \varphi(L) u_t$$

is stat. with mean $\mu_u \sum_{k=-\infty}^{\infty} \varphi_k$ and

$$\Gamma_X(h) = \sum_{j=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} \varphi_j \varphi_k \Gamma_u(h+k-j).$$

If seq. of Γ_u is abs. sum., then so is Γ_X.

Specific processes

MA(2) $X_t = \epsilon_t + \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2}$**:**

$$\Gamma_X(0) = (1+\theta_1^2+\theta_2^2)\sigma^2,$$
$$\Gamma_X(1) = (\theta_1 + \theta_1\theta_2)\sigma^2,$$
$$\Gamma_X(2) = \theta_2\sigma^2, \quad \Gamma_X(h) = 0,\ |h| > 2.$$

AR(1) $X_t = \phi X_{t-1} + \epsilon_t$**:**

Causal repr.:

$$X_t = \sum_{j=0}^{\infty} \phi^j \epsilon_{t-j}.$$

$$\Gamma_X(h) = \frac{\sigma^2 \phi^{|h|}}{1-\phi^2},\ \forall h.$$

AR(2) $X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \epsilon_t$**:**

$$\Gamma_X(0) = \frac{1-\phi_2}{1+\phi_2} \frac{\sigma^2}{(1-\phi_2)^2 - \phi_1^2}.$$

$$\Gamma_X(1) = \frac{\phi_1}{1-\phi_2} \Gamma_X(0)$$

$$\Gamma_X(h) = \phi_1 \Gamma_X(h-1) + \phi_2 \Gamma_X(h-2).$$

ARMA(1,1) $(1+\phi L) X_t = (1+\theta L) u_t$**:**

Causal repr.: $\varphi_0 = 1$ and $\varphi_j = \phi^j - \theta\phi^{j-1}$

$$\Gamma_X(0) = \frac{\theta^2 - 2\phi\theta + 1}{1-\phi^2}\sigma^2,$$

$$\Gamma_X(1) = \frac{(1-\phi\theta)(\theta-\phi)}{1-\phi^2}\sigma^2,$$

$$\Gamma_X(h) = (-\phi)^{h-1}\Gamma_X(1),\ \forall |h| > 1.$$

Martingale Limit Theory

Martingale def.: $\{X_t\}_{t\in\mathbb{Z}}$ is m-g $\Leftrightarrow$

$$\mathbb{E}[X_t|\mathcal{F}_s] = X_s, \quad \forall t > s.$$

M.d.s. def.: $\{u_t\}_{t\in\mathbb{Z}}$ is m.d.s. $\Leftrightarrow$

$$\mathbb{E}[u_t|\mathcal{F}_s] = 0, \quad \forall t > s.$$

By def., $\mathbb{E}[u_t u_s] = 0$. An m.d.s. with finite, constant variance is WN.

M.d.a. def.: $\{u_{t,n}\}_{t=1}^{n}$ is m.d.a. $\Leftrightarrow$

$$\mathbb{E}[u_{t,n}|u_{t-1,n}, u_{t-2,n}, \ldots] = 0, \quad \forall t, n.$$

Marting. converg. thm: Let X_t be a martingale with $\mathbb{V}\text{ar}[X_t] < \Delta$. Then, as $t \to \infty$, $X_t \overset{a.s.}{\to} X_\infty$, where X_∞ is a r.v. with $\mathbb{V}\text{ar}[X_\infty] < \Delta'$.

Marting. LLN: If u_t is m.d.s. with $\mathbb{E}[u_t^2] = \sigma_t^2$ and $\sup_t(\sigma_t^2) = C < \infty$, then

$$\frac{1}{N}\sum_{t=1}^{n} u_t \overset{a.s.}{\to} 0.$$

Marting. CLT: If $X_{t,n}$ is a m.d.a. with bounded $\mathbb{E}[|X_{t,n}|^{2+\delta_0}]$, and $\exists \bar{\sigma}_n^2 > \delta_0 > 0$ s.t. $n^{-1}\sum_{t=1}^{n} X_{t,n}^2 - \bar{\sigma}_n^2 \overset{\mathbb{P}}{\to} 0$, then

$$\frac{n^{-1/2}\sum_{t=1}^{n} X_{t,n}}{\sqrt{\bar{\sigma}_n^2}} \overset{d}{\to} \mathcal{N}(0,1).$$

