Time Series Econometrics
Assignment 1 - Spring 2022

Project Group 5
Fabio Luigi Eugenio Bernasconi (15-942-857)
Maria Kyprouli (21-607-734)
Lauritz Storch (21-607-015)

27 March 2022

1 Stationarity, ergodicity and dependence

(a) The stochastic process $X = (X_t)_{t \in \mathbb{Z}}$ satisfies

$$X_t = cos(t)\xi + sin(t)\eta \tag{1}$$

where ξ is a centered Bernoulli random variable such that $\mathbb{P}[\xi = \pm 1] = 0.5$ and η is a random variable with distribution $N(0, 1)$ independent of ξ.

We start by checking whether the process is covariance stationary, which requires that

(a) $\mathbb{E}[X_t] = \mu$

(b) $\gamma_{t,t} =$ constant

(c) $\gamma_{r,s} = \gamma_{r+t,s+t} \qquad \forall r, s, t$

For $\mathbb{E}[X_t]$ we have:

$$\mathbb{E}[X_t] = cos(t)\,\mathbb{E}[\xi] + sin(t)\,\mathbb{E}[\eta] = 0 \tag{2}$$

For $\gamma_{t,t}$ we obtain:

$$\begin{aligned}\gamma_{t,t} &= Var(X_t) = Var(cos(t)\xi + sin(t)\eta) \\ &\overset{\text{indep.}}{=} cos^2(t)var(\xi) + sin^2 var(\eta) \\ &= cos^2(t) + sin^2(t) \\ &= 1\end{aligned} \tag{3}$$

For $\gamma_{t,t+h}$ we get:

$$\begin{aligned}
\gamma_{t,t+h} &= Cov(x_t, x_{t+h}) \\
&= \mathbb{E}[(X_t - \mu_t)(X_{t+h} - \mu_{t+h})] \\
&= \mathbb{E}[X_t X_{t+h}] \\
&= \mathbb{E}[(cos(t)\xi + sin(t)\eta)(cos(t+h)\xi + sin(t+h)\eta)] \\
&= cos(t)cos(t+h)\,\mathbb{E}[\xi^2] + sin(t)sin(t+h)\,\mathbb{E}[\eta^2] \\
&\qquad + [sin(t)cos(t+h) + cos(t)sin(t+h)]\,\mathbb{E}[\eta\xi] \\
&\overset{\text{indep.}}{=} cos(t)cos(t+h)\,\mathbb{E}[\xi^2] + sin(t)sin(t+h)\,\mathbb{E}[\eta^2] \\
&\qquad + [sin(t)cos(t+h) + cos(t)sin(t+h)]\,\mathbb{E}[\eta]\,\mathbb{E}[\xi] \\
&= cos(t)cos(t+h) + sin(t)sin(t+h) \\
&= cos(h)
\end{aligned} \tag{4}$$

Thus, we see that the process has constant mean and the co-variance depends only on h and not on t. Hence, all the properties for weak stationarity are satisfied and the process is **covariance stationary**.

We then check for strict stationarity. We know that the distribution of X_t is given by

$$p_x(x_t) = \begin{cases} sin(t)\eta + cos(t), & if \quad \xi = 1 \\ sin(t)\eta - cos(t), & if \quad \xi = -1 \end{cases} \tag{5}$$

Hence, in the case where $\xi = 1$ we have $X_t \sim N(cos(t), sin^2(t))$. It follows that

$$\mathbb{P}[X_t|\xi = 1] = \frac{1}{sin(t)\sqrt{2\pi}} \int_{-\infty}^{\eta} e^{-\frac{1}{2}x_t^2} dx_t \tag{6}$$

and

$$\begin{aligned}
&\mathbb{P}[X_1, ..., x_n|\xi = 1] \\
&= \prod_{i=1}^{n} sin^{-1}(i) \left(\frac{1}{\sqrt{2\pi}}\right)^n \int_{-\infty}^{\eta} ... \int_{-\infty}^{\eta} e^{-\frac{1}{2}\sum_{i=1}^{n} x_i} dx_1, ..., dx_n \\
&\neq \prod_{i=1}^{n} sin^{-1}(i+h) \left(\frac{1}{\sqrt{2\pi}}\right)^n \int_{-\infty}^{\eta} ... \int_{-\infty}^{\eta} e^{-\frac{1}{2}\sum_{i=1}^{n} x_{i+h}} dx_1, ..., dx_n \\
&= \mathbb{P}[X_{1+h}, ..., x_{n+h}|\xi = 1]
\end{aligned} \tag{7}$$

The derivation for $\mathbb{P}[X_1, ..., x_n|\xi = -1]$ is analogous.
Hence, we find that

$$\begin{aligned}
&\mathbb{P}[x_1, ..., x_n|\xi = 1] + \mathbb{P}[x_1, ..., x_n|\xi = -1] \\
&\neq \mathbb{P}[x_{1+h}, ..., x_{n+h}|\xi = 1] + \mathbb{P}[x_{1+h}, ..., x_{n+h}|\xi = -1]
\end{aligned} \tag{8}$$

Which is the same as

$$\mathbb{P}[x_1, ..., x_n] \neq \mathbb{P}[x_{1+h}, ..., x_{n+h}] \tag{9}$$

and ultimately indicates that the process **is not strictly stationary**.

