Chapter 4: Statistical Hypothesis Testing

Christophe Hurlin

November 20, 2015

Section 1

Introduction

1. Introduction

The outline of this chapter is the following:

Section 2. Statistical hypothesis testing

Section 3. Tests in the multiple linear regression model

- **Subsection 3.1.** The Student test
- **Subsection 3.2.** The Fisher test

Section 4. MLE and Inference

- **Subsection 4.1.** The Likelihood Ratio (LR) test
- **Subsection 4.2.**The Wald test
- **Subsection 4.3.** The Lagrange Multiplier (LM) test

1. Introduction

References

- Amemiya T. (1985), Advanced Econometrics. Harvard University Press.
- Greene W. (2007), Econometric Analysis, sixth edition, Pearson - Prentice Hil (**recommended**)
- Ruud P., (2000) An introduction to Classical Econometric Theory, Oxford University Press.

1. Introduction

Notations: In this chapter, I will (try to...) follow some conventions of notation.

$f_Y(y)$	probability density or mass function
$F_Y(y)$	cumulative distribution function
$\Pr()$	probability
$\mathbf{y}$	vector
$\mathbf{Y}$	matrix

Be careful: in this chapter, I don't distinguish between a random vector (matrix) and a vector (matrix) of deterministic elements (except in section 2). For more appropriate notations, see:

Abadir and Magnus (2002), Notation in econometrics: a proposal for a standard, Econometrics Journal.

Section 2

Statistical hypothesis testing

2. Statistical hypothesis testing

Objectives

The objective of this section is to define the following concepts:

1. Null and alternative hypotheses
2. One-sided and two-sided tests
3. Rejection region, test statistic and critical value
4. Size, power and power function
5. Uniformly most powerful (UMP) test
6. Neyman Pearson lemma
7. Consistent test and unbiased test
8. p-value

2. Statistical hypothesis testing

Introduction

1. A statistical hypothesis **test** is a method of **making decisions** or a **rule of decision** (as concerned a statement about a **population parameter**) using the data of **sample**.

2. Statistical hypothesis tests define a procedure that **controls (fixes) the probability of incorrectly deciding** that a default position (null hypothesis) is incorrect based on how likely it would be for a set of observations to occur if the null hypothesis were true.

Introduction (cont'd)

In general we distinguish two types of tests:

1. The **parametric tests** assume that the data have come from a type of probability distribution and makes inferences about the parameters of the distribution
2. The **non-parametric tests** refer to tests that do not assume the data or population have any characteristic structure or parameters.

In this course, we only consider the parametric tests.

Introduction (cont'd)

A statistical test is based on three elements:

1. A null hypothesis and an alternative hypothesis
2. A rejection region based on a test statistic and a critical value
3. A type I error and a type II error

Introduction (cont'd)

A statistical test is based on three elements:

1. **A null hypothesis and an alternative hypothesis**
2. A rejection region based on a test statistic and a critical value
3. A type I error and a type II error

2. Statistical hypothesis testing

Definition (Hypothesis)

A **hypothesis** is a statement about a population parameter. The formal testing procedure involves a statement of the hypothesis, usually in terms of a "**null**" or maintained hypothesis and an "**alternative**," conventionally denoted H_0 and H_1, respectively.

2. Statistical hypothesis testing

Introduction

1. The null hypothesis refers to a general or default position: that there is no relationship between two measured phenomena or that a potential medical treatment has no effect.

2. The costs associated to the violation of the null must be higher than the cost of a violation of the alternative.

Example (Choice of the null hypothesis)

In a credit scoring problem, in general we have: H_0 : the client is not risky(acceptance of the loan) versus H_1 : the client is risky (refusal of the loan).

2. Statistical hypothesis testing

Definition (Simple and composite hypotheses)

A **simple hypothesis** specifies the population distribution completely. A **composite** hypothesis does not specify the population distribution completely.

Example (Simple and composite hypotheses)

If $X \sim t(\theta)$, $\mathrm{H}_0 : \theta = \theta_0$ is a simple hypothesis. $\mathrm{H}_1 : \theta > \theta_0$, $\mathrm{H}_1 : \theta < \theta_0$, and $\mathrm{H}_1 : \theta \neq \theta_0$ are composite hypotheses.

2. Statistical hypothesis testing

Definition (One-sided test)

A **one-sided test** has the general form:

$$\begin{aligned} \mathrm{H}_0 &: \theta = \theta_0 \\ \mathrm{H}_1 &: \theta < \theta_0 \end{aligned} \quad \text{or} \quad \begin{aligned} \mathrm{H}_0 : \theta = \theta_0 \\ \mathrm{H}_1 : \theta > \theta_0 \end{aligned}$$

2. Statistical hypothesis testing

Definition (Two-sided test)

A **two-sided test** has the general form:

$$\begin{aligned} H_0 &: \theta = \theta_0 \\ H_1 &: \theta \neq \theta_0 \end{aligned}$$

2. Statistical hypothesis testing

Introduction (cont'd)

A statistical test is based on three elements:

1. A null hypothesis and an alternative hypothesis
2. **A rejection region based on a test statistic and a critical value**
3. A type I error and a type II error

2. Statistical hypothesis testing

Definition (Rejection region)

The **rejection region** is the set of values of the test statistic (or equivalently the set of samples) for which the null hypothesis is rejected. The rejection region is denoted W. For example, a standard rejection region W is of the form:

$$\mathrm{W} = \{x : T(x) > c\}$$

or equivalently

$$\mathrm{W} = \{x_1, .., x_N : T(x_1, .., x_N) > c\}$$

where x denotes a sample $\{x_1, .., x_N\}$, $T(x)$ the realisation of a **test statistic** and c the **critical value**.

2. Statistical hypothesis testing

Remarks

1. A (hypothesis) test is thus a rule that specifies:
    1. For which sample values the decision is made to "**fail to reject H0**" as true;
    2. For which sample values the decision is made to "**reject H0**".
    3. **Never say "Accept H1", "fail to reject H1" etc..**
2. The complement of the rejection region is the non-rejection region.

2. Statistical hypothesis testing

Remark

The rejection region is defined as to be:

$$W = \{x : \underbrace{T(x)}_{\text{test statistic}} \lessgtr \underbrace{c}_{\text{critical value}}\}$$

$T(x)$ is the realisation of the statistic (random variable):

$$T(X) = T(X_1, .., X_N)$$

The test statistic $T(X)$ has an exact or an asymptotic distribution D under the null H_0.

$$T(X) \underset{H_0}{\sim} D \quad \text{or} \quad T(X) \overset{d}{\underset{H_0}{\to}} D$$

Introduction (cont'd)

A statistical test is based on three elements:

1. A null hypothesis and an alternative hypothesis
2. A rejection region based on a test statistic and a critical value
3. **A type I error and a type II error**

2. Statistical hypothesis testing

		Decision	
		Fail to reject H_0	Reject H_0
Truth	H_0	Correct decision	**Type I error**
	H_1	**Type II error**	Correct decision

2. Statistical hypothesis testing

Definition (Size)

The probability of a type I error is the (nominal) **size** of the test. This is conventionally denoted α and is also called the **significance level**.

$$\alpha = \Pr\left(W \vert H_0 \right)$$

Remark

For a simple null hypothesis:

$$\alpha = \Pr\left(\mathrm{W} \right| \mathrm{H}_0)$$

For a composite null hypothesis:

$$\alpha = \sup_{\theta_0 \in \mathrm{H}_0} \Pr\left(\mathrm{W} \right| \mathrm{H}_0)$$

A test is said to have level if its size is less than or equal to α.

2. Statistical hypothesis testing

Definition (Power)

The **power** of a test is the probability that it will correctly lead to rejection of a false null hypothesis:

$$\text{power} = \Pr\left(\left.\mathrm{W}\right|\mathrm{H}_1\right) = 1 - \beta$$

where β denotes the probability of type II error, i.e. $\beta = \Pr\left(\left.\overline{\mathrm{W}}\right|\mathrm{H}_1\right)$ and $\overline{\mathrm{W}}$ denotes the non-rejection region.

2. Statistical hypothesis testing

Example (Test on the mean)

Consider a sequence $X_1, .., X_N$ of $i.i.d.$ continuous random variables with $X_i \sim \mathcal{N}\left(m, \sigma^2\right)$ where σ^2 is known. We want to test

$$\begin{aligned} \mathrm{H}_0 &: \quad m = m_0 \\ \mathrm{H}_1 &: \quad m = m_1 \end{aligned}$$

with $m_1 < m_0$. An econometrician propose the following rule of decision:

$$\mathrm{W} = \left\{ x : \overline{x}_N < c \right\}$$

where $\overline{X}_N = N^{-1} \sum_{i=1}^{N} X_i$ denotes the sample mean and c is a constant (critical value). **Question:** calculate the size and the power of this test.

2. Statistical hypothesis testing

Solution

The rejection region is W$= \{x : \overline{x}_N < c\}$. Under the null $\text{H}_0 : m = m_0$:

$$\overline{X}_N \underset{\text{H}_0}{\sim} \mathcal{N}\left(m_0, \frac{\sigma^2}{N}\right)$$

So, the size of the test is equal to:

$$\begin{aligned}
\alpha &= \Pr\left(\text{W} \middle| \text{H}_0\right) \\
&= \Pr\left(\overline{X}_N < c \middle| \text{H}_0\right) \\
&= \Pr\left(\frac{\overline{X}_N - m_0}{\sigma/\sqrt{N}} < \frac{c - m_0}{\sigma/\sqrt{N}} \middle| \text{H}_0\right) \\
&= \Phi\left(\frac{c - m_0}{\sigma/\sqrt{N}}\right)
\end{aligned}$$

2. Statistical hypothesis testing

Solution (cont'd)

The rejection region is W$= \{x : \overline{x}_N < c\}$. Under the alternative $\mathrm{H}_1 : m = m_1$:

$$\overline{X}_N \underset{\mathrm{H}_1}{\sim} \mathcal{N}\left(m_1, \frac{\sigma^2}{N}\right)$$

So, the power of the test is equal to:

$$\begin{aligned}
\text{power} &= \Pr\left(\mathrm{W} | \mathrm{H}_1\right) \\
&= \Pr\left(\left.\frac{\overline{X}_N - m_1}{\sigma/\sqrt{N}} < \frac{c - m_1}{\sigma/\sqrt{N}} \right| \mathrm{H}_1\right) \\
&= \Phi\left(\frac{c - m_1}{\sigma/\sqrt{N}}\right) \quad \square
\end{aligned}$$

Solution (cont'd)

In conclusion:

$$\alpha = \Phi\left(\frac{c - m_0}{\sigma/\sqrt{N}}\right)$$

$$\beta = 1 - \text{power} = 1 - \Phi\left(\frac{c - m_1}{\sigma/\sqrt{N}}\right)$$

We have a system of two equations with three parameters: α, β (or power) and the critical value c.

1. There is a trade-off between the probabilities of the errors of type I and II, i.e. α and β : if c decreases, α decreases but β increases.
2. A solution is to impose a size α and determine the critical value and the power.

Solution (cont'd)

In order to illustrate the **tradeoff** between α and β given the critical value c, take an example with $\sigma^2 = 1$ and $N = 100$:

$$\text{H}_0 : m = m_0 = 1.2 \qquad \text{H}_1 : m = m_1 = 1$$

$$\overline{X}_N \underset{H_0}{\sim} N\left(m_0, \frac{\sigma^2}{N}\right) \qquad \overline{X}_N \underset{H_1}{\sim} N\left(m_1, \frac{\sigma^2}{N}\right)$$

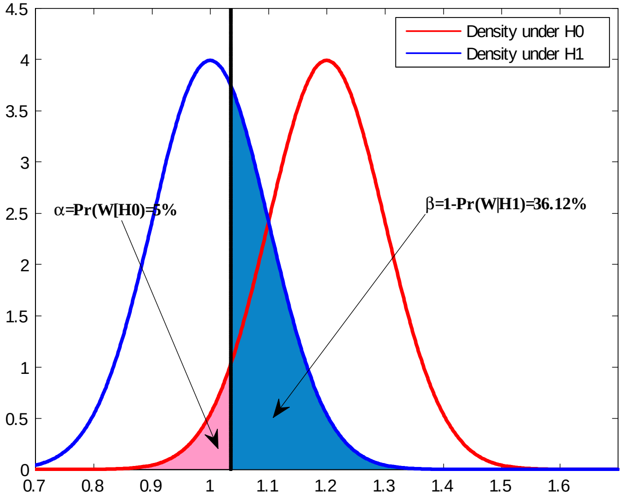
We have

$$\text{W} = \{x : \overline{x}_N < c\}$$

$$\alpha = \Pr\left(\text{W} \middle| \text{H}_0\right) = \Phi\left(\frac{c - m_0}{\sigma / \sqrt{N}}\right) = \Phi\left(10\left(c - 1.2\right)\right)$$

$$\beta = \Pr\left(\overline{\text{W}} \middle| \text{H}_1\right) = 1 - \Phi\left(\frac{c - m_1}{\sigma / \sqrt{N}}\right) = 1 - \Phi\left(10\left(c - 1\right)\right)$$

2. Statistical hypothesis testing



2. Statistical hypothesis testing

Click me!

2. Statistical hypothesis testing

Fact (Critical value)

The (nominal) **size** α *is* **fixed by the analyst** *and the critical value is deduced from* α.

2. Statistical hypothesis testing

Example (Test on the mean)

Consider a sequence $X_1, .., X_N$ of $i.i.d.$ continuous random variables with $X_i \sim \mathcal{N}\left(m, \sigma^2\right)$, $N = 100$ and $\sigma^2 = 1$. We want to test

$$\mathrm{H}_0 : m = 1.2 \qquad \mathrm{H}_1 : m = 1$$

An econometrician propose the following rule of decision:

$$\mathrm{W} = \{x : \overline{x}_N < c\}$$

where $\overline{X}_N = N^{-1} \sum_{i=1}^{N} X_i$ denotes the sample mean and c is a constant (critical value). **Questions:** (1) what is the critical value of the test of size $\alpha = 5\%$? (2) what is the power of the test?

2. Statistical hypothesis testing

Solution

We know that:

$$\alpha = \Pr\left(\left.\mathrm{W}\right|\mathrm{H}_0\right) = \Phi\left(\frac{c - m_0}{\sigma/\sqrt{N}}\right)$$

So, the critical value that corresponds to a significance level of α is:

$$c = m_0 + \frac{\sigma}{\sqrt{N}}\Phi^{-1}\left(\alpha\right)$$

NA: if $m_0 = 1.2$, $m_1 = 1$, $N = 100$, $\sigma^2 = 1$ and $\alpha = 5\%$, then the rejection region is

$$\mathrm{W} = \left\{x : \overline{x}_N < 1.0355\right\}$$

Solution (cont'd)

$$W = \left\{ x : \overline{x}_N < m_0 + \frac{\sigma}{\sqrt{N}} \Phi^{-1}(\alpha) \right\}$$

The power of the test is:

$$\text{power} = \Pr\left(\left. W \right| H_1 \right) = \Phi\left(\frac{c - m_1}{\sigma / \sqrt{N}} \right)$$

Given the critical value, we have:

$$\text{power} = \Phi\left(\frac{m_0 - m_1}{\sigma / \sqrt{N}} + \Phi^{-1}(\alpha) \right) \quad \square$$

NA: if $m_0 = 1.2$, $m_1 = 1$, $N = 100$, $\sigma^2 = 1$ and $\alpha = 5\%$:

$$\text{power} = \Phi\left(\frac{1.2 - 1}{1/\sqrt{100}} + \Phi^{-1}(0.05) \right) = 0.6388 \quad \square$$

2. Statistical hypothesis testing

Example (Test on the mean)

Consider a sequence $X_1, .., X_N$ of *i.i.d.* continuous random variables with $X_i \sim \mathcal{N}\left(m, \sigma^2\right)$ with $\sigma^2 = 1$ and $N = 100$. We want to test

$$\mathrm{H}_0 : m = 1.2 \qquad \mathrm{H}_1 : m = 1$$

The rejection region for a significance level $\alpha = 5\%$ is:

$$\mathrm{W} = \left\{x : \overline{x}_N < 1.0355\right\}$$

where $\overline{X}_N = N^{-1} \sum_{i=1}^{N} X_i$ denotes the sample mean. **Question:** if the realisation of the sample mean is equal to 1.13, what is the conclusion of the test?

2. Statistical hypothesis testing

Solution (cont'd)

For a nominal size $\alpha = 5\%$, the rejection region is:

$$\mathrm{W} = \{x : \overline{x}_N < 1.0355\}$$

If we observe

$$\overline{x}_N = 1.13$$

This realisation does not belong to the rejection region:

$$\overline{x}_N \notin \mathrm{W}$$

For a level of 5%, we do not reject the null hypothesis $\mathrm{H}_0 : m = 1.2$. $\square$

2. Statistical hypothesis testing

Definition (Power function)

In general, the alternative hypothesis is composite. In this case, the power is a **function** of the value of the parameter under the alternative.

$$\text{power} = P(\theta) \qquad \forall \theta \in \text{H}_1$$

2. Statistical hypothesis testing

Example (Test on the mean)

Consider a sequence $X_1, .., X_N$ of $i.i.d.$ continuous random variables with $X_i \sim \mathcal{N}\left(m, \sigma^2\right)$ where σ^2 is known. We want to test

$$\begin{aligned} \mathrm{H}_0 &: \quad m = m_0 \\ \mathrm{H}_1 &: \quad m < m_0 \end{aligned}$$

Consider the following rule of decision:

$$\mathrm{W} = \left\{ x : \overline{x}_N < m_0 + \frac{\sigma}{\sqrt{N}} \Phi^{-1}\left(\alpha\right) \right\}$$

Questions: What is the power function of the test?

