TMA4285 Time series models
Solution to exercise 6, autumn 2020

October 16, 2020

Problem 5.9

We want to construct the likelihood function of $X_1, \dots, X_p, X_{p+1}, \dots, X_n$. We use conditioning to get

$$\begin{aligned} L(\phi, \sigma^2) &= f(X_1, \dots, X_p, X_{p+1}, \dots, X_n | \phi, \sigma^2) \\ &= f(X_{p+1}, \dots, X_n | X_1, \dots, X_p, \phi, \sigma^2) f(X_1, \dots, X_p | \phi, \sigma^2) \end{aligned}$$

where $f(X_{p+1}, \dots, X_n | X_1, \dots, X_p, \phi, \sigma^2) = \Pi_{j=p+1}^{n} f(X_j | X_{j-1}, ..., X_{p+1}, X_1, \dots, X_p, \phi, \sigma^2)$. For $X_1, \dots, X_p$,

$$f(X_1, \dots, X_p | \phi, \sigma^2) = (2\phi\sigma^2)^{-p/2} (detG_p)^{-1/2} \times \exp(-\frac{1}{2\sigma^2} \mathbf{X}_p^T G_p^{-1} \mathbf{X}_p)$$

Next, we need the expected value and covariance for the conditional distribution of $X_{p+1}, \dots, X_n$.

$$\mathrm{E}(X_j | X_{j-1}, ..., X_{p+1}, X_1, \dots, X_p) = \mathrm{E}(X_j | X_{j-1}, ..., X_{j-p}) = \hat{X}_j$$
$$\mathrm{E}((X_j - \hat{X}_j)(X_j - \hat{X}_j) | X_{j-1}, \dots, X_{p+1}, X_1, \dots, X_p) = \mathrm{E}((X_j - \hat{X}_j)^2) = \sigma^2 r_{j-1}.$$

For $j > p$, $r_{j-1} = 1$. This gives

$$f(X_{p+1}, \dots, X_n | X_1, \dots, X_p, \phi, \sigma^2) = (2\phi\sigma^2)^{-(n-p)/2} \times \exp(-\frac{1}{2\sigma^2} \sum_{j=p+1}^{n} (X_j - \hat{X}_j)^2)$$

Finally, we get

$$f(X_1,\dots,X_p|\phi,\sigma^2) = (2\phi\sigma^2)^{-p/2}(detG_p)^{-1/2} \times \exp(-\frac{1}{2\sigma^2}[\mathbf{X}_p^T G_p^{-1}\mathbf{X}_p + \sum_{j=p+1}^{n}(X_j - \phi_1 X_{j-1} - \cdots - \phi_p X_{j-p})^2]),$$

where we have inserted $\hat{X}_j = \phi_1 X_{j-1} - \cdots - \phi_p X_{j-p}$ and used $\mathrm{Var}(X_j - \hat{X}_j) = \sigma^2 forj > p$.

Problem 5.10

Using the likelihood function from problem 5.9 we want to minimize

$$\mathbf{X}_2^T G_2^{-1}\mathbf{X}_2 + \sum_{t=3}^{n}(X_t - \phi_1 X_{t-1} - \phi_2 X_{t-2})^2$$

From example 5.2.1 in the book we find that

$$G_2^{-1} = \begin{bmatrix} 1-\phi_2^2 & -\phi_1(1+\phi_2) \\ -\phi_1(1+\phi_2) & 1-\phi_2^2 \end{bmatrix}$$

which gives

$$\mathbf{X}_2^T G_2^{-1}\mathbf{X}_2 = (X_1^2 + X_2^2)(1-\phi_2^2) - 2X_1X_2\phi_1(1+\phi^2)$$