Asympt. Propert. of LP

Lemma on B-N Decompostion: Let $\varphi(x) = \sum_{k=0}^{\infty}\varphi_k x^k$, where the seq. $\{\varphi_k\}$ is abs. sum. Then

$$\varphi(x) = \varphi(1) - (1-x)\tilde{\varphi}(x), \quad \forall |x| < 1$$

where $\tilde{\varphi}(x) = \sum_{k=0}^{\infty}\tilde{\varphi}_k x^k$ and $\tilde{\varphi}_k = \sum_{s=k+1}^{\infty}\varphi_s$.

If a rand. seq. u_t has $\sup_t \mathbb{E}[|u_t|] < \infty$, then

$$X_t \equiv \sum_{k=0}^{\infty}\varphi_k u_{t-k} = \varphi(L)(u_t)$$

for some seq. φ_k that is abs. sum. If also $\sum_{k=0}^{\infty} k|\varphi_k| < \infty$, then the B-N decomp. of X_t is the RHS of

$$X_t = u_t \sum_{k=0}^{\infty}\varphi_k + \tilde{u}_{t-1} - \tilde{u}_t.$$

Note that $\sum_{k=0}^{\infty}\varphi_k = \varphi(1)$.

A LLN: If u_t has $\sup_t \mathbb{E}[|u_t|] < \infty$ and

$$\frac{1}{n}\sum_{t=1}^{n}(u_t - \mathbb{E}[u_t]) \overset{\mathbb{P}}{\to} 0 \text{ as } n \to \infty,$$

then,

$$\frac{1}{n}\sum_{t=1}^{n}(X_t - \mathbb{E}[X_t]) \overset{\mathbb{P}}{\to} 0,\ \forall t,\ \text{as } n \to \infty$$

for an $X_t = \sum_{k=0}^{\infty}\varphi_k u_{t-k}$ def. by a seq. $\{\varphi_k\}_{k=0}^{\infty}$ s.t. $\sum_{k=0}^{\infty} k|\varphi_k| < \infty$.

A CLT: Take u_t s.t. $\sup_t \mathbb{E}[|u_t|] < \infty$ and $\{\varphi_t\}_{t=0}^{\infty}$ s.t. $\sum_{t=0}^{\infty} k|\varphi_k| < \infty$. Let $\sigma_{u,n}^2 = \mathbb{V}\text{ar}[\frac{1}{\sqrt{n}}\sum_{t=1}^{n}(u_t - \mathbb{E}[u_t])]$ for any $n \geq 1$. If $\lim_{n\to\infty}\sigma_{u,n}^2 > 0$ and

$$\frac{\sum_{t=1}^{n}(u_t - \mathbb{E}[u_t])}{\sqrt{n\sigma_{u,n}^2}} \overset{d}{\to} \mathcal{N}(0,1),$$

then

$$\frac{\sum_{t=1}^{n}(X_t - \mathbb{E}[X_t])}{\sqrt{n\sigma_{u,n}^2}} \overset{d}{\to} \mathcal{N}(0,(\varphi(1))^2),$$

where $X_t = \sum_{k=0}^{\infty}\varphi_k u_{t-k}$ for any t.

LLN of sample auto co-var.: If $u_t \sim iid(0,\sigma^2)$, and $\sum_{k=0}^{\infty} k|\varphi_k| < \infty$, then for $X_t = \sum_{k=0}^{\infty}\varphi_k u_{t-k}$,

$$n^{-1}\sum_{t=1}^{n} X_t X_{t-h} \overset{\mathbb{P}}{\to} \Gamma_X(h).$$

Algebraic tricks

$$\sum_{k=1}^{n} k = \frac{n(n+1)}{2}, \quad \sum_{k=1}^{n} k^2 = \frac{n(n+1)(2n+1)}{6},$$

$$\sum_{k=1}^{n} k^3 = \frac{n^2(n+1)^2}{4}, \quad \sum_{k=0}^{n-1} ax^k = a \cdot \frac{1-x^n}{1-x}$$

