Econometrics 1

Chapter 2: Statistical Uncertainty in Linear Regressions

Xavier D'Haultfœuille and Elia Lapenta

Introduction: Why?

- In Chapter 1 we have described some basic properties of the OLS estimator for n fixed.
- However, we did not consider how to account for uncertainty. Here the uncertainty is due to the fact that n is finite.
- This is key to:
  - understand when the OLS estimator is precise and when it is not;
  - “correctly" select the covariates (to minimize the prediction error, cf. Chapter 3);
  - distinguish those variables that really have an impact on Y from those that do not have an impact
    $\Rightarrow$ *statistical tests;*
  - quantify the uncertainty on the coefficients $\Rightarrow$ *confidence intervals*.

- General philosophy: use the asymptotic results to approximate the estimator's behavior as $n \to \infty$.
- Motivation: obtain results under weak conditions.
- In practice the sample size is often large ($n \geq 1000$), so the asymptotic approximation is often (very) good.
- At the end of this chapter, we will present the results for the exact distribution of the OLS estimator: however, they will be based on strong/restrictive assumptions.

Plan

Limit distribution and accuracy of the OLS estimator

Estimation of the asymptotic variance

Tests and confidence intervals

- Digression on the theory of tests
- Application to linear regressions

Particular cases*

- As well as the sample mean, the OLS estimator is consistent but also asymptotically normal under some moment conditions.
- We consider the following assumptions:

$$(Y_i, X_i')_{i=1,\dots,n} \text{ are i.i.d. with the same law as } (Y, X'). \tag{i.i.d.}$$

$$E(XX') \text{ is invertible }. \tag{Inv th}$$

- Recall that from Chapter 1, $\widehat{\beta} =$ OLS estimator of the regression of Y on $X \in \mathbb{R}^k$, $\beta_0 := E(XX')^{-1}E(XY)$, and $\varepsilon := Y - X^T\beta_0$. We finally assume

$$E(\|X\|^4) < \infty \text{ and } E(\varepsilon^4) < \infty. \tag{Mom}$$

Theorem 1

Under the assumptions (i.i.d.), (Inv th), and (Mom), we have

$$\sqrt{n}\left(\widehat{\beta} - \beta_0\right) \xrightarrow{d} \mathcal{N}\left(0, E(XX')^{-1}E(\varepsilon^2 XX')E(XX')^{-1}\right).$$

Limit distribution of the OLS estimator: Proof

We have $\widehat{\beta} = \left(1/n \sum_{i=1}^{n} X_i X_i'\right)^{-1} \left(1/n \sum_{i=1}^{n} X_i Y_i\right)$. Accordingly,

$$\widehat{\beta} = \beta_0 + \left(\frac{1}{n}\sum_{i=1}^{n} X_i X_i'\right)^{-1} \left(\frac{1}{n}\sum_{i=1}^{n} X_i \varepsilon_i\right). \tag{1}$$

Thus,

$$\sqrt{n}\left(\widehat{\beta} - \beta_0\right) = \left(\frac{1}{n}\sum_{i=1}^{n} X_i X_i'\right)^{-1} \sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n} X_i \varepsilon_i - 0\right).$$

Thanks to the Central Limit Theorem (CLT), since $E(X\varepsilon) = 0$ and $E(\varepsilon^2 ||X||^2) \leq \sqrt{E(\varepsilon^4)\, E(\|X\|^4)} < \infty$,

$$\sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n} X_i \varepsilon_i - 0\right) \xrightarrow{d} \mathcal{N}\left(0, E(\varepsilon^2 XX')\right).$$

By the Law of Large Numbers (LLN),

$$\frac{1}{n}\sum_{i=1}^{n} X_i X_i' \xrightarrow{P} E[XX'].$$

Finally, the result follows from Slutsky's Theorem $\square$

Intuition of the results

- Equation (1) and the LLN imply:

$$\widehat{\beta} \simeq \beta_0 + \frac{1}{n}\sum_{i=1}^{n} E(XX')^{-1} X_i \varepsilon_i.$$

- So, $\widehat{\beta} - \beta_0$ behaves as sample mean of the variable

$$U = E(XX')^{-1} X\varepsilon \ (\in \mathbb{R}^k).$$

$\Rightarrow$ It will be asymptotically normal, as well as a sample mean (and with asymptotic variance $V(U)$).

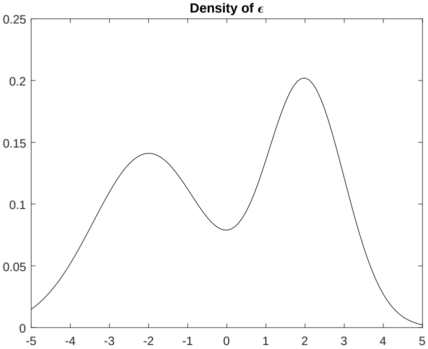
- This is also the case for most non-linear estimators: as $n \to \infty$ they behave as a sample mean.

Example

- Let us assume that $X = (1, D)$ with $D \sim$Poisson(1) and

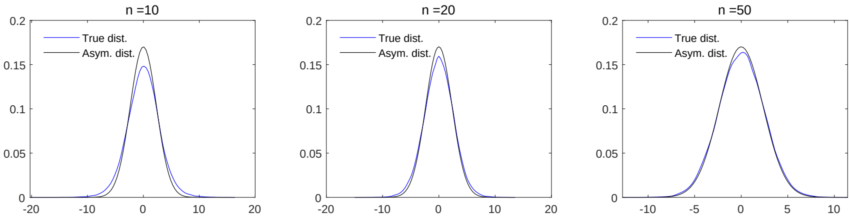
$$Y = -1 + D + \varepsilon, \quad \text{with } f_{\varepsilon|X}(u|x) = \frac{1}{2 \times \sqrt{2}} \varphi\left(\frac{u+2}{\sqrt{2}}\right) + \frac{1}{2}\varphi(u-2)$$

$\Rightarrow \varepsilon$ is independent from X and follows a mixture of normal laws.



Example

- We simulate 10^5 samples $(Y_i, X_i)_{i=1,\dots,n}$ of size $n \in \{10, 20, 50\}$.
- We compare the true distribution of $\sqrt{n}(\widehat{\beta}_D - 1)$ with the asymptotic one.



- The asymptotic law slightly underestimates the true variance: why ?
- But the normal approximation quickly becomes correct as the sample size increases.

- Let us assume the following homoskedasticity condition:

$$E(\varepsilon^2 XX') = E(\varepsilon^2)E(XX'). \tag{Hom}$$

- In this case, the asymptotic variance matrix V_a takes the simple form

$$V_a = \sigma^2 E(XX')^{-1}, \text{ with } \sigma^2 = E(\varepsilon^2).$$

- If in addition to (Hom), we have $X = (1, D)'$ with $D \in \mathbb{R}$ and $\widehat{\beta} = (\widehat{\alpha}, \widehat{\beta}_D)$, the asymptotic variance of $\widehat{\beta}_D$ satisfies

$$V_a(\widehat{\beta}_D) = \frac{\sigma^2}{V(D)}. \tag{2}$$

- Accordingly, the accuracy of $\widehat{\beta}_D$ decreases with σ^2 and increases with $V(D)$.
- Indeed, the precision of $\widehat{\beta}_D$ improves with n.

