TMA4285 Time series models
Solution to exercise 5, autumn 2020

October 11, 2020

Problem A.7

We want to show that $(\mathbf{X}-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\mathbf{X}-\boldsymbol{\mu})$ has a chi square distribution with n degrees of freedom. We have that $\mathbf{Z}=\boldsymbol{\Sigma}^{1/2}(\mathbf{X}-\boldsymbol{\mu})$ is normal with mean 0 and covariance Σ.

$$\begin{aligned}(\mathbf{X}-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\mathbf{X}-\boldsymbol{\mu}) &= (\mathbf{X}-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1/2}\boldsymbol{\Sigma}^{-1/2}(\mathbf{X}-\boldsymbol{\mu}) = [\boldsymbol{\Sigma}^{-1/2}(\mathbf{X}-\boldsymbol{\mu})]^T\boldsymbol{\Sigma}^{-1/2}(\mathbf{X}-\boldsymbol{\mu})\\ &= \mathbf{Z}^T\mathbf{Z} = Z_1^2+Z_2^2+\cdots+Z_n^2 \sim \chi_n^2,\end{aligned}$$

where we have used that the sum of squared standard normal stochastic variables are chi square distributed.

Problem 5.8

We begin by taking the natural logarithm of the likelihood L,

$$\ln L(\boldsymbol{\phi}, \boldsymbol{\theta}, \sigma^2) = -\frac{n}{2}\ln(2\pi\sigma^2) - \frac{1}{2}\ln(r_0 \ldots r_{n-1}) - \frac{1}{2\sigma^2}\sum_{j=1}^{n}\frac{(X_j - \hat{X}_j)^2}{r_{j-1}}.$$

To derive equation (5.2.10), we differentiate $\ln L(\boldsymbol{\phi}, \boldsymbol{\theta}, \sigma^2)$ with respect to σ^2

$$\frac{\partial \ln L(\boldsymbol{\phi}, \boldsymbol{\theta}, \sigma^2)}{\partial \sigma^2} = -\frac{n}{2}\frac{1}{\sigma^2} + \frac{1}{2\sigma^4}\sum_{j=1}^{n}\frac{(X_j - \hat{X}_j)^2}{r_{j-1}}$$

We set this equal to zero and solve for σ^2. This gives

$$\hat{\sigma}^2 = n^{-1}S(\hat{\boldsymbol{\phi}}, \hat{\boldsymbol{\theta}}),$$

where $S(\hat{\boldsymbol{\phi}}, \hat{\boldsymbol{\theta}}) = \sum_{j=1}^{n}\frac{(X_j - \hat{X}_j)^2}{r_{j-1}}$ (equation (5.2.11)).

We insert the estimator for σ^2 into the log likelihood function and get

$$\begin{aligned}\ln L(\boldsymbol{\phi}, \boldsymbol{\theta}, \sigma^2) &= -\frac{n}{2}\ln(2\pi n^{-1}S(\boldsymbol{\phi}, \boldsymbol{\theta})) - \frac{1}{2}\ln(r_0 \ldots r_{n-1}) - \frac{1}{2n^{-1}S(\boldsymbol{\phi}, \boldsymbol{\theta})}\sum_{j=1}^{n}\frac{(X_j - \hat{X}_j)^2}{r_{j-1}} \\ &= -\frac{n}{2}\ln(2\pi n^{-1}S(\boldsymbol{\phi}, \boldsymbol{\theta})) - \frac{1}{2}\ln(r_0 \ldots r_{n-1}) - \frac{n}{2}\end{aligned}$$

We see that in order to maximize the last equation with respect to $\boldsymbol{\phi}$ and $\boldsymbol{\theta}$, we must minimize

$$\ln(n^{-1}S(\hat{\boldsymbol{\phi}}, \hat{\boldsymbol{\theta}})) + n^{-1}\sum_{j=1}^{n}\ln r_{j-1}$$

