Hypothesis testing

Lagrange Multipliers (LM) test

Lagrange multiplier test

In order to derive the ML estimator for the restricted model, we can use the Lagrange function

$$\mathcal{L}(\theta) = l(\theta) + \lambda' h(\theta)$$

where λ is a $(g \times 1)$ vector of Lagrange multipliers.

First order conditions

The first order conditions:

$$\frac{\partial \mathcal{L}(\theta)}{\partial \theta} = DI(\theta) + H'(\theta)\lambda = 0$$

$$h(\theta) = 0$$

So for the restricted estimator $\hat{\theta}_R$

$$DI(\hat{\theta}_R) + H'(\hat{\theta}_R)\hat{\lambda} = 0 \tag{1}$$

$$h(\hat{\theta}_R) = 0$$

First order conditions

The first order conditions:

$$\frac{\partial \mathcal{L}(\theta)}{\partial \theta} = DI(\theta) + H'(\theta)\lambda = 0$$

$$h(\theta) = 0$$

So for the restricted estimator $\hat{\theta}_R$

$$DI(\hat{\theta}_R) + H'(\hat{\theta}_R)\hat{\lambda} = 0 \quad (1)$$

$$h(\hat{\theta}_R) = 0$$

First order conditions

The first order conditions:

$$\frac{\partial \mathcal{L}(\theta)}{\partial \theta} = DI(\theta) + H'(\theta)\lambda = 0$$

$$h(\theta) = 0$$

So for the restricted estimator $\hat{\theta}_R$

$$DI(\hat{\theta}_R) + H'(\hat{\theta}_R)\hat{\lambda} = 0 \tag{1}$$

$$h(\hat{\theta}_R) = 0$$

First order conditions

The first order conditions:

$$\frac{\partial \mathcal{L}(\theta)}{\partial \theta} = Dl(\theta) + H'(\theta)\lambda = 0$$

$$h(\theta) = 0$$

So for the restricted estimator $\hat{\theta}_R$

$$Dl(\hat{\theta}_R) + H'(\hat{\theta}_R)\hat{\lambda} = 0 \tag{1}$$

$$h(\hat{\theta}_R) = 0$$

Taylor approximation of $Dl(\hat{\theta}_R)$

Since

$$H'(\hat{\theta}_R)\hat{\lambda} = -Dl(\hat{\theta}_R)$$

Using the Taylor approximation of $Dl(\hat{\theta}_R)$

$$H(\hat{\theta}_R)Dl^2(\hat{\theta}_R)^{-1}H'(\hat{\theta}_R)\hat{\lambda} = H(\hat{\theta}_R)(\hat{\theta}_R - \hat{\theta})$$

Taylor approximation of $Dl(\hat{\theta}_R)$

Since

$$H'(\hat{\theta}_R)\hat{\lambda} = -Dl(\hat{\theta}_R)$$

Using the Taylor approximation of $Dl(\hat{\theta}_R)$

$$H(\hat{\theta}_R)Dl^2(\hat{\theta}_R)^{-1}H'(\hat{\theta}_R)\hat{\lambda} = H(\hat{\theta}_R)(\hat{\theta}_R - \hat{\theta})$$

LM

Under the null $h(\hat{\theta}_R) = h(\theta_0) = 0$ then

$$h(\hat{\theta}) = H(\hat{\theta}_R)(\hat{\theta} - \hat{\theta}_R)$$

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

and

$$H(\hat{\theta}_R)(\hat{\theta} - \hat{\theta}_R) = H(\theta_0)(\hat{\theta} - \theta_0) \tag{2}$$

LM

Under the null $h(\hat{\theta}_R) = h(\theta_0) = 0$ then

$$h(\hat{\theta}) = H(\hat{\theta}_R)(\hat{\theta} - \hat{\theta}_R)$$

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

and

$$H(\hat{\theta}_R)(\hat{\theta} - \hat{\theta}_R) = H(\theta_0)(\hat{\theta} - \theta_0) \tag{2}$$

