Econometrics Part 3

The linear model: tests

Jose Angel Garcia Sanchez

jagarsanc@gmail.com

Sorbonne Data Analytics

November 2025

UNIVERSITÉ PARIS 1
PANTHÉON SORBONNE

Outline

UNIVERSITÉ PARIS 1
PANTHÉON SORBONNE

Outline

UNIVERSITÉ PARIS 1
PANTHÉON SORBONNE

Usual distributions to know

- Normal
- χ^2
- Student (T)
- Fisher

The Normal distribution

It is the usual "bell-shaped" distribution. It can also be called the Gaussian distribution.
$X \hookrightarrow N(\mu, \sigma^2)$ if:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

Any linear combination of independent normal random variables is a normal random variable

The Chi-squared distribution

It is defined as the sum of squared independent normal variables.
$Y \hookrightarrow \chi^2_n$ if:

$$Y = \sum_{i=1}^{n} X_i^2$$

Where $X_i \hookrightarrow N(0,1)$ and the X_i are independent. Any linear combination of independent χ^2 random variables is a χ^2 random variable

The Student distribution

It is defined as a ratio of a normal and a square root χ^2 that are independent.
$Y \hookrightarrow T_n$ if:

$$Y = \frac{X}{\sqrt{\frac{Z}{n}}}$$

Where $X \hookrightarrow N(0,1)$ and $Z \hookrightarrow \chi^2_n$, and are independent. The Student distribution is bell-shaped, and when $n \to +\infty$ it becomes a normal distribution

The Fisher distribution

It is defined as a ratio of two χ^2 that are independent.
$Y \hookrightarrow F_{p,q}$ if:

$$Y = \frac{Z_1/p}{Z_2/q}$$

Where $Z_1 \hookrightarrow \chi^2_p$ and $Z_2 \hookrightarrow \chi^2_q$, and are independent.

UNIVERSITÉ PARIS 1
PANTHÉON SORBONNE

Reminder: OLS estimators are BLUE

We still consider that the X variables are not random at this stage.

- We have $\hat{b} = (X'X)^{-1}X'y$
- $E(\hat{b}) = b$: unbiased
- $V(\hat{b}) = \sigma^2(X'X)^{-1}$, the smallest possible
- $\hat{\sigma}^2 = (\sum \hat{u}_i^2)/(N-k)$: unbiased
- So $\hat{V}(\hat{b}) = \hat{\sigma}^2(X'X)^{-1}$
- Gauss-Markov theorem: under hypotheses 1 to 4, the OLS estimator is the Best Linear Unbiased Estimator (unbiased with a minimal variance)
- So far, no assumption on the distribution of variables (y, X or u)
- So we don't know yet the distribution of parameters b, which would be needed for tests

UNIVERSITÉ PARIS 1 PANTHÉON SORBONNE

Distribution of estimators in a finite sample (1)

- Finite sample=sample not so large
- We need to assume that $u \hookrightarrow N(0, \sigma^2 I_N)$
- $\hat{b} \hookrightarrow N(b, \sigma^2 (X'X)^{-1})$
- Calling $\hat{b}_j$ the j^{th} element of $\hat{b}$ and α_{jj} the j^{th} element of the diagonal of $(X'X)^{-1}$, $\sigma^2_{\hat{b}_j} = \sigma^2 \alpha_{jj}$:
- $\hat{b}_j \hookrightarrow N(b_j, \sigma^2_{\hat{b}_j})$
- Next, we standardize $\hat{b}_j$ to have it follow a standard Normal distribution:

$$\forall j, \frac{\hat{b}_j - b_j}{\sigma_{\hat{b}_j}} \hookrightarrow N(0, 1)$$

Distribution of estimators in a finite sample (2)

- The previous expression is used to test possible values on b_j, for given $\hat{b}_j$ and $\sigma_{\hat{b}_j}$ found from the sample
- But $\sigma_{\hat{b}_j}$ is still unkown: the standard error of error terms
- Plugging the estimated $\hat{\sigma}_{\hat{b}_j}$ instead of the true $\sigma_{\hat{b}_j}$ will change the distribution from a Normal to a Student
- Why: $\hat{\sigma}_{\hat{b}_j}$ is an imperfect estimate of $\sigma_{\hat{b}_j}$

$$\forall j, t_j = \frac{\hat{b}_j - b_j}{\hat{\sigma}_{\hat{b}_j}} \hookrightarrow T_{N-k}$$

Where $\hat{\sigma}^2_{\hat{b}_j} = \hat{\sigma}^2 \alpha_{jj}$ (cf. *supra*).

Notice that when $N \to \infty$, $T_{N-k} \sim N(0,1)$

UNIVERSITÉ PARIS 1 PANTHÉON SORBONNE

Formal proof

We know that $\hat{\sigma}^2 = \frac{\Sigma \hat{u}_i^2}{N-k}$
We thus get:

$$(N-k)\frac{\hat{\sigma}^2}{\sigma^2} \hookrightarrow \chi^2_{N-k}$$

Proof: using the fact that u is normal and that $\hat{u} = M_X u$ (see Dormont)
Then, using the fact that $\hat{b}_j$'s are normal and $\hat{\sigma}^2$ is a χ^2, we compute their ratio and end up with a Student distribution of parameter $N-k$.