Solution

As in the previous case, we have:

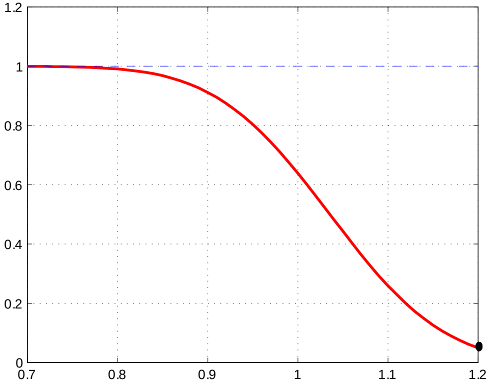
$$\text{power} = P(m) = \Phi\left(\frac{m_0 - m}{\sigma/\sqrt{N}} + \Phi^{-1}(\alpha)\right) \quad \text{with } m < m_0$$

NA: if $m_0 = 1.2$, $N = 100$, $\sigma^2 = 1$ and $\alpha = 5\%$.

$$P(m) = \Phi\left(\frac{1.2 - m}{1/10} - 1.6449\right) \quad \text{with } m < m_0$$

2. Statistical hypothesis testing

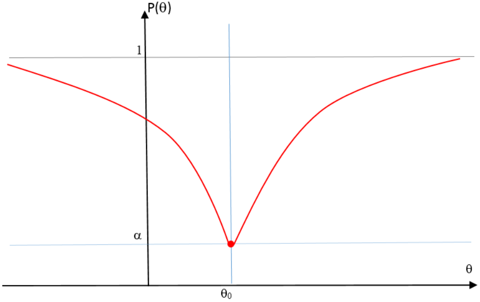
Power function $P(m)$



2. Statistical hypothesis testing

Example (Power function)

Consider a test $H_0 : \theta = \theta_0$ versus $H_1 : \theta \neq \theta_0$, the power function has this general form:



2. Statistical hypothesis testing

Definition (Most powerful test)

A test (denoted A) is **uniformly most powerful (UMP)** if it has greater power than any other test of the same size for all admissible values of the parameter.

$$\alpha_A = \alpha_B = \alpha$$

$$\beta_A \leq \beta_B$$

for any test B of size α.

UMP tests

How to derive the rejection region of the UMP test of size α ?

$=>$ the **Neyman–Pearson lemma**

2. Statistical hypothesis testing

Lemma (Neyman Pearson)

Consider a hypothesis test between two point hypotheses $H_0 : \theta = \theta_0$ *and* $H_1 : \theta = \theta_1$. *The* **uniformly most powerful (UMP)** *test has a rejection region defined by:*

$$W = \left\{ x \middle| \frac{L_N(\theta_0; x)}{L_N(\theta_1; x)} < K \right\}$$

where $L_N(\theta_0; x)$ *denotes the likelihood of the sample* x *and* K *is a constant determined by the size* α *such that:*

$$\Pr\left(\frac{L_N(\theta_0; X)}{L_N(\theta_1; X)} < K \middle| H_0 \right) = \alpha$$

2. Statistical hypothesis testing

Example (Test on the mean)

Consider a sequence $X_1, .., X_N$ of $i.i.d.$ continuous random variables with $X_i \sim \mathcal{N}\left(m, \sigma^2\right)$ where σ^2 is known. We want to test

$$\begin{aligned} \mathrm{H}_0 &: \quad m = m_0 \\ \mathrm{H}_1 &: \quad m = m_1 \end{aligned}$$

with $m_1 > m_0$. **Question:** What is the rejection region of the UMP test of size α?

Solution

Since $X_1, .., X_N$ are $\mathcal{N}.i.d.\left(m, \sigma^2\right)$, the likelihood of the sample $\{x_1, .., x_N\}$ is defined as to be (cf. chapter 2):

$$L_N\left(\theta; x\right) = \frac{1}{\sigma^N \left(2\pi\right)^{N/2}} \exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{N}\left(x_i - m\right)^2\right)$$

Given the Neyman Pearson lemma the rejection region of the UMP test of size α is given by:

$$\frac{L_N\left(\theta_0; x\right)}{L_N\left(\theta_1; x\right)} < K$$

where K is a constant determined by the size α.

2. Statistical hypothesis testing

Solution (cont'd)

$$\frac{\frac{1}{\sigma^N (2\pi)^{N/2}} \exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^N (x_i - m_0)^2\right)}{\frac{1}{\sigma^N (2\pi)^{N/2}} \exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^N (x_i - m_1)^2\right)} < K$$

This expression can rewritten as:

$$\exp\left(\frac{1}{2\sigma^2}\left(\sum_{i=1}^N (x_i - m_1)^2 - \sum_{i=1}^N (x_i - m_0)^2\right)\right) < K$$

$$\Longleftrightarrow \sum_{i=1}^N (x_i - m_1)^2 - \sum_{i=1}^N (x_i - m_0)^2 < K_1$$

where $K_1 = 2\sigma^2 \ln(K)$ is a constant.

Solution (cont'd)

$$\sum_{i=1}^{N}\left(x_i - m_1\right)^2 - \sum_{i=1}^{N}\left(x_i - m_0\right)^2 < K_1$$

$$\Longleftrightarrow 2\left(m_0 - m_1\right)\sum_{i=1}^{N} x_i + N\left(m_1^2 - m_0^2\right) < K_1$$

$$\Longleftrightarrow \left(m_0 - m_1\right)\sum_{i=1}^{N} x_i < K_2$$

where $K_2 = \left(K_1 - N\left(m_1^2 - m_0^2\right)\right)/2$ is a constant.

Solution (cont'd)

$$\left(m_0 - m_1\right) \sum_{i=1}^{N} x_i < K_2$$

Since $m_1 > m_0$, we have

$$\frac{1}{N} \sum_{i=1}^{N} x_i > K_3$$

where $K_3 = K_2 / \left(N \left(m_0 - m_1\right)\right)$ is a constant.

The rejection region of the UMP test for $\mathrm{H}_0 : m = m_0$ against $\mathrm{H}_0 : m = m_1$ with $m_1 > m_0$ has the general form:

$$\mathrm{W} = \left\{ x : \overline{x}_N > A \right\}$$

where A is a constant.

Solution (cont'd)

$$\mathrm{W} = \{x : \overline{x}_N > A\}$$

Determine the critical value A from the nominal size:

$$\begin{aligned}
\alpha &= \Pr\left(\mathrm{W} \middle| \mathrm{H}_0\right) \\
&= \Pr\left(\overline{x}_N > A \middle| \mathrm{H}_0\right) \\
&= 1 - \Pr\left(\frac{\overline{X}_N - m_0}{\sigma/\sqrt{N}} < \frac{A - m_0}{\sigma/\sqrt{N}} \middle| \mathrm{H}_0\right) \\
&= 1 - \Phi\left(\frac{A - m_0}{\sigma/\sqrt{N}}\right)
\end{aligned}$$

Solution (cont'd)

$$\alpha = 1 - \Phi\left(\frac{A - m_0}{\sigma/\sqrt{N}}\right)$$

So, we have

$$A = m_0 + \frac{\sigma}{\sqrt{N}}\Phi^{-1}(1-\alpha)$$

The rejection region of the UMP test of size α for $\text{H}_0 : m = m_0$ against $\text{H}_0 : m = m_1$ with $m_1 > m_0$ is:

$$\text{W} = \left\{x : \overline{x}_N > m_0 + \frac{\sigma}{\sqrt{N}}\Phi^{-1}(1-\alpha)\right\} \quad \square$$

2. Statistical hypothesis testing

Fact (UMP one-sided test)

For a **one-sided test**

$$H_0 : \theta = \theta_0 \quad \textit{against} \quad H_1 : \theta > \theta_0 \quad (\textit{or } H_1 : \theta < \theta_1)$$

the rejection region W of the UMP test is equivalent to the rejection region obtained for the test

$$H_0 : \theta = \theta_0 \quad \textit{against} \quad H_1 : \theta = \theta_1$$

with for $\theta_1 > \theta_0$ *(or* $\theta_1 < \theta_0$*) if this region does not depend on the value of* θ_1.

2. Statistical hypothesis testing

Example (Test on the mean)

Consider a sequence $X_1, .., X_N$ of $i.i.d.$ continuous random variables with $X_i \sim \mathcal{N}\left(m, \sigma^2\right)$ where σ^2 is known. We want to test

$$
\begin{aligned}
\mathrm{H}_0 &: \quad m = m_0 \\
\mathrm{H}_1 &: \quad m > m_0
\end{aligned}
$$

Question: What is the rejection region of the UMP test of size α?

2. Statistical hypothesis testing

Solution

Consider the test:

$$\begin{aligned} H_0 &: \quad m = m_0 \\ H_1 &: \quad m = m_1 \end{aligned}$$

with $m_1 > m_0$. The rejection region of the UMP test of size α is:

$$W = \left\{ x : \overline{x}_N > m_0 + \frac{\sigma}{\sqrt{N}} \Phi^{-1} \left(1 - \alpha \right) \right\}$$

W does not depend on m_1. It is also the rejection region of the UMP one-sided test for

$$\begin{aligned} H_0 &: \quad m = m_0 \\ H_1 &: \quad m > m_0 \quad \square \end{aligned}$$

2. Statistical hypothesis testing

Fact (Two-sided test)

For a two-sided test

$$H_0 : \theta = \theta_0 \quad \text{against} \quad H_1 : \theta \neq \theta_0$$

the non rejection region $\overline{W}$ of the test of size α is the intersection of the non rejection regions of the corresponding one-sided UMP tests of **size $\alpha/2$**

$$\text{Test A:} \quad H_0 : \theta = \theta_0 \quad \text{against} \quad H_1 : \theta > \theta_0$$

$$\text{Test B:} \quad H_0 : \theta = \theta_0 \quad \text{against} \quad H_1 : \theta < \theta_0$$

So, we have:

$$\overline{W} = \overline{W}_A \cap \overline{W}_B$$

Example (Test on the mean)

Consider a sequence $X_1, .., X_N$ of $i.i.d.$ continuous random variables with $X_i \sim \mathcal{N}\left(m, \sigma^2\right)$ where σ^2 is known. We want to test

$$\begin{aligned} \mathrm{H}_0 &: \quad m = m_0 \\ \mathrm{H}_1 &: \quad m \neq m_0 \end{aligned}$$

Question: What is the rejection region of the test of size α?

Solution

Consider the one-sided tests:

$$\text{Test A:} \quad \mathrm{H}_0 : m = m_0 \ \text{ against } \ \mathrm{H}_1 : m < m_0$$

$$\text{Test B:} \quad \mathrm{H}_0 : m = m_0 \ \text{ against } \ \mathrm{H}_1 : m > m_0$$

The rejection regions of the UMP test of size $\alpha/2$ are:

$$\mathrm{W}_{\mathrm{A}} = \left\{ x : \overline{x}_N < m_0 + \frac{\sigma}{\sqrt{N}} \Phi^{-1}\left(\alpha/2\right) \right\}$$

$$\mathrm{W}_{\mathrm{B}} = \left\{ x : \overline{x}_N > m_0 + \frac{\sigma}{\sqrt{N}} \Phi^{-1}\left(1 - \alpha/2\right) \right\}$$

2. Statistical hypothesis testing

Solution (cont'd)

The **non-rejection regions** of the UMP test of size $\alpha/2$ are:

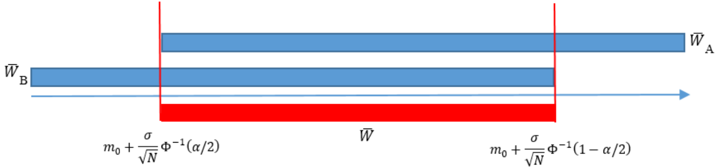
$$\overline{\mathrm{W}}_{\mathrm{A}} = \left\{ x : \overline{x}_N \geq m_0 + \frac{\sigma}{\sqrt{N}} \Phi^{-1}\left(\alpha/2\right) \right\}$$

$$\overline{\mathrm{W}}_{\mathrm{B}} = \left\{ x : \overline{x}_N \leq m_0 + \frac{\sigma}{\sqrt{N}} \Phi^{-1}\left(1-\alpha/2\right) \right\}$$

The non rejection region of the two-sided test corresponds to the intersection of these two regions:

$$\overline{W} = \overline{W}_A \cap \overline{W}_B$$

Solution (cont'd)



So, non rejection region of the two-sided test of size α is:

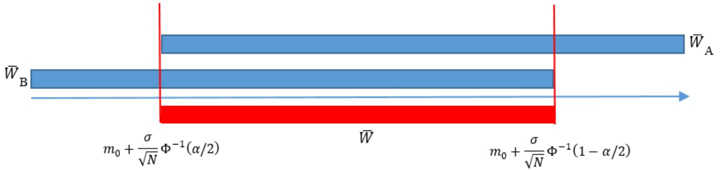
$$\overline{\mathrm{W}} = \left\{ x : m_0 + \frac{\sigma}{\sqrt{N}} \Phi^{-1}\left(\alpha/2\right) \leq \overline{x}_N \leq m_0 + \frac{\sigma}{\sqrt{N}} \Phi^{-1}\left(1-\alpha/2\right) \right\}$$

Since, $\Phi^{-1}\left(\alpha/2\right) = -\Phi^{-1}\left(1-\alpha/2\right)$, this region can be rewritten as:

$$\overline{\mathrm{W}} = \left\{ x : \left| \overline{x}_N - m_0 \right| \leq \frac{\sigma}{\sqrt{N}} \Phi^{-1}\left(1-\alpha/2\right) \right\}$$

2. Statistical hypothesis testing

Solution (cont'd)



$$\overline{W} = \left\{ x : \left| \overline{x}_N - m_0 \right| \leq \frac{\sigma}{\sqrt{N}} \Phi^{-1} \left(1 - \alpha/2 \right) \right\}$$

Finally, the **rejection region** of the two-sided test of size α is:

$$W = \left\{ x : \left| \overline{x}_N - m_0 \right| > \frac{\sigma}{\sqrt{N}} \Phi^{-1} \left(1 - \alpha/2 \right) \right\} \quad \square$$

Solution (cont'd)

$$\mathrm{W}=\left\{x:\left|\overline{x}_N-m_0\right|>\frac{\sigma}{\sqrt{N}}\Phi^{-1}\left(1-\alpha/2\right)\right\}$$

NA: if $m_0=1.2$, $N=100$, $\sigma^2=1$ and $\alpha=5\%$:

$$\mathrm{W}=\left\{x:\left|\overline{x}_N-1.2\right|>\frac{1}{10}\Phi^{-1}\left(0.975\right)\right\}$$

$$\mathrm{W}=\left\{x:\left|\overline{x}_N-1.2\right|>0.1960\right\}$$

If the realisation of $\left|\overline{x}_N-1.2\right|$ is larger than 0.1960, we reject the null $\mathrm{H}_0: m=1.2$ for a significance level of 5%.

Definition (Unbiased Test)

A test is **unbiased** if its power $P(\theta)$ is greater than or equal to its size α for all values of the parameter θ.

$$P(\theta) \geq \alpha \quad \forall \theta \in \mathrm{H}_1$$

By construction, we have $P(\theta_0) = \Pr\left(\mathrm{W} | \, H|_0\right) = \alpha$.

2. Statistical hypothesis testing

Definition (Consistent Test)

A test is **consistent** if its power goes to one as the sample size grows to infinity.

$$\lim_{N \to \infty} P(\theta) = 1 \quad \forall \theta \in \mathrm{H}_1$$

2. Statistical hypothesis testing

Example (Test on the mean)

Consider a sequence $X_1, .., X_N$ of $i.i.d.$ continuous random variables with $X_i \sim \mathcal{N}\left(m, \sigma^2\right)$ where σ^2 is known. We want to test

$$\begin{aligned} \mathrm{H}_0 &: \quad m = m_0 \\ \mathrm{H}_1 &: \quad m < m_0 \end{aligned}$$

The rejection region of the UMP test of size α is

$$\mathrm{W} = \left\{ x : \overline{x}_N < m_0 + \frac{\sigma}{\sqrt{N}} \Phi^{-1}\left(\alpha\right) \right\}$$

Question: show that this test is (1) unbiased and (2) consistent.

Solution

$$\mathrm{W} = \left\{ x : \overline{x}_N < m_0 + \frac{\sigma}{\sqrt{N}} \Phi^{-1}(\alpha) \right\}$$

The power function of the test is defined as to be:

$$\begin{aligned} P(m) &= \Pr(\mathrm{W} | \mathrm{H}_1) \\ &= \Pr\left(\overline{X}_N < m_0 + \frac{\sigma}{\sqrt{N}} \Phi^{-1}(\alpha) \middle| m < m_0 \right) \\ &= \Phi\left(\frac{m_0 - m}{\sigma / \sqrt{N}} + \Phi^{-1}(\alpha) \right) \end{aligned}$$

2. Statistical hypothesis testing

Solution

$$P\left(m\right)=\Phi\left(\frac{m_0-m}{\sigma/\sqrt{N}}+\Phi^{-1}\left(\alpha\right)\right)\quad\forall m<m_0$$

The test is consistent since:

$$\lim_{N\rightarrow\infty}P\left(m\right)=1$$

The test is unbiased since

$$P\left(m\right)\geq\alpha\qquad\forall m<m_0$$

$$\lim_{m\rightarrow m_0}P\left(m\right)=\Phi\left(\Phi^{-1}\left(\alpha\right)\right)=\alpha\ \square$$

Solution

- The decision "Reject H0" or "fail to reject H0" is not so informative!
- Indeed, there is some "arbitrariness" to the choice of α (level).
- Another strategy is to ask, for every α, whether the test rejects at that level.
- Another alternative is to use the so-called **p-value**—the smallest level of significance at which H_0 would be rejected given the value of the test-statistic.

2. Statistical hypothesis testing

Definition (p-value)

Suppose that for every $\alpha \in [0,1]$, one has a size α test with rejection region W_α. Then, the **p-value** is defined to be:

$$\text{p-value} = \inf \left\{ \alpha : T\left(y\right) \in W_\alpha \right\}$$

The p-value is the smallest level at which one can reject H_0.

2. Statistical hypothesis testing

The p-value is a **measure of evidence against** H_0:

p-value	evidence
< 0.01	Very strong evidence against H_0
$0.01 - 0.05$	Strong evidence against H_0
$0.05 - 0.10$	Weak evidence against H_0
> 0.10	Little or no evidence against H_0

2. Statistical hypothesis testing

Remarks

1. A large p-value does not mean "strong evidence in favor of H0".
2. A large p-value can occur for two reasons:
    1. H0 is true;
    2. H0 is false but the test has low power.
3. The p-value is not the probability that the null hypothesis is true!