We take the derivative of

$$(X_1^2 + X_2^2)(1-\phi_2^2) - 2X_1X_2\phi_1(1+\phi^2) + \sum_{t=3}^{n}(X_t - \phi_1 X_{t-1} - \phi_2 X_{t-2})^2$$

with respect to both ϕ_1 and ϕ_2. That gives us

$$X_1X_2(1+\phi_2) + \sum_{t=3}^{n}(X_tX_{t-1} - \phi_1 X_{t-1}^2 - \phi_2 X_{t-2}X_{t-1}) = 0$$

$$\phi_2(X_1^2 + X_2^2) + X_1X_2\phi_1 + \sum_{t=3}^{n}(X_tX_{t-2} - \phi_1 X_{t-1}X_{t-2} - \phi_2 X_{t-2}X_{t-1}^2) = 0$$

which are our two linear equations. If we use the definition of the sample autocovariance we get

$$X_1X_2(1+\phi_2)/n + \hat\gamma(1) - \phi_1\hat\gamma(0) - \phi_2\hat\gamma(1) = 0$$
$$\phi_2(X_1^2 + X_2^2)/n + X_1X_2\phi_1/n + \hat\gamma(2) - \phi_1\hat\gamma(1) - \phi_2\hat\gamma(0) = 0$$

These can be expressed as

$$\begin{bmatrix} \hat\gamma(0) & \hat\gamma(1) - \frac{X_1X_2}{n} \\ \hat\gamma(1) - \frac{X_1X_2}{n} & \hat\gamma(0) - \frac{X_1^2+X_2^2}{n} \end{bmatrix} \begin{bmatrix} \phi_1 \\ \phi_2 \end{bmatrix} = \begin{bmatrix} \hat\gamma(1) + \frac{X_1X_2}{n} \\ \hat\gamma(2) \end{bmatrix}$$

The Yule-Walker equations for the AR(2) model with are derived from

$$X_tX_{t-h} = \phi X_{t-1}X_{t-h} + \phi_2X_{t-2}X_{t-h} + Z_tX_{t-h}$$

If we let $h = 1$ and $h = 2$ and then taking expectations, we get

$$\hat\gamma(1) - \phi_1\hat\gamma(0) - \phi_2\hat\gamma(1) = 0$$
$$\hat\gamma(2) - \phi_1\hat\gamma(1) - \phi_2\hat\gamma(0) = 0$$

which can be expresses as

$$\begin{bmatrix} \hat\gamma(0) & \hat\gamma(1) \\ \hat\gamma(1) & \hat\gamma(0) \end{bmatrix} \begin{bmatrix} \phi_1 \\ \phi_2 \end{bmatrix} = \begin{bmatrix} \hat\gamma(1) \\ \hat\gamma(2) \end{bmatrix}.$$

The least-squares solution is an adjustment of the Yule-Walker equations.

Problem 5.12

The AR(1) model is given by $X_t = \phi X_{t-1} + Z_t$, with autocovariance $\gamma(0) = \frac{\sigma^2}{1-\phi^2}$. We use the likelihood function stated in problem 5.9 with $G_1 = 1/(1-\phi^2)$.

$$L(\phi,\sigma^2) = (2\pi\sigma^2)^{-n/2}(1-\phi^2)^{1/2} \times \exp\left(-\frac{1}{2\sigma^2}\left[X_1^2(1-\phi^2) + \sum_{t=2}^{n}(X_t - \phi X_{t-1})^2\right]\right)$$

By taking the log and taking the derivative with respect to ϕ, we get

$$lnL = (-n/2)ln(2\pi\sigma^2) + (1/2)ln(1-\phi^2) - \frac{1}{2\sigma^2}[X_1^2(1-\phi^2) + \sum_{t=2}^{n}(X_t - \phi X_{t-1})^2]$$

$$\frac{\partial lnL}{\partial\phi} = -\frac{\phi}{1-\phi^2} + \frac{\phi X_1^2}{\sigma^2} + \frac{1}{\sigma^2}\sum_{t=2}^{n}(X_tX_{t-1} - \phi X_{t-1}^2) = 0$$

$$-\phi\sigma^2 + \phi X_1^2(1-\phi^2) + (1-\phi^2)\sum_{t=2}^{n}(X_tX_{t-1} - \phi X_{t-1}^2) = 0.$$

