STAT4005 - Time Series
Tutorial 5 - Inference on Model Parameters
Ma Ting Tin, Martin
Week 9

5 Inference on Model Parameters

Assume $\{Y_t\} \sim \text{ARMA}(p, q)$ being stationary with $\text{E}(Y_t) = \mu$, i.e.

$$\phi(B)(Y_t - \mu) = \theta(B)Z_t,$$

where $p = \deg(\phi)$ and $q = \deg(\theta)$. We will discuss the following.

1. Inference of $(\phi_1, \ldots, \phi_p, \theta_1, \ldots, \theta_q)$ and μ given the order p and q.
2. Estimation of order p and q. (Included in the next tutorial)

5.1 Method of Moment Estimators

Method of Moment Estimators (MM)

Suppose $(X_1, \ldots, X_n) \sim f_\theta$, and for all $i = 1, \ldots, n$, $\text{E}\{g(X_i)\} = T(\theta)$ for some function g and T. Under regularity condition,

$$\overline{g(X)} = \frac{1}{n}\sum_{i=1}^{n} g(X_i) \xrightarrow{\text{pr}} \text{E}\{g(X_1)\} = T(\theta).$$

The method of moment suggest to **equate sample and population quantities** to solve for the estimator $\hat{\theta}_{\text{MM}} = \hat{\theta}_{\text{MM}}(X_1, \ldots, X_n)$, i.e. to solve

$$T(\hat{\theta}_{\text{MM}}) = \frac{1}{n}\sum_{i=1}^{n} g(X_i)$$

Remark 1. *There are several remarks on the method of moment.*

1. *For $\boldsymbol{\theta} \in \mathbb{R}^p$, we need at least p linearly dependent equation to estimate $\boldsymbol{\theta}$, i.e., assume $\text{E}\{g_k(X_i)\} = T_k(\theta)$ for $k = 1, \ldots, p$ and solve $T_k(\hat{\boldsymbol{\theta}}) = n^{-1}\sum_{i=1}^{n} g_k(X_i)$ for $k = 1, \ldots, p$.*
2. *The method of moment estimator is not unique; see exercise 1 and 2. They might be very close to each other in value, but are still distinct, because*
   - ***(Theoretically Biased Building Blocks)*** *$\text{E}(C_k) \neq \gamma(k)$ and $\text{E}(r_k) \neq \rho(k)$ in general.*
   - ***(Lack of "Infinity Sample")*** *$C_k \xrightarrow{\text{pr}} \gamma(k)$ and $r_k \xrightarrow{\text{pr}} \rho(k)$ require "$n \to \infty$"!*

Assume $\{Y_t\} \sim \text{ARMA}(p, q)$ being stationary with ACVF $\gamma(\cdot)$ and ACF $\rho(\cdot)$. The following are some common building blocks for the construction of MM-estimators in the context of time series.

Population Quantity	Estimator
μ	$\bar{Y}_n = \frac{1}{n}\sum_{t=1}^{n} Y_t$
$\gamma(k)$	$C_k = \frac{1}{n}\sum_{t=1}^{n-k}(Y_t - \bar{Y})(Y_{t+k} - \bar{Y})$
$\rho(k)$	$r_k = \frac{C_k}{C_0}$

(★☆☆) MM Estimator - Application in Classical Setting

✎ **Exercise 1.** *Assume* $X_1, \ldots, X_n \overset{iid}{\sim} \text{Po}(\mu)$.

(a) *By considering* $\text{E}(X_1) = \mu$, *suggest a MM-estimator* $\hat{\mu}^{(1)}_{\text{MM}}$ *of* μ.

(b) *By considering* $\text{Var}(X_1) = \mu$, *suggest a MM-estimator* $\hat{\mu}^{(2)}_{\text{MM}}$ *of* μ.

Solution

Recall that $\text{Var}(X) = \text{E}(X^2) - \text{E}(X)^2$, we have

$$\hat{\mu}^{(1)}_{\text{MM}} = \frac{1}{n}\sum_{i=1}^{N} X_i \quad \text{and} \quad \hat{\mu}^{(2)}_{\text{MM}} = \frac{1}{n}\sum_{i=1}^{N} X_i^2 - \left(\frac{1}{n}\sum_{i=1}^{N} X_i\right)^2$$

(★★☆) MM Estimator - Application in Time Series Setting

✎ **Exercise 2.** *Let* $Y_t = Z_t + \phi Z_{t-1} + 2\phi Z_{t-2}$, *where* $Z_t \sim \text{WN}(0, \sigma^2)$.

(a) *Is* $\{Y_t\}$ *stationary? If so, evaluate ACVF* $\gamma(\cdot)$ *in terms of* ϕ *and* σ^2

(b) *Suggest two approaches to construct the MM-estimator for* ϕ *and* σ^2.

Solution

(a) As $\{Y_t\} \sim \text{MA}(2)$, it is causal and therefore stationary, where its ACVF is given by

$$\gamma(k) = \begin{cases} (1+5\phi^2)\sigma^2 & , \text{ if } k = 0 \\ (\phi + 2\phi^2)\sigma^2 & , \text{ if } |k| = 1 \\ 2\phi\sigma^2 & , \text{ if } |k| = 2 \\ 0 & , \text{ if } |k| \geq 3 \end{cases}$$

(b) There is more than one possibility to do so.