LM

Under the null $h(\hat{\theta}_R) = h(\theta_0) = 0$ then

$$h(\hat{\theta}) = H(\hat{\theta}_R)(\hat{\theta} - \hat{\theta}_R)$$

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

and

$$H(\hat{\theta}_R)(\hat{\theta} - \hat{\theta}_R) = H(\theta_0)(\hat{\theta} - \theta_0) \tag{2}$$

LM

Under the null $h(\hat{\theta}_R) = h(\theta_0) = 0$ then

$$h(\hat{\theta}) = H(\hat{\theta}_R)(\hat{\theta} - \hat{\theta}_R)$$

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

and

$$H(\hat{\theta}_R)(\hat{\theta} - \hat{\theta}_R) = H(\theta_0)(\hat{\theta} - \theta_0) \tag{2}$$

Equation for Lagrange multiplier

As the result

$$\sqrt{N}H(\hat{\theta}_R)DI^2(\hat{\theta}_R)^{-1}H'(\hat{\theta}_R)\hat{\lambda} = \sqrt{N}H(\theta_0)(\hat{\theta} - \theta_0)$$

Hence

$$\sqrt{N}H(\hat{\theta}_R)DI^2(\hat{\theta}_R)^{-1}H'(\hat{\theta}_R)\hat{\lambda} \rightarrow_d N(0, H(\theta_0)I^{-1}(\theta_0)H(\theta_0)')$$

Equation for Lagrange multiplier

As the result

$$\sqrt{N}H(\hat{\theta}_R)DI^2(\hat{\theta}_R)^{-1}H'(\hat{\theta}_R)\hat{\lambda} = \sqrt{N}H(\theta_0)(\hat{\theta} - \theta_0)$$

Hence

$$\sqrt{N}H(\hat{\theta}_R)DI^2(\hat{\theta}_R)^{-1}H'(\hat{\theta}_R)\hat{\lambda} \rightarrow_d N(0, H(\theta_0)I^{-1}(\theta_0)H(\theta_0)')$$

LM test statistic

As the result

$$N\hat{\lambda}' H(\hat{\theta}_R) DI^2(\hat{\theta}_R)^{-1} H'(\hat{\theta}_R)\hat{\lambda} \to \chi^2(M)$$

Since

$$H'(\hat{\theta}_R)\hat{\lambda} = -DI(\hat{\theta}_R)$$

Then

$$LM = DI(\hat{\theta}_R)' I^{-1}(\hat{\theta}_R) DI(\hat{\theta}_R) \to \chi^2(M)$$

LM test statistic

As the result

$$N\hat{\lambda}' H(\hat{\theta}_R) DI^2(\hat{\theta}_R)^{-1} H'(\hat{\theta}_R)\hat{\lambda} \rightarrow \chi^2(M)$$

Since

$$H'(\hat{\theta}_R)\hat{\lambda} = -DI(\hat{\theta}_R)$$

Then

$$LM = DI(\hat{\theta}_R)' I^{-1}(\hat{\theta}_R) DI(\hat{\theta}_R) \rightarrow \chi^2(M)$$

LM test statistic

Finally,

$$LM = DI(\hat{\theta}_R)' I^{-1}(\hat{\theta}_R) DI(\hat{\theta}_R) \rightarrow \chi^2(M) \tag{3}$$

So, the LM test statistic depends on the restricted parameters $\hat{\theta}_R$ and on the scores and information matrix of the unrestricted model computed for the restricted parameters.

LM interpretation

The greater the first derivative for the restricted parameters, the greater is the LM statistic and the easier it is to reject the null.

So, the LM statistic describe the curvature of the log likelihood function around $\hat{\theta}_R$. If we are close to the true parameters (and unrestricted estimators) the log likelihood function should be flat and the LM should be close to zero.

