Hypothesis testing

Hypothesis

Suppose, we want to test a hypothesis

$$H_0 : h(\theta) = 0 \tag{1}$$

Against the alternative

$$H_1 : h(\theta) \neq 0 \tag{2}$$

There are two types of estimators:

- Unrestricted estimator of the model, in which the constraint does not need to hold: $\hat{\theta}$.
- Restricted estimator of the model under the null, in which the constraint needs to be fulfilled: $\hat{\theta}_R$.

Example

Suppose

$$y \sim N(\mu, 1)$$

and we want to test the hypothesis

$$H_0 : \mu = 0$$

What are the restricted and unrestricted estimators?

Estimators

Lets denote by θ_0 the true parameter value. Then under the null

$$H_0 : h(\theta_0) = 0$$

there are

- $\hat{\theta}_R \to \theta_0$
- $\hat{\theta} \to \theta_0$
- $h(\hat{\theta}) \to 0$

Hypothesis testing

We will discuss three testing procedures:

- W test statistic (based on the behavior of the restriction function for the unrestricted model)
- LM test statistic (based on the gradient of the log likelihood function for the restricted model parameters)
- LR test statistic (based on the difference between the log likelihood of the unrestricted and restricted model)

Hypothesis testing

We will discuss three testing procedures:

- W test statistic (based on the behavior of the restriction function for the unrestricted model)
- LM test statistic (based on the gradient of the log likelihood function for the restricted model parameters)
- LR test statistic (based on the difference between the log likelihood of the unrestricted and restricted model)

Hypothesis testing

We will discuss three testing procedures:

- W test statistic (based on the behavior of the restriction function for the unrestricted model)
- LM test statistic (based on the gradient of the log likelihood function for the restricted model parameters)
- LR test statistic (based on the difference between the log likelihood of the unrestricted and restricted model)

Further notation

Notation:

- θ is a $(K \times 1)$ vector of parameters.
- $Dl(\theta)$ is a $(K \times 1)$ vector of first derivatives $\partial l(\theta)/\partial\theta$.
- $D^2l(\theta)$ is a $(K \times K)$ matrix of second derivatives of the log likelihood function.
- $I(\theta)$ is a $(K \times K)$ information matrix: $I(\theta) = -E(Dl^2(\theta))$.
- $i(\theta) = I(\theta)/N$ is a $(K \times K)$ matrix..
- $h(\theta)$ is a restriction function (a $(M \times 1)$ vector)
- $H(\theta)$ is a first derivative of $h(\theta)$ (a $(M \times K)$ matrix).

Further notation

Notation:

- θ is a $(K \times 1)$ vector of parameters.
- $Dl(\theta)$ is a $(K \times 1)$ vector of first derivatives $\partial l(\theta)/\partial\theta$.
- $D^2l(\theta)$ is a $(K \times K)$ matrix of second derivatives of the log likelihood function.
- $I(\theta)$ is a $(K \times K)$ information matrix: $I(\theta) = -E(Dl^2(\theta))$.
- $i(\theta) = I(\theta)/N$ is a $(K \times K)$ matrix..
- $h(\theta)$ is a restriction function (a $(M \times 1)$ vector)
- $H(\theta)$ is a first derivative of $h(\theta)$ (a $(M \times K)$ matrix).

Further notation

Notation:

- θ is a $(K \times 1)$ vector of parameters.
- $Dl(\theta)$ is a $(K \times 1)$ vector of first derivatives $\partial l(\theta)/\partial\theta$.
- $D^2 l(\theta)$ is a $(K \times K)$ matrix of second derivatives of the log likelihood function.
- $I(\theta)$ is a $(K \times K)$ information matrix: $I(\theta) = -E(Dl^2(\theta))$.
- $i(\theta) = I(\theta)/N$ is a $(K \times K)$ matrix..
- $h(\theta)$ is a restriction function (a $(M \times 1)$ vector)
- $H(\theta)$ is a first derivative of $h(\theta)$ (a $(M \times K)$ matrix).