Accuracy of the OLS estimator

- Let us consider now the accuracy of $\widehat{\beta}_D$, the coefficient of $D \in \mathbb{R}$ in the regression of Y on $X = (1, D, G')'$.

Proposition 1

Under the assumptions (i.i.d.), (Inv th), (Mom), and (Hom), we have:

$$V_a(\widehat{\beta}_D) = \frac{\sigma^2}{(1 - R^2_\infty)V(D)},$$

where R^2_∞ is the limit in probability of R^2 of the regression of D on G.

- Hence, in a "long" régression of Y on D and G, the accuracy of $\widehat{\beta}_D$ also depends on the dependence between D and G.
- If D is very well predicted by a linear combination $G'b$ of G ($\Rightarrow R^2_\infty$ close to 1), then it will be difficult to estimate $\widehat{\beta}_D$.
- Intuition : it will be difficult to disentangle the effect of D and the effect of $G'b$.
- The most favorable case is when D is uncorrelated/orthogonal to all the components of G ($R^2_\infty = 0$).

Proof of Proposition 1*

- $V_a(\widehat{\beta}_D)$ is the 2nd diagonal term of $V_a = \sigma^2 E(XX')^{-1}$.
- So, $V_a(\widehat{\beta}_D) = e_2' V_a e_2$, with $e_2 = [0\,,\,1\,,\,0]'$.
- Accordingly, $V_a(\widehat{\beta}_D)/\sigma^2$ is the 2nd component (u_2) of the solution u to $E(XX')u = e_2$.
- Now, $e_2 = E[X(\eta/V(\eta))]$, where η is the residual of the theoretical regression of D on G.
- So, u is the coefficient of the theoretical regression of $\eta/V(\eta)$ on X. Then, from Frisch-Waugh Theorem, u_2 is the coeff. of the theoretical regression of $\eta/V(\eta)$ on η.
- So, $u_2 = 1/V(\eta)$ and also $V_a(\widehat{\beta}_D) = \sigma^2/V(\eta)$.
- Since $\widehat{\eta}$ is the residual of the regression of D on G, the R^2 of such a regression will verify

$$\begin{aligned} R^2 =& 1 - \widehat{V}(\widehat{\eta})/\widehat{V}(D) \\ \xrightarrow{p} & R^2_\infty = 1 - V(\eta)/V(D) \quad \square \end{aligned}$$

Homoskedasticity and optimality of the OLS*

- Let us assume that
$$Y = X'\beta_0 + \varepsilon, \text{ with } E(\varepsilon|X) = 0 \text{ and } V(\varepsilon|X) = \sigma^2. \tag{3}$$
- N.B. : these assumptions are stronger than
$$E(X\varepsilon) = 0 \text{ and } E(\varepsilon^2 XX') = E(\varepsilon^2)E(XX').$$
- Question : What is the best available estimator of β_0 ?
- Here we look only at estimators **linear** in $(Y_1, ..., Y_n)$ and **unbiased** :
$$\widetilde{\beta} = \sum_{i=1}^{n} w(X_1, ..., X_n)Y_i, \; E[\widetilde{\beta}|X_1, ..., X_n] = \beta_0 \;\forall\; \beta_0. \tag{4}$$

Theorem 2 (Gauss-Markov Theorem)

Under (i.i.d.), (3), and if $(1/n)\sum_{i=1}^{n} X_iX_i'$ is invertible , any estimator $\widetilde{\beta}$ satisfying (4) is such that
$$V\left(\widehat{\beta}|X_1, ..., X_n\right) \leq V\left(\widetilde{\beta}|X_1, ..., X_n\right).$$

- Careful though: this feature of the OLS is only valid under (3) and if we restrict ourselves to those estimators satisfying (4).

Plan

- To build tests and confidence intervals we need a consistent estimator of the asymptotic variance V_a. We consider:

$$\widehat{V}_a = \widehat{E}(XX')^{-1}\widehat{E}(\widehat{\varepsilon}^2 XX')\widehat{E}(XX')^{-1},$$

where for any random variable A we denote $\widehat{E}(A) = (1/n)\sum_{i=1}^n A_i \, (= \overline{A})$.

- Notice that $\widehat{V}_{a,jj}$, the jth diagonal term of $\widehat{V}_a$, is the estimator of $V_a(\widehat{\beta}_j)$ ($\widehat{V}_{a,jk}$ is the estimator of $\mathrm{Cov}_a(\widehat{\beta}_j, \widehat{\beta}_k)$).
- In practice, statistical software report the *"standard errors"* $\mathrm{se}_j = \widehat{V}_{a,jj}^{1/2}/\sqrt{n}$.

Proposition 2

Under the assumptions (i.i.d.), (Inv th), *and* (Mom), *we have*

$$\widehat{V}_a \xrightarrow{P} V_a.$$

It suffices to show that $\widehat{E}(XX') \xrightarrow{P} E(XX')$ and $\widehat{E}(\widehat{\varepsilon}^2 XX') \xrightarrow{P} E(\varepsilon^2 XX')$.

The first convergence is a direct consequence of the Law of Large Numbers (LLN).

The second term satisfies

$$\begin{aligned}\widehat{E}(\widehat{\varepsilon}^2 XX') =& \widehat{E}(\varepsilon^2 XX') + \widehat{E}((\widehat{\varepsilon}+\varepsilon)(\widehat{\varepsilon}-\varepsilon)XX') \\ =& \widehat{E}(\varepsilon^2 XX') + \widehat{E}((\widehat{\varepsilon}-\varepsilon)^2 XX') + 2\widehat{E}(\varepsilon(\widehat{\varepsilon}-\varepsilon)XX').\end{aligned}$$

By the LLN, $\widehat{E}(\varepsilon^2 XX') \xrightarrow{P} E(\varepsilon^2 XX')$. So, it suffices to prove that

$$\widehat{E}((\widehat{\varepsilon}-\varepsilon)^2 XX') \xrightarrow{P} 0, \quad \widehat{E}(\varepsilon(\widehat{\varepsilon}-\varepsilon)XX') \xrightarrow{P} 0.$$

Let $Q = \widehat{E}((\widehat{\varepsilon} - \varepsilon)^2 XX')$, with a generic term denoted as $Q_{j,\ell}$. Then, $Q_{j,\ell} = (\widehat{\beta} - \beta_0)' \widehat{M}_{j,\ell} (\widehat{\beta} - \beta_0)$ with

$$\widehat{M}_{j,\ell} = \frac{1}{n} \sum_{i=1}^{n} (X_i^j X_i^\ell) X_i X_i',$$

where $X_i^j = j$th component of $X_i \in \mathbb{R}^k$. Since $E(||X||^4) < \infty$, by the LLN,

$$\widehat{M}_{j\ell} \xrightarrow{P} E[(X^j X^\ell) XX'] \quad (\text{ with } X = (X^1, ..., X^k)').$$

So, by the consistency of $\widehat{\beta}$, $\widehat{E}((\widehat{\varepsilon} - \varepsilon)^2 XX') \xrightarrow{P} 0$.