Example 1

Lets consider a linear regression model with normal residuals

$$y_n = x_n\beta + e_t$$

with

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

Suppose, we need to test a hypothesis:

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

Example 1

The score vector

$$DI(\theta) = [\frac{X'e}{\sigma^2}, -0.5\frac{N}{\sigma^2} + 0.5\frac{e'e}{\sigma^4}],$$

where $e = Y - X\beta$
The information matrix

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

Example 1

Score for the restricted parameters is

$$DI(\theta) = [\frac{X'e_R}{\sigma_R^2}, 0]$$

and the information matrix is

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

Example 1

The LM test statistic is

$$LM = DI(\hat{\theta}_R)' I^{-1}(\theta_0) DI(\hat{\theta}_R)$$

Hence in the linear regression model it is equal to

$$LM = e_R' X(X'X)^{-1} X' e_R / \sigma_R^2$$

Example 1

Let $\hat{e}_R$ be a fitted value from the regression

$$e_R = X\alpha + u$$

Hence

$$\hat{e}_R = X(X'X)^{-1}X'e_R$$

Then

$$LM = \hat{e}_R'\hat{e}_R/\sigma_R^2$$

Example 1

Notice that the determination coefficient of the regression

$$e_R = X\alpha + u$$

is

$$R^2 = \hat{e}_R' \hat{e}_R / \sigma_R^2$$

Hence

$$LM = R^2 \to \chi^2(M)$$

Example 1

How LM test can be used to test autocorrelation of residuals?

Example 2

Lets consider the Poisson model of the number of successes y_i. We want to test, if the expected number of successes is constant (do not depend on the exogenous variable x_i).
Model set up

$$f(y_i) = \frac{e^{-\lambda_i}\lambda_i^{y_i}}{y_i!}$$

where

$$\lambda_i = a + bx_i$$

and $\theta = (a, b)$

Example 2

λ_i measures the expected value of the variable y_i. If the expected value does not depend on the exogenous variable, then $\lambda_i = a$ and $b = 0$. Therefore, our hypothesis is

$$H_0 : b = 0$$

against

$$H_1 : b \neq 0$$

Example 2

The unrestricted model does not have a closed form solution and it is much easier to estimate the restricted model:

$$f(y_i) = \frac{e^{-a} a^{y_i}}{y_i!}$$

The restricted estimator is $\hat{a} = \sum_{i=1}^{N} y_i / N$ and

$$\hat{\theta}_R = (\hat{a}, 0)'$$

Example 2

In order to compute the LM test statistic we need to derive expressions for Dl and D^2l.

$$l(\theta) = \sum_{i=1}^{N} \left(y_i \ln(a + bx_i) - (a + bx_i)\right)$$

so

$$Dl(\theta) = \left[\sum_{i=1}^{N} \left(\frac{y_i}{a + bx_i} - 1\right), \sum_{i=1}^{N} \left(\frac{y_i}{a + bx_i} - 1\right)x_i\right]'$$

Example 2

In order to compute the LM test statistic we need to derive expressions for Dl and D^2l.

$$l(\theta) = \sum_{i=1}^{N} (y_i \ln(a + bx_i) - (a + bx_i))$$

so

$$Dl(\theta) = \left[\sum_{i=1}^{N} (\frac{y_i}{a + bx_i} - 1), \sum_{i=1}^{N} (\frac{y_i}{a + bx_i} - 1) x_i \right]'$$

Example 2

$$DI(\hat{\theta}_R) = \left[\sum_{i=1}^{N}(\frac{y_i}{\hat{a}} - 1), \sum_{i=1}^{N}(\frac{y_i}{\hat{a}} - 1)x_i\right]' \quad (4)$$

From the property of the estimator $\hat{a}$ it follows that

$$DI(\hat{\theta}_R) = \left[0, \sum_{i=1}^{N}(\frac{y_i}{\hat{a}} - 1)x_i\right]'$$

Example 2

$$DI(\hat{\theta}_R) = \left[\sum_{i=1}^{N}(\frac{y_i}{\hat{a}} - 1), \sum_{i=1}^{N}(\frac{y_i}{\hat{a}} - 1)x_i\right]' \tag{4}$$