Solving this system gives the estimates $\hat{\phi}_1 = 0.2742$, $\hat{\phi}_2 = 0.3579$ and $\hat{\sigma}^2 = 0.8199$.
c) We construct a 95% confidence interval for μ to test if we can reject the hypothesis that $\mu = 0$. We have that $\overline{X}_{200} \sim \text{AN}(\mu, \nu/n)$ with

$$\nu = \sum_{h=-\infty}^{\infty} \gamma(h) \approx \hat{\gamma}(-3) + \hat{\gamma}(-2) + \hat{\gamma}(-1) + \hat{\gamma}(0) + \hat{\gamma}(1) + \hat{\gamma}(2) + \hat{\gamma}(3) = 3.61.$$

An approximate 95% confidence interval for μ is then

$$I = \overline{x}_n \pm \lambda_{0.025}\sqrt{\nu/n} = 3.82 \pm 1.96\sqrt{3.61/200} = 3.82 \pm 0.263.$$

Since $0 \notin I$ we reject the hypothesis that $\mu = 0$.
d) We have that approximately $\hat{\boldsymbol{\phi}} \sim \text{AN}(\boldsymbol{\phi}, \hat{\sigma}^2\hat{\boldsymbol{\Gamma}}_2^{-1}/n)$. Inserting the observed values we get

$$\frac{\hat{\sigma}^2\hat{\boldsymbol{\Gamma}}_2^{-1}}{n} = \begin{pmatrix} 0.0050 & -0.0021 \\ -0.0021 & 0.0050 \end{pmatrix},$$

and hence $\hat{\phi}_1 \sim \text{AN}(\phi_1, 0.0050)$ and $\hat{\phi}_2 \sim \text{AN}(\phi_2, 0.0050)$. We get the 95% confidence intervals

$$\begin{aligned} I_{\phi_1} &= \hat{\phi}_1 \pm \lambda_{0.025}\sqrt{0.005} = 0.274 \pm 0.139 \\ I_{\phi_2} &= \hat{\phi}_2 \pm \lambda_{0.025}\sqrt{0.005} = 0.358 \pm 0.139. \end{aligned}$$

e) If the data were generated from an AR(2) process, then the PACF would be $\alpha(0) = 1$, $\hat{\alpha}(1) = \hat{\rho}(1) = 0.427$, $\hat{\alpha}(2) = \hat{\phi}_2 = 0.358$ and $\hat{\alpha}(h) = 0$ for $h \geq 3$.

Problem 5.11. To obtain the maximum likelihood estimator we compute as if the process were Gaussian. Then the innovations

$$\begin{aligned} X_1 - \hat{X}_1 &= X_1 \sim N(0, \nu_0), \\ X_2 - \hat{X}_2 &= X_2 - \phi X_1 \sim N(0, \nu_1), \end{aligned}$$

where $\nu_0 = \sigma^2 r_0 = \mathbb{E}[(X_1 - \hat{X}_1)^2]$, $\nu_1 = \sigma^2 r_1 = \mathbb{E}[(X_2 - \hat{X}_2)^2]$. This implies $\nu_0 = \mathbb{E}[X_1^2] = \gamma(0)$, $r_0 = 1/(1-\phi^2)$ and $\nu_1 = \mathbb{E}[(X_2 - \hat{X}_2)^2] = \gamma(0) - 2\phi\gamma(1) + \phi^2\gamma(0)$ and hence

$$r_1 = \frac{\gamma(0)(1+\phi^2) - 2\phi\gamma(1)}{\sigma^2} = \frac{1 + \phi^2 - 2\phi^2}{1 - \phi^2} = 1.$$