(b) The process X is given by

$$X_t = U\epsilon_t \tag{10}$$

where $\epsilon_t \sim iid(0, 1)$ and the iid sequence is independent of the binary random variable U with $\mathbb{P}[U = u] = 0.5$ for $u \in \{0, 1\}$

We start by investigating weak stationarity. For $\mathbb{E}[X_t]$ we have:

$$\mathbb{E}[X_t] = \mathbb{E}[U\epsilon_t] \overset{\text{indep.}}{=} \mathbb{E}[U]\,\mathbb{E}[\epsilon_t] = 0 = \mu \tag{11}$$

For $\gamma_{t,t+h}$ we get:

$$\begin{aligned}\gamma_{t,t+h} &= \mathbb{E}[(X_t - \mu_t)(X_{t+h} - \mu_{t+h})]\\ &= \mathbb{E}[X_t X_{t+h}]\\ &= \mathbb{E}[U^2 \epsilon_t \epsilon_{t+h}]\\ &\overset{\text{indep.}}{=} \mathbb{E}[U^2]\,\mathbb{E}[\epsilon_t \epsilon_{t+h}]\end{aligned} \tag{12}$$

thus, for $h = 0$ we obtain:

$$\mathbb{E}[U^2]\,\mathbb{E}[\epsilon_t \epsilon_{t+h}] = \mathbb{E}[U^2]\,\mathbb{E}[\epsilon_t^2] = 0.5 \tag{13}$$

and for $|h| > 0$ we obtain:

$$\mathbb{E}[U^2]\,\mathbb{E}[\epsilon_t \epsilon_{t+h}] \overset{\text{indep.}}{=} \mathbb{E}[u^2]\,\mathbb{E}[\epsilon_t]\,\mathbb{E}[\epsilon_{t+h}] = 0 \tag{14}$$

All the properties for weak stationarity are satisfied because the average of the process and the auto-covariances for all h are constant. Hence, the process is **covariance stationary** or **weakly stationary**.

We then turn to investigating whether the process is strictly stationary. We know that X_t is distributed according to

$$p_X(x_t) \overset{\text{indep.}}{=} p_U(u)p_\epsilon(\epsilon_t) = 0.5^u 0.5^{1-u} p_\epsilon(\epsilon_t) \tag{15}$$

where $\epsilon_t \sim iid(0, 1)$. Hence, we know that each X_t has the same exact distribution, which makes it clear that X is an *iid* process. The joint probability distribution will then clearly be the product of the marginal probability distributions and will be time-invariant. We can conclude then that the process is **strictly stationary** as well.

Finally, we investigate ergodicity. The process is **ergodic for the mean**. Indeed, the process is covariance-stationary and all auto-covariances for $|h| > 0$ are equal to 0, which implies that

$$\lim_{H\to\infty} \frac{1}{H} \sum_{h=1}^{H} \gamma_h = \lim_{H\to\infty} 0 = 0 \tag{16}$$

which is a necessary and sufficient condition for ergodicity for the mean.

2 Features of ARMA models

(a) The process $z = (z_t)_{t\in\mathbb{Z}}$ satisfies

$$z_t = -0.3z_{t-1} + 0.4z_{t-2} + u_t - 0.5u_{t-1} + 4u_{t-2} - 2u_{t-3} \tag{17}$$

At first glance, this appears to be an **ARMA(2,3)** process.

Derivation of reduced form ARMA process
To check for common roots, we rewrite the process as

$$\Psi(L)z_t = \Theta(L)u_t \tag{18}$$

where

$$\begin{aligned} \Psi(L) &= 1 + 0.3L - 0.4L^2 \\ \Theta(L) &= 1 - 0.5L + 4L^2 - 2L^3 \end{aligned} \tag{19}$$

Of the two, $\Psi(L)$ is the lag polynomial with the smallest order and thus we start with that one.
The lag polynomial $\Psi(L)$ can be factorized as

$$\Psi(z) = (1 + 0.3z - 0.4z^2) = (1 + 0.8z)(1 - 0.5z) \tag{20}$$

Thus, the two roots of the polynomial $\Psi(z)$ are:

$$\begin{aligned} z^*_{1,\Psi} &= -1.25 \\ z^*_{2,\Psi} &= 2 \end{aligned} \tag{21}$$

The lag polynomial $\Theta(L)$, on the other hand, can be factorized as

$$\begin{aligned} \Theta(z) &= (1 - 0.5z + 4z^2 - 2z^3) \\ &= (1 - 0.5z)(1 + 4z^2) \\ &= (1 - 0.5z)(1 - 2iz)(1 + 2iz) \end{aligned} \tag{22}$$

