

Assessment 2

Joana Levtcheva, CID 01252821

Q1

•

$$2X_t = \epsilon_t - X_{t-1} \quad (1)$$

We can rewrite (1) as:

$$X_t = -\frac{1}{2}X_{t-1} + \epsilon_t \quad (2)$$

This is a process of the type $X_t = \delta + \phi X_{t-1} + \epsilon_t$, where ϕ is a constant coefficient and $\phi \neq 0$, and ϵ_t are i.i.d. random variables with mean zero and variance σ_ϵ^2 . Therefore, (2) is a $AR(1)$ process, with $\phi = -\frac{1}{2}$. The $AR(1)$ process is stationary if and only if $|\phi| < 1$, which is true for $\phi = -\frac{1}{2}$. Therefore, (1) is stationary.

•

$$X_t = \epsilon_t + \epsilon_{t-1} + \epsilon_{t-2} \quad (3)$$

This is a process of the type $Y_t = \epsilon_t + \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2}$, where θ_1, θ_2 are constants with $\theta_2 \neq 0$, and $\{\epsilon_t\}_t$ is a white noise process. Therefore (3) a $MA(2)$ process.

The process has mean $\mu = 0$. We are going to check if the process is stationary:

1. For the expectation we have

$$E(X_t) = E(\epsilon_t) + E(\epsilon_{t-1}) + E(\epsilon_{t-2}) = 0 = \mu,$$

which is finite and constant for all t .

2. The variance is

$$\begin{aligned} Var[X_t] &= E[(X_t - \mu)^2] = E[X_t^2] = E[(\epsilon_t + \epsilon_{t-1} + \epsilon_{t-2})^2] = \\ &= E[\epsilon_t^2 + \epsilon_{t-1}^2 + \epsilon_{t-2}^2 + 2\epsilon_t \epsilon_{t-1} + 2\epsilon_t \epsilon_{t-2} + 2\epsilon_{t-1} \epsilon_{t-2}] = \\ &= E[\epsilon_t^2] + E[\epsilon_{t-1}^2] + E[\epsilon_{t-2}^2] + 2E[\epsilon_t \epsilon_{t-1}] + 2E[\epsilon_t \epsilon_{t-2}] + 2E[\epsilon_{t-1} \epsilon_{t-2}] = \\ &= 3\sigma^2, \end{aligned}$$

where σ is the variance of the white noise process. Therefore, $Var[X_t]$ is finite for all t .

3. Finally,

$$\begin{aligned} C_{XX}(t_i, t_i + h) &= E[(X_{t_i} - \mu)(X_{t_i+h} - \mu)] = \\ &= E[(X_{t_i})(X_{t_i+h})] = E[(\epsilon_{t_i} + \epsilon_{t_i-1} + \epsilon_{t_i-2})(\epsilon_{t_i+h} + \epsilon_{t_i+h-1} + \epsilon_{t_i+h-2})] = 0, \end{aligned}$$

which holds for every $h > 0$ and every $t_i \in T$. For $h = 0$ we have

$$C_{XX}(t_i, t_i) = Var[X_{t_i}] = 3\sigma^2.$$

Therefore $C_{XX}(t_1, t_1 + h) = C_{XX}(t_2, t_2 + h)$ for all $t_1, t_2 \in T$.

From 1., 2., and 3. it follows that the process (3) is stationary.

•

$$X_t = \sin\left(\frac{2\pi t}{10}\right) + \epsilon_t \quad (4)$$

First, we are going to show that (4) is an $AR(2)$ process.

Let's observe the process

$$X_{t+1} = \sin(2\pi\phi(t+1)) + \epsilon_{t+1} \quad (5)$$

. We know that

$$\sin(a+b) + \sin(a-b) = 2\cos(b)\sin(a).$$

We have

$$\sin(2\pi\phi(t+1)) = \sin(2\pi\phi t + 2\pi\phi) = 2\cos(2\pi\phi)\sin(2\pi\phi t) - \sin(2\pi\phi t - 2\pi\phi).$$

For (4) we get

$$X_{t+1} = 2\cos\left(\frac{2\pi}{10}\right)\sin\left(\frac{2\pi t}{10}\right) + (-1)\sin\left(\frac{2\pi(t-1)}{10}\right) + \epsilon_{t+1}.$$

Or

$$X_{t+1} = 2\cos\left(\frac{2\pi}{10}\right)X_t + (-1)X_{t-1} + \epsilon_t.$$

This is the same as

$$X_t = 2\cos\left(\frac{2\pi}{10}\right)X_{t-1} + (-1)X_{t-2} + \epsilon_t.$$

This is an $AR(2)$ process with coefficients $\phi_1 = 2\cos(\frac{2\pi}{10}) \neq 0$ and $\phi_2 = -1$.