$$\sum_{k=0}^{n-1} ka^k = \frac{(n-1)a^{n+1} - na^k + a}{(a-1)^2}$$

$$\int \cos^2(x)\ \mathrm{d}x = \frac{1}{2}[x + \sin(x)\cos(x)]$$

$$\int \sin^2(x)\ \mathrm{d}x = \frac{1}{2}[x - \sin(x)\cos(x)]$$

$$\lim_{n\to\infty}(1+x/n)^n = e^x = \sum_{n=0}^{\infty} x^n/(n!)$$

Examples

A simple test from Santos

\# obs. $n = 1$ and $W \sim \mathcal{N}(\mu, 1)$; μ is unknown.

$$H_0: \mu \leq 0 \text{ v.s. } H_1: \mu > 0.$$

Note that $W \overset{d}{=} \mu + Z$, where $Z \sim \mathcal{N}(0,1)$.
E.g., use $\phi(W) = \mathbb{1}(W > c)$. Then,

$$\text{size of } \phi = \sup_{\mathbb{P}\in\mathbf{P}_0} \mathbb{E}_{\mathbb{P}}[\phi(W)] =$$

$$= \sup_{\mu\leq 0} \mathbb{P}(Z > c - \mu) = \mathbb{P}(W > c) \leq \alpha, \ \Leftrightarrow$$

$$\mathbb{P}(z > c_{1-\alpha}) = \alpha \Leftrightarrow c_{1-\alpha} = \Phi^{-1}(1-\alpha).$$

Another test from Santos

$\{W_i\}_{i=1}^{n}$ is i.i.d. with variance 1. $\mathbf{P} = \{\mathbb{P} : \mathbb{E}_{\mathbb{P}}[(W - \mathbb{E}_{\mathbb{P}}[W])^2] = 1\}$. We want to test

$$\mathbf{P}_0 : \{\mathbb{E}_{\mathbb{P}}[W] \leq 0\} \text{ v.s. } \mathbf{P}_1 : \{\mathbb{E}_{\mathbb{P}}[W] > 0\}.$$

For any $\mathbb{P}$, $\sqrt{n}(\bar{W}_n - \mathbb{E}_{\mathbb{P}}[W]) \overset{d}{\to} \mathcal{N}(0,1)$.
Then use test $\phi_n = \mathbb{1}(\sqrt{n}\bar{W}_n > c_{1-\alpha})$.

$$\sup_{\mathbb{P}\in\mathbf{P}_0} \lim_{n\to n} \mathbb{P}(\sqrt{n}\bar{W}_n > c_{1-\alpha}) \leq$$

$$\leq \sup_{\mathbb{P}\in\mathbf{P}_0} \lim_{n\to n} \mathbb{P}(\sqrt{n}(\bar{W}_n - \mathbb{E}_{\mathbb{P}}[W]) > c_{1-\alpha}) =$$

$$\mathbb{P}(Z > c_{1-\alpha}) = \alpha.$$

Linear trend regr.

$X_t = \mu t + u_t$, $u_t \sim iid(0,\sigma_u^2)$, $X_0 = 0$, and $Y_t = X_t\beta + \epsilon_t$, where $\epsilon \sim iid(0,\sigma_\epsilon^2)$, ϵ_t, and u_t are indep.

Derive asympt. distr. of OLS estimator.

$$\hat{\beta}_n = \frac{\sum_{t=1}^{n} Y_t X_t}{\sum_{t=1}^{n} X_t^2} \quad \Leftrightarrow$$

$$\Leftrightarrow \quad \hat{\beta}_n - \beta = \frac{\sum_{t=1}^{n} X_t\epsilon_t}{\sum_{t=1}^{n} X_t^2}.$$