Problem 5.13

The result of Problem A.7: $(\mathbf{X}-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\mathbf{X}-\boldsymbol{\mu})$ has a chi square distribution with n degrees of freedom. We want to use this and the approximate large-sample normal distribution of the maximum likelihood estimator $\hat{\phi}_p$, $\hat{\phi}_p \sim \mathcal{N}(\phi, n^{-1}\sigma^2\Gamma_p^{-1})$, to establish (5.5.1).

$$\mathrm{E}(Y_{n+1} - \hat{\phi}_1 Y_n - \ldots \hat{\phi}_p Y_{n+1-p})^2 = \sigma^2 + \mathrm{E}[(\hat{\phi}_p - \phi_p)^T \Gamma_p (\hat{\phi}_p - \phi_p)]$$
$$= \sigma^2 + \frac{\sigma^2}{n}\mathrm{E}[(\hat{\phi}_p - \phi_p)^T \frac{n}{\sigma^2}\Gamma_p(\hat{\phi}_p - \phi_p)] = \sigma^2 + \frac{\sigma^2}{n}\mathrm{E}[(\hat{\phi}_p - \phi_p)^T (n^{-1}\sigma^2\Gamma_p^{-1})^{-1}(\hat{\phi}_p - \phi_p)]$$
$$= \sigma^2 + \frac{\sigma^2}{n}p = \sigma^2(1 + \frac{p}{n})$$

Chapter 5

Problem 5.1. We begin by writing the Yule-Walker equations. $\{Y_t : t \in \mathbb{Z}\}$ satisfies

$$Y_t - \phi_1 Y_{t-1} - \phi_2 Y_{t-2} = Z_t, \quad \{Z_t : t \in \mathbb{Z}\} \sim \mathrm{WN}(0, \sigma^2).$$

Multiplying this equation with Y_{t-k} and take expectation gives

$$\gamma(k) - \phi_1\gamma(k-1) - \phi_2\gamma(k-2) = \begin{cases} \sigma^2 & k = 0, \\ 0 & k \geq 1. \end{cases}$$

We rewrite the first three equations as

$$\phi_1\gamma(k-1) + \phi_2\gamma(k-2) = \begin{cases} \gamma(k) & k = 1, 2, \\ \gamma(0) - \sigma^2 & k = 0. \end{cases}$$

Introducing the notation

$$\mathbf{\Gamma}_2 = \begin{pmatrix} \gamma(0) & \gamma(1) \\ \gamma(1) & \gamma(0) \end{pmatrix}, \ \boldsymbol{\gamma}_2 = \begin{pmatrix} \gamma(1) \\ \gamma(2) \end{pmatrix}, \ \boldsymbol{\phi} = \begin{pmatrix} \phi_1 \\ \phi_2 \end{pmatrix}$$

we have $\mathbf{\Gamma}_2\boldsymbol{\phi} = \boldsymbol{\gamma}_2$ and $\sigma^2 - \gamma(0) - \boldsymbol{\phi}^T\boldsymbol{\gamma}_2$. We replace $\mathbf{\Gamma}_2$ by $\hat{\mathbf{\Gamma}}_2$ and $\boldsymbol{\gamma}_2$ by $\hat{\boldsymbol{\gamma}}_2$ and solve to get an estimate $\hat{\boldsymbol{\phi}}$ for $\boldsymbol{\phi}$. That is, we solve

$$\hat{\mathbf{\Gamma}}_2\hat{\boldsymbol{\phi}} = \hat{\boldsymbol{\gamma}}_2 \qquad \hat{\sigma}^2 = \hat{\gamma}(0) - \hat{\boldsymbol{\phi}}^T\hat{\boldsymbol{\gamma}}_2.$$