By using $E(|\varepsilon| \times ||X||^3) < \infty$, we also show that

$$\widehat{E}(\varepsilon(\widehat{\varepsilon} - \varepsilon) XX') \xrightarrow{P} 0 \ \square$$

- Careful! Even if $\sqrt{n}\left(\widehat{\beta}_j - \beta_{0j}\right) \xrightarrow{d} \mathcal{N}\left(0, V_{a,jj}\right)$ with $\widehat{V}_{a,jj} \xrightarrow{P} V_{a,jj}$, we do not necessarily have

$$n\frac{V(\widehat{\beta}_j)}{\widehat{V}_{a,jj}} \xrightarrow{P} 1 \quad \text{or} \quad n\frac{V(\widehat{\beta}_j)}{V_{a,jj}} \to 1.$$

- However, $\widehat{V}_{a,jj}$ is enough to build asymptotic tests and confidence intervals for β_{0j}, see below.

$\Rightarrow$ se_j is not necessarily a good estimator of $V(\widehat{\beta}_j)^{1/2}$, but it is enough to conduct a valid inference, as we will see.

Plan

Null and alternative hypotheses

- From a sample, we want to choose between two possibilities: the null hypothesis, denoted H_0, and the alternative hypothesis, denoted H_1.
- Examples:
  - Vaccinated individuals have the same probability of being ill as others (H_0) vs. have a different probability (H_1);
  - Vaccinated individuals have a lower probability (H_0) vs. a higher probability (H_1) of being ill than others;
  - The diploma is independent (H_0) or not (H_1) of the socio-professional category;
  - Individuals form rational expectations (H_0) or not (H_1).
- We will make a decision on H_0 vs. H_1 from the sample.
- Formally, we construct a random variable ϕ, a function of the data, such that $\phi = 1$ if we reject H_0, 0 if we accept it.
- Often, H_0 corresponds to a simple hypothesis, whereas H_1 is composite.

- Two risks associated with the decision:
  - Type I error (test level): probability of rejecting H_0 when H_0 is true;
  - Type II error: probability of accepting H_0 when H_1 is true.
- We also define the power = 1 - type II error.

$\Rightarrow$ Power = probability of correctly accepting H_1.

- Level $= P_{H_0}(\phi = 1)$, where P_{H_0} = probability if H_0 is true.
- Power $= P_{H_1}[\phi = 1]$, where P_{H_1} = probability if H_1 is true.

Example of the vaccine

- Suppose p_0 is known, the probability of being ill without vaccination.
- We have an i.i.d. sample of vaccinated individuals for whom we observe whether they are ill ($X_i = 1$) or not ($X_i = 0$). Let $p_1 = P(X_1 = 1)$.
- We consider here H_0: $p_1 = p_0$ and H_1: $p_1 \neq p_0$ (a two-sided test).
- Intuitively, it makes sense to accept H_0 if $\overline{X}$ is close to p_0, and H_1 otherwise.
- Let $\phi = \mathbb{1}\left\{\overline{X} < p_0 - c_1 \cup \overline{X} > p_0 + c_2\right\}$ for some $c_1, c_2 > 0$. Then:

$$\begin{aligned}\text{Level} =& P_{H_0}(\overline{X} < p_0 - c_1 \cup \overline{X} > p_0 + c_2)\\ =& P(\text{Bin}(n, p_0)/n < p_0 - c_1) + P(\text{Bin}(n, p_0)/n > p_0 + c_2),\\ \text{Power} =& P(\text{Bin}(n, p_1)/n < p_0 - c_1) + P(\text{Bin}(n, p_1)/n > p_0 + c_2).\end{aligned}$$

  where $\text{Bin}(n, p)$ is a binomial random variable with parameters (n, p).
- We can compute the level because p_0 is known.
- But we cannot compute the power: it depends on p_1, which is unknown.

- Ideally, we would like to minimize both risks, but that is impossible.

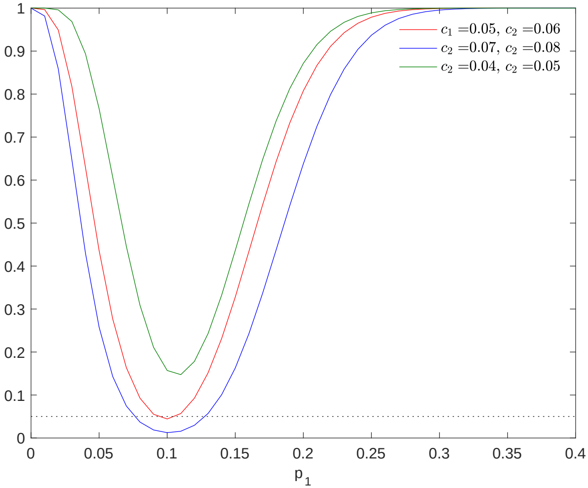
$\Rightarrow$ We restrict ourselves to tests whose level is bounded (by α) (most often, at 1%, 5% or 10%) and we look for the uniformly most powerful test (UMP) among these tests (if it exists).

- In the previous example, we look for a test ϕ such that:
  1. $P_{p_0}(\phi = 1) \leq \alpha$ (with P_p = probability when $P(X = 1) = p$);
  2. For all $p_1 \neq p_0$, $P_{p_1}(\phi = 1) \geq P_{p_1}(\phi' = 1)$ for any other test ϕ' such that $P_{p_0}(\phi' = 1) \leq \alpha$.

- If $\phi = \mathbb{1}\left\{\overline{X} < p_0 - c_1 \cup \overline{X} > p_0 + c_2\right\}$ for some $c_1,\ c_2 > 0$:
  1. The condition $P_{p_0}(\phi = 1) \leq \alpha$ restricts the possible c_1, c_2.
  2. We can then look for (c_1, c_2) that maximize $P_{p_1}(\phi = 1)$ for all p_1.

- We consider here $n = 100$, $\alpha = 5\%$ and $p_0 = 0.1$.
- Power curve for three choices of (c_1, c_2):

  (we speak of a power curve even if we measure the level at $p_1 = p_0$)



Composite null hypotheses

- The level defined previously may be ambiguous if the null hypothesis is composite.
- Example of vaccination: H_0: $p_1 \leq p_0$ vs. H_1: $p_1 > p_0$ (a one-sided test).
- Then under H_0, $P_{p_1}(\phi_n = 1)$ depends on p_1.

$\Rightarrow$ we redefine the level as $\sup_{p_1 : H_0 \text{ holds}} P_{p_1}(\phi = 1)$.