From the property of the estimator $\hat{a}$ it follows that

$$DI(\hat{\theta}_R) = \left[0, \sum_{i=1}^{N}(\frac{y_i}{\hat{a}} - 1)x_i\right]'$$

Example 2

The second derivative is

$$D^2 l(\theta) = - \begin{bmatrix} \sum \frac{y_i}{(a+bx_i)^2} & \sum \frac{y_i x_i}{(a+bx_i)^2} \\ \sum \frac{y_i x_i}{(a+bx_i)^2} & \sum \frac{y_i x_i^2}{(a+bx_i)^2} \end{bmatrix}$$

So for $\hat{\theta}_R$

$$D^2 l(\hat{\theta}_R) = - \begin{bmatrix} \sum \frac{y_i}{\hat{a}^2} & \sum \frac{y_i x_i}{\hat{a}^2} \\ \sum \frac{y_i x_i}{\hat{a}^2} & \sum \frac{y_i x_i^2}{\hat{a}^2} \end{bmatrix}$$

Example 2

The second derivative is

$$D^2 l(\theta) = -\begin{bmatrix} \sum \frac{y_i}{(a+bx_i)^2} & \sum \frac{y_i x_i}{(a+bx_i)^2} \\ \sum \frac{y_i x_i}{(a+bx_i)^2} & \sum \frac{y_i x_i^2}{(a+bx_i)^2} \end{bmatrix}$$

So for $\hat{\theta}_R$

$$D^2 l(\hat{\theta}_R) = -\begin{bmatrix} \sum \frac{y_i}{\hat{a}^2} & \sum \frac{y_i x_i}{\hat{a}^2} \\ \sum \frac{y_i x_i}{\hat{a}^2} & \sum \frac{y_i x_i^2}{\hat{a}^2} \end{bmatrix}$$

Example 2

We may use the estimator of the information matrix

$$I(\hat{\theta}_R) = -D^2 l(\hat{\theta}_R)$$

and

$$LM = Dl(\hat{\theta}_R)' I(\hat{\theta}_R)^{-1} Dl(\hat{\theta}_R)$$

Example 2

We may use the estimator of the information matrix

$$I(\hat{\theta}_R) = -D^2 l(\hat{\theta}_R)$$

and

$$LM = Dl(\hat{\theta}_R)' I(\hat{\theta}_R)^{-1} Dl(\hat{\theta}_R)$$

Example 2

Lets use the data from the Survey file, where x_i is 1 if the person comes from a large town and 0 otherwise. The restricted estimator is

$$\hat{\theta}_R = [3.5455, 0]'$$

and

$$DI(\hat{\theta}_R) = [0, 1.4359]'$$

Example 2

The estimator of the information matrix is

$$I(\hat{\theta}_R) = \begin{bmatrix} 9.3077 & 2.9435 \\ 2.9435 & 2.9435 \end{bmatrix}$$

Example 2

The Lagrange multiplier test statistic is

$$LM = 1.0244$$

The p-value for a χ^2 distribution with one degree of freedom is

$$pvalue = 0.3115$$

So we cannot reject the null of constant expected value of y_i.

Example 2

The Lagrange multiplier test statistic is

$$LM = 1.0244$$

The p-value for a χ^2 distribution with one degree of freedom is

$$pvalue = 0.3115$$

So we cannot reject the null of constant expected value of y_i.