UNIVERSITÉ PARIS 1
PANTHÉON SORBONNE

The test procedure: an example (1)

- Consider the following model:
  $income = a + b.height + c.education$, estimated over N individuals
- Suppose that the estimated parameter $\hat{b}$ is *close* to zero
- We thus infer that variable *height* could be irrelevant: the correlation between income and height could (should) be zero
- The "true" b should be zero
- But even if it is the case, it is very unlikely that we get $\hat{b} = 0$ (due to intrisic randomness).
- Given the computed $\hat{b}$, there should be a way to assess if the "true" b is in fact zero or not

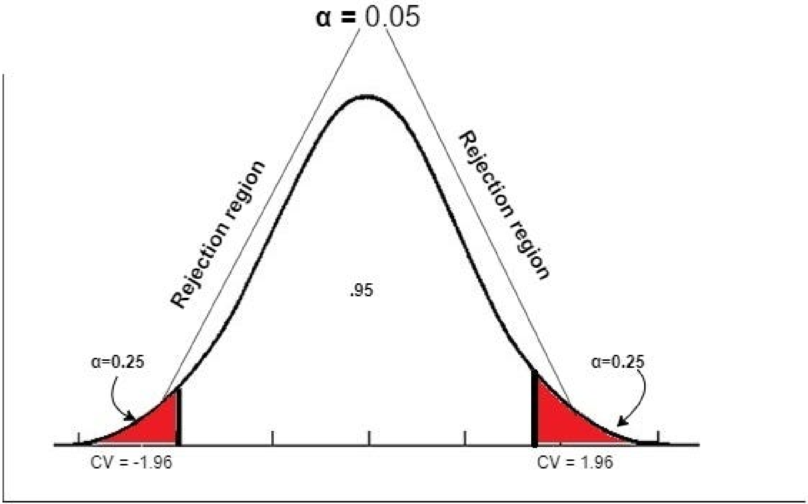
UNIVERSITÉ PARIS 1
PANTHÉON SORBONNE

The test procedure: an example (2)

- Let's call H_0 the hypothesis: $b = 0$, and H_1 the hypothesis: $b \neq 0$
- Should we consider H_0 as true?
- We know that for this model, $t = (\hat{b} - b)/\hat{\sigma}_{\hat{b}} \rightarrow T_{N-3}$
- Is the latter still likely, if we were to assume H_0 is true?
- Assuming H_0 is true means that we assume $b = 0$, so that t simplifies to $t = (\hat{b}/\hat{\sigma}_{\hat{b}})$
- If under H_0, we find this value $t = (\hat{b}/\hat{\sigma}_{\hat{b}})$ to be unlikely to belong to a T_{N-3} distribution, then we will say that H_0 was wrong
- Rejecting $H_0 \Leftrightarrow$ parameter b is significant
- Not rejecting $H_0 \Leftrightarrow$ parameter b is not significant

UNIVERSITÉ PARIS 1 PANTHÉON SORBONNE

Illustration



A general test procedure

- $H_0 : b = 0$
- $H_1 : b \neq 0$

1. Find the value $N - k$
2. Compute $t = \hat{b}/\hat{\sigma}_{\hat{b}}$
3. Compare t with the 2.5% quantile of a T_{N-k} distribution: t^*
4. We take 2.5% because the distribution is symmetric: we want to know if the value t belongs to the 5% less probable values of the distribution, i.e. the first 2.5% or the last 2.5%
5. If $|t| > |t^*|$ then H_0 is rejected: parameter b is significantly different from 0 and thus significant
6. If not, H_0 is not rejected: parameter b is not significantly different from 0 and thus not significant
7. Always mention the level of significance: here (and it is often the case): $\alpha = 5\%$
8. If a variable is found to be non significant, we can remove it from the analysis

The higher the t-statistic, the higher the significance.

An even more general test procedure

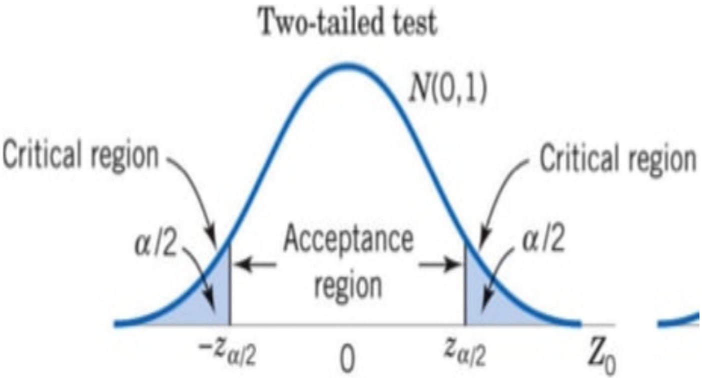
- $H_0 : b = a$
- $H_1 : b \neq a$

1. Find the value $N - k$
2. Compute $t = \frac{\hat{b}-a}{\hat{\sigma}_{\hat{b}}}$
3. Compare t with the 2.5% quantile of a T_{N-k} distribution: t^*
4. If $|t| > |t^*|$ then H_0 is rejected: parameter b is significantly different from a
5. If not, H_0 is not rejected: parameter b is not significantly different from a
6. Always mention the level of significance: here (and it is often the case): $\alpha = 5\%$

The higher the t-statistic, the higher the chance we are right in rejecting H_0.