Hence, the three roots of the polynomial $\Theta(z)$ are:

$$\begin{aligned} z^*_{1,\Theta} &= 2 \\ z^*_{2,\Theta} &= i/2 \\ z^*_{3,\Theta} &= -i/2 \end{aligned} \tag{23}$$

It is now clear that the two polynomials share a common root: $z^*_{2,\Psi} = z^*_{1,\Theta} = 2$.
We could thus define a reduced-form process of the form

$$\tilde{\Psi}(L)z_t = \tilde{\Theta}(L)u_t \tag{24}$$

where

$$\begin{aligned}\tilde{\Psi}(L) &= 1 + 0.8L \\ \tilde{\Theta}(L) &= 1 + 4L^2\end{aligned} \tag{25}$$

This reduced-form process, an **ARMA(1,2)** process, is equivalent to the initial ARMA(2,3) process and doesn't have any common roots anymore.

Check causality and invertibility

We now check whether the reduced form ARMA process is causal and invertible.

To check **invertibility**, we need to consider the polynomial of the MA part of the process $\tilde{\Theta}(L)$.

$$\tilde{\Theta}(z) = (1 + 4z^2) = (1 + 2iz)(1 - 2iz) \overset{!}{=} 0 \tag{26}$$

The roots of the polynomial are:

$$\begin{aligned}\tilde{z}^*_{1,\Theta} &= i/2 \\ \tilde{z}^*_{2,\Theta} &= -i/2\end{aligned} \tag{27}$$

Thus:

$$\begin{aligned}|\tilde{z}^*_{1,\Theta}| &= \sqrt{0^2 + 0.5^2} = 0.5 < 1 \\ |\tilde{z}^*_{2,\Theta}| &= \sqrt{0^2 + (-0.5)^2} = 0.5 < 1\end{aligned} \tag{28}$$

As both roots are within the unit circle, the process **isn't invertible**.

To check **causality**, we need to consider the polynomial of the AR part of the process. All the roots of $\tilde{\Phi}(L)$ must lie outside the unit circle.
We have that:

$$\tilde{\Phi}(z) = (1 + 0.8z) \overset{!}{=} 0 \tag{29}$$

The root of the polynomial is:

$$\tilde{z}^*_{1,\Phi} = -1.25 \tag{30}$$

Thus:

$$|\tilde{z}^*_{1,\Phi}| = 1.25 > 1 \tag{31}$$

Hence, the root lies outside the unit circle and the process is causal.
As a result, the reduced form ARMA(1,2) process is **causal**.

Infer causal and invertible representation

By inverting the roots of the $\tilde{\Theta}(L)$ lag polynomial, we can infer the equivalent causal and invertible representation for the reduced form process

$$\tilde{\Psi}(L)z_t = \hat{\Theta}(L)\tilde{u}_t \tag{32}$$

where

$$\begin{aligned} \tilde{\Psi}(L) &= 1 + 0.8L \\ \hat{\Theta}(L) &= (1 + \frac{i}{2}L)(1 - \frac{i}{2}L) = (1 + 0.25L^2) \end{aligned} \tag{33}$$

and the new iid process $\tilde{u}_t$ is still mean zero but $Var(\tilde{u}_t) = 16$. We know that, by the inversion of the roots of the lag polynomial, this process should be equivalent to the previous reduced form process.

Compare second order properties

In order to compare the second order properties of the reduced form process and of the original process, we derive the second order properties for a **general** stationary model of the form

$$X_t = \phi X_{t-1} + \epsilon + \theta\epsilon_{t-2} \tag{34}$$

For the **mean** of the process we have that

$$\mathbb{E}[X_t] = \mathbb{E}[\phi X_{t-1} + \epsilon_t + \theta\epsilon_{t-2}] = \phi\,\mathbb{E}[X_{t-1}] \tag{35}$$

and since the process is stationary

$$\begin{aligned} \mu &= \phi\mu \\ \Leftrightarrow \mu(1+\phi) &= 0 \\ \Leftrightarrow \mu &= 0 \end{aligned} \tag{36}$$

Hence, the mean of the general stationary model is 0.
The model can be re-written as

$$\Phi(L)X_t = \Theta(L)\epsilon_t \tag{37}$$

where

$$\begin{aligned} \Phi(L) &= 1 - \phi L \\ \Theta(L) &= 1 + \theta L^2 \end{aligned} \tag{38}$$

The stationary solution of the process is then given by

$$X_t = \frac{\Theta(L)}{\Phi(L)}\epsilon_t = \Psi(L)\epsilon_t \tag{39}$$

and we can compute that

$$\begin{aligned}
&\Phi^{-1}(L)\Theta(L) = \Psi(L) \\
\Leftrightarrow &(1-\phi L)^{-1}(1+\theta L^2) = \Psi(L) \\
\Leftrightarrow &\left(\sum_{i=0}^{\infty} \phi^i L^i\right)(1+\theta L^2) = \Psi(L) \\
\Leftrightarrow &\Psi(L) = \sum_{i=0}^{\infty} \phi^i L^i + \theta \sum_{i=0}^{\infty} \phi^i L^{i+2} \\
\Leftrightarrow &\Psi(L) = 1 + \phi L + (\phi^2+\theta) \sum_{i=0}^{\infty} \phi^i L^{i+2}
\end{aligned} \tag{40}$$