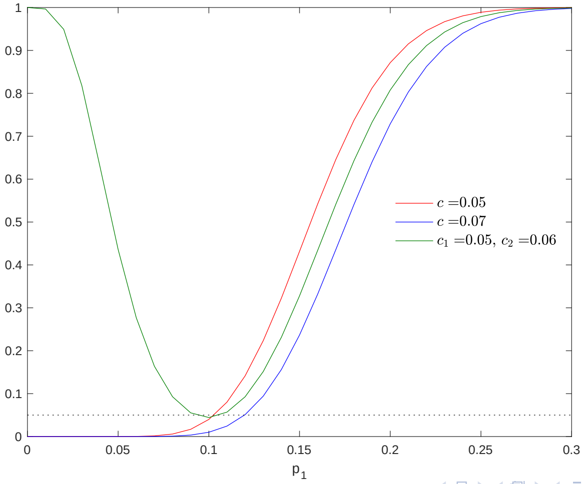
- It is often easy to determine for which value the supremum is reached.
- Example: if $\phi = \mathbb{1}\left\{\overline{X} > p_0 + c\right\}$ for some $c > 0$, one can show that

$$\sup_{p_1 : H_0 \text{ holds}} P_{p_1}(\phi = 1) = P_{p_0}(\phi = 1),$$

i.e., the type I error is maximal when $p_1 = p_0$.

Illustration: one-sided test

- Vaccine example with $n = 100$, $\alpha = 5\%$ and $p_0 = 0.1$.
- Power curve for two choices of c.
  (green curve: power of the test $\phi = \mathbb{1}\left\{\overline{X} < p_0 - c_1 \cup \overline{X} > p_0 + c_2\right\}$)



- In general, tests are written as

$$\phi = \mathbb{1}\left\{T \in W_\alpha\right\},$$

for some variable T and set W_α. Then:

  - T is called the test statistic;
  - W_α is called the critical region.

- N.B.: W_α is often indexed by α to highlight that we determine W_α in such a way that $P_{H_0}(\phi = 1) \leq \alpha$.

- Previous example:

  - Two-sided test: $T = \overline{X}$ and $W_\alpha = [0, p_0 - c_1[\cup]p_0 + c_2, 1]$, where c_1, c_2 such that $P_{H_0}(\phi = 1) \leq \alpha$.
  - One-sided test: $T = \overline{X}$ and $W_\alpha =]p_0 + c, 1]$, where c such that $P_{H_0}(\phi = 1) \leq \alpha$.

P-values

- In general, when $\phi = \mathbb{1}\{T \in W_\alpha\}$, $\alpha \mapsto W_\alpha$ is increasing in the sense of inclusion.
- If we reject H_0 at 5%, $T \in W_{0.05}$ but we do not know if $T \in W_{0.01}$: would we reject H_0 at 1%?
- Similarly, if we accept H_0 at 5%, $T \notin W_{0.05}$, but we do not know if $T \in W_{0.10}$, i.e., whether we would reject H_0 at 10%.
- Richer information: p-value, defined as

$$p := \inf\{\alpha \in]0,1[: \ T \in W_\alpha\}.$$

  It is the minimal level at which H_0 is rejected.
- For example, if the p-value is 2.3%, we reject H_0 at 5% or 10% but not at 1%.
- p-value: a measure of plausibility of H_0 but beware, $\neq P(H_0 \text{ true})$!
- Note: p is a random variable, a function of T.

Asymptotic level of a test

- How can we perform tests in a non-parametric framework?
- Example:
  - Test whether the mean wage of a subgroup equals the mean wage m_0 (assumed known).
  - From an i.i.d. sample $(X_1, ..., X_n)$ of wages from the subgroup, we want to test H_0: $E(X) = m_0$ against H_1: $E(X) \neq m_0$.
  - Intuition: use $\phi = \mathbb{1}\left\{\overline{X} < m_0 - c_1 \cup \overline{X} > m_0 + c_2\right\}$ for some c_1, c_2.
  - Problem: if we do not know the distribution of the X, what is the distribution of $\overline{X}$ under H_0?
- Idea: use limit distributions to simplify the problem.
- We will no longer necessarily have $P_{H_0}(\phi = 1) \leq \alpha$, but only

$$\lim_{n\to\infty} P_{H_0}(\phi_n = 1) \leq \alpha, \tag{5}$$

  where we index ϕ by n to emphasize its dependence on n.
- If (5) holds, we say that the test is of asymptotic level α.

Illustration: test of H_0: $E(X) = m_0$ vs. H_1: $E(X) \neq m_0$.

- According to the CLT and Slutsky's theorem, under H_0:

$$\sqrt{n}\frac{\overline{X} - m_0}{\widehat{\sigma}} \xrightarrow{d} \mathcal{N}(0,1)\,,$$

where $\widehat{\sigma}^2 := \frac{1}{n-1}\sum_{i=1}^n (X_i - \overline{X})^2$.

- Let $T = (\overline{X} - m_0)/(n^{-1/2}\widehat{\sigma})$ and consider the test

$$\phi = \mathbb{1}\left\{|T| > q_{1-\alpha/2}\right\}, \tag{6}$$

where $q_{1-\alpha/2}$ is the quantile of order $1 - \alpha/2$ of a $\mathcal{N}(0,1)$.

- Then, with $Z \sim \mathcal{N}(0,1)$:

$$\begin{aligned} P_{H_0}(\phi = 1) =& P\left(T < -q_{1-\alpha/2}\right) + P\left(T > -q_{1-\alpha/2}\right) \\ \xrightarrow{n\to\infty}& P\left(Z < -q_{1-\alpha/2}\right) + P\left(Z > q_{1-\alpha/2}\right) \\ =& \frac{\alpha}{2} + \frac{\alpha}{2} = \alpha. \end{aligned}$$

- Another important asymptotic property of tests is their consistency:

$$\lim_{n\to\infty} P_{H_1}(\phi = 1) = 1.$$

$\Rightarrow$ If H_1 holds, we reject H_0 with probability $\to 1$.

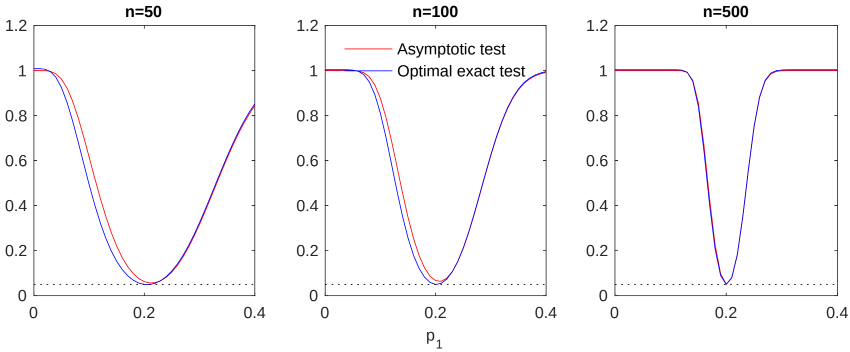
- Most reasonable tests satisfy this property.
- Previous example: the test defined by (6) is consistent.
- Indeed, if for example $E(X) > m_0$,

$$T = \underbrace{\sqrt{n}\frac{\overline{X} - E(X)}{\widehat{\sigma}}}_{\xrightarrow{d}\mathcal{N}(0,1)} + \underbrace{\sqrt{n}\frac{E(X) - m_0}{\widehat{\sigma}}}_{\xrightarrow{P}\infty} \xrightarrow{P} \infty.$$

- Thus:

$$P(|T| > q_{1-\alpha/2}) \geq P(T > q_{1-\alpha/2}) \to 1.$$

Power of tests of H_0: $p_1 = p_0 (= 0.2)$ vs. H_1: $p_1 \neq p_0$.