- Notice that

$$\frac{\gamma(1)}{\gamma(2)} = \frac{1}{2} + \phi \quad \Rightarrow \quad \hat{\phi}^{(1)}_{\text{MM}} = \frac{C_1}{C_2} - \frac{1}{2} = \frac{2C_1 - C_2}{2C_2}$$

and

$$\sigma^2 = \frac{\gamma(2)}{2\phi} = \frac{\gamma^2(2)}{2\gamma(1) - \gamma(2)} \quad \Rightarrow \quad \hat{\sigma}^{(1)2}_{\text{MM}} = \frac{C_2^2}{2C_1 - C_2}.$$

- Notice that

$$\gamma(0) - 2.5\gamma(1) + 1.25\gamma(2) = \sigma^2 \quad \Rightarrow \quad \hat{\sigma}^{(2)2}_{\text{MM}} = C_0 - 2.5C_1 + 1.25C_2$$

and

$$\phi = \frac{\gamma(2)}{2\sigma^2} \quad \Rightarrow \quad \hat{\phi}^{(2)}_{\text{MM}} = \frac{C_2}{C_0 - 2.5C_1 + 1.25C_2}$$

Obviously we can see, $(\hat{\phi}^{(1)}_{\text{MM}}, \hat{\sigma}^{(1)2}_{\text{MM}}) \neq (\hat{\phi}^{(2)}_{\text{MM}}, \hat{\sigma}^{(2)2}_{\text{MM}})$ in general.

Recall for $\{Y_t\} \sim \mathrm{AR}(p)$, we would take $\mathrm{Cov}(\cdot, Y_{t-k})$ to obtain a system of linear equation in $\gamma(k)$'s.

- **(Previous Task)** Given $\phi_1, \ldots, \phi_p$, solve for ACVF $\{\gamma(\cdot)\}$.
- **(Current Task)** Given the sample ACVF $C_1, \ldots, C_p$, estimate $\hat{\phi}_1, \ldots, \hat{\phi}_p$. (Not including σ^2)

We can simply replace (ϕ_i, γ_k) by $(\hat{\phi}_i, C_k)$ in the Yule-Walker equation to obtain the estimators $\hat{\phi}_1, \ldots, \hat{\phi}_p$, and it is known as the **Yule-Walker Estimators**.

Yule-Walker Estimators (YW)

Definition 1. *Assume $\{Y_t\} \sim \mathrm{AR}(p)$ being causal, i.e. $Y_t = \phi_1 Y_{t-1} + \cdots + \phi_p Y_{t-p} + Z_t$, where $\{Z_t\} \sim \mathrm{WN}(0, \sigma^2)$. The Yule-Walker Estimators $\hat{\boldsymbol{\phi}}$ is defined as*

$$\hat{\boldsymbol{\phi}} = \begin{pmatrix} \hat{\phi}_1 \\ \vdots \\ \hat{\phi}_p \end{pmatrix} = \begin{pmatrix} 1 & r_1 & \cdots & r_{p-1} \\ r_1 & 1 & \cdots & r_{p-2} \\ \vdots & \vdots & \ddots & \vdots \\ r_{p-1} & r_{p-2} & \cdots & 1 \end{pmatrix}^{-1} \begin{pmatrix} r_1 \\ \vdots \\ r_p \end{pmatrix} = \begin{pmatrix} C_0 & C_1 & \cdots & C_{p-1} \\ C_1 & C_0 & \cdots & C_{p-2} \\ \vdots & \vdots & \ddots & \vdots \\ C_{p-1} & C_{p-2} & \cdots & C_0 \end{pmatrix}^{-1} \begin{pmatrix} C_1 \\ \vdots \\ C_p \end{pmatrix}$$

Remark 2. *There are several remarks related to Yule-Walker Estimators*

1. *As by construction, Yule-Walker Estimators also belong to the class of MM-estimators.*
2. *The Yule-Walker Estimator does not provide an estimator of σ^2. You can estimate σ^2 by*

$$\mathrm{Cov}(\cdot, Y_t) : \gamma(0) = \sum_{k=1}^{p} \phi_k \gamma(k) + \sigma^2 \quad \Rightarrow \quad \hat{\sigma}^2 = C_0 - \sum_{k=1}^{p} \hat{\phi}_k C_k.$$

3. *You can also use the trick of taking $\mathrm{Cov}(\cdot, Y_{t-k})$ to both sides of the $\mathrm{MA}(q)$ model to obtain the MM-estimators of $\theta_1, \ldots, \theta_q$. However, $\mathrm{Cov}(Y_t, Y_{t-k}) = \sum_{i=0}^{q-k} \theta_i \theta_{i+k}$ adopts a closed-form solution. So we would not say it to be an "equation" as there is nothing to be solved. The trick also works for solving MM-estimators in the ARMA model.*

(★☆☆) Yule-Walker Estimators

Exercise 3. *Assume $Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + Z_t$, where $\{Z_t\} \sim \mathrm{WN}(0, \sigma^2)$. Let $C_0 = 1$, $C_1 = 0.4$, and $C_2 = 0.25$.*

(a) Find the Yule-Walker estimators of $\boldsymbol{\phi} = (\phi_1, \phi_2)^T$.

(b) Suggest a method of moment estimator of σ^2.

Solution

(a) The Yule-Walker estimator is given by

$$\hat{\boldsymbol{\phi}} = \begin{pmatrix} C_0 & C_1 \\ C_1 & C_0 \end{pmatrix}^{-1} \begin{pmatrix} C_1 \\ C_2 \end{pmatrix} = \begin{pmatrix} 0.357 \\ 0.107 \end{pmatrix}$$

(b) $\hat{\sigma}^2 = C_0 - \hat{\phi}_1 C_1 - \hat{\phi}_2 C_2 = 1 - (0.357)(0.4) - (0.107)(0.25) = 0.8304$.