For the **auto-covariances** $\gamma(k)$, $k = 0, 1, 2, ...$ we have that

$$\begin{aligned}
\gamma(k) &= \mathbb{E}[(X_t - \mu_t)(X_{t+k} - \mu_{t+k})] \\
&= \mathbb{E}[X_t X_{t+k}] \\
&= \mathbb{E}[\Psi(L)\epsilon_t \Psi(L)\epsilon_{t+k}] \\
&= \sigma_\epsilon^2 \sum_{j=0}^{\infty} \psi_j \psi_{j+k}
\end{aligned} \tag{41}$$

thus, for $k = 0$ we obtain

$$\begin{aligned}
\gamma(0) &= \sigma_\epsilon^2 \sum_{j=0}^{\infty} \psi_j^2 \\
&= \sigma_\epsilon^2 \left(1 + \phi^2 + (\phi^2+\theta)^2 \sum_{i=0}^{\infty} \phi^{2i}\right) \\
&= \sigma_\epsilon^2 \left(1 + \phi^2 + \frac{(\phi^2+\theta)^2}{1-\phi^2}\right)
\end{aligned} \tag{42}$$

similarly, for $k = 1$ we obtain

$$\begin{aligned}
\gamma(1) &= \sigma_\epsilon^2 \sum_{j=0}^{\infty} \psi_j \psi_{j+1} \\
&= \sigma_\epsilon^2 \left[\phi + \phi(\phi^2+\theta) + \phi(\phi^2+\theta)^2 \sum_{i=0}^{\infty} \phi^{2i}\right] \\
&= \sigma_\epsilon^2 \left[\phi + \phi(\phi^2+\theta) + \frac{\phi(\phi^2+\theta)^2}{1-\phi^2}\right]
\end{aligned} \tag{43}$$

and, similarly, for $k = 2$ and $k = 3$

$$
\begin{aligned}
\gamma(2) &= \sigma_\epsilon^2 \sum_{j=0}^{\infty} \psi_j \psi_{j+2} \\
&= \sigma_\epsilon^2 \left[(\phi^2 + \theta) + \phi^2(\phi^2 + \theta) + \frac{\phi^2(\phi^2 + \theta)^2}{1 - \phi^2} \right] \\
\gamma(3) &= \sigma_\epsilon^2 \left[\phi(\phi^2 + \theta) + \phi^3(\phi^2 + \theta) + \frac{\phi^3(\phi^2 + \theta)^2}{1 - \phi^2} \right]
\end{aligned}
\tag{44}
$$

But, more generally, for $k \geq 2$

$$
\gamma(k) = \sigma_\epsilon^2 \left[\phi^{k-2}(\phi^2 + \theta) + \phi^k(\phi^2 + \theta) + \frac{\phi^k(\phi^2 + \theta)^2}{1 - \phi^2} \right] \tag{45}
$$

For the initial reduced-form process that we derived we have that $\phi = -0.8$, $\theta = 4$ and $\sigma_\epsilon^2 = 1$ and thus:

$$
\begin{aligned}
\gamma(0) &= 1.64 + \frac{4.64^2}{0.36} = \frac{553}{9} \\
\gamma(1) &= -\frac{2356}{45} \\
\gamma(k) &= (-0.8)^{k-2} \frac{10324}{225} \qquad for \quad k \geq 2
\end{aligned}
\tag{46}
$$

For the new causal and invertible reduced-form process we have that $\phi = -0.8$, $\theta = 0.25$ and $\sigma_{\hat{\epsilon}}^2 = 16$ and thus:

$$
\begin{aligned}
\gamma(0) &= 16 \left(1.64 + \frac{0.89^2}{0.36} \right) = \frac{553}{9} \\
\gamma(1) &= -\frac{2356}{45} \\
\gamma(k) &= (-0.8)^{k-2} \frac{10324}{225} \qquad for \quad k \geq 2
\end{aligned}
\tag{47}
$$

Hence, the two reduced-form processes have the same second order properties and the representations are **equivalent**.

3 Estimation of ARMA models

(a) The stochastic process $X = (X_t)_{t \in \mathbb{Z}}$ satisfies

$$
X_t = \omega + \phi X_{t-1} + u_t \tag{48}
$$

where $0 < |\phi| < 1$ and $u_t \sim N(0, \sigma^2)$

The process is an AR(1) process and is stationary since $|\phi| < 1$. We rewrite the process as

$$\Phi(L)X_t = w + u_t \tag{49}$$

where

$$\Phi(L) = (1 - \phi L) \tag{50}$$

The MA(∞) representation of the process is given by

$$\begin{aligned} X_t &= \frac{w}{\Phi(L)} + \frac{u_t}{\Phi(L)} \\ &= \frac{w}{\Phi(1)} + \frac{u_t}{\Phi(L)} \\ &= \frac{w}{1-\phi} + \Psi(L)u_t \end{aligned} \tag{51}$$