2. Statistical hypothesis testing

Dependent Variable: RMSFT
Method: Least Squares
Date: 11/09/13 Time: 21:53
Sample(adjusted): 2 2363
Included observations: 2362 after adjusting endpoints

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.000274	0.000179	1.532829	$\boxed{0.1255}$
RSP500	1.125056	0.025371	44.34419	$\boxed{0.0000}$

R-squared	0.454513	Mean dependent var	0.000617
Adjusted R-squared	0.454282	S.D. dependent var	0.011753
S.E. of regression	0.008682	Akaike info criterion	-6.654227
Sum squared resid	0.177900	Schwarz criterion	-6.649343
Log likelihood	7860.642	F-statistic	1966.407
Durbin-Watson stat	2.028898	Prob(F-statistic)	0.000000

For a nominal size of 5%, we **reject** the null $H_0 : \beta_{SP500} = 0$.

For a nominal size of 5%, we **fail to reject** the null $H_0 : \beta_C = 0$.

2. Statistical hypothesis testing

Summary

Hypothesis testing is defined by the following general procedure

Step 1: State the relevant null and alternative hypotheses (misstating the hypotheses muddies the rest of the procedure!);

Step 2: Consider the statistical assumptions being made about the sample in doing the test (independence, distributions, etc.)—incorrect assumptions mean that the test is invalid!

Step 3: Choose the appropriate test (exact or asymptotic tests) and thus state the relevant test statistic (say, T).

2. Statistical hypothesis testing

Summary (cont'd)

Step 4: Derive the distribution of the test statistic under the null hypothesis (sometimes it is well-known, sometimes it is more tedious!)—for example, the Student t-distribution or the Fisher distribution.

Step 5: Determine the critical value (and thus the critical region).

Step 6: Compute (using the observations!) the observed value of the test statistic T, say t_{obs}.

Step 7: Decide to either fail to reject the null hypothesis or reject in favor of the alternative assumption—the decision rule is to reject the null hypothesis H_0 if the observed value of the test statistic, t_{obs} is in the critical region, and to "fail to reject" the null hypothesis otherwise

2. Statistical hypothesis testing

Key concepts

1. Null and alternative hypotheses
2. Simple and composite hypotheses
3. One-sided and two-sided tests
4. Rejection region, test statistic and critical value
5. Type I and type II errors
6. Size, power and power function
7. Uniformly most powerful (UMP) test
8. Neyman Pearson lemma
9. Consistent test and unbiased test
10. p-value

Section 3

Tests in the multiple linear regression model

3. Tests in the multiple linear regression model

Objectives

In the context of the multiple linear regression model (cf. chapter 3), the objective of this section is to present :

1. the Student test
2. the t-statistic and the z-statistic
3. the Fisher test
4. the global F-test
5. To distinguish the case with normality assumption and the case without any assumption on the distribution of the error term (semi-parametric specification)

3. Tests in the multiple linear regression model

Be careful: in this section, I don't distinguish between a random vector (matrix) and a vector (matrix) of deterministic elements. For more appropriate notations, see:

Abadir and Magnus (2002), Notation in econometrics: a proposal for a standard, Econometrics Journal.

3. Tests in the multiple linear regression model

Model

Consider the (population) **multiple linear regression model**:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$$

where (cf. chapter 3):

- $\mathbf{y}$ is a $N \times 1$ vector of observations y_i for $i = 1, .., N$
- $\mathbf{X}$ is a $N \times K$ matrix of K explicative variables $\mathbf{x}_{ik}$ for $k = 1, ., K$ and $i = 1, .., N$
- $\boldsymbol{\varepsilon}$ is a $N \times 1$ vector of error terms ε_i.
- $\boldsymbol{\beta} = (\beta_1 .. \beta_K)^\top$ is a $K \times 1$ vector of parameters

Assumptions

Fact (Assumptions)

We assume that the multiple linear regression model satisfy the assumptions A1-A5 (cf. chapter 3)

We distinguish two cases:

1. **Case 1:** assumption A6 (Normality) holds and $\varepsilon \sim \mathcal{N}\left(\mathbf{0}, \sigma^2 \mathbf{I}_N\right)$
2. **Case 2:** the distribution of ε is unknown (semi-parametric specification) and $\varepsilon \sim ??$

Parametric tests

The β_k are unknown features of the population, but:

1. One can formulate a **hypothesis about their value**;
2. One can construct a test statistic with a known **finite sample distribution** (case 1) or an **asymptotic distribution** (case 2);
3. One can take a "decision" meaning "reject H0" if the value of the test statistic is too unlikely.

3. Tests in the multiple linear regression model

Three tests of interest:

$$\begin{aligned} \mathrm{H}_0 &: \beta_k = a_k \\ \mathrm{H}_1 &: \beta_k < a_k \end{aligned} \qquad \text{or} \qquad \begin{aligned} \mathrm{H}_0 : \beta_k = a_k \\ \mathrm{H}_1 : \beta_k > a_k \end{aligned}$$

$$\begin{aligned} \mathrm{H}_0 &: \beta_k = a_k \\ \mathrm{H}_1 &: \beta_k \neq a_k \end{aligned}$$

$$\begin{aligned} \mathrm{H}_0 &: \mathbf{R}\boldsymbol{\beta} = \mathbf{q} \\ \mathrm{H}_1 &: \mathbf{R}\boldsymbol{\beta} \neq \mathbf{q} \end{aligned}$$

where $a_k = 0$ or $a_k \neq 0$.

3. Tests in the multiple linear regression model

For that, we introduce two types of test

1. The **Student** test or t-test
2. The **Fisher** test of F-test

Subsection 3.1

The Student test

Case 1: Normality assumption A6

3.1. The Student test

Assumption 6 (normality): the disturbances are normally distributed.

$$\boldsymbol{\varepsilon} | \, \mathbf{X} \sim \mathcal{N}\left(\mathbf{0}_{N \times 1}, \sigma^2 \mathbf{I}_N\right)$$

3.1. The Student test

Reminder (cf. chapter 3)

Fact (Linear regression model)

Under the assumption A6 (normality), the estimators $\widehat{\beta}$ *and* $\widehat{\sigma}^2$ *have a finite sample distribution given by:*

$$\widehat{\beta} \sim \mathcal{N}\left(\beta, \sigma^2 \left(\mathbf{X}^\top \mathbf{X}\right)^{-1}\right)$$

$$\frac{\widehat{\sigma}^2}{\sigma^2}\left(N-K\right) \sim \chi^2\left(N-K\right)$$

Moreover, $\widehat{\beta}$ *and* $\widehat{\sigma}^2$ *are independent. This result holds whether or not the matrix* $\mathbf{X}$ *is considered as random. In this last case, the distribution of* $\widehat{\beta}$ *is conditional to* $\mathbf{X}$.

3.1. The Student test

Remarks

1. Any linear combination of $\widehat{\beta}$ is also normally distributed:

$$\mathbf{A}\widehat{\beta} \sim \mathcal{N}\left(\mathbf{A}\beta,\sigma^2\mathbf{A}\left(\mathbf{X}^\top\mathbf{X}\right)^{-1}\mathbf{A}^\top\right)$$

2. Any subset of $\widehat{\beta}$ has a joint normal distribution.

$$\widehat{\beta}_k \sim \mathcal{N}\left(\beta_k, \sigma^2 m_{kk}\right)$$

where m_{kk} is k^{th} diagonal element of $\left(\mathbf{X}^\top\mathbf{X}\right)^{-1}$.

3.1. The Student test

Reminder

If X and Y are two independent random variables such that

$$X \sim \mathcal{N}(0,1)$$

$$Y \sim \chi^2(\theta)$$

then the variable Z defined as to be

$$Z = \frac{X}{\sqrt{Y/\theta}}$$

has a Student's t-distribution with θ degrees of freedom

$$Z \sim t_{(\theta)}$$

3.1. The Student test

Student test statistic

Consider a test with the null:

$$\mathrm{H}_0 : \beta_k = a_k$$

Under the null H_0 :

$$\frac{\widehat{\beta}_k - a_k}{\sigma\sqrt{m_{kk}}} \underset{\mathrm{H}_0}{\sim} \mathcal{N}(0,1)$$

$$\frac{\widehat{\sigma}^2}{\sigma^2}(N-K) \underset{\mathrm{H}_0}{\sim} \chi^2(N-K)$$

and these two variables are independent...

Student test statistic (cont'd)

$$\frac{\widehat{\beta}_k - a_k}{\sigma\sqrt{m_{kk}}} \underset{H_0}{\sim} \mathcal{N}(0,1)$$

$$\frac{\widehat{\sigma}^2}{\sigma^2}(N-K) \underset{H_0}{\sim} \chi^2(N-K)$$

So, under the null H_0 we have:

$$\frac{\frac{\widehat{\beta}_k - a_k}{\sigma\sqrt{m_{kk}}}}{\sqrt{\frac{\widehat{\sigma}^2}{\sigma^2}\frac{(N-K)}{(N-K)}}} = \frac{\widehat{\beta}_k - a_k}{\widehat{\sigma}\sqrt{m_{kk}}} \underset{H_0}{\sim} t_{(N-K)}$$

3.1. The Student test

Definition (Student t-statistic)

Under the null $H_0 : \beta_k = a_k$, the **Student test-statistic** or **t-statistic** is defined to be:

$$\mathrm{T}_k = \frac{\widehat{\beta}_k - a_k}{\widehat{\mathrm{se}}\left(\widehat{\beta}_k\right)} \underset{H_0}{\sim} t_{(N-K)}$$

where N is the number of observations, K is the number of explanatory variables (including the constant term), $t_{(N-K)}$ is the Student t-distribution with $N-K$ degrees of freedom and

$$\widehat{\mathrm{se}}\left(\widehat{\beta}_k\right) = \widehat{\sigma}\sqrt{m_{kk}}$$

with m_{kk} is k^{th} diagonal element of $\left(\mathbf{X}^\top\mathbf{X}\right)^{-1}$.

3.1. The Student test

Remarks

1. Under the assumption A6 (normality) and under the null $H_0 : \beta_k = a_k$, the Student test-statistic has an **exact (finite sample)** distribution.

$$\mathrm{T}_k \underset{H_0}{\sim} t_{(N-K)}$$

2. The term $\widehat{\mathrm{se}}\left(\widehat{\beta}_k\right)$ denotes the estimator of the standard error of the OLS estimator $\widehat{\beta}_k$ and it corresponds to the square root of the k^{th} diagonal element of $\widehat{\mathbb{V}}\left(\widehat{\boldsymbol{\beta}}\right)$ (cf. chapter 3):

$$\widehat{\mathbb{V}}\left(\widehat{\boldsymbol{\beta}}\right) = \widehat{\sigma}^2 \left(\mathbf{X}^\top \mathbf{X}\right)^{-1}$$

3.1. The Student test

Consider the **one-sided test**:

$$\begin{aligned} \mathrm{H}_0 &: \beta_k = a_k \\ \mathrm{H}_1 &: \beta_k < a_k \end{aligned}$$

The rejection region is defined as to be:

$$\mathrm{W} = \{ y : \mathrm{T}_k(y) < A \}$$

where A is a constant determined by the nominal size α.

$$\alpha = \Pr\left(\mathrm{W} \middle| \mathrm{H}_0\right) = \Pr\left(\mathrm{T}_k(y) < A \middle| \mathrm{T}_k \underset{\mathrm{H}_0}{\sim} t_{(N-K)}\right)$$

3.1. The Student test

$$\alpha = \Pr\left(\mathrm{T}_k \left(y \right) < A \middle| \, \mathrm{T}_k \underset{\mathrm{H}_0}{\sim} t_{(N-K)} \right) = F_{N-K} \left(A \right)$$

where $F_{N-K} \left(. \right)$ denotes the cdf of the Student's t-distribution with $N - K$ degrees of freedom. Denote c_α the α-quantile of this distribution:.

$$A = F_{N-K}^{-1} \left(\alpha \right) = c_\alpha$$

The rejection region of the test of size α is defined as to be:

$$\mathrm{W} = \left\{ y : \mathrm{T}_k \left(y \right) < c_\alpha \right\}$$

3.1. The Student test

Definition (One-sided Student test)

The **critical region** of the Student test is that $\mathrm{H}_0 : \beta_k = a_k$ is rejected in favor of $H_1 : \beta_k < a_k$ at the α (say, 5%) significance level if:

$$\mathrm{W} = \{ y : \mathrm{T}_k (y) < c_\alpha \}$$

where c_α is the α (say, 5%) critical value of a Student t-distribution with $N - K$ degrees of freedom and $\mathrm{T}_k (y)$ is the realisation of the Student test-statistic.

3.1. The Student test

Example (One-sided test)

Consider the CAPM model (cf. chapter 1) and the following results (Eviews). We want to test the beta of MSFT as

$$H_0 : \beta_{MSFT} = 1 \quad \text{against} \quad H_1 : \beta_{MSFT} < 1$$

Question: give a conclusion for a nominal size of 5%.

Dependent Variable: RMSFT
Method: Least Squares
Date: 11/30/13 Time: 17:15
Sample: 2 21
Included observations: 20

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.001189	0.001205	0.986860	0.3368
RSP500	1.989787	0.314210	6.332664	0.0000

R-squared	0.690203	Mean dependent var	-0.000180
Adjusted R-squared	0.672992	S.D. dependent var	0.009272
S.E. of regression	0.005302	Akaike info criterion	-7.546873
Sum squared resid	0.000506	Schwarz criterion	-7.447300
Log likelihood	77.46873	F-statistic	40.10263
Durbin-Watson stat	1.955366	Prob(F-statistic)	0.000006

3.1. The Student test

Solution

Step 1: compute the t-statistic

$$\mathrm{T}_{MSFT}\left(y\right)=\frac{\widehat{\beta}_{MSFT}-1}{\widehat{\mathrm{se}}\left(\widehat{\beta}_{MSFT}\right)}=\frac{1.9898-1}{0.3142}=3.1501$$

Step 2: Determine the rejection region for a nominal size $\alpha = 5\%$.

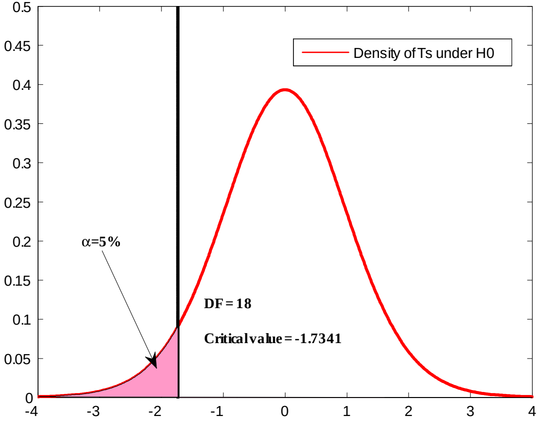
$$\mathrm{T}_{MSFT}\underset{\mathrm{H}_0}{\sim}t_{(20-2)}$$

$$\mathrm{W}=\left\{y:\mathrm{T}_k\left(y\right)<-1.7341\right\}$$

Conclusion: for a significance level of 5%, we fail to reject the null $\mathrm{H}_0 : \beta_{MSFT} = 1$ against $\mathrm{H}_1 : \beta_{MSFT} < 1$ $\square$

3.1. The Student test

Solution (cont'd)



3.1. The Student test

Consider the **one-sided test**

$$\begin{aligned} \mathrm{H}_0 &: \beta_k = a_k \\ \mathrm{H}_1 &: \beta_k > a_k \end{aligned}$$

The rejection region is defined as to be:

$$\mathrm{W} = \{ y : \mathrm{T}_k(y) > A \}$$

where A is a constant determined by the nominal size α.

$$\alpha = \Pr(\mathrm{W} | \mathrm{H}_0) = \Pr\left(\mathrm{T}_k(y) > A \middle| \mathrm{T}_k \underset{\mathrm{H}_0}{\sim} t_{(N-K)} \right)$$

3.1. The Student test

$$\alpha = 1 - \Pr\left(\mathrm{T}_k\left(y\right) < A \middle| \mathrm{T}_k \underset{\mathrm{H}_0}{\sim} t_{(N-K)} \right)$$

or equivalently

$$1 - \alpha = F_{N-K}\left(A\right)$$

where $F_{N-K}\left(.\right)$ denotes the cdf of the Student's t-distribution with $N-K$ degrees of freedom. Denote $c_{1-\alpha}$ the $1-\alpha$ quantile of this distribution:

$$A = F_{N-K}^{-1}\left(1-\alpha\right) = c_{1-\alpha}$$

The rejection region of the test of size α is defined as to be:

$$\mathrm{W} = \left\{ y : \mathrm{T}_k\left(y\right) > c_{1-\alpha} \right\}$$

3.1. The Student test

Definition (One-sided Student test)

The **critical region** of the Student test is that $\text{H}_0 : \beta_k = a_k$ is rejected in favor of $H_1 : \beta_k > a_k$ at the α (say, 5%) significance level if:

$$\text{W} = \{y : \text{T}_k(y) > c_{1-\alpha}\}$$

where $c_{1-\alpha}$ is the $1-\alpha$ (say, 95%) critical value of a Student t-distribution with $N-K$ degrees of freedom and $\text{T}_k(y)$ is the realisation of the Student test-statistic.

3.1. The Student test

Example (One-sided test)

Consider the CAPM model (cf. chapter 1) and the following results (Eviews). We want to test the beta of MSFT as

$$H_0 : \beta_{MSFT} = 1 \quad \text{against} \quad H_1 : \beta_{MSFT} > 1$$

Question: give a conclusion for a nominal size of 5%.