5.2 Least Squares Type Estimators

5.2.1 Unconditional Least Squares (For AR Model)

Unconditional Least-Squares Method

Definition 2. *Assume $\{Y_t\} \sim \mathrm{AR}(p)$, i.e. $Y_t = \phi_1 Y_{t-1} + \cdots + \phi_p Y_{t-p} + Z_t$, where $\{Z_t\} \sim \mathrm{WN}(0, \sigma^2)$. Let $Y_1, \ldots, Y_n$ be the observed time series, define*

$$\boldsymbol{X} = \begin{pmatrix} Y_p & Y_{p-1} & \cdots & Y_1 \\ Y_{p+1} & Y_p & \cdots & Y_2 \\ \vdots & \vdots & \ddots & \vdots \\ Y_{n-1} & Y_{n-2} & \cdots & Y_{n-p} \end{pmatrix} \quad \text{and} \quad \boldsymbol{Y} = \begin{pmatrix} Y_{p+1} \\ Y_{p+2} \\ \vdots \\ Y_n \end{pmatrix}.$$

Then the ***unconditional least-squares estimators*** *of $\boldsymbol{\phi} = (\phi_1, \ldots, \phi_p)$ are given by*

- $\hat{\boldsymbol{\phi}} = (\boldsymbol{X}^{\boldsymbol{T}}\boldsymbol{X})^{-1}\boldsymbol{X}^{\boldsymbol{T}}\boldsymbol{Y}$
- $\hat{\sigma}^2 = (\boldsymbol{Y} - \boldsymbol{X}\hat{\boldsymbol{\phi}})^T(\boldsymbol{Y} - \boldsymbol{X}\hat{\boldsymbol{\phi}})/(n-2p)$.

Remark 3. *There are several remarks concerning unconditional least squares.*

- *The term "unconditional" arises because we do not need to initialize the noise sequence $\{Z_t\}$. Explicitly, we can write $Z_t = Y_t - \phi_1 Y_{t-1} - \cdots - \phi_p Y_{t-p}$ for $t = p+1, \ldots, n$.*
- *As there is a total of $n - (p+1) + 1 = n - p$ estimated noise available and p predictors are used to build the regression, the denominator in $\hat{\sigma}^2$ is thus $(n-p) - p = n - 2p$.*

Alternative Computation Method for ULS Estimators

It is indeed a regression problem. Hence, you can also find the ULS by the following procedure.

1. For $t = p+1, \ldots, n$, write $Z_t = Y_t - \phi_1 Y_{t-1} - \cdots - \phi_p Y_{t-p}$.
2. Write $S(\phi_1, \cdots, \phi_p) = \sum_{t=p+1}^n Z_t^2$ and let $(\hat{\phi}_1, \cdots, \hat{\phi}_p) = \arg\min_{\phi_1, \cdots, \phi_p} S(\phi_1, \cdots, \phi_p)$.
3. Let $\hat{Z}_t = Y_t - \hat{\phi}_1 Y_{t-1} - \cdots - \hat{\phi}_p Y_{t-p}$ for $t = p+1, \ldots, n$. Then $\hat{\sigma}^2 = (n-2p)^{-1}\sum_{t=p+1}^n \hat{Z}_t^2$.

Asymptotic Normality of ULS

Theorem 1. *Assume $\{Y_t\} \sim \mathrm{AR}(p)$, i.e., $Y_t = \phi_1 Y_{t-1} + \cdots + \phi_p Y_{t-p} + Z_t$, where $\{Z_t\} \sim \mathrm{WN}(0, \sigma^2)$. Let the ULS of $\boldsymbol{\phi} = (\phi_1, \ldots, \phi_p)^T$ be $\hat{\boldsymbol{\phi}} = (\hat{\boldsymbol{\phi}}_1, \ldots, \hat{\boldsymbol{\phi}}_p)^T$. Then as $n \to \infty$,*

$$\sqrt{n}(\hat{\boldsymbol{\phi}} - \boldsymbol{\phi}) \xrightarrow{\mathrm{d}} \mathrm{N}_p(0, \sigma^2\Gamma_p^{-1}), \quad \text{where} \quad \Gamma_p = \begin{pmatrix} \gamma(0) & \gamma(1) & \cdots & \gamma(p-1) \\ \gamma(1) & \gamma(0) & \cdots & \gamma(p-2) \\ \vdots & \vdots & \ddots & \vdots \\ \gamma(p-1) & \gamma(p-2) & \cdots & \gamma(0) \end{pmatrix}$$

Remark 4. *There are several remarks on the inference of ULS.*

- *From Theorem 1, we know that for $\boldsymbol{w} = (w_1, \ldots, w_p)^T \in \mathbb{R}^p$, as $n \to \infty$,*

$$\sqrt{n}\Big(\boldsymbol{w}^{\boldsymbol{T}}\boldsymbol{\phi} - \boldsymbol{w}^{\boldsymbol{T}}\hat{\boldsymbol{\phi}}\Big) \xrightarrow{\mathrm{d}} \mathrm{N}\Big(0, \sigma^2\boldsymbol{w}^{\boldsymbol{T}}\boldsymbol{\Gamma}_{\boldsymbol{p}}^{-1}\boldsymbol{w}\Big).$$

  It could be useful in testing a composite hypothesis, for example, $H_0 : \phi_1 = \phi_2$.

- *We can estimate Γ_p by $\hat{\Gamma}_p$ by replacing all $\gamma(k)$ by C_k.*

(★☆☆) Evaluation of ULS Estimators

✎ **Exercise 4.** *Assume that* $\{Y_t\} \sim \text{AR}(2)$, *i.e.,* $Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + Z_t$, *where* $Z_t \sim \text{WN}(0, \sigma^2)$.