where

$$\Psi(L) = \Phi^{-1}(L) = \sum_{i=0}^{\infty} \phi^i L^i \tag{52}$$

Hence, the MA(∞) representation of the process is

$$X_t = \frac{w}{1-\phi} + \sum_{j=0}^{\infty} \phi^j u_{t-j} \tag{53}$$

In order to derive the log likelihood function of the model, we compute first the mean and the variance of the process in order to derive the distribution of X_t.
The mean is given by

$$\mathbb{E}[X_t] = \frac{w}{1-\phi} \tag{54}$$

while the variance is given by

$$\begin{aligned} \gamma(0) &= Var(X_t) \\ &= Var\left(\frac{w}{1-\phi} + \sum_{j=0}^{\infty} \phi^j u_{t-j}\right) \\ &\overset{\text{iid}}{=} \sum_{j=0}^{\infty} \phi^{2j} Var(u_{t-j}) \\ &= \sigma^2 \sum_{j=0}^{\infty} \phi^{2j} \\ &= \frac{\sigma^2}{1-\phi^2} \end{aligned} \tag{55}$$

Hence, the distribution of X_t is

$$X_t \sim N\left(\frac{w}{1-\phi}, \frac{\sigma^2}{1-\phi^2}\right) \tag{56}$$

This follows from the fact that, according to the MA(∞) representation of the process, X_t is the sum of normally distributed random variables and thus normally distributed itself.

Then, we compute the conditional mean and variance to derive the conditional distribution of X_t given X_{t-1} The conditional mean is given by

$$\begin{aligned}
\mathbb{E}[X_t|X_{t-1}] &= \mathbb{E}[w + \phi X_{t-1} + u_t|X_{t-1}] \\
&= w + \phi\,\mathbb{E}[X_{t-1}|X_{t-1}] + \mathbb{E}[u_t|x_{t-1}] \\
&= w + \phi X_{t-1} + \mathbb{E}[u_t] \\
&= w + \phi X_{t-1}
\end{aligned} \tag{57}$$

The conditional variance equals

$$\begin{aligned}
Var(X_t|X_{t-1}) &= Var(w + \phi X_{t-1} + u_t|X_{t-1}) \\
&= Var(\phi X_{t-1} + u_t|X_{t-1}) \\
&\overset{\text{indep.}}{=} Var(\phi X_{t-1}|X_{t-1}) + Var(u_t|X_{t-1}) \\
&= Var(u_t) \\
&= \sigma^2
\end{aligned} \tag{58}$$

Hence, conditional distribution of X_t given X_{t-1} is

$$X_t|X_{t-1} \sim N\left(w + \phi X_{t-1}, \sigma^2\right) \tag{59}$$

and, again, is normally distributed

Now, we want to derive the conditional PDF $p(x_1, ..., x_T|x_0, \eta)$ where $\eta = (\omega, \theta, \sigma^2)$.
The Markov property for an AR(1) process tells us that

$$p(x_t|x_0, ..., x_{t-1}, \eta) = p(x_t|x_{t-1}, \eta) \tag{60}$$

which we know based on the conditional distribution that we have derived above. Hence we obtain

$$
\begin{aligned}
p(x_1,...,x_T|x_0,\eta) &= \prod_{t=1}^{T} p(x_t|x_0,...,x_{t-1},\eta) \\
&= \prod_{t=1}^{T} p(x_t|x_{t-1},\eta) \\
&= \prod_{t=1}^{T} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2\sigma^2}(x_t-(w+\phi x_{t-1}))^2} \\
&= \left(\frac{1}{\sqrt{2\pi}\sigma}\right)^{T} e^{-\frac{1}{2\sigma^2}\sum_{t=1}^{T}(x_t-(w+\phi x_{t-1}))^2} \\
&= \left(\frac{1}{2\pi\sigma^2}\right)^{T/2} e^{-\frac{1}{2\sigma^2}\sum_{t=1}^{T}(x_t-(w+\phi x_{t-1}))^2}
\end{aligned} \tag{61}
$$

Finally, we derive the log-likelihood of the model. We know that

$$
p(x_0) = \frac{1}{\sqrt{2\pi\sigma^2/\sqrt{1-\phi^2}}} e^{-\frac{1}{2\sigma^2/(1-\phi^2)}(x_0-(\frac{w}{1-\phi}))^2} \tag{62}
$$

And by the rule that the *joint pdf = conditional pdf × marginal pdf* we obtain first the likelihood function of the model

$$
\begin{aligned}
\mathcal{L}(\eta) &= p(x_1,...,x_T|x_0)p(x_0) \\
&= \left(\frac{1}{2\pi\sigma^2}\right)^{T/2} e^{-\frac{1}{2\sigma^2}\sum_{t=1}^{T}(x_t-(w+\phi x_{t-1}))^2} \\
&\quad \left(\frac{1}{2\pi\sigma^2/(1-\phi^2)}\right)^{1/2} e^{-\frac{1}{2\sigma^2/(1-\phi^2)}(x_0-(\frac{w}{1-\phi}))^2}
\end{aligned} \tag{63}
$$