Then

$$\frac{1}{n^3}\sum X_t^2 = \overbrace{\frac{\mu^2}{n^3}\sum t^2}^{\to\mu^2/3} + \frac{2\mu}{n^3}\sum t u_t +$$

$$+ \frac{1}{n^3}\sum u_t^2 =$$

$$\mathbb{V}\text{ar}[\frac{2\mu}{n^3}\sum t u_t] = \frac{4\mu^2}{n^6}\sigma_u^2 \sum t^2 \to 0$$

$$\Rightarrow \frac{2\mu}{n^3}\sum t u_t = o_p(1), \text{ by Markov ineq.}$$

By LLN, $\frac{1}{n^3}\sum u_t^2 = o_p(1)$. So $\frac{1}{n^3}\sum X_t^2 \to \frac{\mu^2}{3}$. Then consider the numerator:

$$\frac{1}{n^{3/2}}\sum X_t\epsilon_t = \frac{\mu}{n^{3/2}}\sum t\epsilon_t + \frac{1}{n^{3/2}}\sum u_t\epsilon_t =$$

$$= \mathcal{O}_p(1) + \mathcal{O}_p(n^{-1}), \text{ by Markov ineq.}$$

since

$$\mathbb{V}\text{ar}[\frac{\mu}{n^{3/2}}\sum t\epsilon_t] = \frac{\mu^2\sigma_\epsilon^2}{n^3}\sum t^2 \to \frac{\mu^2\sigma_\epsilon^2}{3},$$

and

$$\mathbb{V}\mathrm{ar}[\frac{1}{n^{3/2}}\sum u_t\epsilon_t] = \frac{\sigma_u^2\sigma_\epsilon^2}{n^2}.$$

Then, by m-g CLT

$$\mathbb{V}\mathrm{ar}[\frac{\mu}{n^{3/2}}\sum t\epsilon_t] \xrightarrow{d} \mathcal{N}(0, \frac{\mu^2\sigma_\epsilon^2}{3}).$$

By Slutsky,

$$n^{3/2}(\hat\beta_n - \beta) \xrightarrow{d} \mathcal{N}(0, \frac{3\sigma_\epsilon^2}{\mu^2}).$$

What if ϵ_t and u_s are correlated for $t = s$ (and not $t \neq s$)? Only difference is that $\mathbb{V}\mathrm{ar}[\frac{1}{n^{3/2}}\sum u_t\epsilon_t] \neq \sigma_u^2\sigma_\epsilon^2$. But this doesn't affect the asympt. distr. of $\hat\beta_n$.

Construct consistent estimates of μ and σ_ϵ^2.

$$\hat\mu_n = \frac{\sum tX_t}{\sum t^2} = \mu + \frac{\sum tu_t}{\sum t^2}$$

Denominator $\to \frac{1}{3}$ and numerator:

$$\frac{1}{n^3}\sum tu_t = o_p(1).$$

So $\hat\mu_n \xrightarrow{\mathbb{P}} \mu$. For $\hat\sigma_\epsilon^2$:

$$\hat\sigma_\epsilon^2 = \frac{1}{n}\sum(Y_t - X_t\hat\beta_n)^2 =$$

$$\frac{1}{n}\sum\epsilon_t^2 - \frac{2}{n}(\hat\beta_n - \beta)\sum X_t\epsilon_t +$$

$$+\frac{1}{n}(\hat\beta_n - \beta)^2\sum X_t^2.$$

$\frac{1}{n}\sum\epsilon_t^2 \xrightarrow{d} \sigma_\epsilon^2 + \mathcal{O}_p(n^{-1/2})$ by iid CLT. From before, $(\hat\beta_n - \beta) = \mathcal{O}_p(n^{-3/2})$ so $(\hat\beta_n - \beta)^2 = \mathcal{O}_p(n^{-3})$ and $\frac{1}{n^3}\sum X_t^2 = \mathcal{O}_p(1)$ implies $\frac{1}{n}\sum X_t^2 = \mathcal{O}_p(n^2)$. Thus, $\frac{1}{n}(\hat\beta_n - \beta)^2\sum X_t^2 = \mathcal{O}_p(n^{-1})$. Also, $\frac{1}{n^{3/2}}\sum X_t\epsilon_t = \mathcal{O}_p(1)$ implies $\frac{1}{n}\sum X_t\epsilon_t = \mathcal{O}_p(n^{1/2})$. Therefore, $\frac{2}{n}(\hat\beta_n - \beta)\sum X_t\epsilon_t = \mathcal{O}_p(n^{-1})$. In summary, $\hat{\sigma_\epsilon}^2) \xrightarrow{\mathbb{P}} \sigma_\epsilon^2$.

