UNIVERSITÀ DEGLI STUDI DI MILANO
Dipartimento di Economia, Management
e Metodi Quantitativi

Academic Year 2019-2020

Time Series Econometics

Fabrizio Iacone

Chapter 7: Asymptotic properties of parametric estimates

Topics: Asymptotic properties of the estimates based on the autocorrelation function; Asymptotic distribution of the OLS/CML estimates in an AR(p); Asymptotic distribution of (Pseudo) Maximum Likelihood estimates.

Estimates limit properties

Limit properties of the Correlogram based estimates, of the Maximum likelihood type estimates (either the exact one, or the "conditional" one) and the minRSS pseudo-maximum likelihood estimate.

Let

$$Y_t = c_0 + \phi_{0;1} Y_{t-1} + \ldots + \phi_{0;p} Y_{t-p} + \varepsilon_t + \theta_{0;1} \varepsilon_{t-1} + \ldots + \theta_{0;q} \varepsilon_{t-q},$$

$$\varepsilon_t \sim iid(0, \sigma_0^2)$$

the roots of $1 - \phi_{0;1} z - \ldots - \phi_{0;p} z^p = 0$ and of $1 + \theta_{0;1} z + \ldots + \theta_{0;q} z^q = 0$ are all outside the unit circle, and there is no common factor.

★For Maximum Likelihood estimates, we also assume $\varepsilon_t \sim Nid(0, \sigma_0^2)$;

★ For Conditional Maximum likelihood we also assume $\varepsilon_t \sim Nid(0, \sigma_0^2)$, $Y_p, \ldots, Y_1$ not random, $\varepsilon_p = 0, \ldots, \varepsilon_{p-q+1} = 0$.

Let

$$\boldsymbol{\beta}_0 = (c_0, \phi_{0;1}, \ldots, \phi_{0;p}, \theta_{0;1}, \ldots, \theta_{0;q})'$$

be the set of parameters of interest (i.e., all the parameters of the model except σ_0^2), and let $\widehat{\boldsymbol{\beta}}$ be one of the following estimates of $\boldsymbol{\beta}_0$

★ Correlogram based ($\widehat{\boldsymbol{\beta}}_C$)
★ Maximum likelihood types ($\widehat{\boldsymbol{\beta}}_{ML}$)
★ Pseudo maximum likelihood ($\widehat{\boldsymbol{\beta}}_{PML}$)

(i.e. $\widehat{\boldsymbol{\beta}}_C$ is the Correlogram based estimate of $\boldsymbol{\beta}_0$, $\widehat{\boldsymbol{\beta}}_{ML}$ is any Maximum likelihood type estimate of $\boldsymbol{\beta}_0$, $\widehat{\boldsymbol{\beta}}_{PML}$ is a minRSS Pseudo-Maximum likelihood type estimate of $\boldsymbol{\beta}_0$).

Limit properties: consistency

Then

$$\hat{\boldsymbol{\beta}} \rightarrow_p \boldsymbol{\beta}_0 \text{ as } T \rightarrow \infty$$

i.e. as $T \rightarrow \infty$, $\hat{\boldsymbol{\beta}}$ (any of $\hat{\boldsymbol{\beta}}_C$, or of $\hat{\boldsymbol{\beta}}_{ML}$ or of $\hat{\boldsymbol{\beta}}_{PML}$) is a consistent estimate of $\boldsymbol{\beta}_0$.

It also holds that $\hat{\boldsymbol{\sigma}}_C^2 \rightarrow_p \sigma_0^2$, $\hat{\boldsymbol{\sigma}}_{ML}^2 \rightarrow_p \sigma_0^2$ and $\hat{\boldsymbol{\sigma}}_{PML}^2 \rightarrow \sigma_0^2$as $T \rightarrow \infty$ (where $\hat{\boldsymbol{\sigma}}_C^2$, $\hat{\boldsymbol{\sigma}}_{ML}^2$ and $\hat{\boldsymbol{\sigma}}_{PML}^2$ are the correlogram based, ML and PML estimates of σ_0^2, respectively).