Dependent Variable: RMSFT
Method: Least Squares
Date: 11/30/13 Time: 17:15
Sample: 2 21
Included observations: 20

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.001189	0.001205	0.986860	0.3368
RSP500	1.989787	0.314210	6.332664	0.0000

R-squared	0.690203	Mean dependent var	-0.000180
Adjusted R-squared	0.672992	S.D. dependent var	0.009272
S.E. of regression	0.005302	Akaike info criterion	-7.546873
Sum squared resid	0.000506	Schwarz criterion	-7.447300
Log likelihood	77.46873	F-statistic	40.10263
Durbin-Watson stat	1.955366	Prob(F-statistic)	0.000006

3.1. The Student test

Solution

Step 1: compute the t-statistic

$$\mathrm{T}_{MSFT}\left(y\right)=\frac{\widehat{\beta}_{MSFT}-1}{\widehat{\mathrm{se}}\left(\widehat{\beta}_{MSFT}\right)}=\frac{1.9898-1}{0.3142}=3.1501$$

Step 2: Determine the rejection region for a nominal size $\alpha = 5\%$.

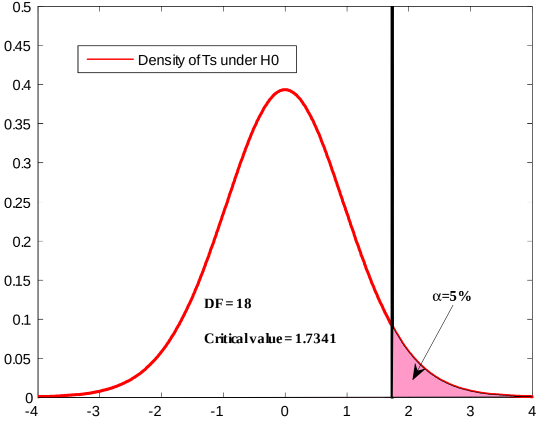
$$\mathrm{T}_{MSFT}\underset{\mathrm{H}_0}{\sim}t_{(20-2)}$$

$$\mathrm{W}=\left\{y:\mathrm{T}_k\left(y\right)>1.7341\right\}$$

Conclusion: for a significance level of 5%, we reject the null $\mathrm{H}_0:\beta_{MSFT}=1$ against $\mathrm{H}_1:\beta_{MSFT}>1$ $\square$

3.1. The Student test

Solution (cont'd)



3.1. The Student test

Consider the **two-sided test**

$$\begin{aligned} \mathrm{H}_0 &: \beta_k = a_k \\ \mathrm{H}_1 &: \beta_k \neq a_k \end{aligned}$$

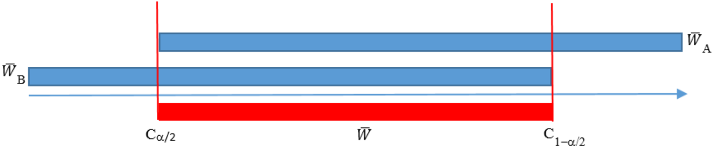
The non-rejection region is defined as the intersection of the two non-rejection regions of the one-sided test of level $\alpha/2$:

$$\overline{\mathrm{W}} = \overline{\mathrm{W}}_A \cap \overline{\mathrm{W}}_B$$

$$\mathrm{H}_0 : \beta_k = a_k \;\; \text{against} \;\; \mathrm{H}_1 : \beta_k < a_k \qquad \overline{\mathrm{W}}_A = \left\{ y : \mathrm{T}_k\left(y\right) > c_{\alpha/2} \right\}$$

$$\mathrm{H}_0 : \beta_k = a_k \;\; \text{against} \;\; \mathrm{H}_1 : \beta_k > a_k \qquad \overline{\mathrm{W}}_B = \left\{ y : \mathrm{T}_k\left(y\right) < c_{1-\alpha/2} \right\}$$

3.1. The Student test



$$\overline{\mathrm{W}} = \left\{ y : c_{\alpha/2} < \mathrm{T}_k\left(y\right) < c_{1-\alpha/2} \right\}$$

Since the Student's t-distribution is symmetric, $c_{\alpha/2} = -c_{1-\alpha/2}$

$$\overline{\mathrm{W}} = \left\{ y : -c_{1-\alpha/2} < \mathrm{T}_k\left(y\right) < c_{1-\alpha/2} \right\}$$

The rejection region is then defined as to be:

$$\mathrm{W} = \left\{ y : \left|\mathrm{T}_k\left(y\right)\right| > c_{1-\alpha/2} \right\}$$

3.1. The Student test

Definition (Two-sided Student test)

The **critical region** of the Student test is that $\mathrm{H}_0 : \beta_k = a_k$ is rejected in favor of $H_1 : \beta_k \neq a_k$ at the α (say, 5%) significance level if:

$$\overline{\mathrm{W}} = \left\{ y : \left| \mathrm{T}_k \left(y \right) \right| > c_{1-\alpha/2} \right\}$$

where $c_{1-\alpha/2}$ is the $1 - \alpha/2$ (say, 97.5%) critical value of a Student t-distribution with $N - K$ degrees of freedom and $\mathrm{T}_k \left(y \right)$ is the realisation of the Student test-statistic.

3.1. The Student test

Example (One-sided test)

Consider the CAPM model (cf. chapter 1) and the following results (Eviews). We want to test the beta of MSFT as

$$\text{H}_0 : \beta_{MSFT} = 1 \quad \text{against} \quad \text{H}_1 : \beta_{MSFT} \neq 1$$

Question: give a conclusion for a nominal size of 5%.

Dependent Variable: RMSFT
Method: Least Squares
Date: 11/30/13 Time: 17:15
Sample: 2 21
Included observations: 20

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.001189	0.001205	0.986860	0.3368
RSP500	1.989787	0.314210	6.332664	0.0000

R-squared	0.690203	Mean dependent var	-0.000180
Adjusted R-squared	0.672992	S.D. dependent var	0.009272
S.E. of regression	0.005302	Akaike info criterion	-7.546873
Sum squared resid	0.000506	Schwarz criterion	-7.447300
Log likelihood	77.46873	F-statistic	40.10263
Durbin-Watson stat	1.955366	Prob(F-statistic)	0.000006

3.1. The Student test

Solution

Step 1: compute the t-statistic

$$\mathrm{T}_{MSFT}(y) = \frac{\widehat{\beta}_{MSFT} - 1}{\widehat{\mathrm{se}}\left(\widehat{\beta}_{MSFT}\right)} = \frac{1.9898 - 1}{0.3142} = 3.1501$$

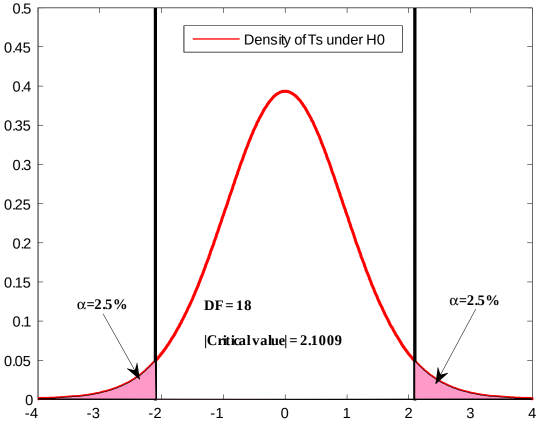
Step 2: Determine the rejection region for a nominal size $\alpha = 5\%$.

$$\mathrm{T}_{MSFT} \underset{\mathrm{H}_0}{\sim} t_{(20-2)}$$

$$\mathrm{W} = \{y : |\mathrm{T}_k(y)| > 2.1009\}$$

Conclusion: for a significance level of 5%, we reject the null $\mathrm{H}_0 : \beta_{MSFT} = 1$ against $\mathrm{H}_1 : \beta_{MSFT} \neq 1$ $\square$

3.1. The Student test



3.1. The Student test

Rejection regions

H_0	H_1	**Rejection region**
$\beta_k = a_k$	$\beta_k > a_k$	$\mathrm{W} = \{y : \mathrm{T}_k(y) > c_{1-\alpha}\}$
$\beta_k = a_k$	$\beta_k < a_k$	$\mathrm{W} = \{y : \mathrm{T}_k(y) < c_{\alpha}\}$
$\beta_k = a_k$	$\beta_k \neq a_k$	$\mathrm{W} = \{y : \lvert\mathrm{T}_k(y)\rvert > c_{1-\alpha/2}\}$

where c_β denotes the β-quantile (critical value) of the Student t-distribution with $N-K$ degrees of freedom.

3.1. The Student test

Definition (P-values)

The **p-values** of Student tests are equal to:

$$\text{Two-sided test: } \text{p-value} = 2 \times F_{N-K}\left(-\left|\mathrm{T}_k\left(y\right)\right|\right)$$

$$\text{Right tailed test: } \text{p-value} = 1 - F_{N-K}\left(\mathrm{T}_k\left(y\right)\right)$$

$$\text{Left tailed test: } \text{p-value} = F_{N-K}\left(-\mathrm{T}_k\left(y\right)\right)$$

where $\mathrm{T}_k\left(y\right)$ is the realisation of the Student test-statistic and $F_{N-K}\left(.\right)$ the cdf of the Student's t-distribution with $N-K$ degrees of freedom.

3.1. The Student test

Example (One-sided test)

Consider the previous CAPM model. We want to test:

$$\text{H}_0 : c = 0 \quad \text{against} \quad \text{H}_1 : c \neq 0$$

$$\text{H}_0 : \beta_{MSFT} = 0 \quad \text{against} \quad \text{H}_1 : \beta_{MSFT} \neq 0$$

Question: find the corresponding p-values.

Dependent Variable: RMSFT
Method: Least Squares
Date: 11/30/13 Time: 18:45
Sample: 2 21
Included observations: 20

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.001189	0.001205	0.986860	
RSP500	1.989787	0.314210	6.332664	

R-squared	0.690203	Mean dependent var	-0.000180
Adjusted R-squared	0.672992	S.D. dependent var	0.009272
S.E. of regression	0.005302	Akaike info criterion	-7.546873
Sum squared resid	0.000506	Schwarz criterion	-7.447300
Log likelihood	77.46873	F-statistic	40.10263
Durbin-Watson stat	1.955366	Prob(F-statistic)	0.000006

3.1. The Student test

Solution

Since we consider two-sided tests with $N = 20$ and $K = 2$:

$$\text{p-value}_c = 2 \times F_{18}\left(-\left|T_c\left(y\right)\right|\right) = 2 \times F_{18}\left(-0.9868\right) = 0.3368$$

$$\text{p-value}_c = 2 \times F_{18}\left(-\left|T_{MSFT}\left(y\right)\right|\right) = 2 \times F_{18}\left(-6.3326\right) = 5.7e^{-006}$$

Dependent Variable: RMSFT
Method: Least Squares
Date: 11/30/13 Time: 18:51
Sample: 2 21
Included observations: 20

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.001189	0.001205	0.986860	$\boxed{0.3368}$
RSP500	1.989787	0.314210	6.332664	$\boxed{0.0000}$

R-squared	0.690203	Mean dependent var	-0.000180
Adjusted R-squared	0.672992	S.D. dependent var	0.009272
S.E. of regression	0.005302	Akaike info criterion	-7.546873
Sum squared resid	0.000506	Schwarz criterion	-7.447300
Log likelihood	77.46873	F-statistic	40.10263
Durbin-Watson stat	1.955366	Prob(F-statistic)	0.000006

3.1. The Student test

Fact (Student test with large sample)

For a **large sample** *size N*

$$T_k \underset{H_0}{\sim} t_{(N-K)} \approx \mathcal{N}(0,1)$$

Then, the rejection region for a Student two-sided test becomes

$$W = \left\{ y : |T_k(y)| > \Phi^{-1}(1-\alpha/2) \right\}$$

where $\Phi(.)$ *denotes the cdf of the standard normal distribution. For* $\alpha = 5\%$, $\Phi^{-1}(0.975) = 1.96$, *so we have:*

$$W = \{ y : |T_k(y)| > 1.96 \}$$

Case 2: Semi-parametric model

3.1. The Student test

Assumption 6 (normality): the distribution of the disturbances is unknown, but satisfy (assumptions A1-A5):

$$\mathbb{E}\left(\left.\boldsymbol{\varepsilon}\right|\mathbf{X}\right)=\mathbf{0}_{N\times 1}$$

$$\mathbb{V}\left(\left.\boldsymbol{\varepsilon}\right|\mathbf{X}\right)=\sigma^2\mathbf{I}_N$$

3.1. The Student test

Problem

1. The exact (finite sample) distribution of $\widehat{\beta}_k$ and $\widehat{\sigma}^2$ are unknown.
2. As a consequence the **finite sample distribution** of $\mathrm{T}_k(y)$ is also **unknown**.
3. But, we can use the asymptotic properties of the OLS estimators (cf. chapter 3). In particular, we have:

$$\sqrt{N}\left(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}\right) \xrightarrow{d} \mathcal{N}\left(\mathbf{0}, \sigma^2 \mathbf{Q}^{-1}\right)$$

where

$$\mathbf{Q} = p\lim \frac{1}{N}\mathbf{X}^\top \mathbf{X} = \mathbb{E}_X\left(\mathbf{x}_i \mathbf{x}_i^\top\right)$$

3.1. The Student test

Definition (Z-statistic)

Under the null $H_0 : \beta_k = a_k$, if the assumptions A1-A5 hold (cf. chapter 3), the **z-statistic** defined by

$$Z_k = \frac{\widehat{\beta}_k - a_k}{\widehat{\text{se}}_{asy}\left(\widehat{\beta}_k\right)} \overset{d}{\underset{H_0}{\longrightarrow}} \mathcal{N}(0,1)$$

where $\widehat{\text{se}}_{asy}\left(\widehat{\beta}_k\right) = \widehat{\sigma}\sqrt{m_{kk}}$ denotes the estimator of the asymptotic standard error of the estimator $\widehat{\beta}_k$ and m_{kk} is k^{th} diagonal element of $\left(\mathbf{X}^\top \mathbf{X}\right)^{-1}$.

3.1. The Student test

Rejection regions

The rejection regions have the same form as for the t-test (except for the distribution)

H_0	H_1	**Rejection region**
$\beta_k = a_k$	$\beta_k > a_k$	$W = \left\{ y : Z_k(y) > \Phi^{-1}(1-\alpha) \right\}$
$\beta_k = a_k$	$\beta_k < a_k$	$W = \left\{ y : Z_k(y) < \Phi^{-1}(\alpha) \right\}$
$\beta_k = a_k$	$\beta_k \neq a_k$	$W = \left\{ y : \lvert Z_k(y) \rvert > \Phi^{-1}(1-\alpha/2) \right\}$

where $\Phi(.)$ denotes the cdf of the standard normal distribution.

3.1. The Student test

Definition (P-values)

The **p-values** of the Z-tests are equal to:

$$\text{Two-sided test:} \;\; \text{p-value} = 2 \times \Phi\left(-\left|Z_k\left(y\right)\right|\right)$$

$$\text{right tailed test:} \;\; \text{p-value} = 1 - \Phi\left(Z_k\left(y\right)\right)$$

$$\text{left tailed test:} \;\; \text{p-value} = \Phi\left(-Z_k\left(y\right)\right)$$

where $Z_k\left(y\right)$ is the realisation of the Z-statistic and $\Phi\left(.\right)$ the cdf of the standard normal distribution.

Summary

	Normality Assumption	Non Assumption
Test-statistic	**t-statistic**	**z-statistic**
Definition	$\mathrm{T}_k = \frac{\widehat{\beta}_k - a_k}{\widehat{\sigma}\sqrt{m_{kk}}}$	$\mathrm{Z}_k = \frac{\widehat{\beta}_k - a_k}{\widehat{\sigma}\sqrt{m_{kk}}}$
Exact distribution	$\mathrm{T}_k \underset{\mathrm{H}_0}{\sim} t_{(N-K)}$	—
Asymptotic distribution	—	$\mathrm{Z}_K \underset{\mathrm{H}_0}{\overset{d}{\rightarrow}} \mathcal{N}(0,1)$

3.1. The Student test

Dependent Variable: RMSFT
Method: Least Squares
Date: 11/30/13 Time: 18:51
Sample: 2 21
Included observations: 20

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.001189	0.001205	0.986860	0.3368
RSP500	1.989787	0.314210	6.332664	0.0000

R-squared	0.690203	Mean dependent var	-0.000180
Adjusted R-squared	0.672992	S.D. dependent var	0.009272
S.E. of regression	0.005302	Akaike info criterion	-7.546873
Sum squared resid	0.000506	Schwarz criterion	-7.447300
Log likelihood	77.46873	F-statistic	40.10263
Durbin-Watson stat	1.955366	Prob(F-statistic)	0.000006

Dependent Variable: Y
Method: ML - Binary Probit
Date: 11/24/13 Time: 18:33
Sample: 1 190
Included observations: 190
Convergence achieved after 3 iterations
Covariance matrix computed using second derivatives

Variable	Coefficient	Std. Error	z-Statistic	Prob.
X	0.215364	0.092715	2.322847	0.0202
C	-0.215364	0.092715	-2.322847	0.0202

Mean dependent var	0.421053	S.D. dependent var	0.495032
S.E. of regression	0.489246	Akaike info criterion	1.353695
Sum squared resid	45.00000	Schwarz criterion	1.387874
Log likelihood	-126.6010	Hannan-Quinn criter.	1.367540
Restr. log likelihood	-129.3196	Avg. log likelihood	-0.666321
LR statistic (1 df)	5.437219	McFadden R-squared	0.021022
Probability(LR stat)	0.019712		

Subsection 3.2

The Fisher test

3.2. The Fisher test

Consider the two-sided test associated to p **linear constraints** on the parameters β_k :

$$\begin{aligned} \mathrm{H}_0 &: \mathbf{R}\beta = \mathbf{q} \\ \mathrm{H}_1 &: \mathbf{R}\beta \neq \mathbf{q} \end{aligned}$$

where $\mathbf{R}$ is a $p \times K$ matrix and $\mathbf{q}$ is a $p \times 1$ vector.