UNIVERSITÉ PARIS 1 PANTHÉON SORBONNE

Illustration



Is it very different?

- If $N \to \infty$, all we've covered so far works well, because it works for *any* N
- But if know in advance we'll only be using large samples, we can directly use large sample tests
- These are called *asymptotic tests*

UNIVERSITÉ PARIS 1
PANTHÉON SORBONNE

Consistency of estimators

We want to know what happens if the sample size becomes very large, i.e. when $N \to \infty$. Call $\hat{b}_N$ the estimate computed on a sample of size N.

- $\hat{b}_N$ is consistent: $plim(\hat{b}_N) = b$
- Proof:
- Consistency in quadratic mean implies consistency in probability
- Do we have $E[(\hat{b}_N - b)(\hat{b}_N - b)'] = V(\hat{b}_N) \to 0$?
- $V(\hat{b}_N) = \sigma^2(X'X)^{-1} = \frac{\sigma^2}{N}(\frac{X'X}{N})^{-1}$
- The limit of which is $0.V_X^{-1} = 0$

So we see a first advantage of working with large samples: estimates are *consistent*.

UNIVERSITÉ PARIS 1
PANTHÉON SORBONNE

Distribution of estimators

Estimators exhibit *asymptotic normality*: when $N \to \infty$, we know that their distribution is a Normal, whatever the distribution of u's (contrary to the finite sample case).
Thanks to the Central Limit Theorem (CLT), we can derive the distribution of estimators when $N \to \infty$. Call $\hat{b}_N$ the estimate computed on a sample of size N.

- $\hat{b}_N$ is asymptotically normal: $\sqrt{N}(\hat{b}_N - b) \to N(0, \sigma^2 V_X^{-1})$
- σ^2 is the variance of error terms u
- "Inflating" by $\sqrt{N}$ is needed, otherwise the variable would be degenerate (zero variance)
- Intuition:
- $\hat{b}_N - b = (X'X)^{-1}u$ so that it is a linear combination of the u's
- So that the CLT applies and gives a normal distribution when $N \to \infty$

Test statistic

- In a large sample, $\hat{\sigma}_{\hat{b}_j}$ is an almost perfect estimate of $\sigma_{\hat{b}_j}$
- Formally, $plim \frac{\hat{\sigma}_{\hat{b}_j}}{\sigma_{\hat{b}_j}} = 1$
- Plugging $\hat{\sigma}_{\hat{b}_j}$ instead of $\sigma_{\hat{b}_j}$ does not change the distribution, it stays a Normal

$$\forall j, t_j = \frac{\hat{b}_j - b_j}{\hat{\sigma}_{\hat{b}_j}} \hookrightarrow N(0,1)$$

So that tests are run directly using the Normal quantiles. That's what happened indeed when we run finite-sample tests with large samples, because the Student distribution becomes a standard Normal when $N \to \infty$.
Notice that in Stata, only finite sample statistics are displayed (t)

Outline

UNIVERSITÉ PARIS 1
PANTHÉON SORBONNE

Size and power of a test

- $\alpha = \textit{Type I error} = P(H_0 \textit{ rejected} | H_0 \textit{ is true})$
- $\beta = \textit{Type II error} = P(H_0 \textit{ accepted} | H_1 \textit{ is true})$
- α is the *size* of a test: we are aware that we'll make the wrong decision $100\alpha\%$ of the time
- α is also called the *significance level*
- $1 - \beta$ is the *power* of a test: it indicates how powerful a test is in finding deviations from the null hypothesis H_0
- $1 - \beta = P(H_0 \textit{ rejected} | H_1 \textit{ is true})$
- Lowering $\alpha \Rightarrow$ increasing β (and reciprocally): there is a trade-off between the two
- Since we cannot minimize both, we set α as fixed (e.g. 5%) and try to find the test that minimizes β for this given α
- Minimizing $\beta \Leftrightarrow$ maximizing the power of a test

UNIVERSITÉ PARIS 1
PANTHÉON SORBONNE

Sample size

- Dropping a useful variable can lead to non consistent estimates, while keeping an unimportant variable only leads to loss in precision
- Say we set $\alpha = 0.01$ with a small sample size: then estimates are likely to have a large variance
- So even if the true parameter is not zero, its t statistic is likely to be small, thus failing to reject H_0 although it is false