Here we have used that $\gamma(1) = \sigma^2\phi/(1-\phi^2)$. Since the distribution of the innovations is normal the density for $X_j - \hat{X}_j$ is

$$f_{X_j - \hat{X}_j} = \frac{1}{\sqrt{2\pi\sigma^2 r_{j-1}}} \exp\left(-\frac{x^2}{2\sigma^2 r_{j-1}}\right)$$

and the likelihood function is

$$\begin{aligned} L(\phi, \sigma^2) = \prod_{j=1}^{2} f_{X_j - \hat{X}_j} &= \frac{1}{\sqrt{(2\pi\sigma^2)^2 r_0 r_1}} \exp\left\{-\frac{1}{2\sigma^2}\left(\frac{(x_1 - \hat{x}_1)^2}{r_0} + \frac{(x_2 - \hat{x}_2)^2}{r_1}\right)\right\} \\ &= \frac{1}{\sqrt{(2\pi\sigma^2)^2 r_0 r_1}} \exp\left\{-\frac{1}{2\sigma^2}\left(\frac{x_1^2}{r_0} + \frac{(x_2 - \phi x_1)^2}{r_1}\right)\right\}. \end{aligned}$$

We maximize this by taking logarithm and then differentiate:

$$
\begin{aligned}
\log L(\phi,\sigma^2) &= -\frac{1}{2}\log(4\pi^2\sigma^4 r_0 r_1) - \frac{1}{2\sigma^2}\big(\frac{x_1^2}{r_0} + \frac{(x_2-\phi x_1)^2}{r_1}\big)\\
&= -\frac{1}{2}\log(4\pi^2\sigma^4/(1-\phi^2)) - \frac{1}{2\sigma^2}\big(x_1^2(1-\phi^2) + (x_2-\phi x_1)^2\big)\\
&= -\log(2\pi) - \log(\sigma^2) + \frac{1}{2}\log(1-\phi^2) - \frac{1}{2\sigma^2}\big(x_1^2(1-\phi^2) + (x_2-\phi x_1)^2\big).
\end{aligned}
$$

Differentiating yields

$$
\begin{aligned}
\frac{\partial l(\phi,\sigma^2)}{\partial\sigma^2} &= -\frac{1}{\sigma^2} + \frac{1}{2\sigma^4}\big(x_1^2(1-\phi^2) + (x_2-\phi x_1)^2\big),\\
\frac{\partial l(\phi,\sigma^2)}{\partial\phi} &= \frac{1}{2}\cdot\frac{-2\phi}{1-\phi^2} + \frac{x_1 x_2}{\sigma^2}.
\end{aligned}
$$

Putting these expressions equal to zero gives $\sigma^2 = \frac{1}{2}\big(x_1^2(1-\phi^2) + (x_2-\phi x_1)^2\big)$ and then after some computations $\phi = 2x_1x_2/(x_1^2+x_2^2)$. Inserting the expression for ϕ is the equation for σ gives the maximum likelihood estimators

$$
\hat\sigma^2 = \frac{(x_1^2-x_2^2)^2}{2(x_1^2+x_2^2)} \text{ and } \hat\phi = \frac{2x_1x_2}{x_1^2+x_2^2}
$$

GT Exercises

Exercise 5

a The process is

$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + Z_t$$

and we will estimate the model parameters through:

(i) Yule-Walker Estimation: By multiplying on both sides of the equation

$$(1 - \phi_1 B - \phi_2 B^2) Y_t = Z_t$$

by Y_t, Y_{t-1} and Y_{t-2}, we get the system of equations:

$$\begin{bmatrix} \gamma(0) & \gamma(1) \\ \gamma(1) & \gamma(0) \end{bmatrix} \begin{bmatrix} \phi_1 \\ \phi_2 \end{bmatrix} = \begin{bmatrix} \gamma(1) \\ \gamma(2) \end{bmatrix}$$

and $\sigma^2 = \gamma(0) - \phi_1 \gamma(1) - \phi_2 \gamma(2)$.