3.2. The Fisher test

Example (Linear constraints)

If $K = 4$ and if we want to test $\mathrm{H}_0 : \beta_1 + \beta_2 = 0$ and $\beta_2 - 3\beta_3 = 4$, then we have $p = 2$ linear constraints with:

$$\underset{(2\times4)}{\mathbf{R}} \; \underset{(4,1)}{\boldsymbol{\beta}} = \underset{(2\times1)}{\mathbf{q}}$$

$$\begin{pmatrix} 1 & 1 & 0 & 0 \\ 0 & 1 & -3 & 0 \end{pmatrix} \begin{pmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \\ \beta_4 \end{pmatrix} = \begin{pmatrix} 0 \\ 4 \end{pmatrix}$$

3.2. The Fisher test

Example (Linear constraints)

If $K = 4$ and if we want to test $\text{H}_0 : \beta_2 = \beta_3 = \beta_4 = 0$, then we have $p = 3$ linear constraints with:

$$\underset{(3\times4)}{\mathbf{R}} \; \underset{(4,1)}{\boldsymbol{\beta}} = \underset{(3\times1)}{\mathbf{q}}$$

$$\begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \\ \beta_4 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$$

Case 1: Normality assumption A6

3.2. The Fisher test

Definition (Fisher test-statistic)

Under assumptions A1-A6 (cf. chapter 3), the **Fisher test-statistic** is defined as to be:

$$\mathrm{F} = \frac{1}{p}\left(\mathbf{R}\widehat{\beta} - \mathbf{q}\right)^{\top}\left(\widehat{\sigma}^{2}\mathbf{R}\left(\mathbf{X}^{\top}\mathbf{X}\right)^{-1}\mathbf{R}^{\top}\right)^{-1}\left(\mathbf{R}\widehat{\beta} - \mathbf{q}\right)$$

where $\widehat{\beta}$ denotes the OLS estimator. Under the null $\mathrm{H}_0 : \mathbf{R}\beta = \mathbf{q}$, the F-statistic has a Fisher exact (finite sample) distribution

$$\mathrm{F} \underset{\mathrm{H}_0}{\sim} F_{(p,N-K)}$$

3.2. The Fisher test

Reminder

If X and Y are two independent random variables such that

$$X \sim \chi^2(\theta_1)$$

$$Y \sim \chi^2(\theta_2)$$

then the variable Z defined by

$$Z = \frac{X/\theta_1}{Y/\theta_2}$$

has a Fisher distribution with θ_1 and θ_2 degrees of freedom

$$Z \sim F_{(\theta_1,\theta_2)}$$

Proof

Under assumption A6, we have the following (conditional to $\mathbf{X}$) distribution

$$\widehat{\boldsymbol{\beta}} \sim \mathcal{N}\left(\boldsymbol{\beta},\sigma^2\left(\mathbf{X}^\top\mathbf{X}\right)^{-1}\right)$$

$$\frac{\widehat{\sigma}^2}{\sigma^2}\left(N-K\right) \sim \chi^2\left(N-K\right)$$

Proof (cont'd)

Consider the vector $\mathbf{m} = \mathbf{R}\widehat{\beta} - \mathbf{q}$. Under the null

$$\mathrm{H}_0 : \mathbf{R}\beta = \mathbf{q}$$

We have

$$\mathbb{E}(\mathbf{m}) = \mathbf{R}\mathbb{E}\left(\widehat{\beta}\right) - \mathbf{q} = \mathbf{R}\beta - \mathbf{q} = \mathbf{0}$$

$$\begin{aligned} \mathbb{V}(\mathbf{m}) &= \mathbb{E}\left(\left(\mathbf{R}\widehat{\beta} - \mathbf{q}\right)\left(\mathbf{R}\widehat{\beta} - \mathbf{q}\right)^{\top}\right) \\ &= \mathbf{R}\mathbb{V}\left(\widehat{\beta}\right)\mathbf{R}^{\top} \\ &= \sigma^2 \mathbf{R}\left(\mathbf{X}^{\top}\mathbf{X}\right)^{-1}\mathbf{R}^{\top} \end{aligned}$$

3.2. The Fisher test

Proof (cont'd)

We can base the test of H_0 on the **Wald criterion**:

$$\underset{(1\times 1)}{W} = \underset{(1\times p)}{\mathbf{m}}^{\top} \Big(\mathbb{V}\underset{p\times p}{(\mathbf{m})}\Big)^{-1} \underset{p\times 1}{\mathbf{m}}$$

$$= \left(\mathbf{R}\widehat{\beta}-\mathbf{q}\right)^{\top}\left(\sigma^2\mathbf{R}\left(\mathbf{X}^{\top}\mathbf{X}\right)^{-1}\mathbf{R}^{\top}\right)^{-1}\left(\mathbf{R}\widehat{\beta}-\mathbf{q}\right)$$

Under assumption A6 (normality)

$$W \underset{\mathrm{H}_0}{\sim} \chi^2(p)$$

$$\frac{\widehat{\sigma}^2}{\sigma^2}(N-K) \sim \chi^2(N-K)$$

These two variables are independent.

3.2. The Fisher test

Proof (cont'd)

$$W \underset{H_0}{\sim} \chi^2(p)$$

$$\frac{\widehat{\sigma}^2}{\sigma^2}(N-K) \sim \chi^2(N-K)$$

So, the ratio of these two variables has a Fisher distribution

$$\mathrm{F} = \frac{\frac{W}{p}}{\frac{\widehat{\sigma}^2}{\sigma^2}\frac{(N-K)}{(N-K)}} \underset{H_0}{\sim} F_{(p,N-K)}$$

3.2. The Fisher test

Proof (cont'd)

$$\mathrm{F} = \frac{\left(\mathbf{R}\widehat{\beta} - \mathbf{q}\right)^{\top} \left(\sigma^2 \mathbf{R} \left(\mathbf{X}^{\top}\mathbf{X}\right)^{-1} \mathbf{R}^{\top}\right)^{-1} \left(\mathbf{R}\widehat{\beta} - \mathbf{q}\right) / p}{\frac{\widehat{\sigma}^2}{\sigma^2} \left(N - K\right) / \left(N - K\right)}$$

After simplification, the F-statistic is defined by:

$$\mathrm{F} = \frac{1}{p} \left(\mathbf{R}\widehat{\beta} - \mathbf{q}\right)^{\top} \left(\widehat{\sigma}^2 \mathbf{R} \left(\mathbf{X}^{\top}\mathbf{X}\right)^{-1} \mathbf{R}^{\top}\right)^{-1} \left(\mathbf{R}\widehat{\beta} - \mathbf{q}\right)$$

Under the null $\mathrm{H}_0 : \mathbf{R}\beta = \mathbf{q}$ **:**

$$\mathrm{F} \underset{\mathrm{H}_0}{\sim} F_{(p, N-K)} \quad \square$$

3.2. The Fisher test

Definition (Fisher test-statistic)

Under assumptions A1-A6 (cf. chapter 3), the **Fisher test-statistic** can be defined as a function of the SSR of the constrained (H_0) and unconstrained model (H_1):

$$\mathrm{F} = \left(\frac{SSR_0 - SSR_1}{SSR_1} \right) \left(\frac{N-K}{p} \right)$$

where SSR_0 denotes the sum of squared residuals of the constrained model estimated under H_0 and SSR_1 denotes the sum of squared residuals of the unconstrained model estimated under H_1.

3.2. The Fisher test

Definition (Fisher test-statistic)

Under assumptions A1-A6 (cf. chapter 3), the **Fisher test-statistic** can be defined as to be:

$$F = \frac{1}{\widehat{\sigma}^2 p} \left(\widehat{\beta}_{H_1} - \widehat{\beta}_{H_0} \right)^\top \left(\mathbf{X}^\top \mathbf{X} \right) \left(\widehat{\beta}_{H_1} - \widehat{\beta}_{H_0} \right)$$

where $\widehat{\beta}_{H_0}$ denotes the OLS estimator obtained in the constrained model (under H_0) and $\widehat{\beta}_{H_1}$ denotes the OLS estimator obtained in the unconstrained model (under H_1).

3.2. The Fisher test

Definition (Constrained OLS estimator)

Under suitable regularity conditions, the constrained OLS estimator $\widehat{\boldsymbol{\beta}}_C$ of $\boldsymbol{\beta}$, obtained under the constraint $\mathbf{R}\boldsymbol{\beta} = \mathbf{q}$, is given by:

$$\widehat{\boldsymbol{\beta}}_C = \widehat{\boldsymbol{\beta}}_{UC} - \left(\mathbf{X}^\top \mathbf{X}\right)^{-1} \mathbf{R}^\top \left(\mathbf{R}\left(\mathbf{X}^\top \mathbf{X}\right)^{-1} \mathbf{R}^\top\right)^{-1} \left(\mathbf{R}\widehat{\boldsymbol{\beta}}_{UC} - \mathbf{q}\right)$$

where $\widehat{\boldsymbol{\beta}}_{UC}$ is the unconstrained OLS estimator.

3.2. The Fisher test

Example (Fisher test and CAPM model)

Consider the extended CAPM model (file: Chapter4_data.xls):

$$r_{MSFT,t} = \beta_1 + \beta_2 r_{SP500,t} + \beta_3 r_{Ford,t} + \beta_4 r_{GE,t} + \varepsilon_t$$

where $r_{MSFT,t}$ is the excess return for Microsoft, $r_{SP500,t}$ for the SP500, $r_{Ford,t}$ for Ford and $r_{GE,t}$ for general electric. We want to test the following linear constraints:

$$H_0 : \beta_2 = 1 \text{ and } \beta_3 = \beta_4$$

Question: write a Matlab code to compute the F-statistic according to the three alternative definitions.

Solution

In this problem, the null $H_0 : \beta_2 = 1$ and $\beta_3 = \beta_4$ can be written as:

$$\underset{(2\times4)}{\mathbf{R}}\ \underset{(4,1)}{\boldsymbol{\beta}} = \underset{(2\times1)}{\mathbf{q}}$$

$$\begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & -1 \end{pmatrix} \begin{pmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \\ \beta_4 \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$$

3.2. The Fisher test

```
%==================================================
% PURPOSE: Chapter 4 Inference - Exercise - Figures 13 and 14
% Lecture: "Advanced Econometrics", HEC Lausanne
%------------------------
% Author: Christophe Hurlin, University of Orleans
% Version: v1. November 2013
%==================================================

clear all ; clc ; close all

data=xlsread('Chapter4_data.xls');
r_MSFT=data(:,1);                   % Excess return for MSFT
r_SP500=data(:,2);                  % Excess return for SP500
r_Ford=data(:,3);                   % Excess return for Ford
r_GE=data(:,4);                     % Excess return for GE
T=length(r_MSFT);                   % Sample size

% Estimation under H1
X=[ones(T,1) r_SP500 r_Ford r_GE];  % Matrix of explicative variables
y=r_MSFT;                           % Dependent variable
beta=X\y;                           % OLS estimator (H1)
res=y-X*beta;                       % Residuals
SSR1=sum(res.^2);                   % SSR of unconstrained model
var_eps=SSR1/(T-4);                 % Estimated variance
disp('beta under H1'),disp(beta')
```

3.2. The Fisher test

```
% Estimation under H0
R=[0 1 0 0 ; 0 0 1 -1];                    % Matrix R
q=[1 ; 0];                                 % Vector q
beta_H0=beta-inv(X'*X)*R'*inv(R*inv(X'*X)*R')*(R*beta-q);
res_H0=y-X*beta_H0;                        % Residuals
SSR0=sum(res_H0.^2);                       % SSR of the constrained model
disp('beta under H0'),disp(beta_H0')

% Fisher test-statistic: first definiton
F1=(1/2)*(R*beta-q)'*inv(var_eps*R*inv(X'*X)*R')*(R*beta-q);
% Fisher test-statistic: second definiton
F2=(SSR0-SSR1)/SSR1*(T-4)/2;
% Fisher test-statistic: second definiton
F3=(1/(2*var_eps))*(beta-beta_H0)'*(X'*X)*(beta-beta_H0);
disp('Fisher test statistics')
disp([F1 F2 F3])
```

3.2. The Fisher test

```
beta under H1
    0.0012    2.7619    0.3131   -0.1391

beta under H0
    0.0007    1.0000    0.4949    0.4949

Fisher test statistics
    4.3406    4.3406    4.3406
```

3.2. The Fisher test

Consider the **Fisher test**

$$\begin{aligned} \mathrm{H}_0 &: \mathbf{R}\beta = \mathbf{q} \\ \mathrm{H}_1 &: \mathbf{R}\beta \neq \mathbf{q} \end{aligned}$$

Since the Fisher test-statistic is always positive, the rejection region is defined as to be:

$$\mathrm{W} = \{ y : \mathrm{F}(y) > A \}$$

where A is a constant determined by the nominal size α.

$$\alpha = \Pr(\mathrm{W} | \mathrm{H}_0) = \Pr\left(\mathrm{F}(y) > A \middle| \mathrm{F} \underset{\mathrm{H}_0}{\sim} F_{(p, N-K)} \right)$$

3.2. The Fisher test

$$\alpha = \Pr\left(\mathrm{W} \middle| \mathrm{H}_0\right) = \Pr\left(\mathrm{F}\left(y\right) > A \middle| \mathrm{F} \underset{\mathrm{H}_0}{\sim} F_{(p,N-K)}\right)$$

or equivalently

$$\alpha = 1 - \Pr\left(\mathrm{F}\left(y\right) < A \middle| \mathrm{F} \underset{\mathrm{H}_0}{\sim} F_{(p,N-K)}\right)$$

Denote $d_{1-\alpha}$ the $1-\alpha$ quantile of the Fisher distribution with p and $N-K$ degrees of freedom.

$$A = d_{1-\alpha}$$

The rejection region of the test of size α is defined as to be:

$$\mathrm{W} = \left\{ y : \mathrm{F}\left(y\right) > d_{1-\alpha} \right\}$$

3.2. The Fisher test

Definition (Rejection region of a Fisher test)

The **critical region** of the Fisher test is that $\text{H}_0 : \mathbf{R}\beta = \mathbf{q}$ is rejected in favor of $H_1 : \mathbf{R}\beta \neq \mathbf{q}$ at the α (say, 5%) significance level if:

$$\text{W} = \{y : \text{F}(y) > d_{1-\alpha}\}$$

where $d_{1-\alpha}$ is the $1-\alpha$ critical value (say 95%) of the Fisher distribution with p and $N-K$ degrees of freedom and $\text{F}_k(y)$ is the realisation of the Fisher test-statistic.

3.2. The Fisher test

Example (Fisher test and CAPM model)

Consider the extended CAPM model (file: Chapter4_data.xls):

$$r_{MSFT,t} = \beta_1 + \beta_2 r_{SP500,t} + \beta_3 r_{Ford,t} + \beta_4 r_{GE,t} + \varepsilon_t$$

where $r_{MSFT,t}$ is the excess return for Microsoft, $r_{SP500,t}$ for the SP500, $r_{Ford,t}$ for Ford and $r_{GE,t}$ for general electric. We want to test the following linear constraints:

$$H_0 : \beta_2 = 1 \text{ and } \beta_3 = \beta_4$$

Question: given the realisation of the Fisher test-statistic (cf. previous example), conclude for a significance level $\alpha = 5\%$.

Solution

Step 1: compute the F-statistic (cf. Matlab code)

$$\mathrm{F}(y) = 4.3406$$

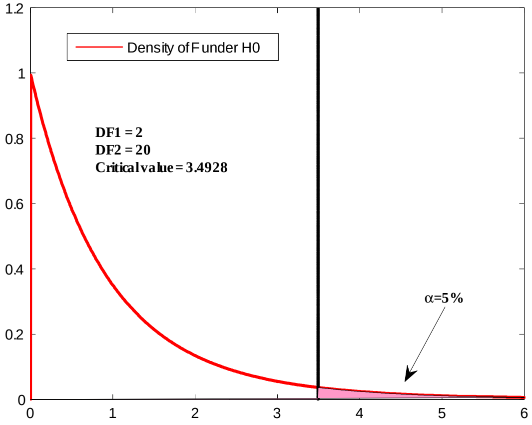
Step 2: Determine the rejection region for a nominal size $\alpha = 5\%$ for $N = 24$, $K = 4$ and $p = 2$

$$\mathrm{F} \underset{\mathrm{H}_0}{\sim} F_{(2,20)}$$

$$\mathrm{W} = \{y : \mathrm{F}(y) > 3.4928\}$$

Conclusion: for a significance level of 5%, we reject the null $\mathrm{H}_0 : \mathbf{R}\beta = \mathbf{q}$ against $\mathrm{H}_1 : \mathbf{R}\beta \neq \mathbf{q}$ $\square$

3.2. The Fisher test



3.2. The Fisher test

Definition (Student test-statistic and Fisher test-statistic)

Consider the test

$$H_0 : \beta_k = a_k \quad \text{versus} \quad H_1 : \beta_k \neq a_k$$

the **Fisher test-statistic** corresponds to the squared of the corresponding **Student's test-statistic**

$$F = T_k^2$$

Proof

Consider the test $\text{H}_0 : \beta_k = a_k$ against $\text{H}_1 : \beta_k \neq a_k$, then we have:

$$\mathbf{R} = \begin{pmatrix} 0 & 0 & .. & \underset{k^{th} \text{ position}}{1} & 0 & 0 \end{pmatrix}$$

$$q = a_k$$

As a consequence :

$$\mathbf{R}\widehat{\beta} - q = \widehat{\beta}_k - a_k$$

$$\widehat{\sigma}^2 \mathbf{R} \left(\mathbf{X}^\top \mathbf{X} \right)^{-1} \mathbf{R}^\top = \widehat{\mathbb{V}} \left(\widehat{\beta}_k \right)$$

3.2. The Fisher test

Proof (cont'd)

So, for a test $\mathrm{H}_0 : \beta_k = a_k$ against $\mathrm{H}_1 : \beta_k \neq a_k$, the Fisher test-statistic becomes

$$\mathrm{F} = \left(\mathbf{R}\widehat{\boldsymbol{\beta}} - q\right)^{\top} \left(\widehat{\sigma}^2 \mathbf{R} \left(\mathbf{X}^{\top}\mathbf{X}\right)^{-1} \mathbf{R}^{\top}\right)^{-1} \left(\mathbf{R}\widehat{\boldsymbol{\beta}} - q\right)$$

So, we have:

$$\mathrm{F} = \frac{\left(\widehat{\beta}_k - a_k\right)^2}{\widehat{\mathbb{V}}\left(\widehat{\beta}_k\right)}$$

and the F test-statistic is equal to the squared t-statistic:

$$\mathrm{F} = \mathrm{T}_k^2 \quad \square$$

3.2. The Fisher test

Definition (P-values)

The **p-value** of the F-test is equal to:

$$\text{p-value} = 1 - F_{p,N-K}\left(\mathrm{F}\left(y\right)\right)$$

where $\mathrm{F}(y)$ is the realisation of the F-statistic and $F_{p,N-K}\left(.\right)$ the cdf of the Fisher distribution with p and $N-K$ degrees of freedom.