Limit properties: asymptotic normality

$$\sqrt{T}\left(\hat{\boldsymbol{\beta}}_C - \boldsymbol{\beta}_0\right) \rightarrow_d N(0, \Sigma_C)$$

$$\sqrt{T}\left(\hat{\boldsymbol{\beta}}_{ML} - \boldsymbol{\beta}_0\right) \rightarrow_d N(0, \Sigma_{ML})$$

$$\sqrt{T}\left(\hat{\boldsymbol{\beta}}_{PML} - \boldsymbol{\beta}_0\right) \rightarrow_d N(0, \Sigma_{ML})$$

as $T \rightarrow \infty$, $\sqrt{T}\left(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\right)$ is asymptotically normally distributed. Notice however the dispersion is, in general, different.

★ Both the matrices Σ_C and Σ_{ML} are positive definite.

★ The ML/PML estimate is at least as efficient the Correlogram based one, i.e. $\Sigma_C - \Sigma_{ML}$ is a positive semidefinite matrix.

★ The Correlogram based estimate and the PML estimates of $\phi_{0,1}$, ..., $\phi_{0,p}$ are as efficient as the ML/PML estimates of them, if $\theta_{0;1} = 0, \ldots, \theta_{0;q} = 0$ (ie the true model is AR(p)).

★ If we are also interested in the estimation of σ_0^2,

$$\boldsymbol{\beta}_0 = (c_0, \phi_{0;1}, \ldots, \phi_{0;p}, \theta_{0;1}, \ldots, \theta_{0;q}, \sigma_0^2)'$$

let $\widehat{\boldsymbol{\beta}}_{ML}$ be the exact ML estimate, then $\sqrt{T}\left(\widehat{\boldsymbol{\beta}}_{ML} - \boldsymbol{\beta}_0\right) \rightarrow_d N(0, \Xi_{ML})$ The matrix Ξ_{ML} is often referred to as $\mathfrak{I}^{-1}$, where

$$\mathfrak{I} = -E\left(\frac{1}{T} \frac{\partial^2 \mathcal{L}(\boldsymbol{\beta})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}'} \bigg|_{\boldsymbol{\beta}=\boldsymbol{\beta}_0} \right)$$

is called information matrix.

As usual, the standard errors can be seen as a measure of the precision of the estimate, and can be also used in testing.

Examples of Σ_{ML}:

$$AR(1): \sqrt{T}\left(\widehat{\phi} - \phi_0\right) \rightarrow_d N(0, 1 - \phi_0^2)$$

$$AR(2): \sqrt{T}\left(\begin{bmatrix} \widehat{\phi}_1 \\ \widehat{\phi}_2 \end{bmatrix} - \begin{bmatrix} \phi_{0;1} \\ \phi_{0;2} \end{bmatrix}\right)$$

$$\rightarrow_d N\left(0, \begin{bmatrix} 1 - \phi_{0;2}^2 & -\phi_{0;1}(1 + \phi_{0;2}) \\ -\phi_{0;1}(1 + \phi_{0;2}) & 1 - \phi_{0;2}^2 \end{bmatrix}\right)$$

$$MA(1): \sqrt{T}\left(\widehat{\theta} - \theta_0\right) \rightarrow_d N(0, 1 - \theta_0^2)$$

$$MA(2): \sqrt{T}\left(\begin{bmatrix} \widehat{\theta}_1 \\ \widehat{\theta}_2 \end{bmatrix} - \begin{bmatrix} \theta_{0;1} \\ \theta_{0;2} \end{bmatrix}\right)$$

$$\rightarrow_d N\left(0, \begin{bmatrix} 1 - \theta_{0;2}^2 & -\theta_{0;1}(1 - \theta_{0;2}) \\ -\theta_{0;1}(1 - \theta_{0;2}) & 1 - \theta_{0;2}^2 \end{bmatrix}\right)$$