Likelihood ratio (LR) test

LR test statistic

Lets notice, that

$$l(\hat{\theta}_R) = l(\hat{\theta}) + Dl(\hat{\theta})(\hat{\theta}_R - \hat{\theta}) + \frac{1}{2}(\hat{\theta}_R - \hat{\theta})' \left[D^2 l(\hat{\theta}) \right] (\hat{\theta}_R - \hat{\theta})$$

Since

$$Dl(\hat{\theta}) = 0$$

then

$$l(\hat{\theta}_R) = l(\hat{\theta}) + \frac{1}{2}(\hat{\theta}_R - \hat{\theta})' \left[D^2 l(\hat{\theta}) \right] (\hat{\theta}_R - \hat{\theta})$$

LR test statistic

Finally,

$$LR = 2\left(l(\hat{\theta}_R) - l(\hat{\theta})\right) = (\hat{\theta}_R - \hat{\theta})' \left[-D^2 l(\hat{\theta})\right] (\hat{\theta}_R - \hat{\theta})$$

It can be shown that

$$LR \to \chi^2(M)$$

Distribution of parameters

In order to show the result, we need to derive the limiting distribution of the difference of parameters.

$$DI(\hat{\theta}_R) = DI(\hat{\theta}) + D^2I(\hat{\theta})(\hat{\theta}_R - \hat{\theta})$$

and

$$(\hat{\theta}_R - \hat{\theta}) = D^2I(\hat{\theta})^{-1}DI(\hat{\theta}_R)$$

Distribution of parameters

Form the derivation of the LM test we know that

$$Dl(\hat{\theta}_R) = -H'(\hat{\theta}_R)\lambda$$

so

$$(\hat{\theta}_R - \hat{\theta}) = -D^2 l(\hat{\theta})^{-1} H'(\hat{\theta}_R)\lambda$$

Moreover

$$N^{-1/2}\hat{\lambda} \to N\left(0, \left[H(\hat{\theta}_R) i^{-1}(\theta_0) H'(\hat{\theta}_R)\right]^{-1}\right)$$

LR test statistic

Since

$$\sqrt{N}(\hat{\theta}_R - \hat{\theta}) = -ND^2 l(\hat{\theta})^{-1} H'(\hat{\theta}_R) N^{-1/2} \lambda$$

Then

$$\sqrt{N}(\hat{\theta}_R - \hat{\theta}) \rightarrow N(0, Q)$$

where

$$Q = i^{-1} H' \left[H i^{-1} H' \right]^{-1} H i^{-1}$$

LR test statistic

It can be shown that the matrix

$$H' \left[Hi^{-1} H' \right]^{-1} H$$

is a generalized inverse of Q

$$QH' \left[Hi^{-1} H' \right]^{-1} HQ = Q$$

Then

$$(\hat{\theta}_R - \hat{\theta})' H' \left[Hi^{-1} H' \right]^{-1} H(\hat{\theta}_R - \hat{\theta}) \to \chi^2(M)$$

LR test statistic

From the LM condition and the distribution of the Lagrange multiplier it follows that

$$N^{-1/2}Dl(\hat{\theta}_R) = N^{-1/2}H'\lambda \to N\left(0, H'\left[Hi^{-1}H'\right]^{-1}H\right)$$

so

$$-\frac{1}{N}D^2l(\hat{\theta}_R) = \frac{1}{N}Dl(\hat{\theta}_R)Dl'(\hat{\theta}_R) \to H'\left[Hi^{-1}H'\right]^{-1}H$$

and

$$-D^2l(\hat{\theta}_R) \to H'\left[HI^{-1}H'\right]^{-1}H$$

LR test statistic

Since

$$(\hat{\theta}_R - \hat{\theta})' H' \left[H i^{-1} H' \right]^{-1} H(\hat{\theta}_R - \hat{\theta}) \to \chi^2(M)$$

and

$$-D^2 l(\hat{\theta}_R) = \rightarrow H' \left[H I^{-1} H' \right]^{-1} H$$

then

$$LR = (\hat{\theta}_R - \hat{\theta})' \left[-D^2 l(\hat{\theta}) \right] (\hat{\theta}_R - \hat{\theta}) \to \chi^2(M)$$

LR test statistic

The LR test statistic is the easiest to compute

$$LR = 2(l(\hat{\theta}) - l(\hat{\theta}_R))$$

The LR is the most popular from the ML tests:

- It has intuitive interpretation: the bigger the difference of the log likelihoods between the unrestricted and restricted models, the easier it is to reject the null.
- It does not require estimation of the information matrix and first derivatives of the log likelihood function.
- Requires estimation of both: restricted and unrestricted model

LR test statistic

The LR test statistic is the easiest to compute

$$LR = 2(l(\hat{\theta}) - l(\hat{\theta}_R))$$

The LR is the most popular from the ML tests:

- It has intuitive interpretation: the bigger the difference of the log likelihoods between the unrestricted and restricted models, the easier it is to reject the null.
- It does not require estimation of the information matrix and first derivatives of the log likelihood function.
- Requires estimation of both: restricted and unrestricted model

LR test statistic

The LR test statistic is the easiest to compute

$$LR = 2(l(\hat{\theta}) - l(\hat{\theta}_R))$$

The LR is the most popular from the ML tests:

- It has intuitive interpretation: the bigger the difference of the log likelihoods between the unrestricted and restricted models, the easier it is to reject the null.
- It does not require estimation of the information matrix and first derivatives of the log likelihood function.
- Requires estimation of both: restricted and unrestricted model

LR test statistic

The LR test statistic is the easiest to compute

$$LR = 2(l(\hat{\theta}) - l(\hat{\theta}_R))$$

The LR is the most popular from the ML tests:

- It has intuitive interpretation: the bigger the difference of the log likelihoods between the unrestricted and restricted models, the easier it is to reject the null.
- It does not require estimation of the information matrix and first derivatives of the log likelihood function.
- Requires estimation of both: restricted and unrestricted model

Tests comparison

	W	LM	LR
Models	unrestricted	restricted	both
Estimates	$I(\hat{\theta})$, $H(\hat{\theta})$, $h(\hat{\theta})$	$I(\hat{\theta}_R)$, $DI(\hat{\theta}_R)$	$\mathrm{l}(\hat{\theta})$, $\mathrm{l}(\hat{\theta}_R)$
Formula	$h'(HI^{-1}H')^{-1}h$	$DI'I^{-1}DI$	$2(l - l_R)$
Asy.distr.	χ^2_M	χ^2_M	χ^2_M

Example

Suppose, that we use a linear model

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

with normally distributed residuals $\varepsilon_t \sim N(\sigma^2)$. We want to test a hypothesis

$$H_0 : \alpha = 0$$

Example

The test statistics take the following form

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

where $R = [1, 0]$ and $X_t = [1, x_t]$.

$$LM = \frac{e_R' X(X'X)^{-1}X' e_R}{sigma_R^2}$$

with $e_R = Y - X\hat{\beta}_R$.

$$LR = T(\ln(\hat{\sigma}_R^2) - \ln(\hat{\sigma}^2))$$

Example

Lets consider a linear model of unemployment, in which a current level of unemployment is a function of its past observation.

$$un_t = \alpha + \beta un_{t-1} + \varepsilon_t$$

We estimate the model for the US data (from Q1 1948 to Q1 2012). Suppose, we want to test if

$$H_0 : \alpha = 0$$

Example

The test statistics have a limiting χ^2_1 distribution.

Test	Value	p-value
W	4.1103	0.0426
LM	4.0458	0.0443
LR	4.0779	0.0434

All the tests allow for rejecting the null for the significance level 5%. So we conclude, that the constant term is statistically significant.

Information criteria

Information criteria

Information criteria helps to compare models with different sets of parameters. It is not a testing procedure, but under certain condition some of them will give a consistent results.
Suppose, we want to choose the explanatory variables. Then the information criteria are:

$$AIC = -2l(\hat{\theta}) + 2K$$

$$BIC = -2l(\hat{\theta}) + K \ln N$$

$$HQ = -2l(\hat{\theta}) + 2K \ln \ln N$$

(AIC - Akaike, BIC - Bayes, HQ - Hannan-Quinn)
We choose the model that minimize the information criteria.