$\Rightarrow$ The power approaches 1 as $n \to \infty$.

- The asymptotic test also approaches the optimal test as $n \to \infty$.

Outline

Two-sided tests

- Consider a test of $H_0 : \beta_{0j} = 0$ against $H_1 : \beta_{0j} \neq 0$.
- Equivalent to: does X^j help to predict Y? If a causal interpretation is possible (cf. Chap. 4), $\beta_{0j} = 0 \Leftrightarrow$ no causal effect of X^j on Y.
- To test H_0 with an asymptotic level α, we consider the Student statistic $t_j = \widehat{\beta}_j / \mathrm{se}_j$ and the critical region

$$W_\alpha^b =]-\infty, \, -q_{1-\alpha/2}[\cup]q_{1-\alpha/2}, \, \infty[,$$

  where $q_{1-\alpha/2}$ is the quantile of order $1-\alpha/2$ of a $\mathcal{N}(0,1)$.
- When we reject (resp. accept) H_0 for a test at the $x\%$ level, we say that X^j is (resp. is not) significant at $x\%$.

Proposition 3

Under assumptions (i.i.d.), (Inv th) *and* (Mom), *the above test has asymptotic level* α *and is consistent:*

$$\lim_{n\to\infty} P_{H_0}(t_j \in W_\alpha^b) = \alpha, \quad \lim_{n\to\infty} P_{H_1}(t_j \in W_\alpha^b) = 1.$$

From Theorem 1,

$$\sqrt{n}\,\frac{\widehat{\beta}_j - \beta_{0j}}{V_{a,jj}^{1/2}} \xrightarrow{d} \mathcal{N}(0,1).$$

From Proposition 2, $\widehat{V}_{a,jj}/V_{a,jj} \xrightarrow{P} 1$. Thus, by Slutsky's theorem,

$$\begin{aligned}
\frac{\widehat{\beta}_j - \beta_{0j}}{\mathrm{se}_j} &= \frac{\widehat{\beta}_j - \beta_{0j}}{(\widehat{V}_{a,jj}/n)^{1/2}} \\
&= \left(\frac{\widehat{V}_{a,jj}}{V_{a,jj}}\right)^{-1/2} \times \sqrt{n}\,\frac{\widehat{\beta}_j - \beta_{0j}}{V_{a,jj}^{1/2}} \\
&\xrightarrow{d} \mathcal{N}(0,1). \qquad (7)
\end{aligned}$$

Therefore, under H_0, $t_j \xrightarrow{d} \mathcal{N}(0,1)$.

The reasoning to obtain $\lim_{n\to\infty} P_{H_0}(t_j \in W_\alpha^b) = \alpha$ (resp. $\lim_{n\to\infty} P_{H_1}(t_j \in W_\alpha^b) = 1$) is detailed in slide 31 (resp. in slide 32).

- Another standard test: the one-sided test of $H_0 : \beta_{0j} = 0$ (or $\beta_{0j} \leq 0$) against $H_1 : \beta_{0j} > 0$.
- Useful when one can assume $\beta_{0j} \geq 0$.
- Example: education coefficient in a wage regression.
- We then consider the critical region $W^u_\alpha =]q_{1-\alpha},\ \infty[$.
- More powerful than the two-sided test if $\beta_{0j} > 0$, since $q_{1-\alpha} < q_{1-\alpha/2}$.

Proposition 4

Under assumptions (i.i.d.), (Inv th) *and* (Mom), *the test* $\phi_n = \mathbb{1}\{t_j \in W^u_\alpha\}$ *has asymptotic level* α *and is consistent:*

$$\lim_{n\to\infty} P_{H_0}(t_j \in W^u_\alpha) = \alpha, \quad \lim_{n\to\infty} P_{H_1}(t_j \in W^u_\alpha) = 1.$$

Proof: similar to that of Proposition 3, now using $P(Z > q_{1-\alpha}) = \alpha$ and the fact that under H_1, $t_j \xrightarrow{P} \infty$ $\square$

P-values in the previous examples

- In the case of the two-sided test, $t_j \in W_\alpha$ as long as $|t_j| \geq q_{1-\alpha/2}$. Thus, the p-value p satisfies $|t_j| = q_{1-p/2}$.

$\Rightarrow$ $p = 2(1 - \Phi(|t_j|))$ or equivalently (letting $Z \sim \mathcal{N}(0,1)$):

$$p = P\left(|Z| > |t_j| \middle| t_j\right) \quad \text{: probability of observing a statistic more extreme in absolute value than } |t_j| \text{ under } H_0.$$

- Similarly, in the case of the one-sided test, the p-value p satisfies $t_j = q_{1-p}$ or equivalently:

$$p = 1 - \Phi(t_j) = P(Z > t_j \mid t_j)$$

$\Rightarrow$ Same interpretation as in the two-sided case.

- In both cases and in many others,

$$p \xrightarrow{d} \left| \begin{array}{ll} \mathcal{U}[0,1] & \text{under } H_0 \\ 0 & \text{under } H_1 \end{array} \right.$$

- Consider the regression of log(wage) on education, experience and job tenure.
- We want to test $\beta_{\text{exp}} = 0$ against $\beta_{\text{exp}} \neq 0$ or against $\beta_{\text{exp}} > 0$.
- Using data from J. Wooldridge, we obtain (standard errors in parentheses):

$$\widehat{\text{lwage}} = \underset{(0.104)}{0.284} + \underset{(0.007)}{0.092}\,\text{educ} + \underset{(0.0017)}{0.0041}\,\text{exp} + \underset{(0.003)}{0.022}\,\text{tenure}.$$

- Given that $t_{\text{exp}} \simeq 2.41$, what are the conclusions of the two-sided and one-sided tests at the 10, 5 and 1% levels?

Table 1: Quantiles of $\mathcal{N}(0,1)$.

α	0.9	0.95	0.975	0.99	0.995
q_α	1.282	1.645	1.960	2.326	2.576

Statistical vs practical significance

- Practical significance: is $X^j\widehat{\beta}_j$ an important component of $X'\widehat{\beta}$ for predicting Y?
- Statistical significance depends on t_j, while practical significance depends on $\widehat{\beta}_j$.
- Important notion because when n is large, t_j can be large even if $\widehat{\beta}_j$ is "small".
- Example: participation rate of U.S. employees in retirement plans, as a function of the company match rate (mrate), plan age and firm size (totemp).
- From the dataset 401k.dta of J. Wooldridge, we obtain:

$$\widehat{\text{prate}} = \underset{(0.79)}{80.29} + \underset{(0.52)}{5.44}\ \text{mrate} + \underset{(0.045)}{0.269}\ \text{age} - \underset{(0.00004)}{0.00013}\ \text{totemp}$$

- totemp is significant at the 1% level and even at the 0.2% level ($t = -3.25$)...
- ... But its effect is very small: an increase in firm size of 10,000 reduces the predicted participation by only 1.3 percentage points.