$$ARMA(1,1): \sqrt{T}\left(\begin{bmatrix} \widehat{\phi} \\ \widehat{\theta} \end{bmatrix} - \begin{bmatrix} \phi_0 \\ \theta_0 \end{bmatrix}\right)$$

$$\rightarrow_d N\left(0, \begin{bmatrix} (1-\phi_0^2)^{-1} & (1+\phi_0\theta_0)^{-1} \\ (1+\phi_0\theta_0)^{-1} & (1-\theta_0^2)^{-1} \end{bmatrix}^{-1}\right)$$

the last variance can be rewritten as

$$\frac{1+\phi_0\theta_0}{(\phi_0+\theta_0)^2} \times \begin{bmatrix} (1-\phi_0^2)(1+\phi_0\theta_0) & -(1-\theta_0^2)(1-\phi_0^2) \\ -(1-\phi_0^2)(1-\theta_0^2) & (1-\theta_0^2)(1+\phi_0\theta_0) \end{bmatrix}$$

★These do not depend on σ_0^2;

★The estimates in the AR(1), MA(1) are more precise the stronger the dependence.

It is easy to derive the limit distribution in the AR models: for example, AR(1), consider the Conditional maximum likelihood estimate (also assume $c_0 = 0$ and it is known)

$$\widehat{\phi} = \frac{\sum_{t=2}^{T} Y_{t-1}Y_t}{\sum_{t=2}^{T} Y_{t-1}^2} = \frac{\sum_{t=2}^{T} Y_{t-1}(\phi_0 Y_{t-1} + \varepsilon_t)}{\sum_{t=2}^{T} Y_{t-1}^2}$$

$$= \phi_0 + \frac{\frac{1}{T-1}\sum_{t=2}^{T} Y_{t-1}\varepsilon_t}{\frac{1}{T-1}\sum_{t=2}^{T} Y_{t-1}^2}$$

then look at

$$\sqrt{T}\left(\widehat{\phi} - \phi_0\right) = \frac{\sqrt{T}\frac{1}{T-1}\sum_{t=2}^{T} Y_{t-1}\varepsilon_t}{\frac{1}{T-1}\sum_{t=2}^{T} Y_{t-1}^2}$$

Clearly (by a Law of Large Number)

$$\frac{1}{T-1}\sum_{t=2}^{T} Y_{t-1}^2 \to_p E(Y_{t-1}^2) = \frac{\sigma_0^2}{1-\phi_0^2};$$

in $\sqrt{T}\frac{1}{T-1}\sum_{t=2}^{T} Y_{t-1}\varepsilon_t$, $Y_{t-1}\varepsilon_t$ is not actually independent, but it has similar properties, so, upon noticing that

$$E(Y_{t-1}\varepsilon_t) = E(Y_{t-1})E(\varepsilon_t) = 0,$$

$$V(Y_{t-1}\varepsilon_t) = V(Y_{t-1})V(\varepsilon_t) = \frac{\sigma_0^2}{1-\phi_0^2}\sigma_0^2,$$

by a Central Limit Theorem

$$\sqrt{T-1}\,\frac{1}{T-1}\sum_{t=2}^{T} Y_{t-1}\varepsilon_t \;\to_d\; N\left(0, \frac{\sigma_0^2}{1-\phi_0^2}\sigma_0^2\right)$$

combining the two, (using also the fact that $\sqrt{T}/\sqrt{T-1} \;\to\; 1$)

$$\sqrt{T}\left(\widehat{\phi} - \phi_0\right) \;\to_d\; N\left(0, \frac{\frac{\sigma_0^2}{1-\phi_0^2}\sigma_0^2}{\left(\frac{\sigma_0^2}{1-\phi_0^2}\right)^2}\right) = N(0, 1-\phi_0^2)$$