Hence

$$\begin{aligned} \hat{\boldsymbol{\phi}} = \hat{\mathbf{\Gamma}}_2^{-1}\hat{\boldsymbol{\gamma}}_2 &= \frac{1}{\hat{\gamma}(0)^2 - \hat{\gamma}(1)^2} \begin{pmatrix} \hat{\gamma}(0) & -\hat{\gamma}(1) \\ -\hat{\gamma}(1) & \hat{\gamma}(0) \end{pmatrix} \begin{pmatrix} \hat{\gamma}(1) \\ \hat{\gamma}(2) \end{pmatrix} \\ &= \frac{1}{\hat{\gamma}(0)^2 - \hat{\gamma}(1)^2} \begin{pmatrix} \hat{\gamma}(0)\hat{\gamma}(1) & -\hat{\gamma}(1)\hat{\gamma}(2) \\ -\hat{\gamma}(1)^2 & \hat{\gamma}(0)\hat{\gamma}(2) \end{pmatrix}. \end{aligned}$$

We get that

$$\begin{aligned} \hat{\phi}_1 &= \frac{(\hat{\gamma}(0) - \hat{\gamma}(2))\hat{\gamma}(1)}{\hat{\gamma}(0)^2 - \hat{\gamma}(1)^2} = 1.32 \\ \hat{\phi}_2 &= \frac{\hat{\gamma}(0)\hat{\gamma}(2) - \hat{\gamma}(1)^2}{\hat{\gamma}(0)^2 - \hat{\gamma}(1)^2} = -0.634 \\ \hat{\sigma}^2 &= \hat{\gamma}(0) - \hat{\phi}_1\hat{\gamma}(1) - \hat{\phi}_2\hat{\gamma}(2) = 289.18. \end{aligned}$$

We also have that $\hat{\boldsymbol{\phi}} \sim \mathrm{AN}(\boldsymbol{\phi}, \sigma^2\mathbf{\Gamma}_2^{-1}/n)$ and approximately $\hat{\boldsymbol{\phi}} \sim \mathrm{AN}(\boldsymbol{\phi}, \hat{\sigma}^2\hat{\mathbf{\Gamma}}_2^{-1}/n)$. Here

$$\hat{\sigma}^2\hat{\mathbf{\Gamma}}_2^{-1}/n = \frac{289.18}{100} \begin{pmatrix} 0.0021 & -0.0017 \\ -0.0017 & 0.0021 \end{pmatrix} = \begin{pmatrix} 0.0060 & -0.0048 \\ -0.0048 & 0.0060 \end{pmatrix}$$

So we have approximately $\hat{\phi}_1 \sim N(\phi_1, 0.0060)$ and $\hat{\phi}_2 \sim N(\phi_2, 0.0060)$ and the confidence intervals are

$$\begin{aligned} I_{\phi_1} &= \hat{\phi}_1 \pm \lambda_{0.025}\sqrt{0.006} = 1.32 \pm 0.15 \\ I_{\phi_2} &= \hat{\phi}_2 \pm \lambda_{0.025}\sqrt{0.006} = -0.634 \pm 0.15. \end{aligned}$$

Problem 5.3. a) $\{X_t : t \in \mathbb{Z}\}$ is causal if $\phi(z) \neq 0$ for $|z| \leq 1$ so let us check for which values of ϕ this can happen. $\phi(z) = 1 - \phi z - \phi^2 z^2$ so putting this equal to zero implies

$$z^2 + \frac{z}{\phi} - \frac{1}{\phi^2} = 0 \Rightarrow z_1 = -\frac{1-\sqrt{5}}{2\phi} \text{ and } z_2 = -\frac{1+\sqrt{5}}{2\phi}$$

Furthermore $|z_1| > 1$ if $|\phi| < (\sqrt{5}-1)/2 = 0.61$ and $|z_2| > 1$ if $|\phi| < (1+\sqrt{5})/2 = 1.61$. Hence, the process is causal if $|\phi| < 0.61$.
b) The Yule-Walker equations are

$$\gamma(k) - \phi\gamma(k-1) - \phi^2\gamma(k-2) = \begin{cases} \sigma^2 & k = 0, \\ 0 & k \geq 1. \end{cases}$$