Statistical vs practical significance

- Conversely, one may have a small t but a large estimated coefficient, especially if the sample is small.
- Example: job training and worker productivity (measured here by the proportion of defective pieces in factories, variable scrap).
- From 29 Michigan firms in 1987, defining training = annual hours of job training per employee, we obtain:

$$\widehat{\text{log scrap}} = \underset{(5.69)}{12.46} - \underset{(0.023)}{0.029}\ \text{training} + \ldots\text{(other covariates)}$$

$\Rightarrow$ 10 additional hours of training predict a 29% decrease in the proportion of defective pieces. Substantial effect!

- But since the sample is very small, the test statistic is relatively weak in absolute value ($t \simeq -1.26$), and the p-value of a one-sided test remains fairly large (0.11).

$\Rightarrow$ hard to conclude here without additional data.

Multiple hypothesis tests: definition and idea

- Consider now the test of $H_0 : R\beta_0 = b$ against $H_1 : R\beta_0 \neq b$, with R a full rank $r \times k$ matrix and b an $r \times 1$ column vector.
- Example: $\beta_{0k-r+1} = \beta_{0k-r+2} = ... = \beta_{0k} = 0$. Joint nullity test.
- Idea: under H_0, we have

$$\sqrt{n}(R\widehat{\beta} - b) = R \times \sqrt{n}(\widehat{\beta} - \beta_0) \xrightarrow{d} \mathcal{N}\left(0, RV_aR'\right).$$

- Thus, since $\widehat{V}_a \xrightarrow{P} V_a$,

$$n(R\widehat{\beta} - b)'(R\widehat{V}_aR')^{-1}(R\widehat{\beta} - b) \xrightarrow{d} \chi^2(r). \tag{8}$$

- We define the Fisher statistic $F = n(R\widehat{\beta} - b)'(R\widehat{V}_a R')^{-1}(R\widehat{\beta} - b)/r$ and, for a test of level α, the critical region

$$W_\alpha^m =]q_{1-\alpha}(r)/r, \infty[,$$

with $q_{1-\alpha}(r)$ the quantile of order $1-\alpha$ of a $\chi^2(r)$ distribution.

Proposition 5

Under assumptions (i.i.d.), (Inv th) *and* (Mom), *the above test (called the Fisher test) has asymptotic level* α *and is consistent:*

$$\lim_{n\to\infty} P_{H_0}(F \in W_\alpha^m) = \alpha, \quad \lim_{n\to\infty} P_{H_1}(F \in W_\alpha^m) = 1.$$

Proof: the first result follows from (8). For the second, note that under H_1, $R\widehat{\beta} \xrightarrow{P} \widetilde{b} \neq b$. Thus:

$$(R\widehat{\beta} - b)'(R\widehat{V}_a R')^{-1}(R\widehat{\beta} - b) \xrightarrow{P} (\widetilde{b} - b)'(RV_a R')^{-1}(\widetilde{b} - b) > 0.$$

Hence, $F \to \infty$ under H_1 and $\lim_{n\to\infty} P_{H_1}(F \in W_\alpha^m) = 1$ $\square$

Example of multiple tests

- Where does the wage dispersion of baseball players come from? In particular, do players' usual game statistics predict their salary?
- Consider the regression:

$$\begin{aligned}\log(\text{salary}) &= \beta_0+\beta_1\text{years}+\beta_2\text{matches\_per\_year}\\ &\quad +\beta_3\text{bavg}+\beta_4\text{hrunsyr}+\beta_5\text{rbisyr}+u \qquad (9)\end{aligned}$$

- We then want to test:

$$H_0:\beta_3=\beta_4=\beta_5=0 \text{ versus } H_1:\beta_3\neq 0 \text{ or } \beta_4\neq 0 \text{ or } \beta_5\neq 0.$$

- From the dataset mlb1.dta of J. Wooldridge, we obtain $F\simeq 9.97$. Now

$$W^m_{0.1}=]2.08,+\infty[,\;\; W^m_{0.05}=]2.67,+\infty[\text{ and } W^m_{0.01}=]3.78,+\infty[.$$

  Conclusion?

- We also obtain:

$$\begin{aligned}\widehat{\log(\text{salary})} &= \underset{(0.29)}{11.19}+\underset{(0.0157)}{0.0689}\text{ years}+\underset{(0.0026)}{0.0126}\text{ matches\_per\_year}\\ &\quad +\underset{(0.0008)}{0.00098}\text{ bavg}+\underset{(0.0166)}{0.0144}\text{ hrunsyr}+\underset{(0.0072)}{0.0108}\text{ rbisyr}.\end{aligned}$$

- Conclusion on the simple tests?

- Remark 1: in fact, one can use Fisher's tests to test $H_0 : \beta_{0j} = 0$ against $H_1 : \beta_{0j} \neq 0$.
- Then we get $F = t_j^2$. The Student's t-test and Fisher's test are therefore identical.
- But one cannot use Fisher's test to perform a one-sided test.
- Remark 2: one may reject no simple hypothesis underlying the multiple test but still reject the multiple test (see earlier).
- This occurs when the corresponding variables are highly correlated.
- Often (but not always!) if at least one variable is significant in a simple test, a joint test including it will also be significant.

- Idea: quantify the uncertainty on $\widehat{\beta}_j$ by considering an interval that contains it.
- Formally, for any $\alpha \in (0,1)$, we construct an interval $\mathrm{IC}_{1-\alpha}$ such that

$$\lim_{n\to\infty} P(\mathrm{IC}_{1-\alpha} \ni \beta_{0j}) = 1-\alpha. \tag{10}$$

- To do this, let $q_\tau = \Phi^{-1}(\tau)$ and note that

$$\begin{aligned} 1-\alpha &= \lim_{n\to\infty} P\left(-q_{1-\alpha/2} \le \frac{\widehat{\beta}_j - \beta_{0j}}{\mathrm{se}(\widehat{\beta}_j)} \le q_{1-\alpha/2}\right) \\ &= \lim_{n\to\infty} P\left(\widehat{\beta}_j - q_{1-\alpha/2}\,\mathrm{se}(\widehat{\beta}_j) \le \beta_{0j} \le \widehat{\beta}_j + q_{1-\alpha/2}\,\mathrm{se}(\widehat{\beta}_j)\right) \end{aligned}$$

- Therefore, the interval $[\widehat{\beta}_j - q_{1-\alpha/2}\,\mathrm{se}(\widehat{\beta}_j), \widehat{\beta}_j + q_{1-\alpha/2}\,\mathrm{se}(\widehat{\beta}_j)]$ satisfies (10).
- Relation with tests: we reject the test $H_0 : \beta_{0j} = a_j$ against $H_1 : \beta_{0j} \neq a_j$ at level x% if and only if a_j is not in the (100-x)% confidence interval.