- *Let* $Y_1 = -1,\ Y_2 = 1,\ Y_3 = 0,\ Y_4 = 4,\ Y_5 = -1,\ Y_6 = 3.$
- *The parameters* $\phi_1,\ \phi_2$ *and* σ^2 *are unknown.*

(a) *Find the ULS estimator of* $\phi_1,\ \phi_2$ *and* σ^2 *according to the definition 2.*

(b) *Find the ULS estimator of* $\phi_1,\ \phi_2$ *and* σ^2 *by directly minimizing the sum of squared error, i.e., by the alternative computation method.*

Solution

(a) As the model is AR(2), write

$$\boldsymbol{X} = \begin{pmatrix} Y_2 & Y_1 \\ Y_3 & Y_2 \\ Y_4 & Y_3 \\ Y_5 & Y_4 \end{pmatrix} = \begin{pmatrix} 1 & -1 \\ 0 & 1 \\ 4 & 0 \\ -1 & 4 \end{pmatrix} \quad \text{and} \quad \boldsymbol{Y} = \begin{pmatrix} Y_3 \\ Y_4 \\ Y_5 \\ Y_6 \end{pmatrix} = \begin{pmatrix} 0 \\ 4 \\ -1 \\ 3 \end{pmatrix}$$

$\Rightarrow \hat{\boldsymbol{\phi}} = (\boldsymbol{X}^{\boldsymbol{T}}\boldsymbol{X})^{-1}\boldsymbol{X}^{\boldsymbol{T}}\boldsymbol{Y} = (-2/13, 11/13)^T$, $\hat{\sigma}^2 = (\boldsymbol{Y} - \boldsymbol{X}\hat{\boldsymbol{\phi}})^T(\boldsymbol{Y} - \boldsymbol{X}\hat{\boldsymbol{\phi}})/(n-2p) = 74/13.$

(b) As $Z_t = Y_t - \phi_1 Y_{t-1} - \phi_2 Y_{t-2}$, we have

$$Z_3 = Y_3 - \phi_1 Y_2 - \phi_2 Y_1 = -\phi_1 + \phi_2, \qquad Z_4 = Y_4 - \phi_1 Y_3 - \phi_2 Y_2 = 4 - \phi_2,$$
$$Z_5 = Y_5 - \phi_1 Y_4 - \phi_2 Y_3 = -1 - 4\phi_1, \qquad Z_6 = Y_6 - \phi_1 Y_5 - \phi_2 Y_4 = 3 + \phi_1 - 4\phi_2.$$

Define

$$S(\phi_1, \phi_2) = \sum_{k=3}^{6} Z_t^2 = (-\phi_1 + \phi_2)^2 + (4 - \phi_2)^2 + (-1 - 4\phi_1)^2 + (3 + \phi_1 - 4\phi_2)^2.$$

Taking partial derivative on S w.r.t. ϕ_1 and ϕ_2 yields

$$\frac{\partial}{\partial \phi_1} S(\phi_1, \phi_2) = 2(-\phi_1 + \phi_2)(-1) + 2(-1 - 4\phi_1)(-4) + 2(3 + \phi_1 - 4\phi_2);$$
$$\frac{\partial}{\partial \phi_2} S(\phi_1, \phi_2). = 2(-\phi_1 + \phi_2) + 2(4 - 4\phi_2)(-1) + 2(3 + \phi_1 - 4\phi_2)(-4).$$

Solving $\dfrac{\partial}{\partial \phi_1} S(\phi_1, \phi_2) = \dfrac{\partial}{\partial \phi_2} S(\phi_1, \phi_2) = 0$ gives $\hat{\phi}_1 = -2/13$ and $\hat{\phi}_2 = 11/13$. Define

$$\hat{Z}_t^2 = Y_t - \hat{\phi}_1 Y_{t-1} - \hat{\phi}_2 Y_{t-2}$$

for $t = 3, \ldots, 6$. Then we have

$$\hat{\sigma}^2 = (6 - 2 \times 2)^{-1} \sum_{t=3}^{6} \hat{Z}_t^2 = 74/13.$$

(★★☆) Inference on ULS Estimators

✎ **Exercise 5.** *Let* $\{Y_t\} \sim \text{AR}(2)$ *with* $n = 200$, $C_0 = 0.3$ *and* $C_1 = 0.2$. *The unconditional least square estimators are given by* $\hat{\phi}_1 = 0.5$, $\hat{\phi}_2 = 0.1$ *and* $\hat{\sigma}^2 = 0.25$,

(a) Mathan claimed that an AR(1) *model should be used instead, do you agree with him?*

(b) Martin claimed that Y_{t-1} *and* Y_{t-2} *have the same amount of impact on* Y_t, *do you agree with him?*

Remark: Assume that we are conducting tests under the 95% confidence level.