Rewriting the first 3 equations and using $\gamma(k) = \gamma(-k)$ gives

$$\begin{aligned} \gamma(0) - \phi\gamma(1) - \phi^2\gamma(2) &= \sigma^2 \\ \gamma(1) - \phi\gamma(0) - \phi^2\gamma(1) &= 0 \\ \gamma(2) - \phi\gamma(1) - \phi^2\gamma(0) &= 0. \end{aligned}$$

Multiplying the third equation by ϕ^2 and adding the first gives

$$\begin{aligned} -\phi^3\gamma(1) - \phi\gamma(1) - \phi^4\gamma(0) + \gamma(0) &= \sigma^2 \\ \gamma(1) - \phi\gamma(0) - \phi^2\gamma(1) &= 0. \end{aligned}$$

We solve the second equation to obtain

$$\phi = -\frac{1}{2\rho(1)} \pm \sqrt{\frac{1}{4\rho(1)^2} + 1}.$$

Inserting the estimated values of $\hat{\gamma}(0)$ and $\hat{\gamma}(1) = \hat{\gamma}(0)\hat{\rho}(1)$ gives the solutions $\hat{\phi} = \{0.509, -1.965\}$ and we choose the causal solution $\hat{\phi} = 0.509$. Inserting this value in the expression for σ^2 we get

$$\hat{\sigma}^2 = -\hat{\phi}^3\hat{\gamma}(1) - \hat{\phi}\hat{\gamma}(1) - \hat{\phi}^4\hat{\gamma}(0) + \hat{\gamma}(0) = 2.985.$$

Problem 5.4. a) Let us construct a test to see if the assumption that $\{X_t - \mu : t \in \mathbb{Z}\}$ is $\text{WN}(0, \sigma^2)$ is reasonable. To this end suppose that $\{X_t - \mu : t \in \mathbb{Z}\}$ is $\text{WN}(0, \sigma^2)$. Then, since $\rho(k) = 0$ for $k \geq 1$ we have that $\hat{\rho}(k) \sim \text{AN}(0, 1/n)$. A 95% confidence interval for $\rho(k)$ is then $I_{\rho(k)} = \hat{\rho}(k) \pm \lambda_{0.025}/\sqrt{200}$. This gives us

$$\begin{aligned} I_{\rho(1)} &= 0.427 \pm 0.139 \\ I_{\rho(2)} &= 0.475 \pm 0.139 \\ I_{\rho(3)} &= 0.169 \pm 0.139. \end{aligned}$$

Clearly $0 \notin I_{\rho(k)}$ for any of the observed $k = 1, 2, 3$ and we conclude that it is not reasonable to assume that $\{X_t - \mu : t \in \mathbb{Z}\}$ is white noise.
b) We estimate the mean by $\hat{\mu} = \overline{x}_{200} = 3.82$. The Yule-Walker estimates is given by

$$\hat{\boldsymbol{\phi}} = \hat{\mathbf{R}}_2^{-1}\hat{\boldsymbol{\rho}}_2, \quad \hat{\sigma}^2 = \hat{\gamma}(0)(1 - \hat{\boldsymbol{\rho}}_2{}^T\hat{\mathbf{R}}_2^{-1}\hat{\boldsymbol{\rho}}_2),$$

where

$$\hat{\boldsymbol{\phi}} = \begin{pmatrix} \hat{\phi}_1 \\ \hat{\phi}_2 \end{pmatrix}, \; \hat{\mathbf{R}}_2 = \begin{pmatrix} \hat{\rho}(0) & \hat{\rho}(1) \\ \hat{\rho}(1) & \hat{\rho}(0) \end{pmatrix}, \; \hat{\boldsymbol{\rho}}_2 = \begin{pmatrix} \hat{\rho}(1) \\ \hat{\rho}(2) \end{pmatrix}.$$

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 4

a Let

$$Y_t = \mu + \phi Y_{t-1} + Z_t$$

It can be rewritten as:

$$Y_t^* = \phi Y_{t-1}^* + Z_t \tag{1}$$

with $Y_t^* = Y_t - \mu$.