Choosing the right test size

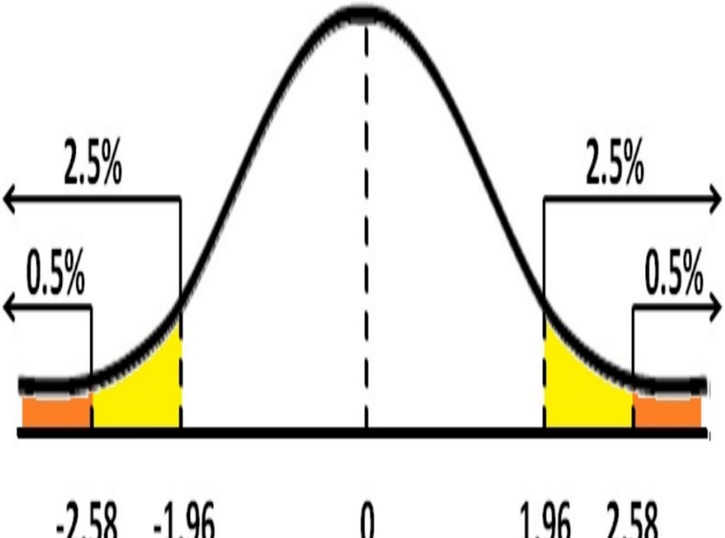
- In that case, we might remove from the analysis a relevant variable simply because we've been too stringent about the size of the test
- Since it is easier to control α than β, we usually set the value of α with respect to the sample size in order to have enough power (i.e. ability to discriminate H_0 from the true value of the parameter)

Example

- Suppose we are testing the hypothesis $b = 0$, while the true value is $b = 0.1$
- The probability that we reject the null (H_0) depends on the standard error of $\hat{b}$, thus on the sample size
- The larger the sample, the smaller the standard error so the more likely we are to reject H_0
- Type II errors thus become increasingly unlikely when sample size increases
- In a big sample; we can thus decrease the size of the test α: e.g. 1%
- Similarly, we can choose a size of 10% in small samples

UNIVERSITÉ PARIS 1
PANTHÉON SORBONNE

Illustration



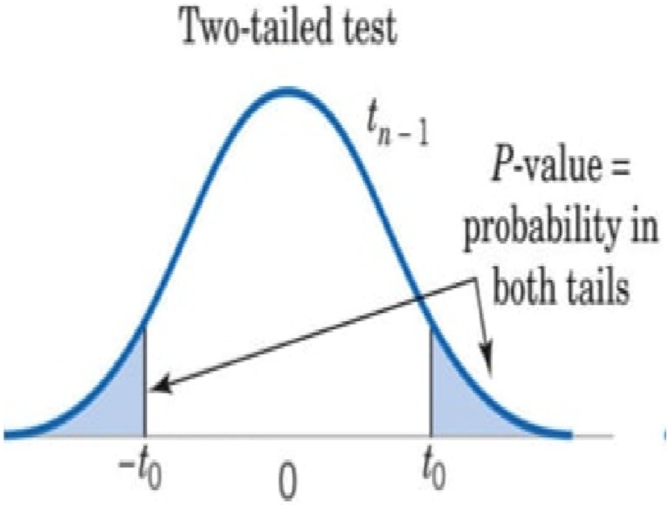
UNIVERSITÉ PARIS 1
PANTHÉON SORBONNE

The p-value

- It is defined with respect to the test statistic t that we just computed
- Its definition is: $pvalue = P(|X| > |t|)$ with $X \hookrightarrow T_{T-k}$
- It tells where t is located in the Student distribution T_{T-k} (more precisely, at which quantile)
- If $pvalue < 0.05$, it means that $|t|$ lies above $t^*_{2.5\%}$
- It means that H_0 is to be rejected
- Statsmodels OLS fit provide p-values for all tests so that decision is made easy

UNIVERSITÉ PARIS 1
PANTHÉON SORBONNE

Illustration



Remark

- Say we are testing $b = 0$, i.e. no impact of some variable on the outcome
- If we don't reject H_0, it simply means that there is not enough evidence to reject H_0
- But another study could possibly find enough evidence to reject it
- It's why we must write: H_0 can or cannot be rejected

- Using the model $y = Xb + u$ and estimating it on the period $[1, N]$, how could we predict y_T using any X_T?
- A natural prediction is: $\hat{y}_T = X'_T \hat{b}$
- Notice that X'_T is a line vector, since in the matrix of observations, one line = one observation
- And the forecast error is: $\hat{e} = y_T - \hat{y}_T$

The forecast error

- The "true" y_T is: $y_T = X_T' b + u_T$
- The forecast error is thus: $\hat{e} = X_T'(b - \hat{b}) + u_T$
- u_T is uncorrelated to the $(u_i)_{i \in [1,N]}$ and to the X's, so:
- $E(\hat{e}) = 0$: the forecast is unbiased
- $V(\hat{e}) = E(\hat{e}^2) = \sigma^2 X_T'(X'X)^{-1} X_T + \sigma^2$
- Thus depends on:
  - one term linked to model inaccuracy (σ^2)
  - another linked to the accuracy of $\hat{b}$ (its variance $\sigma^2(X'X)^{-1}$) and the distance between X_T and the X's used for the estimation