Solution

(a) Consider testing $H_0 : \phi_2 = 0$ against $H_1 : \phi_2 \neq 0$. Noticing that

$$\hat{\sigma}^2\hat{\Gamma}_2^{-1} = \begin{pmatrix} C_0 & C_1 \\ C_1 & C_0 \end{pmatrix}^{-1} = \frac{0.25}{0.3^2 - 0.2^2}\begin{pmatrix} 0.3 & -0.2 \\ -0.2 & 0.3 \end{pmatrix} = \begin{pmatrix} 1.5 & -1 \\ -1 & 1.5 \end{pmatrix}$$

Recall from Theorem 1 that $\sqrt{n}(\hat{\boldsymbol{\phi}} - \boldsymbol{\phi}) \xrightarrow{\text{d}} \text{N}_2(0, \sigma^2\Gamma_2^{-1})$, then under $H_0 : \phi_2 = 0$,

$$T_1 := \sqrt{\frac{n}{n\widehat{\text{Var}}(\hat{\phi}_2)}}\hat{\phi}_2 \xrightarrow{\text{d}} \text{N}(0,1), \quad \text{where} \quad n\widehat{\text{Var}}(\hat{\phi}_2) = \hat{\sigma}^2\hat{\Gamma}_2^{-1}(2,2) = 1.5.$$

As we are conducting a two-sided test, we should reject H_0 when $|T_1| \geq z_{0.975} \approx 1.96$, where $z_{0.975}$ is the 97.5% quantile of a standard normal random variable. Notice that $T_1 = 1.1547$ and hence H_0 is NOT rejected. We do not have sufficient evidence to claim that Mathan is wrong.

(b) Consider testing $H_0 : \phi_1 = \phi_2$ against $H_1 : \phi_1 \neq \phi_2$. According to remark 4, let $\boldsymbol{w} = (1, -1)^T$, we have

$$\sqrt{n}\Big\{(\hat{\phi}_1 - \hat{\phi}_2) - (\phi_1 - \phi_2)\Big\} = \sqrt{n}(\boldsymbol{w^T}\hat{\boldsymbol{\phi}} - \boldsymbol{w^T}\boldsymbol{\phi}) \xrightarrow{\text{d}} \text{N}(0, \sigma^2\boldsymbol{w^T}\boldsymbol{\Gamma_2^{-1}}\boldsymbol{w})$$

Then under $H_0 : \phi_1 - \phi_2 = 0$,

$$T_2 := \sqrt{\frac{n}{n\widehat{\text{Var}}(\hat{\phi}_1 - \hat{\phi}_2)}}(\hat{\phi}_1 - \hat{\phi}_2) \xrightarrow{\text{d}} \text{N}(0,1), \quad \text{where} \quad n\widehat{\text{Var}}(\hat{\phi}_1 - \hat{\phi}_2) = \hat{\sigma}^2\boldsymbol{w^T}\hat{\boldsymbol{\Gamma}}_{\boldsymbol{2}}^{\boldsymbol{-1}}\boldsymbol{w} = 5.$$

Notice that $|T_2| = 2.5298 > z_{0.975}$ and therefore H_0 is rejected. We have sufficient evidence to claim that Martin is wrong.

5.2.2 Conditional Least Squares (For MA/ARMA Model)

Recall that Y_t are **observable** for $t = 1, , n$, while $\{Z_t\}$ is **unobservable**. For $\{Y_t\} \sim \text{AR}(p)$, we can write

$$Z_t = Y_t - \phi_1 Y_{t-1} - \cdots - \phi_p Y_{t-p}$$

for $t = p + 1, \ldots, n$, which could be computed once the values of $\phi_1, \ldots, \phi_p$ are known. In contrast, for $\{Y_t\} \sim \text{ARMA}(p, q)$, i.e.,

$$Z_t = Y_t - \phi_1 Y_{t-1} - \cdots - \phi_p Y_{t-p} + \theta_1 Z_{t-1} + \cdots + \theta_p Z_{t-p}.$$

As only $Y_1, \ldots, Y_n$ are known to us, it is impossible for us to get a value of Z_t at any time t even though $\phi_1, \ldots, \phi_p, \theta_1, \ldots, \theta_p$ are known. Therefore, we could NOT directly apply the least-square method. It motivates us to consider the **conditional-least square method**.

Conditional Least Squares (CLS) Method

Definition 3. *Let $\{Y_t\} \sim \text{ARMA}(p, q)$ be* ***invertible***. *Then*

1. ***(Initialization)*** *Assume $\tilde{Z}_s = Y_s = 0$ for all $s \leq 0$.*
2. ***(Sequential Estimation of Noise)*** *Let $\tilde{Z}_t = Y_t - \phi_1 Y_{t-1} - \cdots - \phi_p Y_{t-p} + \theta_1 \tilde{Z}_{t-1} + \cdots + \theta_p \tilde{Z}_{t-p}$.*
3. ***(Quantification of Error)*** *Define the sum of squared error as $S_*(\boldsymbol{\phi}, \boldsymbol{\theta}) = \sum_{t=1}^{n} \tilde{Z}_t^2$.*

Then the ***Conditional-Least Squares Estimators*** *are given by the minimizer of $S_*(\boldsymbol{\phi}, \boldsymbol{\theta})$, i.e.*

$$(\hat{\boldsymbol{\phi}}, \hat{\boldsymbol{\theta}}) = \arg\min_{\boldsymbol{\phi}, \boldsymbol{\theta}} S_*(\boldsymbol{\phi}, \boldsymbol{\theta})$$

❗ **Remark 5.** *There are several remarks on the CLS method.*

- *The invertibility assumption is necessary for the following reasons.*
  1. *The CLS method assumed that Z_t can be estimated by past observations $\{Y_{t-k}\}_{k=1}^{\infty}$.*
  2. *Invertibility requires that the coefficient ψ_i satisfies $\sum_{k=0}^{\infty} |\psi_k| < \infty$, implying $\lim_{k\to\infty} \psi_k = 0$, i.e. the effect of the past observation on current noise would decay as the time lag increases. Hence, the error of assuming $\tilde{Z}_s = Y_s = 0$ for $s < 0$ would fade away as the sample size grows.*
- *There may not be an analytical/closed-form solution for the CLS estimator. Some common numerical methods, eg: the Newton-Raphson method or Expectation-Maximization (EM) algorithm, may help to give an estimated optimizer of $S_*(\boldsymbol{\phi}, \boldsymbol{\theta})$.*