Two-sided hypoth. w/ unknown σ^2

Let $X_1, \ldots, X_n$ be iid $\mathcal{N}(\mu, \sigma^2)$ where σ^2 is unknown. Consider $H_0 : \mu = \mu_0$ v.s. $H_1 : \mu \neq \mu_0$.

Define LR test for these hypotheses.

$$LR = \frac{\max\ f(\bar X, \mu, \sigma^2)}{\max\ f(\bar X, \mu_0, \sigma^2)} =$$

$$= \frac{f(\bar X, \mu_{\mathrm{MLE}}, \hat\sigma^2_{\mathrm{MLE}})}{\max_{\sigma^2}\ f(\bar X, \mu_0, \sigma^2)} \quad \Rightarrow$$

$$2\ln(LR) = n\ln\left(\frac{\sum(X_i - \mu_0)^2}{\sum(X_i - \bar X_n)^2}\right) \xrightarrow{d} \chi^2(1).$$

So, the (asymptotic) LR test is to reject H_0 iff

$$2\ln(LR) = n\ln\left(\frac{\sum(X_i - \mu_0)^2}{\sum(X_i - \bar X_n)^2}\right) > \chi^2_\alpha(1),$$

where $\chi^2_\alpha(1)$ is the $1-\alpha$ percentile of the χ^2 distribution with df. 1.

Whats the critical region of the likelihood ratio test with size $\alpha = 5\%$? The critical region is

$$\{\bar X : n\ln\left(\frac{\sum(X_i - \mu_0)^2}{\sum(X_i - \bar X_n)^2}\right) > \chi^2_{.05}(1)\}.$$

Estimate ARMA(1,1)

$X_t = \theta X_{t-1} + \epsilon_t + \epsilon_{t-1}$ where $\epsilon_t \sim iid(0, \sigma_\epsilon^2)$ and $|\theta| < 1$.

Derive the distribution of the estimator

$$\hat\theta_n = \frac{\sum X_tX_{t-2}}{\sum X_{t-1}X_{t-2}}.$$

Estimator is equivalent to

$$\sqrt{n}(\hat\theta_n - \theta) = \frac{\frac{1}{\sqrt{n}}\sum(\epsilon_t + \epsilon_{t-1})X_{t-2}}{\frac{1}{n}\sum X_{t-1}X_{t-2}}.$$

Denominator:

Use LLN for sample auto-covar. $\Rightarrow \Gamma_X(1)$.

Numerator:

$$\sum(\epsilon_t + \epsilon_{t-1})X_{t-2} =$$

$$= \frac{1}{\sqrt{n}}\sum\epsilon_t(X_{t-2} + X_{t-1}) + o_p(1) + o_p(1).$$

The remaining term is a m.d.s. Take

$$\bar\sigma_n = \mathbb{E}[\epsilon_t^2(X_{t-2} + X_{t-1})^2] =$$

$$= 2(\Gamma_X(0) + \Gamma_X(1))\sigma_\epsilon^2.$$

Then use m-g CLT and Slutsky.

Provide a test with size α for $\theta = 0$. Use Wald test. Under $H_0 : \theta = 0$, $\sqrt{n}\hat\theta_n \xrightarrow{d} \mathcal{N}(0, K)$. So, critical region is $C = \{\hat\theta_n > Z_{1-\frac{\alpha}{2}}\frac{K}{n}\} \cup \{\hat\theta_n < -Z_{1-\frac{\alpha}{2}}\frac{K}{n}\}$.

AR(1) error term

The model is $Y_t = \theta + u_t$ with $u_t = \rho u_{t-1} + \epsilon_t$ where $|\rho| < 1$, $\epsilon_t \sim iid\ \mathcal{N}(0, \sigma^2)$. All coefficients θ, ρ, and σ^2 are unknown.

$$\hat\theta_n = \frac{1}{T}\sum_{t=1}^T Y_t$$

$$\hat u_t \equiv Y_t - \hat\theta.$$

Show that

$$\hat\rho_T = \frac{\sum_{t=2}^T \hat u_t\hat u_{t-1}}{\sum_{t=2}^T \hat u_t^2}$$

is a consistent estimator of ρ and that $\hat\rho_T - \rho = \mathcal{O}_p(T^{-1/2})$.