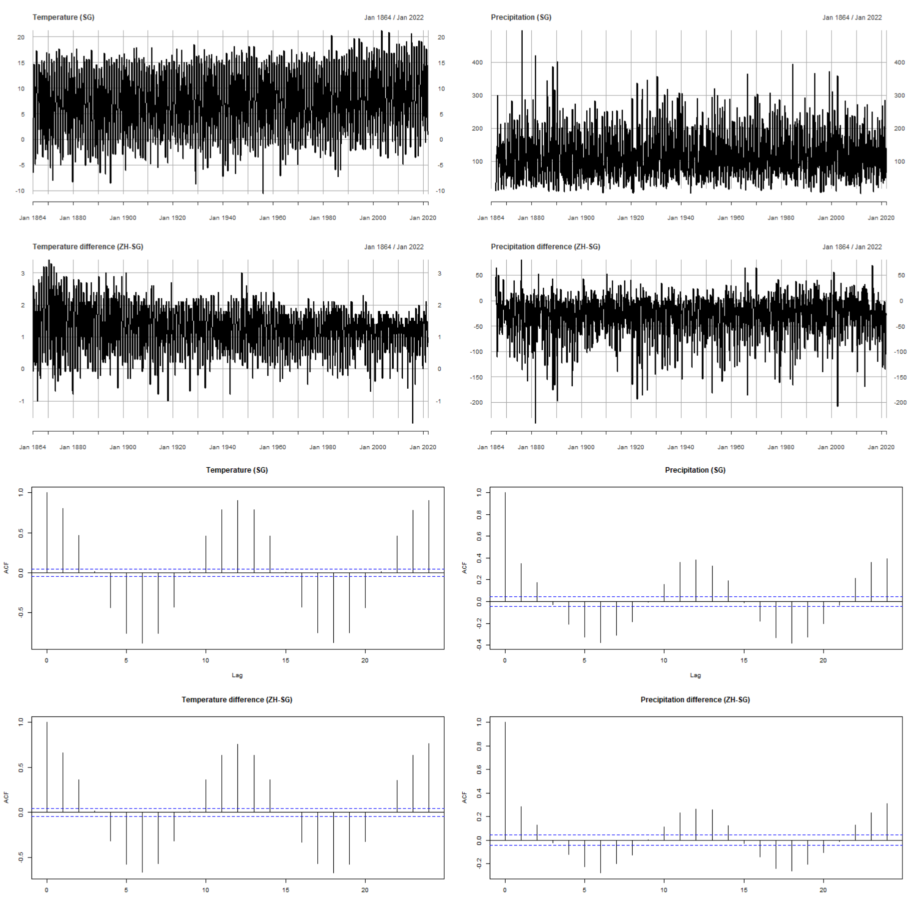
And the log-likelihood is then

$$
\begin{aligned}
log(\mathcal{L}(\eta)) &= log\left(p(x_1,...,x_T|x_0)p(x_0)\right) \\
&= -\frac{T}{2}log(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_{t=1}^{T}(x_t-(w+\phi x_{t-1}))^2 \\
&\quad -\frac{1}{2}log\left(2\pi\sigma^2/(1-\phi^2)\right) - \frac{1}{2\sigma^2/(1-\phi^2)}\left(x_0-\frac{w}{1-\phi}\right)^2
\end{aligned} \tag{64}
$$

4 Empirical exercise using R

(a) *See R code*

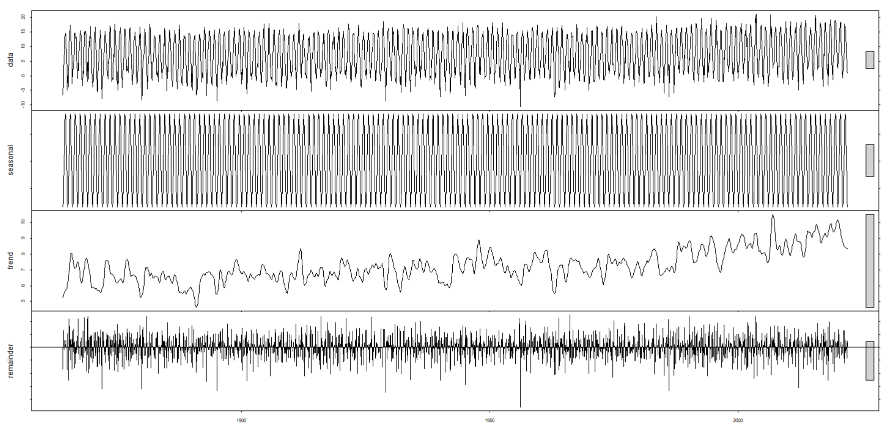
(b) The plots for the temperatures and precipitation data for SG and the respective differences are shown below. Further below, the ACF for the same time series are shown.