Confidence region (vector parameter)

- Confidence region for $\theta_0 = R\beta_0$, with R a full-rank $r \times k$ matrix?
- To do this, we use

$$n(R\widehat{\beta} - \theta_0)'(R\widehat{V}_a R')^{-1}(R\widehat{\beta} - \theta_0) \xrightarrow{d} \chi^2(r). \tag{11}$$

- Consider:

$$R_{1-\alpha} = \left\{\theta : n(R\widehat{\beta} - \theta)'(R\widehat{V}_a R')^{-1}(R\widehat{\beta} - \theta) \leq q_{1-\alpha}(r)\right\}.$$

- Using (11), we obtain:

$$\begin{aligned} P(R_{1-\alpha} \ni \theta_0) &= P\left(n(R\widehat{\beta} - \theta_0)'(R\widehat{V}_a R')^{-1}(R\widehat{\beta} - \theta_0) \leq q_{1-\alpha}(r)\right) \\ &\to 1 - \alpha. \end{aligned}$$

- Note: $R_{1-\alpha}$ is an ellipsoid.

Plan

- The condition (Hom) $E(\varepsilon^2 XX') = E(\varepsilon^2)E(XX')$ simplifies the asymptotic variance into
$$V_a = \sigma^2 E(XX')^{-1}, \quad \text{with } \sigma^2 = E(\varepsilon^2).$$
- In this case we can estimate V_a by $\widehat{V}_a^h = \widehat{\sigma}^2 \widehat{E}(XX')^{-1}$, with
$$\widehat{\sigma}^2 = \frac{1}{n-k} \sum_{i=1}^{n} \widehat{\varepsilon}_i^2.$$
Note: $\widehat{\sigma}^2$ unbiasedly estimates σ^2 if $E(\varepsilon|X) = 0$ and $V(\varepsilon|X) = \sigma^2$.
- Warning! By default, standard errors computed by statistical software correspond to $\text{se}_j^h = (\widehat{V}_{a,jj}^h)^{1/2}/\sqrt{n}$.
- To obtain the more general estimators se_j with R, we can for instance use the command `lm_robust` instead of `lm`.
- In practice, se_j and se_j^h are often close.

Homoskedasticity: example

- Do students attending private schools perform better than others?
- We regress the test score at the beginning of the 3rd grade on the test score at the beginning of the 1st grade, parental education, and being in a private school or not
- The p-values computed under homoskedasticity are slightly larger, but the difference with the standard errors robust to heteroskedasticity is negligible.

Table 2: Private schools and students' performance

Response Variable	OLS estimator	St. Err. robust	p-val. robuste	St.Err. (homosk.)	p-val. (homosk.)
test score in math	-1,15	0,48	0,016	0,49	0,019
Test score in French	0,17	0,41	0,68	0,44	0,70

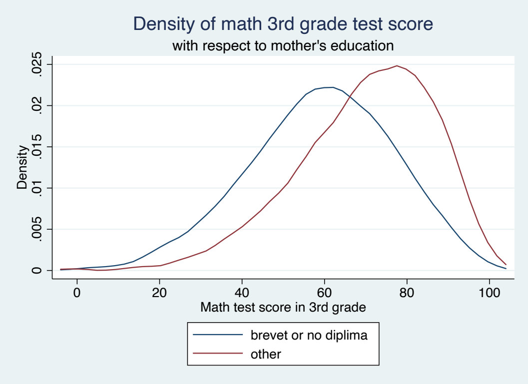
Notes : French DEPP 1997. 5 809 observations. X also includes 1st grade test scores and parental education.

Exact tests under normality of the ε's

- Up to now, we have only considered asymptotic tests and confidence intervals.
- When *n* is small, their true level can be very different from the *nominal* level (i.e., the α we fix).
- We can obtain exact test for *n* fixed, but under the very restrictive following assumption:
$$\varepsilon|X \sim \mathcal{N}(0, \sigma^2). \tag{Nor}$$
- This assumption imposes that $Y|X \sim \mathcal{N}(X'\beta_0, \sigma^2)$. But most variables are not normally distributed.
- (Nor) also imposes that ε be independent from X.
- N.B.: under (Nor), $\widehat{\beta}$ is the maximum likelihood estimator of β_0.

Normality of the ε's : counter-examples

- We consider math test scores in 3rd grade as a function of mother's education.
- Then $Y|X$ is not normal (normality tests are rejected).
- Also, $V(Y|X = x)$ is not constant with x. Skewness also varies with x.



Normality of the ε's: distribution of $(\widehat{\beta}, \widehat{\sigma}^2)$

- But under (Nor), we can get the following finite sample result:

Theorem 3

Assume that (i.i.d.), $\widehat{E}(XX')$ *is invertible, and* (Nor). *Then, letting* $\mathcal{X} = (X_1, ..., X_n)$, *we get:*

$$\widehat{\beta}|\mathcal{X} \sim \mathcal{N}\left(\beta_0, \frac{\sigma^2}{n}\widehat{E}(XX')^{-1}\right), \quad (n-k)\frac{\widehat{\sigma}^2}{\sigma^2}|\mathcal{X} \sim \chi^2(n-k).$$

Moreover, conditional on $\mathcal{X}$, $\widehat{\beta}$ *and* $\widehat{\sigma}^2$ *are independent.*

Sketch of proof: $\widehat{\beta} = \beta_0 + \widehat{E}(XX')^{-1}\widehat{E}(X\varepsilon)$. From (Nor), we get:

$$\widehat{E}(X\varepsilon)|\mathcal{X} \sim \mathcal{N}\left(0, \frac{\sigma^2}{n^2}\sum_{i=1}^{n} X_i X_i'\right).$$

The first result follows since $\sum_{i=1}^{n} X_i X_i'/n^2 = \widehat{E}(XX')/n$.

For the 2nd formula, let $\mathcal{E} = (\varepsilon_1, ..., \varepsilon_n)'$. Then $(n-k)\widehat{\sigma}^2 = \mathcal{E}' N \mathcal{E}$ with N an orthogonal projection matrix, function of $\mathcal{X}$ only.

Then $N = P'\Delta P$ with P an orthogonal matrix ($P'P = Id_n$) and Δ a diagonal matrix with $\text{rank}(N) = n-k$ diagonal coeffs=1 and k diagonal coeffs=0.

Let $\widetilde{\mathcal{E}} = P\mathcal{E}$. Then $\widetilde{\mathcal{E}}|\mathcal{X} \sim \mathcal{N}(0, \sigma^2 Id_n)$. Thus:

$$(n-k)\frac{\widehat{\sigma}^2}{\sigma^2} = \frac{\widetilde{\mathcal{E}}'\Delta\widetilde{\mathcal{E}}}{\sigma^2} = \sum_{i=1}^{n-k} \frac{\widetilde{\mathcal{E}}_i^2}{\sigma^2}.$$

The 2nd result follows by definition of the $\chi^2(n-k)$ distribution.

For conditional independence, we use $\widehat{\beta} - \beta_0 = Q\mathcal{E}$ with Q an orthogonal projection matrix, function of $\mathcal{X}$, such that

$$Q \times N = 0.$$

By Cochran's Theorem, conditionally on $\mathcal{X}$, $\widehat{\beta} - \beta_0$ is independent from $N\mathcal{E}$, and hence independent from $(n-k)\widehat{\sigma}^2 = \mathcal{E}' N \mathcal{E}$.