This system is solved, replacing $\gamma(k)$ by $\hat{\gamma}(k)$, by

$$\hat{\phi}_1 = \frac{\hat{\gamma}(1)[\hat{\gamma}(0) - \hat{\gamma}(2)]}{\hat{\gamma}^2(0) - \hat{\gamma}^2(1)}$$

$$\hat{\phi}_2 = \frac{\hat{\gamma}(2)\hat{\gamma}(0) - \hat{\gamma}^2(1)}{\hat{\gamma}^2(0) - \hat{\gamma}^2(1)}, \text{ and}$$

$$\hat{\sigma}^2 = (\hat{\gamma}(0) - \hat{\gamma}(2)) \left[\frac{\hat{\gamma}(0)(\hat{\gamma}(0) + \hat{\gamma}(2)) - 2\hat{\gamma}^2(1)}{\hat{\gamma}^2(0) - \hat{\gamma}^2(1)} \right]$$

(ii) Durbin-Levinson Algorithm: According to this algorithm the process can be expressed as:

$$Y_t - \hat{\phi}_{21} Y_{t-1} - \hat{\phi}_{22} Y_{t-2} = Z_t, \qquad \{Z_t\} \sim WN(0, \hat{\nu}_2)$$

with:

$$\hat{\nu}_0 = \hat{\gamma}(0)$$

$$\hat{\phi}_{11} = \frac{\hat{\gamma}(1)}{\hat{\gamma}(0)}$$

$$\hat{\nu}_1 = \hat{\gamma}(0)[1 - \hat{\phi}_{11}^2]$$

$$\hat{\phi}_{22} = \frac{1}{\hat{\nu}_1}[\hat{\gamma}(2) - \hat{\phi}_{11}\hat{\gamma}(1)] = \hat{\phi}_2 \text{ (as in Yule-Walker)}$$

$$\hat{\phi}_{21} = \hat{\phi}_{11} - \hat{\phi}_{22}\hat{\phi}_{11} = \hat{\phi}_1 \text{ (as in Yule-Walker), and}$$

$$\hat{\nu}_2 = \hat{\nu}_1(1 - \hat{\phi}_{22}^2) = \hat{\sigma}^2 \text{ (as in Yule-Walker)}$$

(iii) Hannan-Rissanen Algorithm - Step 2: All we need to do in this case is to regress Y_t onto $(Y_{t-1}, Y_{t-2}), t = 3, \ldots, n$ as Ordinary Least Squares. It means the vector $[\phi_1, \phi_2]^T$ is estimated by $(Z^T Z)^{-1} Z^T \mathbf{Y}_n$, with:

$$Z = \begin{bmatrix} Y_3 & Y_2 \\ Y_4 & Y_3 \\ \vdots & \vdots \\ Y_{n-1} & Y_{n-2} \end{bmatrix} \qquad \text{and} \qquad \mathbf{Y}_n = \begin{bmatrix} Y_4 \\ Y_5 \\ \vdots \\ Y_n \end{bmatrix}$$

Finally,

$$\hat{\sigma}^2 = \frac{1}{n-3} \sum_{t=4}^{n} (Y_t - \hat{\phi}_1 Y_{t-1} - \hat{\phi}_2 Y_{t-2})^2$$

b (i) Yule-Walker Estimation: We aim to find the $P_n Y_{n+1} = a_0 + \sum_{i=1}^{n} a_i Y_{n+1-i}$ so that

$$E\left[\left(Y_{n+1} - a_0 - \sum_{i=1}^{n} a_i Y_{n+1-i}\right)^2\right]$$

is minimized. Taking partial derivatives with respect to $a_0, a_1, \ldots, a_n$ we find

$$a_0 = 0$$

$$E\left[\left(Y_{n+1} - \sum_{i=1}^{n} a_i Y_{n+1-i}\right) Y_{n+1-j}\right] = 0, \qquad j = 1, \ldots, n$$

Which becomes the Yule-Walker system of equations:

$$\Gamma_n \mathbf{a}_n = \boldsymbol{\gamma}(n)$$

with $\Gamma_{n_{ij}} = [\gamma(i-j)]_{i,j=1}^{n}$, $\mathbf{a}_n = [a_1, \ldots, a_n]^T$ and $\boldsymbol{\gamma}_n = [\gamma(1), \ldots, \gamma(n)]^T$. This system is solved by $\mathbf{a}_n = \Gamma_n^{-1}\boldsymbol{\gamma}_n$. Finally, its mean square prediction error is given by:

$$\begin{aligned} MSPE &= E\left[\left(Y_{n+1} - \sum_{i=1}^{n} a_i Y_{n+1-i}\right)^2\right] \\ &= \gamma(0) - \mathbf{a}_n^T \boldsymbol{\gamma}_n \\ &= \gamma(0) - \boldsymbol{\gamma}_n^T \Gamma_n^{-1} \boldsymbol{\gamma}_n \end{aligned}$$