(c) No, the time series are clearly not stationary. Seasonality is evident for all 4 series in the ACF plots, and the periodicity corresponds to 12 months. The seasonality is less evident from the time series plots directly because of the large number of data points. Still, we can recognize peaks and throughs at what seems to be a regular frequency. There might also be a trend component, in particular for the SG temperature data, which appears to be trending higher over time and, as a result, possibly for the time series of the temperature difference as well (downward-trending). The trend is not that pronounced, however, and is not visible from the ACF plots. Finally, in particular the temperature difference time series appears to be more volatile in the early portion of the sample compared to the latter portion of the sample. This would also indicate the absence of stationarity and would be even more problematic for the analysis of that particular time series.

(d) The SG temperature time series can be decomposed in a trend component and a seasonal component.

(e) The result of the decomposition of the SG temperature time series using the STL method is shown below. The *s.window* parameter, which requires a value set by the user, was set to 'periodic' since we expect the kind of seasonal component in temperatures data to be periodic (i.e., identical across years).



What we clearly observe is (from top to bottom):

- The original data does not appear stationary, as was discussed above. Temperatures appear to be tending higher and there clearly seem to be some sort of seasonality.
- The seasonal component picked up by the STL decomposition is pretty clear and appears to have an annual frequency, as one would presumably expect in the case of monthly temperature data.
- The STL decomposition makes the trend component, which picks up in more recent years, very visible.
- The remainder series appears stationary. Neither a trend nor a seasonal component is visible anymore.

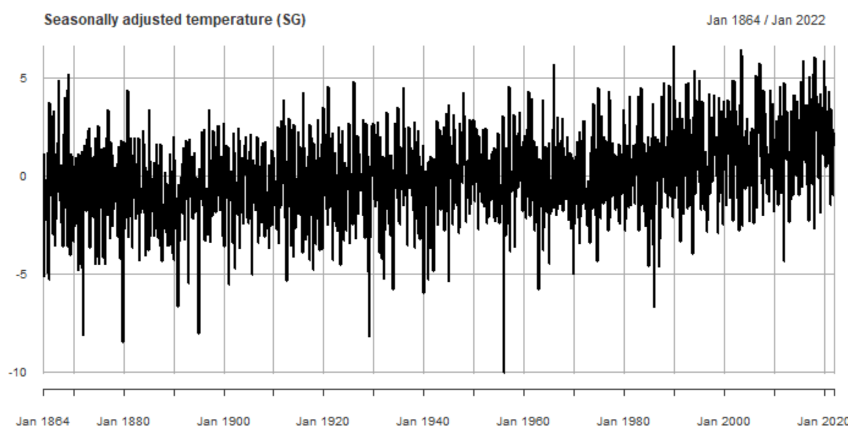
(f) Since the seasonality component in the SG temperature data appears to be deterministic, the alternative way we used to remove seasonality from the time series was based on trigonometric functions. We set the frequency S to 12, since it was apparent from the ACF plots of the time series that the periodicity of the seasonal component corresponded to 12 months. Hence, we estimated the component s_t using OLS with a regression of the form

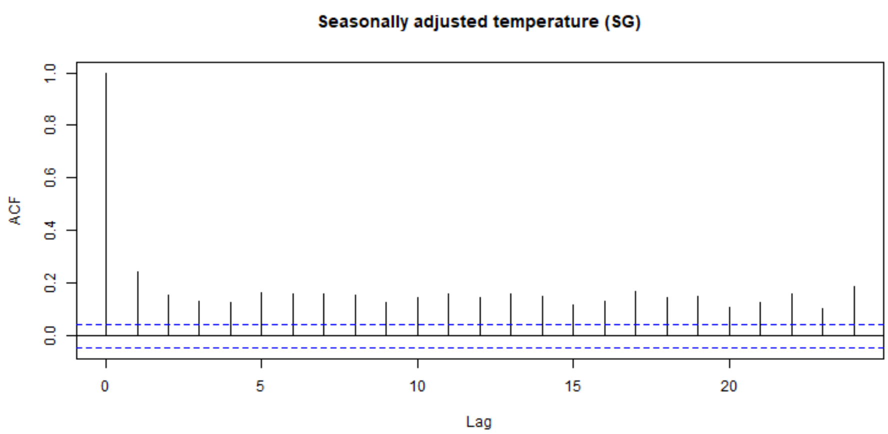
$$s_t = \beta_1 cos(\omega_1) + \gamma_1 sin(\omega_1 t) \tag{65}$$

where $\omega_1 = 2\pi/12$. We then removed the estimated seasonal component from the original time series X_t according to