3.2. The Fisher test

Definition (Global F-test)

In a multiple linear regression model with a constant term

$$y_i = \beta_1 + \sum_{k=2}^{K} \beta_k x_{ik} + \varepsilon_i$$

the **global F-test** corresponds to the test of significance of all the explicative variables:

$$\mathrm{H}_0 : \beta_2 = .. = \beta_K = 0$$

Under the assumption A6 (normality), the global F-test-statistic satisfies:

$$\mathrm{F} \underset{\mathrm{H}_0}{\sim} F_{(K-1,N-K)}$$

3.2. The Fisher test

Remarks

1. The global F-test is a test designed to see if the model is useful overall.

2. The null $H_0 : \beta_2 = .. = \beta_K = 0$ can be written as:

$$\underset{(K-1\times K)}{\mathbf{R}} \underset{(K,1)}{\boldsymbol{\beta}} = \underset{(K-1\times 1)}{\mathbf{q}}$$

$$\begin{pmatrix} 0 & 1 & 0 & 0 & .. & 0 \\ 0 & 0 & 1 & 0 & .. & 0 \\ .. & .. & 0 & 1 & .. & .. \\ .. & .. & .. & .. & .. & .. \\ 0 & .. & 0 & 0 & .. & 1 \end{pmatrix} \begin{pmatrix} \beta_1 \\ .. \\ .. \\ \beta_K \end{pmatrix} = \begin{pmatrix} 0 \\ .. \\ .. \\ 0 \end{pmatrix}$$

3.2. The Fisher test

Corollary (Global F-test)

In a multiple linear regression model with a constant term

$$y_i = \beta_1 + \sum_{k=2}^{K} \beta_k x_{ik} + \varepsilon_i$$

the global F-test-statistic can also be defined as:

$$F = \left(\frac{R^2}{1 - R^2} \right) \left(\frac{N - K}{K - 1} \right)$$

where R^2 denotes the (unadjusted) coefficient of determination.

Example (Global F-test and CAPM model)

Consider the extended CAPM model (file: Chapter4_data.xls):

$$r_{MSFT,t} = \beta_1 + \beta_2 r_{SP500,t} + \beta_3 r_{Ford,t} + \beta_4 r_{GE,t} + \varepsilon_t$$

Question: write a Matlab code to compute the global F-test, the critical value for $\alpha = 5\%$ and the p-value. Compare your results with Eviews.

Dependent Variable: R_MSFT
Method: Least Squares
Date: 11/30/13 Time: 22:37
Sample: 2 25
Included observations: 24

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.001219	0.000974	1.250453	0.2256
R_SP500	2.761927	0.629752	4.385734	0.0003
R_FORD	0.313054	0.174803	1.790895	0.0885
R_GE	-0.139065	0.287520	-0.483672	0.6339

R-squared	0.722707	Mean dependent var	0.000978
Adjusted R-squared	0.681113	S.D. dependent var	0.008312
S.E. of regression	0.004694	Akaike info criterion	-7.734203
Sum squared resid	0.000441	Schwarz criterion	-7.537861
Log likelihood	96.81044	F-statistic	
Durbin-Watson stat	2.036200	Prob(F-statistic)	

3.2. The Fisher test

```
data=xlsread('Chapter4_data.xls');
r_MSFT=data(:,1);                        % Excess return for MSFT
r_SP500=data(:,2);                       % Excess return for SP500
r_Ford=data(:,3);                        % Excess return for Ford
r_GE=data(:,4);                          % Excess return for GE
T=length(r_MSFT);                        % Sample size

% Estimation under H1
X=[ones(T,1) r_SP500 r_Ford r_GE];       % Matrix of explicative variables
y=r_MSFT;                                % Dependent variable
beta=X\y;                                % OLS estimator (H1)
res=y-X*beta;                            % Residuals
SSR1=sum(res.^2);                        % SSR of unconstrained model
var_eps=SSR1/(T-4);                      % Estimated variance

% Estimation under H0
R=[zeros(3,1) eye(3)];                  % Matrix R
q=zeros(3,1);                           % Vector q
beta_H0=beta-inv(X'*X)*R'*inv(R*inv(X'*X)*R')*(R*beta-q);
res_H0=y-X*beta_H0;                     % Residuals
SSR0=sum(res_H0.^2);                    % SSR of the constrained model

% Fisher test-statistic: second definiton
F=(SSR0-SSR1)/SSR1*(T-4)/3;
critical=finv(0.95,3,T-4);
pvalue=1-fcdf(F,3,T-4);
```

3.2. The Fisher test

Dependent Variable: R_MSFT
Method: Least Squares
Date: 12/01/13 Time: 00:03
Sample: 2 25
Included observations: 24

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.001219	0.000974	1.250453	0.2256
R_SP500	2.761927	0.629752	4.385734	0.0003
R_FORD	0.313054	0.174803	1.790895	0.0885
R_GE	-0.139065	0.287520	-0.483672	0.6339

R-squared	0.722707	Mean dependent var	0.000978
Adjusted R-squared	0.681113	S.D. dependent var	0.008312
S.E. of regression	0.004694	Akaike info criterion	-7.734203
Sum squared resid	0.000441	Schwarz criterion	-7.537861
Log likelihood	96.81044	F-statistic	17.37532
Durbin-Watson stat	2.036200	Prob(F-statistic)	0.000009

```
F =

   17.3753

critical =

    3.0984

pvalue =

  8.5996e-006
```

Case 2: Semi-parametric model

3.2. The Fisher test

Assumption 6 (normality): the distribution of the disturbances is unknown, but satisfy (assumptions A1-A5):

$$\mathbb{E}\left(\varepsilon \mid \mathbf{X}\right) = \mathbf{0}_{N\times 1}$$

$$\mathbb{V}\left(\varepsilon \mid \mathbf{X}\right) = \sigma^2 \mathbf{I}_N$$

3.2. The Fisher test

Problem

1. The exact (finite sample) distribution of $\widehat{\beta}_k$ and $\widehat{\sigma}^2$ are unknown. As a consequence the **finite sample distribution** of $\mathrm{F}(y)$ is also **unknown**.
2. But, we can express the F-statistic as a linear function of the **Wald statistic**.
3. The Wald statistic has a chi-squared asymptotic distribution (cf. next section)

3.2. The Fisher test

Definition (F-test-statistic and Wald statistic)

The Fisher test-statistic can expressed as a linear function of the **Wald test-statistic** as

$$\mathrm{F} = \frac{1}{p}\mathrm{Wald}$$

$$\mathrm{Wald} = \frac{1}{p}\left(\mathbf{R}\widehat{\beta} - \mathbf{q}\right)^{\top}\left(\mathbf{R}\left(\mathbb{V}_{asy}\left(\widehat{\beta}\right)\right)^{-1}\mathbf{R}^{\top}\right)^{-1}\left(\mathbf{R}\widehat{\beta} - \mathbf{q}\right)$$

Under assumptions A1-A5, the Wald test-statistic converges to a chi-squared distribution

$$\mathrm{Wald} \underset{\mathrm{H}_0}{\overset{d}{\rightarrow}} \chi^2\left(p\right)$$

Key concepts of Section 3

1. Student test
2. Fisher test
3. t-statistic and z-statistic
4. Global F-test
5. Exact (finite sample) distribution under the normality assumption
6. Asymptotic distribution

Section 4

MLE and Inference

Introduction

- Consider a parametric model, **linear** or **nonlinear** (GARCH, probit, logit, etc.), with a vector of parameters $\boldsymbol{\theta} = (\theta_1 : .. : \theta_K)^\top$
- We assume that the problem is regular (cf. chapter 2) and we consider a ML estimator $\widehat{\boldsymbol{\theta}}$
- The **finite sample distribution** of $\widehat{\boldsymbol{\theta}}$ is unknown, but $\widehat{\boldsymbol{\theta}}$ is **asymptotically** normally distributed (cf. chapter 2).
- We want to test a set of **linear** or **nonlinear** constraints on the true parameters (population) $\theta_1, .., \theta_K$.

Definition (Null hypothesis)

Consider a null hypothesis of p **linear** and/or **nonlinear** constraints

$$\mathrm{H}_0 : \underbrace{\mathbf{c}(\boldsymbol{\theta})}_{p \times 1} = \mathbf{0}_{p \times 1}$$

where $\mathbf{c}(\boldsymbol{\theta})$ is a vectorial function defined as:

$$\begin{aligned} \mathbf{c} : \quad & \mathbb{R}^K \rightarrow \mathbb{R}^p \\ & \boldsymbol{\theta} \mapsto \mathbf{c}(\boldsymbol{\theta}) \end{aligned}$$

Notations

1. $\mathbf{c}(\boldsymbol{\theta})$ is a $p \times 1$ vector of functions $c_1(\boldsymbol{\theta}), .., c_p(\boldsymbol{\theta})$:

$$\mathbf{c}(\boldsymbol{\theta}) = \begin{pmatrix} c_1(\boldsymbol{\theta}) \\ c_2(\boldsymbol{\theta}) \\ .. \\ c_p(\boldsymbol{\theta}) \end{pmatrix}$$

2. In the case of p **linear** constraints, we have:

$$\mathrm{H}_0 : \mathbf{c}(\boldsymbol{\theta}) = \mathbf{R}\boldsymbol{\theta} - \mathbf{q} = \mathbf{0}$$

4. MLE and inference

Example (Linear constraints)

Consider the two **linear** constraints $\theta_1 = \theta_2 + \theta_3$ and $\theta_2 + \theta_4 = 1$. We have $p = 2$ constraints such that:

$$\mathrm{H}_0 : \underset{(2,1)}{\mathbf{c}(\boldsymbol{\theta})} = \begin{pmatrix} \theta_1 - \theta_2 - \theta_3 \\ \theta_2 + \theta_4 - 1 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

The function $\mathbf{c}(\boldsymbol{\theta})$ can be written as $\mathbf{R}\boldsymbol{\theta} - \mathbf{q}$. For instance if $K = 4$ and $\boldsymbol{\theta} = (\theta_1 \ \theta_2 \ \theta_3 \ \theta_4)^{\top}$, we have

$$\mathbf{c}(\boldsymbol{\theta}) = \mathbf{R}\boldsymbol{\theta} - \mathbf{q} = \begin{pmatrix} 1 & -1 & -1 & 0 \\ 0 & 1 & 0 & 1 \end{pmatrix} \begin{pmatrix} \theta_1 \\ \theta_2 \\ \theta_3 \\ \theta_4 \end{pmatrix} - \begin{pmatrix} 0 \\ 1 \end{pmatrix}$$

Example (Nonlinear constraints)

Consider the **linear** and **nonlinear** constraints

$$\theta_1 - \theta_2 = 0 \qquad \theta_1^2 - \theta_3 = 0$$

We have $p = 2$ constraints such that:

$$\mathrm{H}_0 : \underset{(2,1)}{\mathbf{c}(\boldsymbol{\theta})} = \begin{pmatrix} \theta_1 - \theta_2 \\ \theta_1^2 - \theta_3 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

Assumptions

1. The functions $c_1(\boldsymbol{\theta}), .., c_p(\boldsymbol{\theta})$ are **differentiable**.

2. There is no **redundant** constraint (identification assumption). Formally, we have

$$(\text{row}) \text{ rank}\left(\frac{\partial \mathbf{c}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^{\top}}\right) = p \quad \forall \boldsymbol{\theta} \in \boldsymbol{\Theta}$$

with

$$\underset{(p,K)}{\frac{\partial \mathbf{c}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^{\top}}} = \begin{pmatrix} \frac{\partial c_1(\boldsymbol{\theta})}{\partial \theta_1} & \frac{\partial c_1(\boldsymbol{\theta})}{\partial \theta_2} & .. & \frac{\partial c_1(\boldsymbol{\theta})}{\partial \theta_K} \\ \frac{\partial c_2(\boldsymbol{\theta})}{\partial \theta_1} & \frac{\partial c_2(\boldsymbol{\theta})}{\partial \theta_2} & .. & \frac{\partial c_2(\boldsymbol{\theta})}{\partial \theta_K} \\ .. & .. & .. & .. \\ \frac{\partial c_p(\boldsymbol{\theta})}{\partial \theta_1} & \frac{\partial c_p(\boldsymbol{\theta})}{\partial \theta_2} & .. & \frac{\partial c_p(\boldsymbol{\theta})}{\partial \theta_K} \end{pmatrix}$$

4. MLE and inference

Consider the **two-sided** test

$$H_0 : \mathbf{c}(\boldsymbol{\theta}) = \mathbf{0} \quad \text{versus} \quad H_1 : \mathbf{c}(\boldsymbol{\theta}) \neq \mathbf{0}$$

We introduce three different asymptotic tests (*the trilogy*..)

1. The **Likelihood Ratio (LR)** test
2. The **Wald** test
3. The **Lagrance Multiplier (LM)** test

4. MLE and inference

For each of the three tests, we will present:

1. the test-statistic
2. its asymptotic distribution under the null
3. the (asymptotic) rejection region
4. the (asymptotic) p-value

Subsection 4.1

The Likelihood Ratio (LR) test

4.1. The Likelihood Ratio (LR) test

Definition (Likelihood Ratio (LR) test statistic)

The **likelihood ratio (LR)** test-statistic is defined by as to be:

$$\mathrm{LR} = -2\left(\ell_N\left(\widehat{\boldsymbol{\theta}}_{\mathrm{H}_0}; y|\, x\right) - \ell_N\left(\widehat{\boldsymbol{\theta}}_{\mathrm{H}_1}; y|\, x\right)\right)$$

where $\ell_N(\boldsymbol{\theta}; y|\, x)$ denotes the (conditional) log-likelihood of the sample y, $\widehat{\boldsymbol{\theta}}_{\mathrm{H}_0}$ and $\widehat{\boldsymbol{\theta}}_{\mathrm{H}_1}$ are respectively the maximum likelihood estimator of $\boldsymbol{\theta}$ under the alternative and the null hypothesis.

4.1. The Likelihood Ratio (LR) test

Comments

Consider the ratio of likelihoods under H_1 (no constraint) and under H_0 (with $\mathbf{c}(\boldsymbol{\theta}) = \mathbf{0}$).

$$\lambda = \frac{L_N\left(\widehat{\boldsymbol{\theta}}_{\text{H}_0}; y|\, x\right)}{L_N\left(\widehat{\boldsymbol{\theta}}_{\text{H}_1}; y|\, x\right)}$$

1. $\lambda > 0$ since both likelihood are positive.
2. $\lambda < 1$ since $L_N(\text{H}_0)$ cannot be larger than $L_N(\text{H}_1)$. A restricted optimum is never superior to an unrestricted one.
3. If λ is too small, then doubt is cast on the restrictions $\mathbf{c}(\boldsymbol{\theta}) = \mathbf{0}$.
4. Consider the statistic LR$= 2\ln(\lambda)$: if λ is "too small", then LR is large (rejection of the null)...

4.1. The Likelihood Ratio (LR) test

Definition (Asymptotic distribution and critical region)

Under some regularity conditions (cf. chapter 2) and under the null $\mathrm{H}_0 : \mathbf{c}(\boldsymbol{\theta}) = \mathbf{0}$, the LR test-statistic **converges** to a chi-squared distribution with p degrees of freedom (the number of restrictions imposed):

$$\mathrm{LR} \underset{\mathrm{H}_0}{\overset{d}{\rightarrow}} \chi^2(p)$$

The (asymptotic) **critical region** for a significance level of α is:

$$\mathrm{W} = \left\{ y : \mathrm{LR}(y) > \chi^2_{1-\alpha}(p) \right\}$$

where $\chi^2_{1-\alpha}(p)$ is the $1-\alpha$ critical value of the chi-squared distribution with p degrees of freedom and $\mathrm{LR}(y)$ is the realisation of the LR test-statistic.

4.1. The Likelihood Ratio (LR) test

Definition (p-value of the LRT test)

The **p-value** of the LR test is equal to:

$$\text{p-value} = 1 - G_p\left(\text{LR}\left(y\right)\right)$$

where $\text{LR}(y)$ is the realisation of the LR test-statistic and $G_p\left(.\right)$ is the cdf of the chi-squared distribution with p degrees of freedom.

4.1. The Likelihood Ratio (LR) test

Example (LRT and Poisson distribution)

Suppose that $X_1, X_2, \cdots, X_N$ *are i.i.d. discrete random variables, such that* $X_i \sim Pois(\theta)$ *with a* pmf (probability mass function) defined as:

$$\Pr(X_i = x_i) = \frac{\exp(-\theta)\,\theta^{x_i}}{x_i!}$$

where θ is an unknown parameter to estimate. We have a sample (realisation) of size $N = 10$ given by $\{5, 0, 1, 1, 0, 3, 2, 3, 4, 1\}$. **Question:** use a LR test to test the null $\mathrm{H}_0 : \theta = 1.8$ against $H_1 : \theta \neq 1.8$ and give a conclusion for significance level of 5%.