UNIVERSITÉ PARIS 1
PANTHÉON SORBONNE

The forecast interval

$$IC_{95\%} = [\hat{y}_T - t\sqrt{\hat{\sigma}^2 X_T'(X'X)^{-1}X_T + \hat{\sigma}^2}, \hat{y}_T + t\sqrt{\hat{\sigma}^2 X_T'(X'X)^{-1}X_T + \hat{\sigma}^2}]$$

With t being the 2.5% quantile of the T_{N-k} distribution (if N is large, it is approximately the 2.5% quantile of the normal distribution: 1.96).

Outline

UNIVERSITÉ PARIS 1 PANTHÉON SORBONNE

Testing a linear combination of parameters

- A T-test can be used to test a linear restriction of parameters, e.g. $b_1 + 2b_2 = a$
- For more than one linear restriction, we would need the Fisher test (see below)
- Variable $(\hat{b}_1 + 2\hat{b}_2)$ does follow a Student distribution, since $\hat{b}_1$ and $\hat{b}_1$ do
- The test statistic would be: $t = \frac{(\hat{b}_1 + 2\hat{b}_2) - a}{\hat{\sigma}_{\hat{b}_1 + 2\hat{b}_2}}$
- With $\hat{\sigma}_{\hat{b}_1 + 2\hat{b}_2}$ being computed as the usual standard error of a sum of variables
- Hint: $V(aX + bY) = a^2 V(X) + b^2 V(Y) + 2abcov(X, Y)$

UNIVERSITÉ PARIS 1
PANTHÉON SORBONNE

Example: Cobb Douglas function (1)

Say we use a Cobb Douglas production function to describe the technology of N different firms. y is the output, K and L the inputs (capital and labor respectively). α and β are the output elasticities of K and L respectively. A_i is the total factor productivity of each firm (comprises firm-specific efficiency). Our goal is to estimate α and β. For each firm i:

$$y_i = A_i K_i^{\alpha} L_i^{\beta}$$

Taking the log on both sides (assuming no variable ever has value 0), we get:

$$log(y_i) = log(A_i) + \alpha log(K_i) + \beta log(L_i)$$

Example: Cobb Douglas function (2)

Say A is the average total factor productivity, with $E(A_i) = A$, we can then write:

$$log(y_i) = log(A_i) - log(A) + log(A) + \alpha log(K_i) + \beta log(L_i)$$

And calling $u_i = log(A_i) - log(A)$, we have $E(u_i) = 0$ and:

$$log(y_i) = log(A) + \alpha log(K_i) + \beta log(L_i) + u_i$$

Which amounts to a classical linear model of the form $y = a + bx + cz + u$.

Example: Cobb Douglas function (3)

- Data taken from Greene, on 27 firms
- Estimation of a Cobb-Douglas production function
- Parameters of interest: α and β, the output elasticities
- If $\alpha = \beta \Rightarrow$ same elasticity for both factors
- If $\alpha + \beta = 1 \Rightarrow$ constant returns to scale

You can use the T-test to test for a single linear combination. For any number of linear combinations (including one), use the Fisher test, see below.

One-tailed tests

- We saw we could test $b = \textit{any value } a$ and follow the same procedure, that is not restricted to testing $b = 0$
- By default, softwares give the $b = 0$ test, but we can ask for any test we like
- So far, we implemented *two-sided tests*: $b = 0$ vs. $b > 0$ or $b < 0$
- We can run *one-sided tests* if relevant: e.g. $b = 0$ vs. $b > 0$, or $b = a$ vs. $b > a$
- One-tailed tests are used if the results are interesting only if they turn out in one direction
- In practice, we would be using only one half of the test distribution
- One-tailed tests still posit $H_0 : b = a$ because the t statistic must be computed under some simple assumption H_0 (assuming $H_0 : b < a$ would be too vague).

One-tailed tests: procedure

- Say we wish to compare parameter b to value a
- While testing $H_0 : b = a$ in the two-tailed test procedure, we get $pvalue = x$
- One-tailed tests still posit $H_0 : b = a$, while H_1 can either be $b > a$ or $b < a$
- If $\hat{b} > a$, then:
  - $H_1 : b > a$ leads to $pvalue = \frac{x}{2}$
  - $H_1 : b < a$ leads to $pvalue = 1 - \frac{x}{2}$
- If $\hat{b} < a$, then:
  - $H_1 : b > a$ leads to $pvalue = 1 - \frac{x}{2}$
  - $H_1 : b < a$ leads to $pvalue = \frac{x}{2}$

All this can be easily generalized to a linear restriction involving 2 parameters, e.g. $b_1 + b_2 > 0$