An $AR(2)$ process is stationary provided all the roots of the following polynomial equation have an absolute value greater than 1, they are outside of the unit circle:

$$1 - \phi_1 z - \phi_2 z^2 = 0.$$

We have

$$1 - 2\cos\left(\frac{2\pi}{10}\right)z + z^2 = 0$$

with solutions $z_{1,2} = \frac{1+\sqrt{5}}{4} \pm i\frac{\sqrt{10-2\sqrt{5}}}{4}$ having an absolute value of 1. Therefore, (4) is not stationary.

Q2

Part a The Kwiatkowski-Phillips-Schmidt-Shin (KPSS) test is a test of stationarity. The KPSS test assumes that the time series of interest can be decomposed to a deterministic trend, random walk and stationary process, i.e. $\{X_t\}_{t>0}$ with

$$\begin{aligned} X_t &= \delta_t + r_t + u_t, \\ r_t &= r_{t-1} + w_t, \end{aligned}$$

where δ_t denotes the deterministic trend, r_t denotes the random walk, and u_t denotes the stationary process, and w_t is an i.i.d. process with mean 0, and variance σ^2 .

The KPSS test can test for trend-stationarity against the null hypothesis,

$$H_0 : \text{The time series is trend-stationary, or } \sigma^2 = 0.$$

and with an alternative hypothesis

$$H_1 : \text{The time series contains a unit root, or } \sigma^2 > 0.$$

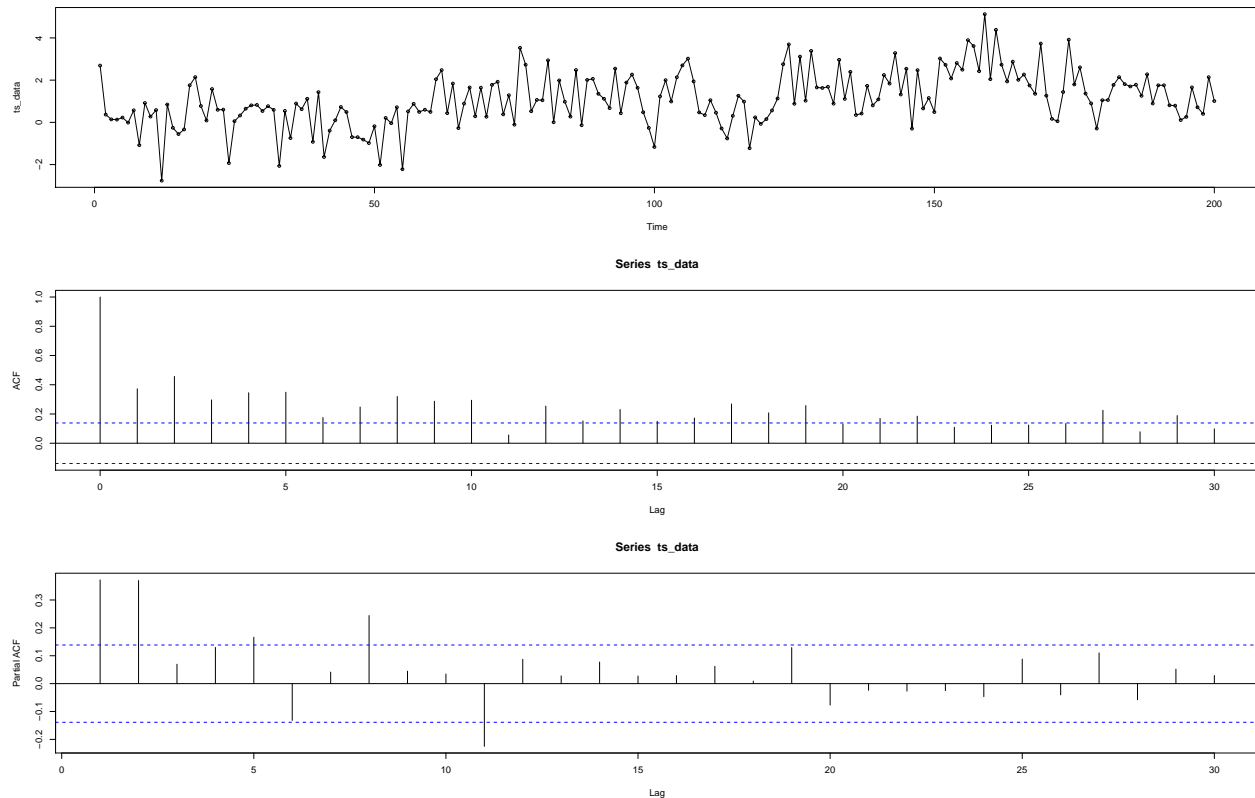
Note: The KPSS test can also test for level-stationarity.

Part b Loading data, plotting data and ACF:

```
par(mfrow = c(3, 1))
ts_data <- read.table("time-series.csv", header = T, sep = ",")
ts_data <- as.ts(unlist(ts_data[, -1]))
str(ts_data)
```

```
## Time-Series [1:200] from 1 to 200: 2.69 0.368 0.141 0.129 0.227 ...
```

```
plot(ts_data, type='o')
acf(ts_data, lag.max = 30)
pacf(ts_data, lag.max = 30)
```



Based on the ACF the process seems to have both (slight) trend and seasonality.