Observe that $\hat u_t = u_t - (\hat\theta_T - \theta)$.

Denominator:

$$\frac{1}{T}\sum_{t=2}^T\hat u_t^2 =$$

$$= \frac{1}{T}\sum u_t^2 + (\hat\theta_T - \theta)^2 - \frac{2(\hat\theta_T - \theta)}{T}\sum u_t =$$

$$= T^{-1}\sum u_t^2 + \mathcal{O}_p(T) + \mathcal{O}_p(T^{-1}) =$$

$$= \mathbb{E}[u_t^2] + \mathcal{O}_p(T^{-1/2}) \xrightarrow{\mathbb{P}} \frac{\sigma^2}{1-\rho^2}.$$

Numerator:

$$\frac{1}{T}\sum_{t=2}^T\hat u_t\hat u_{t-1} =$$

$$= \frac{1}{T}\sum[u_t - (\hat\theta_T - \theta)][u_{t-1} - (\hat\theta_T - \theta)] =$$

$$= \frac{1}{T}\sum u_tu_{t-1} - \frac{\hat\theta_T - \theta}{T}\sum_{t=2}^T u_{t-1} -$$

$$-\frac{\hat\theta_T - \theta}{T}\sum_{t=2}^T u_t + \frac{T-1}{T}(\hat\theta_T - \theta)^2 =$$

$$= \frac{1}{T}\sum_{t=2}^T u_tu_{t-1} + \mathcal{O}_p(T^{-1}).$$

By def. of u_t

$$= \frac{\rho}{T}\sum u_{t-1}^2 + \frac{1}{T}\sum\epsilon_tu_{t-1} =$$

$$= \rho\mathbb{E}[u_t^2] + \mathcal{O}_p(T^{-1/2}).$$

Thus,

$$\hat\rho_T - \rho = \frac{T^{-1}\sum\hat u_t\hat u_{t-1}}{T^{-1}\sum\hat u_t^2} - \rho =$$

$$= \frac{\mathcal{O}_p(T^{-1/2})}{\mathbb{E}[u_t^2] + \mathcal{O}_p(T^{-1/2})} = \mathcal{O}_p(T^{-1/2}).$$

Which is what we sought.

Derive the prob. limit of

$$\hat\sigma^2_{u,T} = \frac{1}{T}\sum\hat u_t^2.$$

and construct a root-T consistent estimator of σ^2:

From before, $\hat\sigma^2_{u,T} \xrightarrow{\mathbb{P}} \frac{\sigma^2}{1-\rho^2}$.

Given that $\hat\rho_T$ is consistent implies for the following estimator

$$\hat\sigma_T^2 = (1 - \hat\rho_T^2)\hat\sigma^2_{u,T} =$$

$$= (1 - \rho^2 + \mathcal{O}_p(T^{-1/2}))\cdot$$

$$\left(\frac{\sigma^2}{1-\rho^2} + \mathcal{O}_p(T^{-1/2})\right) =$$

$$\sigma^2 + \mathcal{O}_p(T^{-1/2}).$$

Sample size AR(1)

For a model $Y_t = 4\theta_0^2Y_{t-1} + \epsilon_t$, by LLN for mixed processes,

$$\lim_{n\to\infty} n\mathbb{V}\mathrm{ar}[\bar Y_n] = \frac{\sigma_\epsilon^2}{(1 - 4\theta_0^2)^2}.$$

How large a sample would we need in order to have 95% CI s.t. $\bar Y_n$ differed from the true value zero by no more than 0.1?

The 95% CI for the true value is

$$\{\bar Y_n \pm Z_{1-\frac{\alpha}{2}}\sqrt{\mathbb{V}\mathrm{ar}[\bar Y_n]}\} =$$

$$= \{\bar Y_n \pm 1.96\cdot\sqrt{\frac{\sigma_\epsilon^2}{n(1 - 4\theta_0^2)^2}}\}.$$

So we need $1.96\cdot\sqrt{\frac{\sigma_\epsilon^2}{n(1-4\theta_0^2)^2}} \leq 0.1 \Leftrightarrow n \geq \frac{19.6^2\sigma_\epsilon^2}{(1-4\theta_0^2)^2}$.