Making use of the Durbin-Levinson algorithm for Y_t with

$$\hat{\nu}_0 = E[Y_1^* + \hat{\mu} - P_0 Y_1^* - \hat{\mu}]^2 = \hat{\gamma}_{Y^*}(0) = \hat{\gamma}_Y(0) - \hat{\mu}^2 = \frac{\hat{\sigma}^2}{1 - \hat{\phi}^2}$$

$$\hat{\phi} = \hat{\phi}_{11} = \frac{\hat{\gamma}_{Y^*}(1)}{\hat{\gamma}_{Y^*}(0)} = \frac{\hat{\gamma}_Y(1) - \hat{\mu}^2}{\hat{\gamma}_Y(0) - \hat{\mu}^2}$$

Let's remember that according to the properties of the prediction operator, $P_n Y_{n+1} = P_n Y_{n+1}^* + \mu$. Based on the expression for $\hat{\gamma}_Y(0)$ we find

$$\hat{\sigma}^2 = \hat{\nu}_1 = \hat{\nu}_0(1 - \hat{\phi}^2) = (\hat{\gamma}_Y(0) - \hat{\mu}^2)\left[1 - \left(\frac{\hat{\gamma}_Y(1) - \hat{\mu}^2}{\hat{\gamma}_Y(0) - \hat{\mu}^2}\right)^2\right]$$

b We start by removing μ from Y_t for all t, i.e. $Y_t^* = Y_t - \mu$. It means that for every all t the innovations

$$\begin{aligned} Y_t - \hat{Y}_t &= Y_t - P_{t-1}Y_t \\ &= Y_t^* + \mu - P_{t-1}Y_t^* - \mu \\ &= Y_t^* - P_{t-1}Y_t^* \\ &= Y_t^* - \hat{Y}_t^* \end{aligned}$$

Before stating the likelihood function when Gaussian errors are assumed, let's compute some useful terms for it. Based on the innovations algorithm $\hat{Y}_{t+1}^* = \phi Y_t^*, t \geq 1$ and (remind that for AR(1), $\gamma(0) = \frac{\sigma^2}{1-\phi^2}$ and $\gamma(1) = \frac{\sigma^2\phi}{1-\phi^2}$)

$$\nu_0 = \gamma(0) = r_0 \cdot \sigma^2 \rightarrow r_0 = \frac{1}{1-\phi^2}$$

$$\nu_1 = E[(Y_2^* - \hat{Y}_2^*)^2] = \gamma(0) - 2\phi\gamma(1) + \phi^2\gamma(0) = r_1 \cdot \sigma^2 \rightarrow r_1 = \frac{\sigma^2}{\sigma^2}\left[\frac{1-\phi^2}{1-\phi^2}\right] = 1$$

$$\vdots$$

$$\nu_n = E[(Y_n^* - \hat{Y}_n^*)^2] = \gamma(0) - 2\phi\gamma(1) + \phi^2\gamma(0) = r_n \cdot \sigma^2 \rightarrow r_n = \frac{\sigma^2}{\sigma^2}\left[\frac{1-\phi^2}{1-\phi^2}\right] = 1$$

Then, $r_0 = \frac{1}{1-\phi^2}$ and $r_t = 1$, $t \geq 1$. The log-likelihood is given by:

$$l(\phi, \sigma^2) = -\frac{n}{2}\ln(2\pi\sigma^2) - \frac{1}{2}\ln(r_0 \cdot r_1 \cdot \cdots \cdot r_{n-1}) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}\frac{(Y_i^* - \hat{Y}_i^*)^2}{r_{i-1}}$$