Further notation

Notation:

- θ is a $(K \times 1)$ vector of parameters.
- $Dl(\theta)$ is a $(K \times 1)$ vector of first derivatives $\partial l(\theta)/\partial\theta$.
- $D^2 l(\theta)$ is a $(K \times K)$ matrix of second derivatives of the log likelihood function.
- $I(\theta)$ is a $(K \times K)$ information matrix: $I(\theta) = -E(Dl^2(\theta))$.
- $i(\theta) = I(\theta)/N$ is a $(K \times K)$ matrix..
- $h(\theta)$ is a restriction function (a $(M \times 1)$ vector)
- $H(\theta)$ is a first derivative of $h(\theta)$ (a $(M \times K)$ matrix).

Further notation

Notation:

- θ is a $(K \times 1)$ vector of parameters.
- $Dl(\theta)$ is a $(K \times 1)$ vector of first derivatives $\partial l(\theta)/\partial\theta$.
- $D^2 l(\theta)$ is a $(K \times K)$ matrix of second derivatives of the log likelihood function.
- $I(\theta)$ is a $(K \times K)$ information matrix: $I(\theta) = -E(Dl^2(\theta))$.
- $i(\theta) = I(\theta)/N$ is a $(K \times K)$ matrix..
- $h(\theta)$ is a restriction function (a $(M \times 1)$ vector)
- $H(\theta)$ is a first derivative of $h(\theta)$ (a $(M \times K)$ matrix).

Further notation

Notation:

- θ is a $(K \times 1)$ vector of parameters.
- $Dl(\theta)$ is a $(K \times 1)$ vector of first derivatives $\partial l(\theta)/\partial\theta$.
- $D^2 l(\theta)$ is a $(K \times K)$ matrix of second derivatives of the log likelihood function.
- $I(\theta)$ is a $(K \times K)$ information matrix: $I(\theta) = -E(Dl^2(\theta))$.
- $i(\theta) = I(\theta)/N$ is a $(K \times K)$ matrix..
- $h(\theta)$ is a restriction function (a $(M \times 1)$ vector)
- $H(\theta)$ is a first derivative of $h(\theta)$ (a $(M \times K)$ matrix).

Further notation

Notation:

- θ is a $(K \times 1)$ vector of parameters.
- $Dl(\theta)$ is a $(K \times 1)$ vector of first derivatives $\partial l(\theta)/\partial\theta$.
- $D^2 l(\theta)$ is a $(K \times K)$ matrix of second derivatives of the log likelihood function.
- $I(\theta)$ is a $(K \times K)$ information matrix: $I(\theta) = -E(Dl^2(\theta))$.
- $i(\theta) = I(\theta)/N$ is a $(K \times K)$ matrix..
- $h(\theta)$ is a restriction function (a $(M \times 1)$ vector)
- $H(\theta)$ is a first derivative of $h(\theta)$ (a $(M \times K)$ matrix).

Wald (W) test

Wald test statistic

In the Wald test statistic we examine, how far is $h(\hat{\theta})$ from zero.

Under the null, the restriction function should converge to zero with growing sample size.