(★☆☆) Computation of CLS Estimator

✎ **Exercise 6.** *Assume $Y_t = Z_t - \theta Z_{t-1}$ to be invertible and $\{Z_t\} \sim \text{WN}(0, 1)$. Given $Y_1 = 0, Y_2 = 4, Y_3 = 5$, find the CLS estimator of θ.*

Solution

Let $\tilde{Z}_0 = 0$, then $\tilde{Z}_1 = Y_1 + \theta\tilde{Z}_0 = 0$, $\tilde{Z}_2 = Y_2 + \theta\tilde{Z}_1 = 4$, and $\tilde{Z}_3 = Y_3 + \theta\tilde{Z}_2 = 5 + 4\theta$. We have $S_*(\theta) = \sum_{t=1}^{3} \tilde{Z}_t^2 = 4^2 + (5 + 4\theta)^2$, which is minimized at $\hat{\theta}_{\text{CLS}} = -5/4$.

5.3 Maximum Likelihood Estimators

Let $X_1, \dots, X_n$ be random variables with joint-density $f_\theta(x_1, \dots, x_n)$ and likelihood L.

- $f_\theta(x_1, \dots, x_n)$: A measurement of the "probability" of the event "$X_1 = x_1, \dots, X_n = x_n$" when the distribution is characterized by the parameter θ.
- $L(\theta|x_1, \dots, x_n)$: A measurement of "how likely" the true parameter would be θ when the observed data are given by $x_1, \dots, x_n$.

Remark 6. *In likelihood, we would NOT use the wording "probability of the parameter being θ" because θ is a fixed parameter without randomness (under frequentist setting).*

Their value agrees, i.e. $f_\theta(x_1, \dots, x_n) = L(\theta|x_1, \dots, x_n)$, but differs in statistical meaning.

Maximum Likelihood Estimators

Definition 4. *Let $(Y_1, \dots, Y_n)$ be time series observations with joint distribution $\boldsymbol{f_\theta(y_1, \dots, y_n)}$, then the **Maximum Likelihood Estimator** of $\boldsymbol{\theta}$ is defined as*

$$\hat{\boldsymbol{\theta}}_{\text{MLE}} = \arg\max_{\theta} L(\theta|Y_1, \dots, Y_n)$$

In order to evaluate the MLE, we have to be able to write the joint density of a time series. There are two standard approaches to writing down the joint density.

Method I: Iterative Conditioning (More General)

Let $Y_1, \dots, Y_n$ be the observed time series data, then we can write the joint density as

$$f(Y_1, \dots, Y_n) = \left\{ \prod_{t=2}^{n} f(Y_t|Y_{t-1}, \dots, Y_1) \right\} f(Y_1)$$

Remark 7. *There are several remarks on Method I.*

- *It requires specification of marginal distribution of Y_1, see Exercise 7.*
- *It provides flexibility for considering alternative noise other than normal.*

(★☆☆) Initialization of AR process

Exercise 7. *Let $Y_t = \phi Y_{t-1} + Z_t$ for $t \geq 2$ and $Y_1 = cZ_1$ for some c, where $Z_t \sim \mathrm{WN}(0, \sigma^2)$. Suppose that there exists a constant c such that $\{Y_t\}_{t\geq 1}$ is weakly stationary, find value(s) of c.*

Solution

If $\{Y_t\}_{t\geq 1}$ is weakly stationary, $\mathrm{Var}(Y_1) = \mathrm{Var}(Y_2)$, then

$$c^2\sigma^2 = \mathrm{Var}(Y_1) = \mathrm{Var}(Y_2) = \mathrm{Var}(\phi Y_1 + Z_2) = \mathrm{Var}(\phi c Z_1 + Z_2) = (1 + \phi^2 c^2)\sigma^2.$$

It follows that $c = \pm 1/\sqrt{1 - \phi^2}$.

Takeaway 1. *If we initialize a $\mathrm{AR}(1)$ process by $Y_1 = Z_1$, the process would be nonstationary. However, you should justify the existence of c formally by showing that $\mathrm{Cov}(Y_t, Y_{t+k})$ is free of t.*

(★★☆) MLE under AR(1) Model - Iterative Conditioning

✎ **Exercise 8.** *Assume* $Y_t = \phi Y_{t-1} + Z_t$ *to be causal, where* $\{Z_t\} \overset{iid}{\sim} \mathrm{N}(0, \sigma^2)$.

(a) Derive the log-likelihood function of $Y_1, \ldots, Y_n$, *i.e.,* $\ell(\phi, \sigma^2 | Y_1, \ldots, Y_n) := \ln L(\phi, \sigma^2 | Y_1, \ldots, Y_n)$.

(b) Find the MLE of ϕ *and* σ^2 *by iterative conditioning.*

(c) Given $Y_1 = 3$ *and* $Y_2 = 4$, *find the value of* $\hat{\phi}_{\text{MLE}}$ *and* $\hat{\sigma}^2_{\text{MLE}}$.