Solution

The log-likelihood function is defined as to be:

$$\ell_N(\theta; x) = -\theta N + \ln(\theta) \sum_{i=1}^{N} x_i - \ln\left(\prod_{i=1}^{N} x_i!\right)$$

In the chapter 2, we found that the ML estimator of θ is the sample mean:

$$\widehat{\theta} = \frac{1}{N} \sum_{i=1}^{N} X_i$$

Given the sample $\{5, 0, 1, 1, 0, 3, 2, 3, 4, 1\}$, the estimate of θ (under H_1, with non constraint) is $\widehat{\theta}_{\mathrm{H}_1} = 2$, and the corresponding log-likelihood is equal to:

$$\ell_N\left(\widehat{\theta}_{\mathrm{H}_1}; x\right) = \ln(0.104)$$

Solution (cont'd)

Under the null $\text{H}_0 : \theta = 1.8$, we don't need to estimate θ and the log-likelihood is equal to:

$$\ell_N \left(\theta_{\text{H}_0}; x \right) = -1.8N + \ln \left(1.8 \right) \sum_{i=1}^{N} x_i - \ln \left(\prod_{i=1}^{N} x_i ! \right) = \ln \left(0.0936 \right)$$

The LR test-statistic is equal to:

$$\text{LR} \left(y \right) = -2 \ln \left(\frac{0.0936}{0.104} \right) = 0.21072$$

Solution (cont'd)

$$\text{LR}(y) = 0.21072$$

For $N = 10$, $p = 1$ (one restriction) and $\alpha = 0.05$, the critical region is:

$$\text{W} = \left\{ y : \text{LR}(y) > \chi^2_{0.95}(1) = 3.8415 \right\}$$

and the p-value is

$$\text{pvalue} = 1 - G_1(0.21072) = 0.6462$$

where $G_1(.)$ is the cdf of the $\chi^2(1)$ distribution.

Conclusion: for a significance level of 5%, we fail to reject the null $\text{H}_0 : \theta = 1.8$. $\square$

Subsection 4.2

The Wald test

4.2. The Wald test

Definition (Wald test-statistic)

The **Wald** test-statistic associated to the test of $\mathrm{H}_0 : \mathbf{c}(\boldsymbol{\theta}) = \mathbf{0}$ is defined as to be:

$$\mathrm{Wald} = \mathbf{c}\left(\widehat{\boldsymbol{\theta}}_{\mathrm{H}_1}\right)^{\top} \left(\frac{\partial \mathbf{c}}{\partial \boldsymbol{\theta}^{\top}} \left(\widehat{\boldsymbol{\theta}}_{\mathrm{H}_1}\right) \widehat{\mathbb{V}}_{asy}\left(\widehat{\boldsymbol{\theta}}_{\mathrm{H}_1}\right) \frac{\partial \mathbf{c}}{\partial \boldsymbol{\theta}^{\top}} \left(\widehat{\boldsymbol{\theta}}_{\mathrm{H}_1}\right)^{\top} \right)^{-1} \mathbf{c}\left(\widehat{\boldsymbol{\theta}}_{\mathrm{H}_1}\right)$$

where $\widehat{\boldsymbol{\theta}}_{\mathrm{H}_1}$ is the maximum likelihood estimator of $\boldsymbol{\theta}$ under the alternative hypothesis (unconstrained model) and $\widehat{\mathbb{V}}_{asy}\left(\widehat{\boldsymbol{\theta}}_{\mathrm{H}_1}\right)$ is an estimator of its asymptotic variance covariance matrix.

4.2. The Wald test

Remark

$$\text{Wald} = \underbrace{\mathbf{c}\left(\widehat{\boldsymbol{\theta}}_{\mathrm{H}_1}\right)^{\top}}_{1\times p}\left(\underbrace{\frac{\partial\mathbf{c}}{\partial\boldsymbol{\theta}^{\top}}\left(\widehat{\boldsymbol{\theta}}_{\mathrm{H}_1}\right)}_{p\times K}\underbrace{\widehat{\mathbb{V}}_{asy}\left(\widehat{\boldsymbol{\theta}}_{\mathrm{H}_1}\right)}_{K\times K}\underbrace{\frac{\partial\mathbf{c}}{\partial\boldsymbol{\theta}^{\top}}\left(\widehat{\boldsymbol{\theta}}_{\mathrm{H}_1}\right)^{\top}}_{K\times p}\right)^{-1}\underbrace{\mathbf{c}\left(\widehat{\boldsymbol{\theta}}_{\mathrm{H}_1}\right)}_{p\times 1}$$

4.2. The Wald test

Example (Wald test-statistic)

Consider a model with $K = 3$ parameters $\theta = (\theta_1 : \theta_2 : \theta_3)^\top$ with

$$\theta_1 - \theta_2 = 0 \qquad \theta_1^2 - \theta_3 = 0$$

We have two constraints ($p = 2$) and:

$$\mathrm{H}_0 : \underset{(2,1)}{\mathbf{c}(\boldsymbol{\theta})} = \begin{pmatrix} \theta_1 - \theta_2 \\ \theta_1^2 - \theta_3 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

Denote $\widehat{\boldsymbol{\theta}}_{\mathrm{H}_1} = (\theta_1 : \theta_2 : \theta_3)^\top$ the ML estimator of $\boldsymbol{\theta}$ under the alternative hypothesis and $\widehat{\mathbb{V}}_{asy}\left(\widehat{\boldsymbol{\theta}}_{\mathrm{H}_1}\right)$ the estimator of its asymptotic variance covariance matrix. **Question:** write the Wald test-statistic.

4.2. The Wald test

Solution

Here we have $K = 3$ and $p = 2$

$$\mathbf{c}\left(\widehat{\boldsymbol{\theta}}_{\mathrm{H}_1}\right) = \begin{pmatrix} \widehat{\theta}_1 - \widehat{\theta}_2 \\ \widehat{\theta}_1^2 - \widehat{\theta}_3 \end{pmatrix}$$

$$\frac{\partial \mathbf{c}}{\partial \boldsymbol{\theta}^\top}\left(\widehat{\boldsymbol{\theta}}_{\mathrm{H}_1}\right) = \begin{pmatrix} 1 & -1 & 0 \\ 2\widehat{\theta}_1 & 0 & -1 \end{pmatrix}$$

$$\text{Wald} = \mathbf{c}\left(\widehat{\boldsymbol{\theta}}_{\mathrm{H}_1}\right)^\top \left(\frac{\partial \mathbf{c}}{\partial \boldsymbol{\theta}^\top}\left(\widehat{\boldsymbol{\theta}}_{\mathrm{H}_1}\right) \widehat{\mathbb{V}}_{asy}\left(\widehat{\boldsymbol{\theta}}_{\mathrm{H}_1}\right) \frac{\partial \mathbf{c}}{\partial \boldsymbol{\theta}^\top}\left(\widehat{\boldsymbol{\theta}}_{\mathrm{H}_1}\right)^\top\right)^{-1} \mathbf{c}\left(\widehat{\boldsymbol{\theta}}_{\mathrm{H}_1}\right)$$

Remark

In the case of **linear constraints**

$$H_0 : \mathbf{R}\theta - \mathbf{q} = \mathbf{0}$$

we have

$$H_0 : \mathbf{c}\left(\boldsymbol{\theta}\right) = \mathbf{0}$$

with

$$\mathbf{c}\left(\boldsymbol{\theta}\right) = \mathbf{R}\theta - \mathbf{q}$$

$$\frac{\partial \mathbf{c}}{\partial \boldsymbol{\theta}^{\top}}\left(\boldsymbol{\theta}\right) = \mathbf{R}$$

4.2. The Wald test

Definition (Wald test-statistic and linear constraints)

Consider the test of **linear constraints** $\text{H}_0 : \mathbf{c}(\boldsymbol{\theta}) = \mathbf{R}\boldsymbol{\theta} - \mathbf{q} = \mathbf{0}$. The **Wald** test-statistic is defined as to be:

$$\text{Wald} = \left(\mathbf{R}\widehat{\boldsymbol{\theta}}_{\text{H}_1} - \mathbf{q}\right)^{\top} \left(\mathbf{R}\ \widehat{\mathbb{V}}_{asy}\left(\widehat{\boldsymbol{\theta}}_{\text{H}_1}\right)\ \mathbf{R}^{\top}\right)^{-1} \left(\mathbf{R}\widehat{\boldsymbol{\theta}}_{\text{H}_1} - \mathbf{q}\right)$$

where $\widehat{\boldsymbol{\theta}}_{\text{H}_1}$ is the maximum likelihood estimator of $\boldsymbol{\theta}$ under the alternative hypothesis (unconstrained model) and $\widehat{\mathbb{V}}_{asy}\left(\widehat{\boldsymbol{\theta}}_{\text{H}_1}\right)$ is an estimator of its asymptotic variance covariance matrix.

4.2. The Wald test

Definition (Asymptotic distribution and critical region)

Under some regularity conditions (cf. chapter 2) and under the null $\mathrm{H}_0 : \mathbf{c}(\boldsymbol{\theta}) = \mathbf{0}$, the Wald test-statistic **converges** to a chi-squared distribution with p degrees of freedom (the number of restrictions imposed):

$$\text{Wald} \xrightarrow[\mathrm{H}_0]{d} \chi^2(p)$$

The (asymptotic) **critical region** for a significance level of α is:

$$\mathrm{W} = \left\{ y : \text{Wald}(y) > \chi^2_{1-\alpha}(p) \right\}$$

where $\chi^2_{1-\alpha}(p)$ is the $1-\alpha$ critical value of the chi-squared distribution with p degrees of freedom and Wald(y) is the realisation of the Wald test-statistic.

Proof

Under some regularity conditions, we have

$$\sqrt{N}\left(\widehat{\boldsymbol{\theta}}_{\mathrm{H}_1}-\boldsymbol{\theta}_0\right)\xrightarrow{d}\mathcal{N}\left(0,I^{-1}\left(\boldsymbol{\theta}_0\right)\right)$$

We use the **delta method** for the function $\mathbf{c}\left(.\right)$. The function $\mathbf{c}\left(.\right)$ is a continuous and continuously differentiable function not involving N, then

$$\sqrt{N}\left(\mathbf{c}\left(\widehat{\boldsymbol{\theta}}_{\mathrm{H}_1}\right)-\mathbf{c}\left(\boldsymbol{\theta}_0\right)\right)\xrightarrow{d}\mathcal{N}\left(0,\frac{\partial\mathbf{c}}{\partial\boldsymbol{\theta}^{\top}}\left(\boldsymbol{\theta}_0\right)I^{-1}\left(\boldsymbol{\theta}_0\right)\frac{\partial\mathbf{c}}{\partial\boldsymbol{\theta}^{\top}}\left(\boldsymbol{\theta}_0\right)^{\top}\right)$$

Under the null $\mathrm{H}_0:\mathbf{c}\left(\boldsymbol{\theta}_0\right)=\mathbf{0}$, we have

$$\left(\frac{\partial\mathbf{c}}{\partial\boldsymbol{\theta}^{\top}}\left(\boldsymbol{\theta}_0\right)I^{-1}\left(\boldsymbol{\theta}_0\right)\frac{\partial\mathbf{c}}{\partial\boldsymbol{\theta}^{\top}}\left(\boldsymbol{\theta}_0\right)^{\top}\right)^{-1/2}\sqrt{N}\mathbf{c}\left(\widehat{\boldsymbol{\theta}}_{\mathrm{H}_1}\right)\xrightarrow{d}\mathcal{N}\left(0,\mathbf{I}_p\right)$$

where $\mathbf{I}_p$ is the identity matrix of size p.

Proof (cont'd)

The **Wald criteria** is defined as to be:

$$
\begin{aligned}
& \text{Wald criteria} \\
= \quad & N \times \mathbf{c}\left(\widehat{\boldsymbol{\theta}}_{\mathrm{H}_1}\right)^\top \left(\left(\frac{\partial \mathbf{c}}{\partial \boldsymbol{\theta}^\top} \left(\boldsymbol{\theta}_0\right) I^{-1}\left(\boldsymbol{\theta}_0\right) \frac{\partial \mathbf{c}}{\partial \boldsymbol{\theta}^\top} \left(\boldsymbol{\theta}_0\right)^\top \right)^{-1/2} \right)^\top \\
& \times \left(\frac{\partial \mathbf{c}}{\partial \boldsymbol{\theta}^\top} \left(\boldsymbol{\theta}_0\right) I^{-1}\left(\boldsymbol{\theta}_0\right) \frac{\partial \mathbf{c}}{\partial \boldsymbol{\theta}^\top} \left(\boldsymbol{\theta}_0\right)^\top \right)^{-1/2} \times \mathbf{c}\left(\widehat{\boldsymbol{\theta}}_{\mathrm{H}_1}\right) \\
= \quad & N \times \mathbf{c}\left(\widehat{\boldsymbol{\theta}}_{\mathrm{H}_1}\right)^\top \times \left(\frac{\partial \mathbf{c}}{\partial \boldsymbol{\theta}^\top} \left(\boldsymbol{\theta}_0\right) I^{-1}\left(\boldsymbol{\theta}_0\right) \frac{\partial \mathbf{c}}{\partial \boldsymbol{\theta}^\top} \left(\boldsymbol{\theta}_0\right)^\top \right)^{-1} \times \mathbf{c}\left(\widehat{\boldsymbol{\theta}}_{\mathrm{H}_1}\right)
\end{aligned}
$$

So, under the null $\mathrm{H}_0 : \mathbf{c}\left(\boldsymbol{\theta}_0\right) = \mathbf{0}$,, we have

$$
\text{Wald criteria} \underset{\mathrm{H}_0}{\overset{d}{\longrightarrow}} \chi^2\left(p\right)
$$

4.2. The Wald test

Proof (cont'd)

$$
\begin{aligned}
\text{Wald Criteria} &= N \times \mathbf{c}\left(\widehat{\boldsymbol{\theta}}_{\mathrm{H}_1}\right)^{\top} \\
&\times \left(\frac{\partial \mathbf{c}}{\partial \boldsymbol{\theta}^{\top}} \left(\boldsymbol{\theta}_0\right) I^{-1}\left(\boldsymbol{\theta}_0\right) \frac{\partial \mathbf{c}}{\partial \boldsymbol{\theta}^{\top}} \left(\boldsymbol{\theta}_0\right)^{\top} \right)^{-1} \times \mathbf{c}\left(\widehat{\boldsymbol{\theta}}_{\mathrm{H}_1}\right)
\end{aligned}
$$

A feasible **Wald test-statistic** is given by

$$
\begin{aligned}
\text{Wald} &= N \times \mathbf{c}\left(\widehat{\boldsymbol{\theta}}_{\mathrm{H}_1}\right)^{\top} \\
&\times \left(\frac{\partial \mathbf{c}}{\partial \boldsymbol{\theta}^{\top}} \left(\widehat{\boldsymbol{\theta}}_{\mathrm{H}_1}\right) \widehat{I}^{-1}\left(\widehat{\boldsymbol{\theta}}_{\mathrm{H}_1}\right) \frac{\partial \mathbf{c}}{\partial \boldsymbol{\theta}^{\top}} \left(\widehat{\boldsymbol{\theta}}_{\mathrm{H}_1}\right)^{\top} \right)^{-1} \times \mathbf{c}\left(\widehat{\boldsymbol{\theta}}_{\mathrm{H}_1}\right)
\end{aligned}
$$

4.2. The Wald test

Proof (cont'd)

Since

$$\widehat{\mathbb{V}}_{asy}\left(\widehat{\boldsymbol{\theta}}_{\mathrm{H}_1}\right) = N^{-1}\widehat{I}^{-1}\left(\widehat{\boldsymbol{\theta}}_{\mathrm{H}_1}\right)$$

We have finally

$$\text{Wald} = \mathbf{c}\left(\widehat{\boldsymbol{\theta}}_{\mathrm{H}_1}\right)^{\top} \times \left(\frac{\partial \mathbf{c}}{\partial \boldsymbol{\theta}^{\top}}\left(\widehat{\boldsymbol{\theta}}_{\mathrm{H}_1}\right)\widehat{\mathbb{V}}_{asy}\left(\widehat{\boldsymbol{\theta}}_{\mathrm{H}_1}\right)\frac{\partial \mathbf{c}}{\partial \boldsymbol{\theta}^{\top}}\left(\widehat{\boldsymbol{\theta}}_{\mathrm{H}_1}\right)^{\top}\right)^{-1} \times \mathbf{c}\left(\widehat{\boldsymbol{\theta}}_{\mathrm{H}_1}\right)$$

and

$$\text{Wald} \underset{\mathrm{H}_0}{\overset{d}{\to}} \chi^2(p) \quad \square$$

4.2. The Wald test

Definition (p-value of the Wald test)

The **p-value** of the Wald test is equal to:

$$\text{p-value} = 1 - G_p\left(\text{Wald}\left(y\right)\right)$$

where $\text{Wald}(y)$ is the realisation of the Wald test-statistic and $G_p\left(.\right)$ is the cdf of the chi-squared distribution with p degrees of freedom.

4.2. The Wald test

Definition (z-statistic)

Consider the test $\mathrm{H}_0 : \theta_k = a_k$ versus $\mathrm{H}_1 : \theta_k \neq a_k$. The **z-statistic** corresponds to the square root of the **Wald test-statistic** and satisfies

$$\mathrm{Z}_k = \frac{\left(\widehat{\theta}_k - a_k\right)}{\sqrt{\widehat{\mathbb{V}}_{asy}\left(\widehat{\theta}_k\right)}} \overset{d}{\underset{\mathrm{H}_0}{\rightarrow}} \mathcal{N}\left(0,1\right)$$

where $\widehat{\theta}_k$ is the ML estimator of θ_k obtained under H_1 (unconstrained model). The critical region for a significance level of α is:

$$\mathrm{W} = \left\{ y : \left|\mathrm{Z}_k\left(y\right)\right| > \Phi^{-1}\left(1 - \frac{\alpha}{2}\right)\right\}$$

where $\Phi\left(.\right)$ denotes the cdf of the standard normal distribution.

4.2. The Wald test

Computational issues

The Wald test-statistic depends on the **estimator of the asymptotic variance covariance matrix**:

$$\text{Wald} = \mathbf{c}\left(\widehat{\boldsymbol{\theta}}_{\mathrm{H}_1}\right)^{\top}\left(\frac{\partial \mathbf{c}}{\partial \boldsymbol{\theta}^{\top}}\left(\widehat{\boldsymbol{\theta}}_{\mathrm{H}_1}\right)\widehat{\mathbb{V}}_{asy}\left(\widehat{\boldsymbol{\theta}}_{\mathrm{H}_1}\right)\frac{\partial \mathbf{c}}{\partial \boldsymbol{\theta}^{\top}}\left(\widehat{\boldsymbol{\theta}}_{\mathrm{H}_1}\right)^{\top}\right)^{-1}\mathbf{c}\left(\widehat{\boldsymbol{\theta}}_{\mathrm{H}_1}\right)$$

$$\widehat{\mathbb{V}}_{asy}\left(\widehat{\boldsymbol{\theta}}_{\mathrm{H}_1}\right) = N^{-1}\widehat{I}^{-1}\left(\widehat{\boldsymbol{\theta}}_{\mathrm{H}_1}\right)$$

where $I\left(\widehat{\boldsymbol{\theta}}_{\mathrm{H}_1}\right)$ denotes the average Fisher information matrix.