Wald statistic

From the Taylor approximation of $h(\hat{\theta})$ it follows that

$$h(\hat{\theta}) = h(\theta_0) + H(\theta_0)(\hat{\theta} - \theta_0) \tag{3}$$

Since under the null

$$h(\theta_0) = 0$$

then (4) takes the form

$$h(\hat{\theta}) = H(\theta_0)\left(\hat{\theta} - \theta_0\right) \tag{4}$$

Wald statistic

From the Taylor approximation of $h(\hat{\theta})$ it follows that

$$h(\hat{\theta}) = h(\theta_0) + H(\theta_0)(\hat{\theta} - \theta_0) \tag{3}$$

Since under the null

$$h(\theta_0) = 0$$

then (4) takes the form

$$h(\hat{\theta}) = H(\theta_0)\left(\hat{\theta} - \theta_0\right) \tag{4}$$

Wald statistic

From the Taylor approximation of $h(\hat{\theta})$ it follows that

$$h(\hat{\theta}) = h(\theta_0) + H(\theta_0)(\hat{\theta} - \theta_0) \tag{3}$$

Since under the null

$$h(\theta_0) = 0$$

then (4) takes the form

$$h(\hat{\theta}) = H(\theta_0)\left(\hat{\theta} - \theta_0\right) \tag{4}$$

Wald statistic

The asymptotic distribution of the parameters

$$\sqrt{N}\left(\hat{\theta}-\theta_0\right) \to N\left(0, i^{-1}(\theta_0)\right)$$

so

$$\sqrt{N}h(\hat{\theta}) \to N\left(0, H(\theta_0)i^{-1}(\theta_0)H'(\theta_0)\right) \tag{5}$$

Wald statistic

The asymptotic distribution of the parameters

$$\sqrt{N}\left(\hat{\theta}-\theta_0\right) \rightarrow N\left(0, i^{-1}(\theta_0)\right)$$

so

$$\sqrt{N}h(\hat{\theta}) \rightarrow N\left(0, H(\theta_0)i^{-1}(\theta_0)H'(\theta_0)\right) \tag{5}$$

Wald statistic

All together

$$W = h'(\hat{\theta}) \left(H(\theta_0) I^{-1}(\theta_0) H'(\theta_0) \right)^{-1} h(\hat{\theta}) \rightarrow \chi^2(M) \qquad (6)$$

The Wald statistic depends on the way, we formulate the restrictions.

Example 1 - Model setup

Suppose we want to test linear restrictions of a form

$$H_0 : R\theta = r$$

then

$$h(\theta) = R\theta - r$$

$$H(\theta) = R'$$

Example 1

The Wald statistic is

$$W = [R\hat{\theta} - r]'[RI^{-1}(\hat{\theta})R']^{-1}[R\hat{\theta} - r]$$

Remark:

- It can be computed on the basis of the unrestricted model
- It is useful if we want to procedure a few statistical tests for the same data and the model structure.

Example 1

Suppose, that the model is linear

$$y_t = x_t\beta + \varepsilon_t$$

with normal residuals

$$\varepsilon_t \sim N(0, \sigma^2)$$

Then

$$I(\theta_0) = \begin{bmatrix} \frac{X'X}{\sigma^2} & 0 \\ 0 & \frac{N}{2\sigma^4} \end{bmatrix}$$

Notice: the information matrix consists of two blocks. Therefore, it is easy to compute its inverse.

Example 1

Suppose, that the model is linear

$$y_t = x_t\beta + \varepsilon_t$$

with normal residuals

$$\varepsilon_t \sim N(0, \sigma^2)$$

Then

$$I(\theta_0) = \begin{bmatrix} \frac{X'X}{\sigma^2} & 0 \\ 0 & \frac{N}{2\sigma^4} \end{bmatrix}$$

Notice: the information matrix consists of two blocks. Therefore, it is easy to compute its inverse.

Example 1

Suppose, that the model is linear

$$y_t = x_t\beta + \varepsilon_t$$

with normal residuals

$$\varepsilon_t \sim N(0, \sigma^2)$$

Then

$$I(\theta_0) = \begin{bmatrix} \frac{X'X}{\sigma^2} & 0 \\ 0 & \frac{N}{2\sigma^4} \end{bmatrix}$$

Notice: the information matrix consists of two blocks. Therefore, it is easy to compute its inverse.