Information criteria for Gaussian models

For normally distributed residuals

$$-\frac{2}{N} l(\hat{\theta}_k) = \ln 2\pi + \ln \hat{\sigma}_k^2 + \frac{\frac{1}{N}\sum_{i=1}^{N} \hat{e}_{ki}^2}{\hat{\sigma}_k^2}$$

From the first order conditions it follows that

$$\hat{\sigma}_k^2 = \frac{1}{N}\sum_{i=1}^{N} \hat{e}_{ki}^2$$

so

$$-\frac{2}{N} l(\hat{\theta}_k) = \ln 2\pi + \ln \hat{\sigma}_k^2 + 1$$

Information criteria for Gaussian models

If we drop all components that does not depend on the choice of variables then

$$AIC = \ln \hat{\sigma}_k^2 + \frac{2K}{N}$$

$$BIC = \ln \hat{\sigma}_k^2 + \frac{K \ln N}{N}$$

$$HQ = \ln \hat{\sigma}_k^2 + \frac{2K \ln \ln N}{N}$$

It can be shown that

- The criteria BIC and HQ are consistent.
- The AIC criterium is not consistent but may produce superior forecasts

Information criteria

Information criteria are often used to choose

- Number of lags in the autoregressive models (AR, VAR models)
- Number of components in the mixture models
- Subsets of variables (although, if possible, a more formal testing procedure is preferred)

Information criteria prise for a increase of the log likelihoods and punish for the increase of the rise of the number of parameters (K)

Information criteria

Information criteria are often used to choose

- Number of lags in the autoregressive models (AR, VAR models)
- Number of components in the mixture models
- Subsets of variables (although, if possible, a more formal testing procedure is preferred)

Information criteria prise for a increase of the log likelihoods and punish for the increase of the rise of the number of parameters (K)

Information criteria

Information criteria are often used to choose

- Number of lags in the autoregressive models (AR, VAR models)
- Number of components in the mixture models
- Subsets of variables (although, if possible, a more formal testing procedure is preferred)

Information criteria prise for a increase of the log likelihoods and punish for the increase of the rise of the number of parameters (K)

Information criteria

Information criteria are often used to choose

- Number of lags in the autoregressive models (AR, VAR models)
- Number of components in the mixture models
- Subsets of variables (although, if possible, a more formal testing procedure is preferred)

Information criteria prise for a increase of the log likelihoods and punish for the increase of the rise of the number of parameters (K)

Information criteria

Information criteria are often used to choose

- Number of lags in the autoregressive models (AR, VAR models)
- Number of components in the mixture models
- Subsets of variables (although, if possible, a more formal testing procedure is preferred)

Information criteria prise for a increase of the log likelihoods and punish for the increase of the rise of the number of parameters (K)

Example

Suppose, we want to model the unemployment rate with the $AR(p)$ model. It means that

$$un_t = \alpha + \sum_{i=1}^{p} \beta_i un_{t-i} + \varepsilon_t$$

where p is the number of lags.
In order to choose the number of lags we may use the information criteria. It is important to choose the maximum number of lags and estimate models with the same sample length.

Exmaple

p	AIC	BIC	HQ
0	1.0365	1.0507	1.0422
1	-1.9105	-1.8820	-1.8990
2	-2.4325	-2.3898	-2.4153
3	-2.4323	-2.3755	-2.4095
4	-2.4260	-2.3549	-2.3974
5	-2.4248	-2.3395	-2.3905
6	-2.4262	-2.3267	-2.3861
7	-2.4219	-2.3083	-2.3762
8	-2.4177	-2.2898	-2.3662
9	-2.4319	-2.2898	-2.3747
10	-2.4729	-2.3166	-2.4100
11	-2.4654	-2.2949	-2.3967
12	-2.4672	-2.2825	-2.3928

Example

The information criteria indicate the following lags order p

	AIC	BIC	HQ
p	10	2	2

Thus, we may conclude that there are two lags.