This can be generalised to AR(p),

$$\sqrt{T}\left(\begin{bmatrix} \widehat{\phi}_1 \\ \dots \\ \widehat{\phi}_p \end{bmatrix} - \begin{bmatrix} \phi_{0;1} \\ \dots \\ \phi_{0;p} \end{bmatrix}\right) \;\to_d\; N(0, V_p^{-1})$$

To prove the general result for the limit distribution, consider an approximate Taylor expansion of $\widehat{\sigma^2}\frac{1}{\sqrt{T}}g\left(\hat{\boldsymbol{\beta}}\right)$ in $\boldsymbol{\beta}_0$,

$$\widehat{\sigma^2}\frac{1}{\sqrt{T}}g\left(\hat{\boldsymbol{\beta}}\right)$$
$$\approx \sigma_0^2\frac{1}{\sqrt{T}}g(\boldsymbol{\beta}_0) - \sigma_0^2\frac{1}{T}H(\boldsymbol{\beta}_0)\sqrt{T}\left(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\right)$$

We know that

$$g\left(\hat{\boldsymbol{\beta}}\right) = 0;$$

$$\sigma_0^2\frac{1}{\sqrt{T}}g(\boldsymbol{\beta}_0) = -\frac{1}{\sqrt{T}}\sum_{t=p+1}^{T}\varepsilon_t(\boldsymbol{\beta})\frac{\partial\varepsilon_t(\boldsymbol{\beta})}{\partial\boldsymbol{\beta}}\bigg|_{\boldsymbol{\beta}=\boldsymbol{\beta}_0}$$
$$\rightarrow_d N\left(0, \sigma_0^2 E\left(\frac{\partial\varepsilon_t(\boldsymbol{\beta})}{\partial\boldsymbol{\beta}}\frac{\partial\varepsilon_t(\boldsymbol{\beta})}{\partial\boldsymbol{\beta}}'\bigg|_{\boldsymbol{\beta}=\boldsymbol{\beta}_0}\right)\right)$$

$$\sigma_0^2 \frac{1}{T} H(\boldsymbol{\beta}_0)$$

$$= -\frac{1}{T} \sum_{t=p+1}^{T} \left(\frac{\partial \varepsilon_t(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} \frac{\partial \varepsilon_t(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}}' + \varepsilon_t(\boldsymbol{\beta}) \frac{\partial^2 \varepsilon_t(\boldsymbol{\beta})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}'} \right) \Bigg|_{\boldsymbol{\beta}=\boldsymbol{\beta}_0}$$

$$\rightarrow_p E\left(\frac{\partial \varepsilon_t(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} \frac{\partial \varepsilon_t(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}}' + \varepsilon_t(\boldsymbol{\beta}) \frac{\partial^2 \varepsilon_t(\boldsymbol{\beta})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}'} \right) \Bigg|_{\boldsymbol{\beta}=\boldsymbol{\beta}_0}$$

$$= E\left(\frac{\partial \varepsilon_t(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} \frac{\partial \varepsilon_t(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}}' \right) \Bigg|_{\boldsymbol{\beta}=\boldsymbol{\beta}_0}$$

so rearranging terms

$$\sqrt{T}\left(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\right) \approx \left(\frac{1}{T} H(\boldsymbol{\beta}_0)\right)^{-1} \frac{1}{\sqrt{T}} g(\boldsymbol{\beta}_0)$$

so

$$\sqrt{T}\left(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\right)$$

$$\rightarrow_d N\left(0, \sigma_0^2 \left(E\left(\frac{\partial \varepsilon_t(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} \frac{\partial \varepsilon_t(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}}' \right) \Bigg|_{\boldsymbol{\beta}=\boldsymbol{\beta}_0} \right)^{-1} \right)$$

Example

MA(1) ($\mu_0 = 0$ and known)

$$Y_t = \varepsilon_t + \theta_0 \varepsilon_{t-1},\ \varepsilon_t \ N.i.d.(0, \sigma_0^2)$$

Compute $\sigma_0^2 \left(E\left(\frac{\partial \varepsilon_t(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} \frac{\partial \varepsilon_t(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}}' \right) \Big|_{\boldsymbol{\beta}=\boldsymbol{\beta}_0} \right)^{-1}$.