UNIVERSITÉ PARIS 1
PANTHÉON SORBONNE

One-tailed tests: example

- Say we run an estimation on a large sample, and we want to compare parameter b to value 4
- $\hat{b} = 3.8$ and $\hat{\sigma}_{\hat{b}} = 0.1$
- The Student t-value is: $\frac{3.8-4}{0.1} = -2$
- Testing $H_0 : b = 4$ with the two-tailed procedure, we get $pvalue = 0.046 = 4.6\%$
- Since $\hat{b} < 4$, then:
    - Testing $H_1 : b > 4$ leads to $pvalue = 1 - \frac{0.046}{2} = 0.977 = 97.7\%$
    - Testing $H_1 : b < 4$ leads to $pvalue = \frac{0.046}{2} = 0.023 = 2.3\%$
- So b can be said to be not significantly greater than 4

UNIVERSITÉ PARIS 1
PANTHÉON SORBONNE

Simultaneous testing

- We have $\hat{b} \hookrightarrow N(b, \sigma^2(X'X)^{-1})$
- Thus: $\hat{b} - b \hookrightarrow N(0, \sigma^2(X'X)^{-1})$
- Taking the quadratic form of these, "divided" by their variance matrix, we get a χ^2_k:
- $(\hat{b} - b)'[\sigma^2(X'X)^{-2}]^{-1}(\hat{b} - b)$
- But σ is unknown
- In a large sample, no problem: we plug-in the estimated $\hat{\sigma}$, and it still is a χ^2_k

UNIVERSITÉ PARIS 1
PANTHÉON SORBONNE

Finite vs. large sample

- In a finite sample, replacing σ by its estimate $\hat{\sigma}$ will change the distribution from a χ^2_k to a Fisher $F_{(k,N-k)}$
- Formally: we use $(N-k)\frac{\hat{\sigma}^2}{\sigma^2}$ which is a χ^2_{N-k}
- We compute their ratio, and get to a Fisher distribution

F-test continued

$$f = \frac{(\hat{b} - b)'(X'X)(\hat{b} - b)}{(k\hat{\sigma}^2)} \hookrightarrow F_{(k,N-k)}$$

- The testing procedure is the same as for the T-test:
- Assume H_0 is true: that $\hat{b} = b$
- Compute f
- Check if it is likely that f belongs to a $F_{(k,N-k)}$ distribution
- If it is not, e.g. at the 5% level, say that H_0 is not rejected at the 5% level

Remark on a single-parameter Fischer

- For a test for a single linear combination after a regression, we should use the Student test
- Instead, we could practically use a Fisher test
- This is because Fisher tests could be used to test any number of linear constraints, so they can be used even if there is only one
- The square of a variable following a student T distribution is equal to a variable following a Fisher distribution, so Student and Fisher procedures are equivalent here
- $Y_1 \hookrightarrow F_{p,q}$ if: $Y_1 = \frac{Z_1/p}{Z_2/q}$ with $Z_1 \hookrightarrow \chi^2_p$ and $Z_2 \hookrightarrow \chi^2_p$
- Say $p = q = 1$: Y is the ratio of two normal variables squared
- $Y_2 \hookrightarrow T_n$ if: $Y_2 = \frac{X}{\sqrt{\frac{Z}{n}}}$ with $X \hookrightarrow N(0,1)$ and $Z \hookrightarrow \chi^2_n$
- Say $n = 1$: Y_2 is the ratio of two normal variables
- Hence $Y_1 = Y_2^2$

UNIVERSITÉ PARIS 1
PANTHÉON SORBONNE

About simultaneous testing

- A F-test is about testing a group of parameters (here k parameters)
- Why couldn't we use k T-tests instead?
- First, the global F-test takes into account possible correlation between parameters
- Second, even assuming that parameters are independent (which never happens), see the following example:
- Consider we have 20 parameters b_i to test, each using a T-test of size $a = 0.05$ testing $h_0^i : b_i = 0$ vs. $h_1^i : b_i \neq 0$
- Assume we want to run a global test, testing H_0: all the parameters are equal to 0 vs. H_1: at least one of them is not equal to 0, using the results of the previous T-tests
- What is the actual size of this global test?
- $\alpha = Type\ I\ error = P(H_0\ rejected | H_0\ is\ true)$
- $\alpha = 1 - P(H_0\ accepted | H_0\ is\ true)$
- $\alpha = 1 - P(all\ the\ h_0^i\ are\ accepted | they\ are\ true)$
- $\alpha = 1 - (1 - a)^{20} = 1 - (1 - 0.05)^{20} \approx 0.64$

A generalization of the F-test (1)

- We may want to test only for a subset of parameters. In that case, we simply take the preceding formula with what concerns the q parameters in question:
- $f = (\hat{\beta} - \beta)'[Z/(q\hat{\sigma}^2)](\hat{\beta} - \beta) \hookrightarrow F_{(q,N-k)}$
- β: subset of parameters tested
- q: number of parameters tested (= no. of constraints)
- Z^{-1}: submatrix of $(X'X)^{-1}$ corresponding to the q coefficients
- The default test $b = 0$ (except constant term) is always given by statsmodels: it is a general test of the model, called the global F-test