Solution

(a) From Exercise 7, we should set $Y_1 \sim \mathrm{N}(0, \sigma^2/(1-\phi^2))$. Also, for $t \geq 2$, we have $[Y_t | Y_{t-1}, \ldots, Y_1] \sim \mathrm{N}(\phi_1 Y_{t-1}, \sigma^2)$. It follows that

$$\begin{aligned} L(\phi, \sigma^2 | Y_1, \ldots, Y_n) &= \left\{ \prod_{t=2}^{n} f(Y_t | Y_{t-1}, \ldots, Y_1) \right\} f(Y_1) \\ &= \left(\frac{1}{2\pi\sigma^2} \right)^{n/2} (1-\phi^2)^{1/2} \exp\left\{ -\frac{1}{2\sigma^2} \left[\sum_{t=2}^{n} (Y_t - \phi Y_{t-1})^2 + Y_1^2 (1-\phi^2) \right] \right\} \end{aligned}$$

Then the log-likelihood is given by

$$\ell(\phi, \sigma^2 | Y_1, \ldots, Y_n) = C - \frac{n}{2} \ln(\sigma^2) + \frac{1}{2} \ln(1-\phi^2) - \frac{1}{2\sigma^2} \left[\sum_{t=2}^{n} (Y_t - \phi Y_{t-1})^2 + Y_1^2 (1-\phi^2) \right],$$

where $C = -(n/2)\ln(2\pi)$ is constant w.r.t the parameters ϕ and σ^2.

(b) Taking partial-derivative w.r.t. ϕ gives

$$\begin{aligned} \frac{\partial \ell}{\partial \phi} &= -\frac{\phi}{1-\phi^2} - \frac{1}{2\sigma^2} \left[\sum_{t=2}^{n} -2Y_{t-1}(Y_t - \phi Y_{t-1}) - 2\phi Y_1^2 \right] \\ &= -\frac{\phi}{1-\phi^2} + \frac{1}{\sigma^2} \sum_{t=2}^{n} Y_t Y_{t-1} + \frac{\phi}{\sigma^2} \left(Y_1^2 - \sum_{t=2}^{n} Y_{t-1}^2 \right). \end{aligned}$$

Taking partial-derivative w.r.t. σ^2 gives

$$\frac{\partial \ell}{\partial \sigma^2} = -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4} \left[\sum_{t=2}^{n} (Y_t - \phi Y_{t-1})^2 + Y_1^2 (1-\phi^2) \right],$$

(c) From (b), it suffices to solve

$$\frac{\partial \ell}{\partial \phi} = -\frac{\phi}{1-\phi^2} + \frac{(3)(4)}{\sigma^2} + \frac{\phi}{\sigma^2} \left(3^2 - 3^2 \right) = -\frac{\phi}{1-\phi^2} + \frac{12}{\sigma^2} = 0$$

and

$$\frac{\partial \ell}{\partial \sigma^2} = -\frac{2}{2\sigma^2} + \frac{1}{2\sigma^4} \left[(4 - 3\phi)^2 + 3^2 (1-\phi^2) \right] = 0,$$

On solving, we have $\hat{\phi}_{\text{MLE}} = 24/25$ and $\hat{\sigma}^2_{\text{MLE}} = 49/50$.

Email contact: martinmtt@link.cuhk.edu.hk ; Personal website: https://sites.google.com/view/martinmtt/home

Under regularity conditions, a linear combination of uncorrelated random variables follows normal distribution. We can model the joint distribution of causal time series by multivariate normal distribution.

Method II: Normality Assumption

Let $\{Y_t\} \sim \text{ARMA}(p,q)$ being causal and $\{Z_t\} \sim \text{N}(0,\sigma^2)$, $(Y_1,\ldots,Y_n) \sim \text{N}_n(0,\mathbf{\Sigma})$, with density

$$f(y_1,\ldots,y_n) = \frac{1}{(2\pi)^{n/2}|\mathbf{\Sigma}|^{1/2}} \exp\left\{-\frac{1}{2}\boldsymbol{y}^{\boldsymbol{T}}\mathbf{\Sigma}^{\mathbf{-1}}\boldsymbol{y}\right\}, \text{ where } \mathbf{\Sigma} = \begin{pmatrix} \gamma(0) & \gamma(1) & \cdots & \gamma(n-1) \\ \gamma(1) & \gamma(0) & \cdots & \gamma(n-2) \\ \vdots & \vdots & \ddots & \vdots \\ \gamma(n-1) & \gamma(n-2) & \cdots & \gamma(0) \end{pmatrix}$$

and $\boldsymbol{y} = (y_1,\ldots,y_n)$. In general, $\gamma(k)$ can be expressed in terms of σ^2, ϕ's and θ's. Define

$$\left(\hat{\boldsymbol{\phi}}_{\text{MLE}}, \hat{\boldsymbol{\theta}}_{\text{MLE}}, \hat{\boldsymbol{\sigma}}^{\mathbf{2}}_{\text{MLE}}\right) = \underset{\boldsymbol{\phi},\boldsymbol{\theta},\sigma^2}{\arg\max}\, L(\boldsymbol{\phi},\boldsymbol{\theta},\sigma^2|Y_1,\ldots,Y_n) = \underset{\boldsymbol{\phi},\boldsymbol{\theta},\sigma^2}{\arg\max}\, f(Y_1,\ldots,Y_n)$$

Remark 8. *Method II is simpler as it provides an explicit form of joint density, without the need for initialization of the process as in Exercise 7.*

(★★☆) MLE under AR(1) model - Multivariate Normal Assumption

Exercise 9. *Assume $Y_t = \phi Y_{t-1} + Z_t$ to be causal, where $\{Z_t\} \sim \text{N}(0,\sigma^2)$.*

(a) By noticing that (Y_1,Y_2) follows multivariate normal distribution, specify the joint-density of (Y_1,Y_2).

(b) Given $Y_1 = 3$ and $Y_2 = 4$, find the log-likelihood of (Y_1,Y_2).