In this case,

$$\frac{\partial \varepsilon_t(\boldsymbol{\beta})}{\partial \theta} = -\varepsilon_{t-1}(\boldsymbol{\beta}) - \theta \frac{\partial \varepsilon_{t-1}(\boldsymbol{\beta})}{\partial \theta}$$

Introduce

$$z_t(\boldsymbol{\beta}) = -\frac{\partial \varepsilon_t(\boldsymbol{\beta})}{\partial \theta},$$

then the iteration above is

$$z_t(\boldsymbol{\beta}) = \varepsilon_{t-1}(\boldsymbol{\beta}) - \theta z_{t-1}(\boldsymbol{\beta})$$

and

$$z_t(\boldsymbol{\beta}_0) = \varepsilon_{t-1} - \theta_0 z_{t-1}(\boldsymbol{\beta}_0)$$

This is an AR(1) for $z_t(\boldsymbol{\beta}_0)$, so, using the Variance of an AR(1),

$$E(z_t(\boldsymbol{\beta}_0))^2 = \frac{\sigma_0^2}{1-\theta_0^2}$$

and

$$\sigma_0^2\left(E\left(\frac{\partial \varepsilon_t(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}}\frac{\partial \varepsilon_t(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}}'\right)\Bigg|_{\boldsymbol{\beta}=\boldsymbol{\beta}_0}\right)^{-1} = \frac{\sigma_0^2}{\frac{\sigma_0^2}{1-\theta_0^2}} = 1-\theta_0^2$$

Appendix

- Properties of the Correlogram Based estimates and Maximum Likelihood estimates

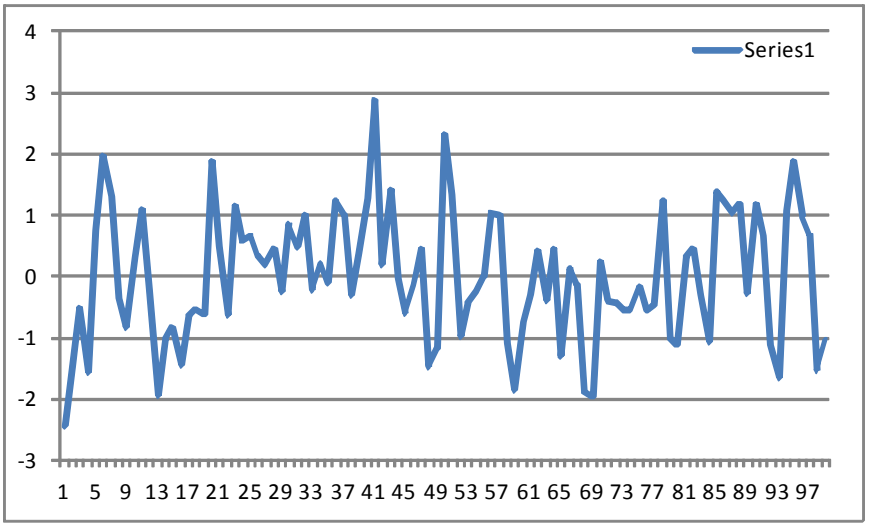
- Interpretation of the standard errors

Properties of the Correlogram Based estimate and Maximum Likelihood estimate

What does it mean to say that the Maximum Likelihood estimate is more precise than the Correlogram based estimate?

✠ Example 1. MA(1).

The series



was generated as MA(1) with $\theta = 0.5$.