UNIVERSITÉ PARIS 1
PANTHÉON SORBONNE

A generalization of the F-test (2)

- We may want to test for some linear constraints (ex: $b_1 = b_2$ and $b_3 = b_4$)
- We have to rewrite these q constraints in a matrix form: $Rb = r$

The test statistic is the following:

$$f = \frac{1}{q}(R\hat{b} - r)'(\hat{\sigma}^2 R(X'X)^{-1}R')^{-1}(R\hat{b} - r) \hookrightarrow F_{(q,N-k)}$$

UNIVERSITÉ PARIS 1
PANTHÉON SORBONNE

How to implement this? (1)

- In fact, all this amounts to testing whether the full model is better or not than the model with constraints
- Let's call R^2_{nc} the R^2 of the non constrained model, and R^2_c the R^2 of the constrained model
- It can be shown that:

$$f = \frac{(R^2_{nc} - R^2_c)}{1 - R^2_{nc}} \cdot \frac{N-k}{q} \hookrightarrow F_{(q,N-k)}$$

With N: sample size, k: no. of variables, q: no. of constraints

UNIVERSITÉ PARIS 1
PANTHÉON SORBONNE

How to implement this? (2)

- Call SSR the sum of the squared residuals
- df the degrees of freedom of models
- c: constrained, nc: non constrained models
- It can be shown that:

$$f = \frac{(SSR_c - SSR_{nc})}{SSR_{nc}} \cdot \frac{df_{nc}}{df_c - df_{nc}} \hookrightarrow F_{(df_c - df_{nc}, df_{nc})}$$

UNIVERSITÉ PARIS 1 PANTHÉON SORBONNE

How to select variables

- Choose to use variables in levels, logs or growth rates using economic theory, variable plots, tests or common sense
- Detect perfect multicollinearity and remove it
- Then, 3 options:
- Backward elimination: estimate the model with all the variables and remove those that are not significant (T then F tests)
- Forward regression: begin with only one explanatory variable, the one that is the most correlated to the dependent variable, then introduce the other ones one by one according to their partial correlation coefficient with y, while they are significant
- Stepwise regression: same process, and at each step variables that cease to be significant are removed
- No automatic procedure can replace common sense

UNIVERSITÉ PARIS 1
PANTHÉON SORBONNE

Outline

UNIVERSITÉ PARIS 1 PANTHÉON SORBONNE

The Chow test

- It could be relevant to test whether parameters are the same on 2 subsamples of the data
- Ex: are consumer behaviors the same before and after the German reunification? Idea: test the temporal stability of the regression
- Ex: are the parameters of a wage equation the same for men and women? Idea: test homogeneity of behaviors
- Since we are testing a set of parameters simultaneously, we use a F test
- The constrained model is the one where parameters are forced to be the same on the two subsamples
- The non-constrained model is the one where parameters are allowed to differ across the two subsamples
- The test will tell us if we can or cannot reject homogeneity of behaviors

1st method

Let's call SSR the sum of squared residuals

- SSR_c: of the constrained model
- SSR_1: of the model estimated on the 1st subsample
- SSR_2: of the model estimated on the 2nd subsample
- Consequence: $SSR_{nc} = SSR_1 + SSR_2$ is in fact SSR of the non constrained model

$$F = \frac{(SSR_c - SSR_{nc})/k}{SSR_{nc}/(N-2k)}$$

And $F \hookrightarrow F(k, N-2k)$

- Rmk 1: this method requires the estimation of 3 models
- Rmk 2: this is the same as the formula with the *constrained* and *non constrained* R^2

UNIVERSITÉ PARIS 1
PANTHÉON SORBONNE

2nd method

We can run the Chow test by estimating one single model, using interacted variables

- Say we use the following model, for each observation i:
  $y_i = b_0 + b_1 x_{1,i} + b_2 x_{2,i} + ... + b_{k-1} x_{k-1,i} + u_i$
- Create a dummy I_i with value 1 if observation i belongs to the 1st subsample, and 0 otherwise
- For each explanatory variable $x_{j,i}$, create a variable $z_{j,i} = x_{j,i} * I_i$
- In other words, if i belongs to the 1st subsample, $z_j = x_j$ and otherwise $z_j = 0$
- Run the following "augmented" regression:
  $y_i = b_0 + b_1 x_{1,i} + b_2 x_{2,i} + ... + b_{k-1} x_{k-1,i} + a_0 I + a_1 z_{1,i} + a_2 z_{2,i} + ... + a_{k-1} z_{k-1,i} + u_i$
- Parameters are the same across the 2 subsamples if adding these new variables doesn't improve the initial model
- In other words, the Chow test amounts here to testing if the new parameters a_j are all simultaneously zero