Example 1

The inverse of the information matrix

$$I(\theta_0)^{-1} = \begin{bmatrix} \sigma^2(X'X)^{-1} & 0 \\ 0 & \frac{2\sigma^4}{N} \end{bmatrix}$$

$$W = [R\hat{\theta} - r]'[R\hat{\sigma}^2(X'X)^{-1}R']^{-1}[R\hat{\theta} - r]$$

Example 1

The inverse of the information matrix

$$I(\theta_0)^{-1} = \begin{bmatrix} \sigma^2(X'X)^{-1} & 0 \\ 0 & \frac{2\sigma^4}{N} \end{bmatrix}$$

$$W = [R\hat{\theta} - r]'[R\hat{\sigma}^2(X'X)^{-1}R']^{-1}[R\hat{\theta} - r]$$

Example 1

Finally,

$$W = [R\hat{\theta} - r]'[R(X'X)^{-1}R']^{-1}[R\hat{\theta} - r]/\hat{\sigma}^2$$

Suppose, we have a following type of restrictions

$$R\theta = 0$$

then

$$W = \hat{\theta}'R'[R(X'X)^{-1}R']^{-1}R\hat{\theta}/\hat{\sigma}^2$$

Example 1

Finally,

$$W = [R\hat{\theta} - r]'[R(X'X)^{-1}R']^{-1}[R\hat{\theta} - r]/\hat{\sigma}^2$$

Suppose, we have a following type of restrictions

$$R\theta = 0$$

then

$$W = \hat{\theta}'R'[R(X'X)^{-1}R']^{-1}R\hat{\theta}/\hat{\sigma}^2$$

Example 1

Finally,

$$W = [R\hat{\theta} - r]'[R(X'X)^{-1}R']^{-1}[R\hat{\theta} - r]/\hat{\sigma}^2$$

Suppose, we have a following type of restrictions

$$R\theta = 0$$

then

$$W = \hat{\theta}'R'[R(X'X)^{-1}R']^{-1}R\hat{\theta}/\hat{\sigma}^2$$

Example 1

Suppose, there are three variables in the model and $\beta = [b_1, b_2, b_3]'$. What is the Wald statistic for hypothesis:

$$H_0 : b_1 = b_2 = 0$$

Lets denote $b = [b_1, b_2]'$ and

$$R = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}$$

Then

$$W = b'(R(X'X)^{-1}R')^{-1}b/\hat{\sigma}^2$$

Example 1

Suppose, there are three variables in the model and $\beta = [b_1, b_2, b_3]'$. What is the Wald statistic for hypothesis:

$$H_0 : b_1 = b_2 = 0$$

Lets denote $b = [b_1, b_2]'$ and

$$R = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}$$

Then

$$W = b'(R(X'X)^{-1}R')^{-1}b/\hat{\sigma}^2$$

Example 1

Suppose, there are three variables in the model and $\beta = [b_1, b_2, b_3]'$. What is the Wald statistic for hypothesis:

$$H_0 : b_1 = b_2 = 0$$

Lets denote $b = [b_1, b_2]'$ and

$$R = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}$$

Then

$$W = b'(R(X'X)^{-1}R')^{-1}b/\hat{\sigma}^2$$

Example 1

$R(X'X)^{-1}R'$ is the upper block of the matrix $(X'X)^{-1}$. Does it mean that $(R(X'X)^{-1}R')^{-1}$ is the upper block of the matrix $X'X$? NO

For

$$M = \begin{bmatrix} A & B \\ B' & C \end{bmatrix}$$

there is

$$M^{-1} = \begin{bmatrix} (A - BC^{-1}B')^{-1} & * \\ * & * \end{bmatrix}$$

Example 1

$R(X'X)^{-1}R'$ is the upper block of the matrix $(X'X)^{-1}$. Does it mean that $(R(X'X)^{-1}R')^{-1}$ is the upper block of the matrix $X'X$? NO
For