(ii) Durbin Levinson algorithm: The one-step predictor can be expressed as:

$$P_n Y_{n+1} = \hat{\phi}_{n1} Y_n + \cdots + \hat{\phi}_{nn} Y_1$$

with

$$\hat{\phi}_{nn} = \left[\hat{\gamma}(n) - \sum_{j=1}^{n-1} \hat{\phi}_{n-1,j} \hat{\gamma}(n-j)\right] \hat{\nu}_{n-1}^{-1}$$

$$\begin{bmatrix} \hat{\phi}_{n1} \\ \vdots \\ \hat{\phi}_{n,n-1} \end{bmatrix} = \begin{bmatrix} \hat{\phi}_{n-1,1} \\ \vdots \\ \hat{\phi}_{n-1,n-1} \end{bmatrix} - \hat{\phi}_{nn} \begin{bmatrix} \hat{\phi}_{n-1,n-1} \\ \vdots \\ \hat{\phi}_{n-1,1} \end{bmatrix}$$

In this case the prediction error is $\nu_n = \nu_{n-1}[1 - \phi_{nn}^2]$, which is exactly the same as for the Yule-Walker estimation as proven in proof 1, page 61 of the book.

(iii) Hannan-Rissanen Algorithm - Step 2: Given that what is performed in step 2 of this algorithm is a linear regression fitted by OLS, we can find $P_n Y_{n+1}$ as:

$$P_n Y_{n+1} = \hat{\phi}_1 Y_n + \hat{\phi}_2 Y_{n-1}$$

Its associated prediction error is

$$E(Y_{n+1} - \phi_1 Y_n - \phi_2 Y_{n-1})^2 = (1+\phi_1^2+\phi_2^2)\gamma(0) - 2\phi_1(1-\phi_2)\gamma(1) - 2\phi_2\gamma(2)$$

c Given that all the one-step predictors are linear combinations of $\{Y_n, \ldots, Y_1\}$, and the process is zero-mean, then all the one-step predictors are unbiased.

Now we'll try to compare the three predictors in terms of mean square error. As mentioned in part b the Yule-Walker estimation and thee Durbin-Levinson algorithm produce the same one-step predictor with the same mean square error, which is the minimum for a linear predictor. It also means that the MSE of the one-step predictor obtained through the Hannan-Rissanen algorithm is larger than the one obtained through the other two approaches.

d For the AR(2) process we model Y_t in function of Y_{t-1} and Y_{t-2}. That is, the likelihood we expect to maximize is

$$L_2(\boldsymbol{\theta}) = f(Y_3, \ldots, Y_n | Y_1, Y_2, \boldsymbol{\theta})$$

which resembles the likelihood in a regular regression model.

Making use of the Bayes' theorem and the law of total probability, we can see that the full likelihood $f(Y_1, Y_2, \ldots, Y_n, \boldsymbol{\theta})$ can be obtained from the conditional likelihood through

$$f(Y_1, Y_2, \ldots, Y_n, \boldsymbol{\theta}) = f(Y_3, \ldots, Y_n | Y_1, Y_2, \boldsymbol{\theta}) \cdot f(Y_1, Y_2, \boldsymbol{\theta})$$

If it is taken to the logarithmic scale we can say that the full likelihood can be computed as the addition of the conditional likelihood and the marginal likelihood of the initial values.

The estimates of the parameters in the AR(2) process can be obtained using the conditional or the full likelihood. As n increases there is no big difference in the estimates, given that the estimates based on these likelihoods have the same limiting distribution.