$$\hat{u}_t = X_t - \hat{s}_t \tag{66}$$

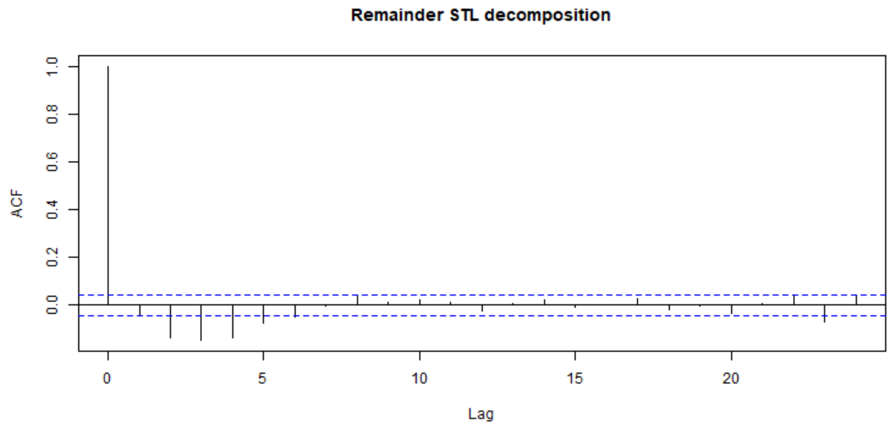
where $\hat{u}_t$ is the SG temperature data time series without the seasonal component.
Below, we plot $\hat{u}_t$ and its ACF.

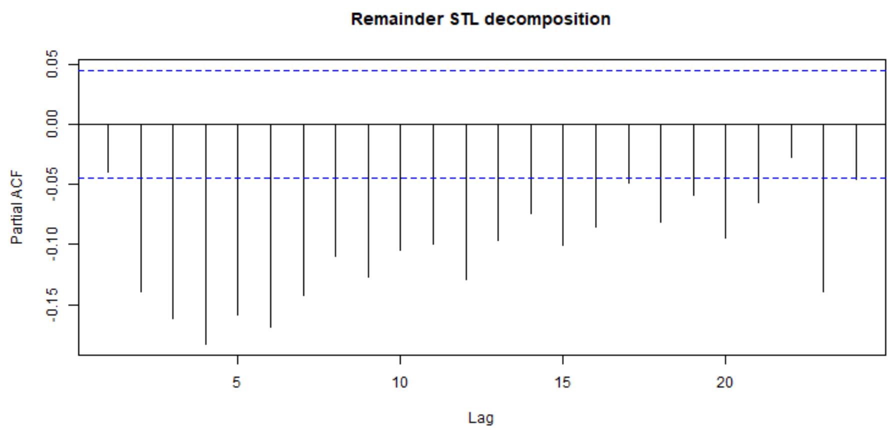




We observe that removing the seasonality made it even more clear that the original time series has a trend component as well, which is visible now in both plots. Remaining seasonality, on the other hand, is not recognizable.

(g) The charts below show the ACF and PACF for the remainder after the STL decomposition





Based on the ACF structure we observe an AR component, possibly an AR(3) or even AR(4). There is no sharp decrease in the ACF after a couple of lags thus we would exclude that the process is simply a MA process.
There is no sharp decrease in the PACF, hence we would choose either an MA model or a ARMA model. It must be said that the PACF, in particular, potentially shows us that the remainder that we obtained the STL decomposition might not be stationary.
Combining the information from the two charts, we would preferably fit an ARMA(p, q) model. We can indicate which ARMA model better based on the significant spikes from the PACF plot. It is observed that the PACF plot has significant spikes at lag 4 because of the significant PACF value.

On the other hand we can select the order q for model MA(q) from ACF if this plot has a sharp cut-off after lag q. This can be seen in lag of 3 or 4. Combining both it is reasonable to argue that we would preferably fit an ARMA(4,4).

(h) *See R code*

(i) The table below shows the results of applying our routine to the remainder time series of the STL decomposition of the SG temperature data. Our routine fits to the time series a selection of ARMA(p, q) models with lag combinations $p, q \in \{0, 1, 2\}$ and reports the AIC and BIC for each combination of p and q as well as returns the best model according to each information criteria. The combination of p and q that delivers the best model in terms of the AIC and BIC is shown in bold in the table.

p	q	AIC	BIC
0	0	7481.43	7492.53
1	0	7480.41	7497.06
2	0	7445.25	7467.44
0	1	7479.15	7495.79
1	1	7036.85	7059.04
2	1	6944.55	6972.29
0	2	7099.14	7121.33
1	2	6975.86	7003.60
2	**2**	**6788.09**	**6821.38**

In this case, it is the ARMA(2,2) model that minimizes both information criterias and, thus, both AIC and BIC agree that the model that best fits the STL-decomposed time series is an ARMA(2,2) model.
As a side note, we tried to check our results using R's *auto.arima* function but, for some reason, the function always ended up selecting an AR(2) model, even though the ARMA(2,2) model has lower AIC and BIC than the AR(2). We could not really make sense of why the *auto.arima* function did not pick the model that actually minimized the criteria. Indeed, the AR(2) model has an AIC of 7'443.25 and a BIC of 7'459.89
We then also fit an ARMA(4,4) to the data, which has a AIC of 6'778.94 and BIC of 6'834.42. According to both information criteria it should be preferred to the ARMA(2,2) model. Hence, we would prefer the ARMA(4,4) model over the previously selected one.