★ If we pretend not to know θ, and we estimate it as correlogram based or maximum likelihood estimate,

$$\hat{\theta}_C = 0.35, \hat{\theta}_{ML} = 0.43$$

so in this particular example $\hat{\theta}_{ML}$ got closer to θ (so, it worked better).

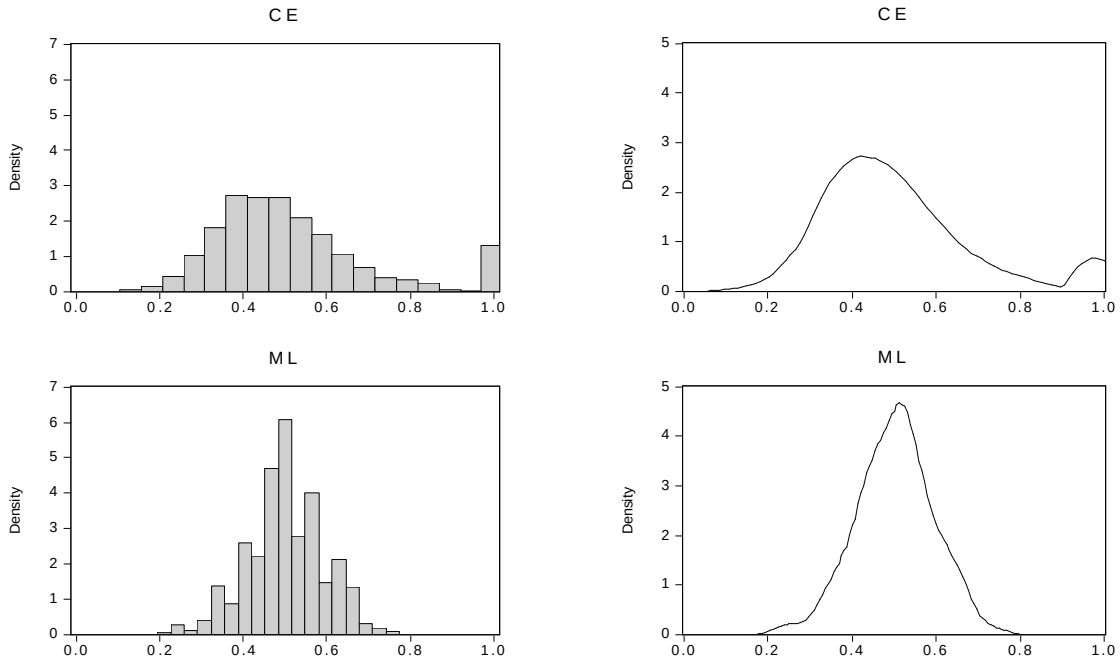
✠ Example 2. 1000s MA(1), an experiment.

I took 1000 random series from the same process:

★ the estimate $\widehat{\theta}_{ML}$ gets closer to 0.5 than $\widehat{\theta}_{C}$ does in 68.5% of the cases;

★ the standard error of the estimated values $\widehat{\theta}_{ML}$ is 0.075, the standard error of the estimated values $\widehat{\theta}_{C}$ is 0.104.

★ We can look at the whole sample distribution of the estimates (there are two ways to represent it, with histograms or with smooth functions). $\widehat{\theta}_{ML}$ clusters more estimated values around 0.5, and much less in points away from it.



All this means that $\widehat{\theta}_{ML}$ is more precise than $\widehat{\theta}_{C}$ in a statistical sense.

Interpretation of the standard errors and application to testing

The standard errors can be seen as a measure of the precision of the estimate, and can be also used in testing.

★ Example 1 (MA(1)). Consider the estimation of the parameter θ assuming that the true model is an (invertible) MA(1). Compare the asymptotic variance when a MA(1), a MA(2) and ARMA(1,1) are used. Notice that $\theta_{0;2}$ in the MA(2) is 0, and ϕ_0 in the ARMA(1,1) is 0.