Performing ADF test:

```
adf.test(ts_data)

##
## Augmented Dickey-Fuller Test
##
## data: ts_data
## Dickey-Fuller = -4.4946, Lag order = 5, p-value = 0.01
## alternative hypothesis: stationary
```

The p-value is 0.01 with the default lag order $k = 5$, and at the 5 significance level we have that it is smaller than 0.05. Therefore, there is evidence against non-stationarity.

The k parameter is a set of lags added to address serial correlation and it can vary depending on the order p . The value of k should be chosen such that $p \leq k - 1$. The default lag order in `adf.test` is calculated by $\text{trunc}((\text{length}(x) - 1)^{\frac{1}{3}})$. In our case the default lag order is $k = 5$, which leads p to ideally be $p \leq 4$.

Therefore, if the true value of p is less than 5 the test would perform as expected and would reject the null hypothesis as well, whereas if it is significantly bigger the test would return the opposite result.

```
adf.test(ts_data, k = 2)
```

```
##  
## Augmented Dickey-Fuller Test  
##  
## data: ts_data  
## Dickey-Fuller = -5.7439, Lag order = 2, p-value = 0.01  
## alternative hypothesis: stationary
```

```
adf.test(ts_data, k = 3)
```

```
##  
## Augmented Dickey-Fuller Test  
##  
## data: ts_data  
## Dickey-Fuller = -4.8482, Lag order = 3, p-value = 0.01  
## alternative hypothesis: stationary
```

```
adf.test(ts_data, k = 7)
```

```
##  
## Augmented Dickey-Fuller Test  
##  
## data: ts_data  
## Dickey-Fuller = -3.0712, Lag order = 7, p-value = 0.1275  
## alternative hypothesis: stationary
```

```
adf.test(ts_data, k = 15)
```

```
##  
## Augmented Dickey-Fuller Test  
##  
## data: ts_data  
## Dickey-Fuller = -2.4554, Lag order = 15, p-value = 0.3854  
## alternative hypothesis: stationary
```

```
adf.test(ts_data, k = 30)
```

```
##  
## Augmented Dickey-Fuller Test  
##  
## data: ts_data  
## Dickey-Fuller = -1.8567, Lag order = 30, p-value = 0.6361  
## alternative hypothesis: stationary
```

The results are different depending on whether the lag order is bigger or smaller than the default lag order. Depending on the true value of p they can be right or wrong. For example, if the true value of p is less than or equal to the default value of k the correct result would be of rejecting the null hypothesis, but if the true value of p turns out to be somewhat bigger than the default value of k the correct result would be the opposite.

Part c Performing KPSS test for trend-stationarity:

```
kpss.test(ts_data, null = "Trend")
```

```
##  
## KPSS Test for Trend Stationarity
```

```
##  
## data:  ts_data  
## KPSS Trend = 0.13481, Truncation lag parameter = 4, p-value = 0.07073
```

The p-value is 0.07073 with the default lag order $k = 4$. Since we are evaluating at the 5% significance level and this value is not less than 0.05, we fail to reject the null hypothesis of the KPSS test. This means we can assume that the time series is trend stationary.

Q3

Wickham, Hadley., “A Layered Grammar of Graphics” (2010), Journal of Computational and Graphical Statistics, vol. 19, no. 1, pp. 3–28

Keywords: grammar of graphics, ggplot2, statistical graphics, R

In “A Layered Grammar of Graphics” the author presents a graphical grammar which allows us to move beyond the famous named graphics (such as scatterplot) and enables us to gain insight into the composition of complicated statistical graphics and their underlying deep structure, making it possible to describe their components; the grammar also reveals “unexpected connections between seemingly different graphics”.

The author is comparing and contrasting the layered grammar to Wilkinson’s grammar (“The Grammar of Graphics”) and walks us through introducing the grammar (showing how to build a basic plot, and a more complicated plot), defining the components of the layered grammar (such as layers (data and mapping, statistical transformation, geometric object, position adjustment), scales coordinate system, faceting), describing a hierarchy of defaults, mentioning embedding the grammar in the programming language R, and showing some implications of the layered grammar (histograms, polar coordinates, transformations), and finally Wickham discusses “some perceptual issues and thinking about how we can build on the grammar to learn how to create graphical ‘poems’”.

In an attempt to create a grammar which can be used to build higher level tools for data analysis, Wickham proposes an alternative parametrization of Wilkinson’s grammar “based around the idea of building up a graphic from multiple layers of data” by creating a strong foundation for understanding a diverse range of graphics and defining what a well-formed or correct graphics looks like.

The author adopts a formal and technical tone, relatively easy to be understood by his audience, the readers of “A Layered Grammar of Graphics” (2010) published in Journal of Computational and Graphical Statistics, vol. 19, others interested in the topic of building a graphical grammar, the users of the ggplot2 library, and people interested in further building on the grammar Wickham proposed.