UNIVERSITÉ PARIS 1 PANTHÉON SORBONNE

Two remarks

- The 2 methods presented here are totally equivalent
- Implicitly, this test assumes that the variance of the error term is the same in the two subsamples, while it is not always the case (see later)
- If the second subsample is too small (less observations than variables), we can't run this test because there is no way to estimate the model on the small subsample. In that case, we use the CUSUM test (see later)

Variance of error terms not constant

- Parameters could differ across 2 sub periods, so could the variance of the error terms
- If it's the case, then it can be shown that we overestimate the size of the test, i.e. we reject too often the hypothesis of homogeneity of behaviors
- The Chow test needs to be adjusted in that case
- The fact that error terms can have a non constant variance is called *heteroskedasticity*, and will be studied later

UNIVERSITÉ PARIS 1
PANTHÉON SORBONNE

Test of equality of variances

- We deal with subsamples 1 and 2, calling T_1 and T_2 their size and k the number of variables in the model
- Assume $u_1 \hookrightarrow N(0, \sigma_1^2)$ and $u_2 \hookrightarrow N(0, \sigma_2^2)$
- We assume a constant variance within each subsample, while σ_1 et σ_2 are not necessarily equal
- We thus have: $\frac{\hat{u}_1' \hat{u}_1}{\sigma_1^2} \hookrightarrow \chi^2_{T_1 - k}$ and $\frac{\hat{u}_2' \hat{u}_2}{\sigma_2^2} \hookrightarrow \chi^2_{T_2 - k}$
- We know that the ratio of two χ^2, each divided by its parameter, gives a Fisher distribution
- Under the null hypothesis H_0 where $\sigma_1 = \sigma_2$, this ratio is thus:

$$C = \frac{\hat{u}_1' \hat{u}_1}{T_1 - k} / \frac{\hat{u}_2' \hat{u}_2}{T_2 - k} = \frac{\hat{\sigma}_1^2}{\hat{\sigma}_2^2} \hookrightarrow F(T_1 - k, T_2 - k)$$

Remark

- Subsamples 1 and 2 are apparently interchangeable
- In practice, we can compute two test statistics: one > 1 and the other < 1, because there will always be one variance greater than the other
- We commonly use the statistic that is > 1 (otherwise we'll never reject H_0)

Consequence on the Chow test

- Assume we proved that $\sigma_1 \neq \sigma_2$
- Let's use the ratio $w = \frac{\sigma_1}{\sigma_2}$
- If we multiply the second sample's model by w, we manage to make the variances of the error terms the same for the two models
- Ex: $y_2 = a_2 + b_2 x_2 + u_2$ with $V(u_2) = \sigma_2^2$ becomes:
- $w.y_2 = w.a_2 + w.b_2 x_2 + w.u_2$, with the variance of the new error term: $V(w.u_2) = (\frac{\sigma_1}{\sigma_2})^2 \sigma_2^2 = \sigma_1^2$
- Since we don't know the "true" w because we don't know the "true" σ_1 and σ_2, we replace them by their estimate, computed when running the variance equality test
- I.e.: $\hat{w} = \frac{\hat{\sigma}_1}{\hat{\sigma}_2}$

UNIVERSITÉ PARIS 1 PANTHÉON SORBONNE

What if the second subsample is too small

Ex: the second subperiod is too short ($T_2 < k$). We can still run this test, with the following adjustments:

- The model is estimated on the first subperiod ($y_1 = X_1.b_1 + u_1$) to get parameters $\hat{b}_1$
- Comparing prediction $X_2.\hat{b}_1$ and observation y_2 is a way of testing whether model 1 is still valid on period 2
- More precisely, we test if the prediction error has expectancy zero
- Calling SSR_c and SSR_1 the sum of squared residuals on model constrained and 1 respectively, and T_1 and T_2 the sizes of samples 1 and 2, the following statistic can be used to run a Chow test:

$$(\frac{SSR_c - SSR_1}{T_2})/(\frac{SSR_1}{T_1 - k}) \hookrightarrow F(T_2, T_1 - k)$$

UNIVERSITÉ PARIS 1
PANTHÉON SORBONNE

Remark

- The idea of the Chow test is to compare the constrained vs. the non constrained model
- Estimating a model with more variables than observations provides non unique parameters
- If $N < k$, then the projection of y is itself, so all residuals are zero
- That's why we eventually get a formula close to the usual Chow test, where SSR_2 would be zero

UNIVERSITÉ PARIS 1
PANTHÉON SORBONNE

Choosing a breakpoint in a time series

Unless we know a specific event happened, finding a breakpoint is not obvious

- This is the purpose of the cumulated sum test
- Method: analyze the prediction of y_t computed with the parameters of the model estimated on the first $t - 1$ periods
- The test will "begin" at date t=k+1 at the earliest
- If the regression is stable with time, then the expectancy of the forecast error is 0
- We plot the graph of the cumulated sum of these so-called *recursive residuals* against time, and if the model is stable with time it must stay within some confidence bounds

For more details, see Cadoret et al., chapter 3.