Trend regr. #2 (fall comp 2018)

$$Y_t = \theta_0\rho^t + \epsilon_t,$$

$$\rho^n(\hat\theta_n - \theta_0) \xrightarrow{d} \mathcal{N}(0, \frac{\sigma^2(\rho^2 - 1))}{\rho^2})$$

$$\rho^{-2n}\sum_{t=1}^n\rho^{2t} \to \frac{\rho^2}{\rho^2 - 1}$$

Exp. trend regr. (HW4)

Show consistency and limit. distr. for below estimator of following model.

$$Y_t = \rho^t\theta_o + u_t,\ u_t \sim \mathcal{N}(0, \sigma^2).$$

$$\hat\theta_n = \left(\sum_{t=1}^n\rho^{2t}\right)^{-1}\left(\sum_{t=1}^n\rho^tY_t\right)$$

Use that $\xrightarrow{L^2}$ implies $\xrightarrow{\mathbb{P}}$.

$$\mathbb{E}[(\hat\theta_n - \theta_o)^2] = \mathbb{E}[(\frac{\sum_{t=1}^n\rho^tY_t}{\sum_{t=1}^n\rho^{2t}} - \theta_o)^2] =$$

$$= \frac{\sigma^2}{\rho^2}\frac{\rho - 1}{\rho^{2n} - 1} \to 0 \text{ as } n \to \infty.$$

So, $\hat{\theta}_n \xrightarrow{L^2} \theta_o$, which implies $\hat{\theta} \xrightarrow{\mathbb{P}} \theta_o$, i.e., that the estimator is consistent.

Deriving limit. distr.: Rewriting as

$$\rho^n(\hat{\theta}_n - \theta_o) = \frac{\rho^{-n}\sum_{t=1}^n \rho^t u_t}{\frac{1}{\rho^{2n}}\sum_{t=1}^n \rho^{2t}}.$$

The denominator:

$$\rho^{-2n}\sum_{t=1}^n \rho^{2t} = \rho^{-2n}\frac{\rho^{2n+2}-1}{\rho^2-1} =$$

$$= \frac{\rho^2 - \rho^{-2n}}{\rho^2-1} \to \frac{\rho^2}{\rho^2-1}.$$

Numerator: Define $Z_n \equiv \sum_{t=1}^n \rho^{t-n}u_t = \sum_{t'=0}^{n-1}\rho^{-t'}u_{n-t'}$. Then, by i.i.d. $u_t \sim \mathcal{N}(0,\sigma^2)$,

$$Z_n \sim \mathcal{N}(0, \sum_{t'=0}^{n-1}(\rho^{-2})^{t'}\sigma^2) =$$

$$= \mathcal{N}(0, \sigma^2\frac{1-\rho^{-2n}}{1-\rho^{-2}}).$$

Then the MGF $M_n(t)$ of Z_n is

$$M_n(t) = \exp(\sigma^2\frac{1-\rho^{-2n}}{1-\rho^{-2}}t^2).$$

By Dominated convergence theorem, then

$$\lim_{n\to\infty} M_n(t) = \exp(\frac{\sigma^2}{1-\rho^{-2}}t^2).$$

Conv. in MGF $\Leftrightarrow$ conv. in distribution.

$$Z_\infty \sim \mathcal{N}(0, \frac{\rho^2\sigma^2}{\rho^2-1}).$$

Then, by Slutsky,

$$\rho^n(\hat{\theta}_n - \theta_o) \xrightarrow{d} \mathcal{N}(0, \sigma^2\frac{\rho^2-1}{\rho^2}).$$

Prob. 2 on HW4

Part (a): $\exists\mathbb{E}[X_t^4] \Rightarrow \exists\mathbb{E}[X_t^2]$ by Jensen. $u_t \equiv X_t\epsilon_t$. This is a m.d.s.

$$\frac{1}{n}\sum_{t=1}^n u_{t,n}^2 - \bar{\sigma}_n^2 = \frac{1}{n}\sum_{t=1}^n (X_t\epsilon_t)^2 - \bar{\sigma}_n^2.$$