$$M = \begin{bmatrix} A & B \\ B' & C \end{bmatrix}$$

there is

$$M^{-1} = \begin{bmatrix} (A - BC^{-1}B')^{-1} & * \\ * & * \end{bmatrix}$$

Model of unemployment

Lets model the dependance of unemployment on the GDP growth and the interest rate. US data for Q1.1954-Q1.2012. The model is linear and we assume normality of the residuals

$$u_t = \alpha + \beta_1 \Delta GDP_t + \beta_2 i_t + \varepsilon_t$$

We want to test if the unemployment rate depends on the exogenous variables

$$H_0 : \beta_1 = \beta_2 = 0$$

and

$$H_0 : \beta_2 = 0$$

Model of unemployment

Lets denote $\theta = [\alpha, \beta_1, \beta_2]'$. Then for

$$h(\theta) = R\theta$$

there is

$$R_1 = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

and

$$R_2 = \begin{bmatrix} 0 & 0 & 1 \end{bmatrix}$$

Model of unemployment - results

parameter	Estimate	St.dev
α	6.4242	0.2014
β_1	-0.2193	0.0384
β_2	0.0386	0.0283
H_0	W	p-value
$\beta_1 = \beta_2 = 0$	32.8524	0
$\beta_2 = 0$	1.8624	0.1723

So, we can reject the null $H_0 : \beta_1 = \beta_2 = 0$ but can not reject $H_0 : \beta_2 = 0$.

Model of unemployment - results

parameter	Estimate	St.dev
α	6.4242	0.2014
β_1	-0.2193	0.0384
β_2	0.0386	0.0283

H_0	W	p-value
$\beta_1 = \beta_2 = 0$	32.8524	0
$\beta_2 = 0$	1.8624	0.1723

So, we can reject the null $H_0 : \beta_1 = \beta_2 = 0$ but can not reject $H_0 : \beta_2 = 0$.

Example 2

Suppose, we want to test, if in a linear model with normally distributed residuals

$$y_t = \beta_0 + \beta_1 x_t + \varepsilon_t$$

the parameters satisfy the condition:

$$H_0 : \beta_0\beta_1 = 1$$

Example 2

In a linear model, the first block of the information matrix is

$$I^{-1}(\beta) = \sigma^2(X'X)^{-1}$$

The restriction function

$$h(\beta) = \beta_0\beta_1 - 1$$

and therefore

$$H(\beta) = [\beta_1, \beta_0]$$

Example 2

The Wald statistic is

$$W = \frac{1}{\hat{\sigma}^2} \frac{(\hat{\beta}_0 \hat{\beta}_1 - 1)^2}{[\hat{\beta}_0, \hat{\beta}_1](X'X)^{-1}[\hat{\beta}_0, \hat{\beta}_1]'}$$

Example 2

If we formulate the restriction differently

$$h(\beta) = \beta_0 - \frac{1}{\beta_1}$$

then

$$H(\beta) = \left[1, \frac{1}{\beta_1^2}\right]$$

and

$$W = \frac{1}{\hat{\sigma}^2} \frac{\left(\hat{\beta}_0 - \frac{1}{\hat{\beta}_1}\right)^2}{\left[1, \frac{1}{\hat{\beta}_1{}^2}\right] (X'X)^{-1} \left[1, \frac{1}{\hat{\beta}_1{}^2}\right]'}$$

Example 2

If we formulate the restriction differently

$$h(\beta) = \beta_0 - \frac{1}{\beta_1}$$

then

$$H(\beta) = \left[1, \frac{1}{\beta_1^2}\right]$$

and

$$W = \frac{1}{\hat{\sigma}^2} \frac{\left(\hat{\beta}_0 - \frac{1}{\hat{\beta}_1}\right)^2}{\left[1, \frac{1}{\hat{\beta}_1{}^2}\right] (X'X)^{-1} \left[1, \frac{1}{\hat{\beta}_1{}^2}\right]'}$$

Wald statistic

Notice

- the formula of the W test depends on the form of the restriction function
- asymptotically, the results are equivalent but in small samples may differ significantly.