Normality of the ε's: simple tests

- We consider the test of $H_0 : \beta_{0j} = 0$ vs $H_1 : \beta_j \neq 0$.
- Let $t_j^h = \widehat{\beta}_j / se_j^h$, with $se_j^h = (\widehat{V}_{a,jj}^h)^{1/2}/\sqrt{n}$.
- Let $q_{1-\alpha/2}(T_{n-k})$ the quantile of order $1-\alpha/2$ of T_{n-k}, the t (or "Student") distribution with $(n-k)$ degrees of freedom.
- T_{n-k} is the distribution of $Z/\sqrt{W_{n-k}/(n-k)}$, with $Z \sim \mathcal{N}(0,1)$, $W_{n-k} \sim \chi^2(n-k)$ and Z independent of W_{n-k}.
- Then, to test H_0 at the level α, we consider the critical region

$$W_\alpha^{b,n} =]-\infty, -q_{1-\alpha/2}(T_{n-k})] \cup [q_{1-\alpha/2}(T_{n-k}), \infty[.$$

Corollary 1

Under (i.i.d.), (Inv th), (Nor), *and if $\widehat{E}(XX')$ is invertible with probability 1, then :*

$$P_{H_0}(t_j^h \in W_\alpha^{b,n}) = \alpha, \quad \lim_{n\to\infty} P_{H_1}(t_j^h \in W_\alpha^{b,n}) = 1.$$

Proof of Corollary 1*

Proof: Under (Nor) we have, by Theorem 3 and by definition of $\widehat{V}_a^h$,

$$\frac{\widehat{V}_{a,jj}^h}{nV(\widehat{\beta}_j)} = \frac{\widehat{\sigma}^2}{\sigma^2}.$$

Thus,

$$t_j^h = \frac{\widehat{\beta}_j/V(\widehat{\beta}_j)^{1/2}}{\{\widehat{V}_{a,jj}^h/[nV(\widehat{\beta}_j)]\}^{1/2}} = \frac{\widehat{\beta}_j/V(\widehat{\beta}_j)^{1/2}}{\sqrt{\frac{1}{n-k}\frac{(n-k)\widehat{\sigma}^2}{\sigma^2}}}.$$

Thus, by Theorem 3 again, under H_0,

$$t_j^h = Z/\sqrt{W_{n-k}/n-k}$$

with $Z \sim \mathcal{N}(0,1)$ and $W_{n-k} \sim \chi^2(n-k)$, Z and W_{n-k} independent conditional on $\mathcal{X}$.

Thus, under H_0, $t_j^h|\mathcal{X} \sim T_{n-k}$. Since T_{n-k} is independent of $\mathcal{X}$, $t_j^h \sim T_{n-k}$.

The two results of the corollary can then be obtained as those of Prop. 3 □

Comparison with the asymptotic approach

- Compared to Proposition 3, we have just replaced $q_{1-\alpha/2}$ (quantile of a $\mathcal{N}(0,1)$) by $q_{1-\alpha/2}(T_{n-k})$ in W_α^b.
- No contradiction if (Nor) holds, since when $n \to \infty$, $T_{n-k} \xrightarrow{d} \mathcal{N}(0,1)$ and thus $q_{1-\alpha/2}(T_{n-k}) \to q_{1-\alpha/2}$.
- Advantage: $q_{1-\alpha/2}(T_{n-k}) > q_{1-\alpha/2}$ so for a given n, the tests are more conservative, with a true level likely closer to α.

- We consider the test of $H_0 : R\beta_0 = b$ vs $H_1 : R\beta_0 \neq b$, with R a $r \times k$ full-rank matrix.
- We let $F^h = n(R\widehat{\beta} - b)'(R\widehat{V}_a^h R')^{-1}(R\widehat{\beta} - b)/r$. Identical to F but with $\widehat{V}_a$ replaced by $\widehat{V}_a^h$.
- Let $q_{1-\alpha}(F_{r,n-k})$ be the quantile of order $1-\alpha$ of $F_{r,n-k}$, a Fisher (F) distribution with $(r, n-k)$ degrees of freedom.
- $F_{r,n-k}$ is the distribution of $(W_r/r)/(W_{n-k}/(n-k))$, with $W_r \sim \chi^2(r)$, $W_{n-k} \sim \chi^2(n-k)$ and W_r independent of W_{n-k}.
- Then, to test H_0 at the level α, we consider the critical region

$$W_\alpha^{m,n} = [q_{1-\alpha}(F_{r,n-k}) \, , \, \infty[\, .$$

Corollary 2

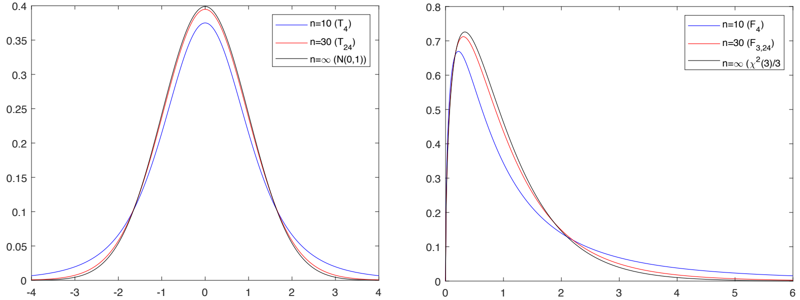
Under (i.i.d.), (Inv th), (Nor) *and if* $\widehat{E}(XX')$ *is invertible with probability 1, then:*

$$P_{H_0}(F^h \in W_\alpha^{m,n}) = \alpha, \qquad \lim_{n\to\infty} P_{H_1}(F^h \in W_\alpha^{m,n}) = 1.$$

Comparison with the asymptotic approach (2)

- Compared to Proposition 5, we just replaced $q_{1-\alpha}(r)/r$, quantile of a $\chi^2(r)/r$, by $q_{1-\alpha}(F_{r,n-k})$ in W_α^m.
- No contradiction since when $n \to \infty$, $F_{r,n-k} \xrightarrow{d} \chi^2(r)/r$ and thus $q_{1-\alpha}(F_{r,n-k}) \to q_{1-\alpha}(r)/r$.
- Advantage: $q_{1-\alpha}(F_{r,n-k}) > q_{1-\alpha}(r)/r$ so for a given n, the tests are more conservative and a priori with level closer to α.

Student vs $\mathcal{N}(0,1)$ and Fisher vs $\chi^2(r)/r$



T_{n-k} vs $\mathcal{N}(0,1)$, for $k=6$. $F_{3,n-k}$ vs $\chi^2(3)/3$, for $k=6$.

Figure 1: Comparaison between Student's and normal law, and Fisher's and $\chi^2(r)/r$ law

- Remarkable difference when n is small, but negligible difference when $n \geq 30$.

Summary

- Asymptotic normality of $\widehat{\beta}$.
- Factors of the accuracy of $\widehat{\beta}$.
- Consistent estimators of the asymptotic variance.
- Simple and Multiple Tests asymptotically valid.
- Homoskedasticity : definition.
- Student and Fisher Tests: exact tests under the error normality.