By LLN and Slutsky, $\frac{1}{n}\sum_{t=1}^n X_t^2\epsilon_t^2 \xrightarrow{\mathbb{P}} \mathbb{E}[X_t^2\epsilon_t^2] = \mathbb{E}[X_t^2]\Delta^2$. The last step follows from independence. Choose the sequence $\bar{\sigma}_n^2 = \mathbb{E}[X_t^2]\Delta^2$. Then,

$$\frac{1}{n}\sum_{t=1}^n u_{t,n}^2 - \bar{\sigma}_n^2 \xrightarrow{\mathbb{P}} 0.$$

Then we can use the Martingale CLT, i.e.,

$$\frac{\sum_{t=1}^n X_t\epsilon_t}{\sqrt{n}\sqrt{\mathbb{E}[X_t^2]\Delta^2}} \xrightarrow{d} \mathcal{N}(0,1) \qquad \Leftrightarrow$$

$$\frac{\frac{\sum_{t=1}^n X_t\epsilon_t}{\sum_{t=1}^n X_t^2}}{\sqrt{n}\frac{\sqrt{\mathbb{E}[X_t^2]\Delta^2}}{\sum_{t=1}^n X_t^2}} \xrightarrow{\mathbb{P}} \frac{\hat{\beta}_n - \beta_o}{\frac{1}{\sqrt{n\mathbb{E}[X_t^2]}}\Delta} \xrightarrow{d} \mathcal{N}(0,1).$$

Use Slutsky to get the above. This is the sought result,

$$\frac{\sqrt{n\mathbb{E}[X_t^2]}(\hat{\beta}_n - \beta_o)}{\Delta} \xrightarrow{d} \mathcal{N}(0,1).$$

Part (b):

$$\hat{\Delta}_n^2 = \frac{1}{n}\sum_{t=1}^n \left(X_t\beta_o + \epsilon_t - X_t\hat{\beta}_n\right)^2 =$$

$$= \frac{1}{n}\sum_{t=1}^n [\epsilon_t^2 + 2\epsilon X_t(\beta_o - \hat{\beta}_n) +$$

$$+ X_t^2(\beta_o - \hat{\beta}_n)^2].$$

First term $\xrightarrow{\mathbb{P}} \Delta^2$ by LLN; the middle term $o_p(1)$ by LLN; for the last term we use the result in (a) to note that $(\beta_o - \hat{\beta}_n)^2$ is $\mathcal{O}_p(n^{-1})$ while $\frac{1}{n}X_t^2 \xrightarrow{\mathbb{P}} \mathbb{E}[X_t^2]$ by LLN. The product of the two is $o_p(1)$ as $n \to \infty$. Thus,

$$\hat{\Delta}_n^2 \xrightarrow{\mathbb{P}} \Delta^2.$$

Part (c): Note that

$$\frac{1}{n}\sum_{t=1}^n X_t^2 \xrightarrow{\mathbb{P}} \mathbb{E}[X_t^2]$$

by LLN. Combine this with (a) and (b), then by Slutsky,

$$\sqrt{\frac{\sum_{t=1}^n X_t^2}{\hat{\Delta}_n^2}}(\hat{\beta}_n - \beta_o) \xrightarrow{d} \mathcal{N}(0,1).$$

MLE

$\mu_{MLE} = \bar{X}_n$ and $\sigma_{MLE}^2 = \frac{1}{n}\sum_{i=1}^n(x_i - \mu)^2$, which are derived a follows if $X_i, \ldots, X_n$ are i.i.d and $\mathcal{N}(\mu,\sigma^2)$:

$$\mathcal{L}(X_1,\ldots,X_n) =$$

$$= \prod_{i=1}^n \frac{1}{\sqrt{2\pi\sigma^2}}\exp(\frac{-(x_i-\mu)^2}{2\sigma^2}) =$$

$$= (2\pi\sigma^2)^{-n/2}\exp\left(\frac{-\sum_{i=1}^n(x_i-\mu)^2}{2\sigma^2}\right),$$

then take logs, then take FOC w.r.t. μ and σ^2 to obtain the result.