For this case it becomes

$$l(\phi, \sigma^2) = -\frac{n}{2}\ln(2\pi\sigma^2) + \ln(1-\phi^2) - \frac{1}{2\sigma^2}\left(Y_1^{*^2}(1-\phi^2) + \sum_{i=2}^{n}(Y_i^* - \phi Y_{i-1}^*)^2\right)$$

Now we compute the estimates of σ^2, ϕ and μ

$$\frac{\partial l}{\partial \sigma^2} = -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4}\left(Y_1^{*^2}(1-\phi^2) + \sum_{i=2}^{n}(Y_i^* - \phi Y_{i-1}^*)^2\right) = 0$$

$$\frac{\partial l}{\partial \phi} = -\frac{\phi}{1-\phi^2} + \frac{1}{\sigma^2}\left(\phi Y_1^{*^2} - \sum_{i=2}^{n}(Y_i^* - \phi Y_{i-1}^*)(-Y_{i-1}^*)\right) = 0$$

$$\frac{\partial l}{\partial \mu} = (Y_1-\mu)(1-\phi^2) + \sum_{i=2}^{n}(Y_i - \phi Y_{i-1} - \mu(1-\phi))(1-\phi) = 0$$

Then,

$$\sigma^2 = \frac{Y_1^2(1-\phi^2) + \sum_{i=2}^{n}(Y_i - \phi Y_{i-1})^2}{n}, \quad ,$$

$$\mu = \frac{(1-\phi^2)Y_1 + \sum_{i=2}^{n}(1-\phi)(Y_i - \phi Y_{i-1})}{n(1-\phi)^2 - 2\phi}$$

and $\hat{\phi}$ is the solution of the cubic equation:

$$\phi^3\left[(1-n)\sum_{i=2}^{n-1} Y_i^{*^2}\right] - \phi^2\left[(2-n)\sum_{i=2}^{n} Y_i^* Y_{i-1}^*\right] + \phi\left[Y_1^{*^2} + (n+1)\sum_{i=2}^{n-1} Y_i^{*^2} + Y_n^{*^2}\right] - n\left[\sum_{i=2}^{n} Y_i^* Y_{i-1}^*\right] = 0 \quad (2)$$

Obtained from

$$\frac{\phi}{1-\phi^2} = n\frac{\phi Y_1^{*^2} - \sum_{i=2}^{n}(Y_i^* - \phi Y_{i-1}^*)(-Y_{i-1}^*)}{Y_1^{*^2}(1-\phi^2) + \sum_{i=2}^{n}(Y_i^* - \phi Y_{i-1}^*)^2}$$

Finally,

$$\hat{\sigma}^2 = \frac{Y_1^{*^2}(1-\hat{\phi}^2) + \sum_{i=2}^{n}(Y_i^* - \hat{\phi} Y_{i-1}^*)^2}{n}$$

c Now let's explore the implications of having $\phi = 1$. Were that the case, the AR(1) model would become:

$$Y_t^* = Y_{t-1}^* + Z_t$$

It has an effect on the ACVF of Y_t^* since for every k:

$$E(Y_t Y_{t+k}) = E(Y_t^* Y_{t+k}^*) + \mu^2 = E\left(Y_t^*, Y_t^* + \sum_{i=1}^{k} Z_{t+i}\right) + \mu^2$$

$$= \gamma(0) + \mu^2$$

In addition to it,

$$E(Y_t) = E(Y_1) + \sum_{i=1}^{t-1} E(Z_{1+i}) = E(Y_1) = \mu$$

Thus, the information of the process is reduced to $E(Y_1)$ and $\gamma(0)$, estimated by $\hat{\mu}$ and $\hat{\gamma}(0)$.