Model	MA(1)	MA(2)	ARMA(1,1)
as. Var.	$(1-\theta_0^2)\times 1/T$	$1/T$	$\frac{1}{\theta_0^2}(1-\theta_0^2)\times 1/T$

The asymptotic variance in the MA(1) model is smaller. Heuristically, we may think that the information is used only to estimate θ, instead of dispersing it to estimate also θ_2 or ϕ.

✠ Example 2 (MA(1)).

Suppose that a MA(1) model is estimated (via ML/CML), with 100 observations, and $\widehat{\theta}$ takes value 0.8.

The standard error, $\sqrt{\frac{1-\theta_0^2}{T}}$ is not observable (because we do not know θ_0). The estimate takes value $\sqrt{\frac{1-0.8^2}{100}} = 0.06$.

If we want to test $H_0 : \{\theta_0 = \theta\}$ we use

$$\sqrt{T}\frac{\left(\widehat{\theta} - \theta_0\right)}{\sqrt{1-\theta_0^2}} \rightarrow_d N(0,1)$$

so for example, to test

$$H_0 : \{\theta_0 = 0.7\} \text{ vs } H_A : \{\theta_0 \neq 0.7\}$$

the test statistic under the null hypothesis takes value 1.4003, so the null hypothesis is not rejected.

✠ Example 3 (MA(2)).

Suppose that a MA(2) model is estimated (via ML/CML), with 100 observations, and $\widehat{\theta}_1$ takes value 0.8, $\widehat{\theta}_2$ takes value 0.05.

The standard error, $\sqrt{\frac{1-\theta_{0,2}^2}{T}}$ is not observable (because we do not know $\theta_{0;2}$). The estimate takes value $\sqrt{\frac{1-0.05^2}{100}} = 09.9875$.

If we want to test $H_0 : \{\theta_{0;1} = \theta\}$ we use

$$\sqrt{T}\frac{\left(\widehat{\theta} - \theta_{0;1}\right)}{\sqrt{1-\theta_{0;2}^2}} \rightarrow_d N(0,1)$$

Notice that this require knowledge of $\theta_{0;2}^2$, and this not know not even under H_0: we can, however, replace it by a consistent estimate ($\widehat{\theta}_2$).

So for example, to test

$$H_0 : \{\theta_{0;1} = 0.7\} \text{ vs } H_A : \{\theta_{0;1} \neq 0.7\}$$

the test statistic under the null hypothesis takes value 1.0013, so the null hypothesis is not rejected.

✠ Example 4 (ARMA(1,1)). Suppose that an ARMA(1,1) model is estimated (via ML/CML), with 100 observations, and $\widehat{\phi}$ takes value 0.8, $\widehat{\theta}$ takes value 0.05.

If we want to test $H_0 : \{\phi_0 = \phi, \theta_0 = \theta\}$ we use the Wald test statistic

$$T\begin{pmatrix} \widehat{\phi} - \phi_0 & \widehat{\theta} - \theta_0 \end{pmatrix} \times$$

$$\left(\begin{bmatrix} (1-\phi_0^2)^{-1} & (1+\phi_0\theta_0)^{-1} \\ (1+\phi_0\theta_0)^{-1} & (1-\theta_0^2)^{-1} \end{bmatrix}^{-1} \right)^{-1} \times$$

$$\begin{pmatrix} \widehat{\phi} - \phi_0 \\ \widehat{\theta} - \theta_0 \end{pmatrix} \rightarrow_d \chi^2_2$$

(i.e., the Wald test statistic is asymptotically χ^2_k distributed, with k equal to the number of parameters being tested).

So for example, to test

$$H_0 : \{\phi_0 = 0.7, \theta_0 = 0.2\}$$

vs

$$H_A : \left\{\phi_0 \neq 0.7, \text{ \&/or } \theta_0 = 0.2\right\}$$

the test statistic takes value 1.6730, so the null hypothesis is not rejected with size 5% (c.v. 5.99).