4.2. The Wald test

Computational issues (cont'd)

Three estimators are available for the average Fisher information matrix:

$$\textbf{Actual Average Fisher Matrix}: \; \widehat{I}_{\mathrm{A}}\left(\widehat{\boldsymbol{\theta}}\right) = \frac{1}{N}\sum_{i=1}^{N}\widehat{I}_{i}\left(\widehat{\boldsymbol{\theta}}\right)$$

$$\textbf{BHHH estimator}: \; \widehat{I}_{\mathrm{B}}\left(\widehat{\boldsymbol{\theta}}\right) = \frac{1}{N}\sum_{i=1}^{N}\left(\left.\frac{\partial \ell_i\left(\boldsymbol{\theta};\, y_i\,|\,x_i\right)}{\partial \boldsymbol{\theta}}\right|_{\widehat{\boldsymbol{\theta}}} \left.\frac{\partial \ell_i\left(\boldsymbol{\theta};\, y_i\,|\,x_i\right)}{\partial \boldsymbol{\theta}}\right|_{\widehat{\boldsymbol{\theta}}}^{\top}\right)$$

$$\textbf{Hessian based estimator:} \; \widehat{I}_{\mathrm{c}}\left(\widehat{\boldsymbol{\theta}}\right) = \frac{1}{N}\sum_{i=1}^{N}\left(-\left.\frac{\partial^2 \ell_i\left(\boldsymbol{\theta};\, y_i\,|\,x_i\right)}{\partial \boldsymbol{\theta}\partial \boldsymbol{\theta}^{\top}}\right|_{\widehat{\boldsymbol{\theta}}}\right)$$

4.2. The Wald test

Computational issues (cont'd)

1. These estimators are asymptotically equivalent, but the corresponding estimates may be **very different** in small samples.

2. Thus, we can obtain **three different values** for the Wald statistic given the choice of the estimator for $\mathbb{V}_{asy}\left(\widehat{\theta}_{\mathrm{H}_1}\right)$ (cf. exercises).

3. In general, the estimator A is rarely available and the estimator B (BHHH) gives erratic results.

4. Most of the software use the **estimator C** (Hessian based estimator).

4.2. The Wald test

Computational issues (cont'd)

Dependent Variable: Y
Method: ML - Binary Probit
Date: 11/24/13 Time: 18:33
Sample: 1 190
Included observations: 190
Convergence achieved after 3 iterations
Covariance matrix computed using second derivatives

Variable	Coefficient	Std. Error	z-Statistic	Prob.
X	0.215364	0.092715	2.322847	0.0202
C	-0.215364	0.092715	-2.322847	0.0202

Mean dependent var	0.421053	S.D. dependent var	0.495032
S.E. of regression	0.489246	Akaike info criterion	1.353695
Sum squared resid	45.00000	Schwarz criterion	1.387874
Log likelihood	-126.6010	Hannan-Quinn criter.	1.367540
Restr. log likelihood	-129.3196	Avg. log likelihood	-0.666321
LR statistic (1 df)	5.437219	McFadden R-squared	0.021022
Probability(LR stat)	0.019712		

Obs with Dep=0	110	Total obs	190
Obs with Dep=1	80		

Subsection 4.3

The Lagrange Multiplier (LM) test

4.3. The Lagrange Multiplier (LM) test

Introduction

Consider the set of constraints $\mathbf{c}(\boldsymbol{\theta}) = \mathbf{0}$. Let $\boldsymbol{\lambda}$ be a vector of Lagrange multipliers and define the Lagrangian function

$$\ell_N(\boldsymbol{\theta}^*; y|x) = \ell_N(\boldsymbol{\theta}; y|x) + \boldsymbol{\lambda}\mathbf{c}(\boldsymbol{\theta})$$

The solution to the constrained maximization problem is the root of

$$\frac{\partial \ell_N(\boldsymbol{\theta}^*; y|x)}{\partial \boldsymbol{\theta}} = \frac{\partial \ell_N(\boldsymbol{\theta}; y|x)}{\partial \boldsymbol{\theta}} + \left(\frac{\partial \mathbf{c}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^\top}\right)^\top \boldsymbol{\lambda}$$

$$\frac{\partial \ell_N(\boldsymbol{\theta}^*; y|x)}{\partial \boldsymbol{\lambda}} = \mathbf{c}(\boldsymbol{\theta})$$

4.3. The Lagrange Multiplier (LM) test

Introduction (cont'd)

$$\frac{\partial \ell_N \left(\boldsymbol{\theta}^*; y|\, x\right)}{\partial \boldsymbol{\theta}} = \frac{\partial \ell_N \left(\boldsymbol{\theta}; y|\, x\right)}{\partial \boldsymbol{\theta}} + \left(\frac{\partial \mathbf{c}\left(\boldsymbol{\theta}\right)}{\partial \boldsymbol{\theta}^\top}\right)^\top \boldsymbol{\lambda}$$

1. If the restrictions are valid, then imposing them will not lead to a significant difference in the maximized value of the likelihood function. In the first-order conditions, the meaning is that the second term in the derivative vector will be small. In particular, $\boldsymbol{\lambda}$ **will be small**.
2. We could test this directly, that is, test

$$\mathrm{H}_0 : \boldsymbol{\lambda} = \mathbf{0}$$

which leads to the **Lagrange multiplier test.**

4.3. The Lagrange Multiplier (LM) test

Introduction (cont'd)

There is an equivalent simpler formulation, however. If the restrictions $\mathbf{c}(\boldsymbol{\theta}) = \mathbf{0}$ are valid, the derivatives of the log-likelihood of the **unconstrained model** evaluated at the **restricted parameter vector** will be approximately zero.

$$\left. \frac{\partial \ell_N (\boldsymbol{\theta}; y | x)}{\partial \boldsymbol{\theta}} \right|_{\widehat{\boldsymbol{\theta}}_{H_0}} = \mathbf{0}$$

The vector of first derivatives of the log-likelihood is the vector of (efficient) **scores**.

4.3. The Lagrange Multiplier (LM) test

Definition (LM or score test)

For these reasons, this test is called the **score test** as well as the **Lagrange multiplier test**.

4.3. The Lagrange Multiplier (LM) test

Guess

Let us assume that θ is scalar, i.e. $K = 1$, then the LM statistic is simply defined as:

$$\text{LM} = \frac{s_N\left(\widehat{\boldsymbol{\theta}}_{\text{H}_0};\, Y \middle|\, x\right)^2}{\mathbb{V}\left(s_N\left(\widehat{\boldsymbol{\theta}}_{\text{H}_0};\, Y \middle|\, x\right)\right)}$$

Since $\widehat{I}_N\left(\widehat{\boldsymbol{\theta}}_{\text{H}_0}\right) = \mathbb{V}\left(s_N\left(\widehat{\boldsymbol{\theta}}_{\text{H}_0};\, Y \middle|\, x\right)\right)$, we have:

$$\text{LM} = \frac{s_N\left(\widehat{\boldsymbol{\theta}}_{\text{H}_0};\, Y \middle|\, x\right)^2}{\widehat{I}_N\left(\widehat{\boldsymbol{\theta}}_{\text{H}_0}\right)}$$

4.3. The Lagrange Multiplier (LM) test

Definition (LM or score test)

The **LM** test-statistic or **score** test associated to the test of H_0 : $\mathbf{c}(\boldsymbol{\theta}) = \mathbf{0}$ is defined as to be:

$$\text{LM} = s_N\left(\widehat{\boldsymbol{\theta}}_{\text{H}_0} ;\, Y \middle|\, x\right)^{\top} \widehat{I}_N^{-1}\left(\widehat{\boldsymbol{\theta}}_{\text{H}_0}\right) s_N\left(\widehat{\boldsymbol{\theta}}_{\text{H}_0} ;\, Y \middle|\, x\right)$$

where $\widehat{\boldsymbol{\theta}}_{\text{H}_0}$ is the maximum likelihood estimator of $\boldsymbol{\theta}$ under the null hypothesis (**constrained** model), $s_N(\boldsymbol{\theta};\, Y|\, x)$ is the score vector of the **unconstrained** model and $\widehat{I}_N\left(\widehat{\boldsymbol{\theta}}_{\text{H}_0}\right)$ is an estimator of the Fisher information matrix of the sample evaluated at $\widehat{\boldsymbol{\theta}}_{\text{H}_0}$.

Remark

Since:

$$\widehat{\mathbb{V}}_{asy}\left(\widehat{\boldsymbol{\theta}}_{\mathrm{H}_0}\right)=\widehat{I}_N^{-1}\left(\widehat{\boldsymbol{\theta}}_{\mathrm{H}_0}\right)$$

there is another expression for the LM statistic.

4.3. The Lagrange Multiplier (LM) test

Definition (LM or score test)

The **LM** test-statistic or **score** test associated to the test of $\mathrm{H}_0 : \mathbf{c}(\boldsymbol{\theta}) = \mathbf{0}$ is defined as to be:

$$\mathrm{LM} = s_N\left(\widehat{\boldsymbol{\theta}}_{\mathrm{H}_0};\, Y\middle|\, x\right)^\top \widehat{\mathbb{V}}_{asy}\left(\widehat{\boldsymbol{\theta}}_{\mathrm{H}_0}\right) s_N\left(\widehat{\boldsymbol{\theta}}_{\mathrm{H}_0};\, Y\middle|\, x\right)$$

where $\widehat{\boldsymbol{\theta}}_{\mathrm{H}_0}$ is the maximum likelihood estimator of $\boldsymbol{\theta}$ under the null hypothesis (**constrained** model), $s_N\left(\boldsymbol{\theta};\, Y\middle|\, x\right)$ is the score vector of the **unconstrained** model and $\widehat{\mathbb{V}}_{asy}\left(\widehat{\boldsymbol{\theta}}_{\mathrm{H}_0}\right)$ is an estimator of the asymptotic variance covariance matrix of $\widehat{\boldsymbol{\theta}}_{\mathrm{H}_0}$.

Remark

The **LM** test-statistic can also be defined by:

$$\mathrm{LM} = \boldsymbol{\lambda}^{\top} \frac{\partial \mathbf{c}}{\partial \boldsymbol{\theta}^{\top}} \left(\widehat{\boldsymbol{\theta}}_{\mathrm{H}_0} \right) \widehat{\mathbb{V}}_{asy} \left(\widehat{\boldsymbol{\theta}}_{\mathrm{H}_0} \right) \left(\frac{\partial \mathbf{c}}{\partial \boldsymbol{\theta}^{\top}} \left(\widehat{\boldsymbol{\theta}}_{\mathrm{H}_0} \right) \right)^{\top} \boldsymbol{\lambda}$$

where $\boldsymbol{\lambda}$ denotes the Lagrange Multiplier associated to the constraints $\mathbf{c}(\boldsymbol{\theta}) = \mathbf{0}$.

4.3. The Lagrange Multiplier (LM) test

The LM test-statistic can be obtained from the following **auxiliary procedure**:

Step 1: Estimate the constrained model and obtain $\widehat{\boldsymbol{\theta}}_{\mathrm{H}_0}$.

Step 2: Form the gradients for each observation of the **unrestricted** model evaluated at $\widehat{\boldsymbol{\theta}}_{\mathrm{H}_0}$

$$g_i\left(\widehat{\boldsymbol{\theta}}_{\mathrm{H}_0}; y_i \middle| x_i\right) \quad \forall i = 1, ..N$$

Step 3: Run the regression of a vector of 1 on the variables $g_i\left(\widehat{\boldsymbol{\theta}}_{\mathrm{H}_0}; y_i \middle| x_i\right)$ $\forall i = 1, ..N$, then

$$\mathrm{LM} = N \times R^2$$

where R^2 denotes the (unadjusted) coefficient of determination of this auxiliary regression.

4.3. The Lagrange Multiplier (LM) test

Computational issues

1. The LM test-statistic depends on the **estimator of the asymptotic variance covariance matrix**:

$$\text{LM} = s_N \left(\widehat{\boldsymbol{\theta}}_{\text{H}_0}; \, Y | \, x \right)^\top \widehat{\mathbb{V}}_{asy} \left(\widehat{\boldsymbol{\theta}}_{\text{H}_0} \right) s_N \left(\widehat{\boldsymbol{\theta}}_{\text{H}_0}; \, Y | \, x \right)$$

$$\widehat{\mathbb{V}}_{asy} \left(\widehat{\boldsymbol{\theta}}_{\text{H}_0} \right) = N^{-1} \widehat{I}^{-1} \left(\widehat{\boldsymbol{\theta}}_{\text{H}_0} \right)$$

where $I \left(\widehat{\boldsymbol{\theta}}_{\text{H}_0} \right)$ denotes the average Fisher information matrix.

2. Thus, we can obtain **three different values** for the LM statistic given the choice of the estimator for $\mathbb{V}_{asy} \left(\widehat{\boldsymbol{\theta}}_{\text{H}_0} \right)$ (cf. exercises).

4.3. The Lagrange Multiplier (LM) test

Definition (Asymptotic distribution and critical region)

Under some regularity conditions (cf. chapter 2) and under the null $\mathrm{H}_0 : \mathbf{c}(\boldsymbol{\theta}) = \mathbf{0}$, the LM test-statistic **converges** to a chi-squared distribution with p degrees of freedom (the number of restrictions imposed):

$$\mathrm{LM} \xrightarrow[\mathrm{H}_0]{d} \chi^2(p)$$

The (asymptotic) **critical region** for a significance level of α is:

$$\mathrm{W} = \left\{ y : \mathrm{LM}(y) > \chi^2_{1-\alpha}(p) \right\}$$

where $\chi^2_{1-\alpha}(p)$ is the $1-\alpha$ critical value of the chi-squared distribution with p degrees of freedom and $\mathrm{LM}(y)$ is the realisation of the LM test-statistic.

4.3. The Lagrange Multiplier (LM) test

Definition (p-value of the LM test)

The **p-value** of the LM test is equal to:

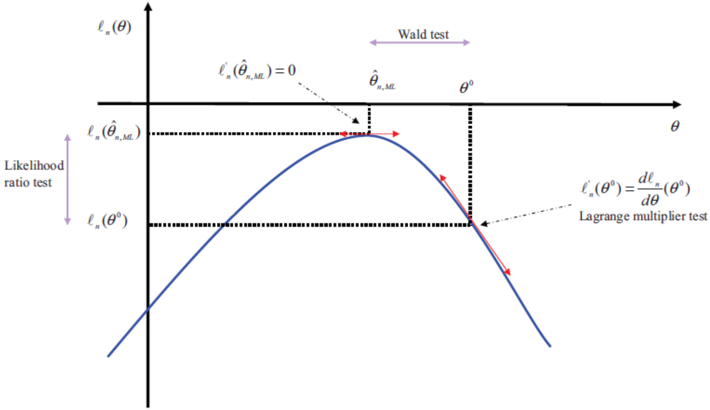
$$\text{p-value} = 1 - G_p\left(\text{LM}\left(y\right)\right)$$

where $\text{LM}(y)$ is the realisation of the LM test-statistic and $G_p\left(.\right)$ is the cdf of the chi-squared distribution with p degrees of freedom.

Subsection 4.4

A comparison of the three tests

4.4. A comparison of the three tests



Source: Pelgrin (2010), Lecture notes, Advanced Econometrics

4.4. A comparison of the three tests

Summary

Test	Requires estimation under
LRT	H_0 and H_1
Wald	H_1
LM	H_0

Computational problems

- If the ML maximisation problem is complex (with local extrema) and if it has no closed form solution (nonlinear models: GARCH, Markov Switching models etc.), it may be particularly difficult to get a ML estimates $\widehat{\boldsymbol{\theta}}$ through a numerical optimisation of the log-likelihood.
- If the constraints $\mathbf{c}(\boldsymbol{\theta}) = \mathbf{0}$ are not valid in the data, the (numerical) convergence of the optimisation algorithm may be very problematic under the null H_0.

4.4. A comparison of the three tests

Asymptotic comparison

The three tests have the same asymptotic distribution under the null $\mathrm{H}_0 : \mathbf{c}(\boldsymbol{\theta}) = \mathbf{0}$:

$$\mathrm{LRT} \underset{\mathrm{H}_0}{\overset{d}{\longrightarrow}} \chi^2(p)$$

$$\mathrm{Wald} \underset{\mathrm{H}_0}{\overset{d}{\longrightarrow}} \chi^2(p)$$

$$\mathrm{LM} \underset{\mathrm{H}_0}{\overset{d}{\longrightarrow}} \chi^2(p)$$

4.4. A comparison of the three tests

Theorem (Asymptotic comparison)

The three tests are **asymptotically equivalent**. *Under some regularity conditions and under the null* $H_0 : \mathbf{c}(\boldsymbol{\theta}) = \mathbf{0}$, *the differences between the three test statistics converge to 0 as* N *tends to infinity:*

$$LRT - LM \underset{H_0}{\overset{p}{\longrightarrow}} 0$$

$$LRT - Wald \underset{H_0}{\overset{p}{\longrightarrow}} 0$$

$$LM - Wald \underset{H_0}{\overset{p}{\longrightarrow}} 0$$

4.4. A comparison of the three tests

Fact (Finite sample properties)

The **finite sample properties** *of the three tests may be different, especially in small samples. For small sample size, they can lead to opposite conclusion about the rejection of the null hypothesis.*

Key concepts of Section 4

1. Likelihood Ratio (LR) test
2. Wald test
3. Lagrange Multiplier (LM) test
4. Computational issues
5. Comparison of the three tests (the trilogy) in finite samples

End of Chapter 4

Christophe Hurlin (University of Orléans)