Solution

(a) By considering $\text{Cov}(\cdot,Y_t)$ and $\text{Cov}(\cdot,Y_{t-1})$, we have

$$\gamma(0) = \phi\gamma(1) + \sigma^2 \quad \text{and} \quad \gamma(1) = \phi\gamma(0) \quad \Rightarrow \quad \gamma(0) = \frac{\sigma^2}{1-\phi^2} \quad \text{and} \quad \gamma(1) = \frac{\phi\sigma^2}{1-\phi^2}$$

$$\Rightarrow \quad \mathbf{\Sigma} = \frac{\sigma^2}{1-\phi^2}\begin{pmatrix} 1 & \phi \\ \phi & 1 \end{pmatrix} \quad \Rightarrow \quad |\mathbf{\Sigma}|^{1/2} = \frac{\sigma^2}{(1-\phi^2)^{1/2}}, \quad \Sigma^{-1} = \frac{1}{\sigma^2}\begin{pmatrix} 1 & -\phi \\ -\phi & 1 \end{pmatrix}$$

Then the joint density is given by

$$f(y_1,y_2) = \frac{(1-\phi^2)^{1/2}}{2\pi\sigma^2} \exp\left\{-\frac{1}{2\sigma^2}\left(y_1^2 - 2\phi y_1 y_2 + y_2^2\right)\right\}.$$

(b) The log-likelihood is given by

$$\begin{aligned} \ell(\phi,\sigma^2|y_1,y_2) &= -\ln(2\pi) - \frac{1}{2}\ln(1-\phi^2) - \ln(\sigma^2) - \frac{1}{2\sigma^2}\left(y_1^2 - 2\phi y_1 y_2 + y_2^2\right) \\ &= -\ln(2\pi) + \frac{1}{2}\ln(1-\phi^2) - \ln(\sigma^2) - \frac{\sigma^2}{2(1-\phi^2)}\left(25 - 24\phi\right). \end{aligned}$$

Takeaway 2. *You can solve $\frac{\partial \ell}{\partial \phi} = 0$ and $\frac{\partial \ell}{\partial \sigma^2} = 0$ for $\hat{\phi}_{\text{MLE}}$ and $\hat{\sigma}^2_{\text{MLE}}$. However, it does not have a closed-form solution. You may find the optimizer through numerical methods. Therefore, Method II provides an explicit form of joint density, but the evaluation of the estimator could be harder.*

5.4 Summary of Estimation Methods

The following is a brief summary to the applicability of methods.

Method	AR(p) model	MA(q) or ARMA(p, q) model ($q \neq 0$)
Method of Moment (MM)	✔	✔
Yule-Walker (YW)	✔	✘
Unconditional Least-Squares (ULS)	✔	✘
Conditional Least-Squares (CLS)	✘	✔
Maximum Likelihood Estimator (MLE)	✔	✔

5.5 M-estimation

In general, most of the estimation methods mentioned above belong to the class of M-estimation.

M-estimator

Definition 5. *Let $\theta \in \mathbb{R}^p$ be the estimand. Given the observations $(Y_1, \ldots, Y_n)$. If there exists some function M such that the equation*

$$M(\theta) = M(\theta, Y_{1:n}) = 0$$

has root $\hat{\theta}_M = \hat{\theta}_M(Y_1, \ldots, Y_n)$, i.e., $M(\hat{\theta}_M) = 0$. Then, $\hat{\theta}_M$ is said to be a M estimator.

1. **Method of moment estimator:** Suppose that $\mathrm{E}\{g_k(Y_1)\} = T_k(\theta)$ for $k = 1, \ldots, p$, where $g_k(\cdot)$ and $T_k(\cdot)$ are univariate function. Then $\hat{\theta}_{\mathrm{MM}}$ is the root to

$$M(\theta) = \begin{pmatrix} T_1(\theta) - \frac{1}{n}\sum_{t=1}^{n} g_1(Y_t) \\ \vdots \\ T_p(\theta) - \frac{1}{n}\sum_{t=1}^{n} g_p(Y_t) \end{pmatrix}.$$

2. **Unconditional least-square estimator:** $[\theta = (\phi_1, \ldots, \phi_p)]$ For $\{Y_t\} \sim \mathrm{AR}(p)$, write $Z_t(\phi_1, \ldots, \phi_p) = Y_t - \sum_{k=1}^{p} \phi_k Y_{t-k}$ and $S_0(\theta) = \sum_{t=p+1}^{n} Z_t^2(\phi_1, \ldots, \phi_p)$. Then $\hat{\theta}_{\mathrm{ULS}}$ is the root to

$$M(\theta) = \nabla S_0(\theta).$$

3. **Conditional least-square estimator:** $[\theta = (\boldsymbol{\phi}, \boldsymbol{\theta})]$ Let $S(\theta) = \sum_{t=1}^{n} \tilde{Z}_t^2(\boldsymbol{\phi}, \boldsymbol{\theta})$. Then $\hat{\theta}_{\mathrm{CLS}}$ is the root to

$$M(\theta) = \nabla S(\theta).$$

4. **Maximum likelihood estimator:** Let $L(\theta)$ be the likelihood function and assume it to be differentiable. Then $\hat{\theta}_{\mathrm{MLE}}$ is the root to

$$M(\theta) = \nabla L(\theta).$$

Asymptotic normality of M-estimator

Theorem 2. *Let θ_0 be the true value of the parameter and $\hat{\theta}$ as the M-estimator. Then*

$$\sqrt{n}(\hat{\theta} - \theta_0) \approx \sqrt{n}\{M'(\theta_0)\}^{-1} M(\theta_0) \xrightarrow{\mathrm{d}} \mathrm{N}(0, \Sigma).$$

❗ **Remark 9.** *M-estimator (in this course) is commonly refered to Z-estimator instead in the literature. The M-estimator (in the literature) have another statistical meaning.*