d Our process is:

$$Y_t^* = Z_t + \theta Z_{t-1}$$

Now, we need to follow the innovations algorithm to both fit the process and to express the likelihood function in terms of θ and σ^2. Let's begin the algorithm with $\nu_0 = \gamma(0)$. Then,

$$\theta_{11} = \nu_0^{-1}[\gamma(1)] = \rho(1)$$

$$\nu_1 = \gamma(0) - \theta^2 \nu_0$$

$$\vdots$$

$$\theta_{n,n-k} = \nu_k^{-1}\left(\gamma(n-k) - \sum_{j=0}^{k-1} \theta_{k,k-j}\theta_{n,n-j}\nu_j\right), \qquad 0 \le k < n$$

$$\nu_n = \gamma(0) - \sum_{j=0}^{n-1} \theta_{n,n-j}^2 \nu_j$$

Given that for an MA(1) process $\gamma(k) = 0$ for $k \geq 2$, then for each n only $\theta_{n1} \neq 0$. Thus, for all $n \geq 1$:

$$\theta_{n1} = \nu_{n-1}^{-1}[\gamma(1)]$$

$$\nu_n = \gamma(0) - \theta_{n1}^2 \nu_1$$

We will focus on estimating θ and σ^2 using the innovations algorithm. Let

$$Y_t^* = Z_t + \hat{\theta}_{11} Z_{t-1}; \quad \{Z_t\} \sim WN(0, \hat{\nu}_1)$$

Then,

$$\hat{\theta} = \hat{\theta}_{11} = \hat{\rho}_{Y^*}(1) = \frac{\hat{\gamma}_Y(1) - \hat{\mu}^2}{\hat{\gamma}_Y(0) - \hat{\mu}^2}$$

and

$$\hat{\sigma}^2 = \hat{\nu}_1 = \frac{\hat{\gamma}_{Y^*}^2(0) - \hat{\gamma}_{Y^*}^2(1)}{\hat{\gamma}_{Y^*}(0)} = \frac{(\hat{\gamma}_Y(0) - \hat{\mu}^2)^2 - (\hat{\gamma}_Y(1) - \hat{\mu}^2)^2}{\hat{\gamma}_Y(0) - \hat{\mu}^2}$$

Based on the expression for the likelihood in part b and on the independence of r_i and $(Y_i^* - \hat{Y}_i^*)$ from σ^2, we get:

$$\sigma^2 = \sum_{i=1}^{n} \frac{(Y_i^* - \hat{Y}_i^*)^2}{n \cdot r_{i-1}}$$

On the other hand, given that r_i is not as easy to express as in the AR(1) case, we can say that θ is the value that minimizes (see exercise 5.8)

$$l(\theta) = \ln\left(\sum_{i=1}^{n} \frac{(Y_i^* - \hat{Y}_i^*)^2}{n \cdot r_{i-1}}\right) + \frac{\sum_{i=1}^{n} \ln(r_{i-1})}{n}$$

with $\hat{Y}_i^* = \theta_{i-1,1}(Y_{i-1}^* - \hat{Y}_{i-1}^*), i \geq 1$ and $r_{i-1} = \frac{\nu_{i-1}}{\sigma^2}$. This problem can be solved numerically.

Finally, in case $\theta = 1$, our process becomes:

$$Y_t^* = Z_t + Z_{t-1}$$

meaning this that the process is the addition of two zero-mean white noise terms with variance $2\sigma^2$. The ACVF of Y_t is:

$$E(Y_t Y_{t+k}) = \begin{cases} 2\sigma^2 + \mu^2 & k = 0 \\ \sigma^2 + \mu^2 & k = 1 \end{cases}$$

And

$$E(Y_t) = E(Y_t^* + \mu) = \mu$$

Then, all the information of the process ends up contained in $\hat{\mu}$ and σ^2. It also affects the innovations algorithm since every θ_{i1} and ν_1 become functions of $\gamma(0)$. Thus, $\hat{\sigma}^2$ depends on $\gamma(0)$ as well. Making all the information of the process contained in $\hat{\mu}$ and $\hat{\gamma}(0)$