Prediction and Time Series Final Exam

Exercise 1

1. True or False (4 p.)

   Determine whether the statement is true or false. You do not have to justify your answers. Simply state wether the statement is true or false. (Every correct answer +1 p., every wrong answer -1 p., no answer 0 p.)

   (a) In the context of linear regression, the traditional least-squares estimators are sensitive to outlying observations.

   (b) In the context of linear regression, the variance inflation factors (VIF) are calculated in order to detect multicollinearity.

   (c) A moving average process of order > 1 is never stationary.

   (d) In simple exponential smoothing, the value of $\hat{x}_{t+1|t}$ is predicted using a weighted sum of the previous observation $x_t, x_{t-1}, x_{t-2}, \dots$

Nguyen Xuan Binh 887799

Exercise 1: True or False ?

a) True

b) True

c) False

d) True

Exercise 2

2. Linear regression (4 p.)

   Consider the linear regression model

$$y_i = \beta_0 + \beta_1 x_{i1} + \epsilon_i.$$

   You have a sample of 862 iid observations and you estimate the parameters using the traditional least squares estimator. The estimated values are $\hat{\beta}_0 = 3.5$ and $\hat{\beta}_1 = -12.3$.

   (a) Explain, step by step, how to construct a 95% bootstrap confidence interval for the parameter β_1. How many bootstrap samples do you take? What are the sample sizes of your bootstrap samples? How do you use the bootstrap samples in constructing the confidence interval? (3 p.)

   (b) Your estimated 95% bootstrap confidence interval for β_1 is $(-17.3, -7.2)$. Based on that, does it seem that the parameter β_1 is significant? (1 p.)

Nguyen Xuan Binh 887799 Binh

a) A 95 % bootstrap confidence interval for the parameter β_1 can be constructed as we follow the steps below

Step 1: Choose 862 data points randomly (replacement is allowed) from the original data points (y_i, x_{i1}), $i = 1, 2, \dots, 862$. Each observation (y_i, x_{i1}) can be chosen once, many times or not chosen.

Step 2: Calculate a new estimate parameter β_1 from the new constructed bootstrap sample in step 1

Step 3: Repeat the steps 1-2 for 999 times

Step 4: Order the obtained parameter estimates β_1, including the original parameter, from smallest to largest. The total number of parameter β_1 should be 1000

Step 5: Set the lower end of the bootstrap confidence interval to be smaller than or equal to the 25^{th} ordered estimate β_1 in the ordered list in step 4 and set the upper end of the bootstrap confidence interval to be larger than or equal to the 975^{th} ordered estimate in the ordered list in step 4

=> A 95 % bootstrap confidence interval for the parameter β_1 is in the range between the 25^{th} and 975^{th} values of β_1 from the bootstrapping method, because $(975 - 25)/1000 \times 100\% = 95\%$. This is how the bootstrap samples are used in constructing the confidence interval for parameter β_1

* How many bootstrap samples do you take: 999 bootstrap samples
* Sample sizes of your bootstrap samples: each sample has exactly 862 observations

b) Based on the information, it seems that the parameter β_1 is significant

H_0 hypothesis: β_1 parameter is significant

If values near 0 (including 0) are included in the 95% interval, there are chances that value β_1 can be null. In other words, β_1 is insignificant. However, the obtained 95% interval $(-17.3, -7.2)$ does not contain 0 nor values near 0. Moreover the original $\beta_1 = -12.3$ is also included in the interval

=> Null hypothesis is accepted: parameter β_1 is significant and should not be omitted from the linear regression model

Exercise 3

3. Stationarity (4 p.)

   Let x_t and z_t be weakly stationary stochastic processes such that, for all $t, s \in \mathbb{Z}$, we have that

$$E[x_t] = 2,\ E[z_t] = 1,\ E[x_t^2] = 6,\ E[z_t^2] = 4,$$

$$E[(x_t - E[x_t])(x_s - E[x_s])] = \frac{2}{1+|t-s|},$$

$$E[(z_t - E[z_t])(z_s - E[z_s])] = \frac{3}{1+(t-s)^2},$$

   and

$$E[(x_t - E[x_t])(z_s - E[z_s])] = 0.$$

   Let $y_t = x_t + z_t$. Show that the process y_t is weakly stationary.

Nguyen Xuan Binh 887799

- $E[x_t] = 2,\ E[z_t] = 1,\ E[x_t^2] = 6,\ E[z_t^2] = 4$
- $E[(x_t - E[x_t])(x_s - E[x_s])] = Cov(x_t, x_s) = \frac{2}{1+|t-s|} \quad \forall t,s \in Z$
- $E[(z_t - E[z_t])(z_s - E[z_s])] = Cov(z_t, z_s) = \frac{3}{1+(t-s)^2} \quad \forall t,s \in Z$
- $E[(x_t - E[x_t])(z_s - E[z_s])] = 0 \quad \forall t,s \in Z$

x_t and z_t are weakly stationary. Prove that $y_t = x_t + z_t$ is stationary

* Expected value of y_t

$E[y_t] = E[x_t + z_t] = E[x_t] + E[z_t] = 2 + 1 = 3$ independent of t

* Variance of y_t

$$Var(y_t) = E[y_t^2] - E[y_t]^2 = E[y_t]^2 - 9$$
$$= E[(x_t + z_t)^2] - 9 = E[x_t^2 + 2x_t z_t + z_t^2] - 9$$
$$= 2E[x_t z_t] + 6 + 4 - 9 = 2E[x_t z_t] + 1$$

We have $E[(x_t - E[x_t])(z_s - E[z_s])] = 0$

$\Rightarrow E[(x_t - 2)(z_s - 1)] = 0 \Rightarrow E[x_t z_s - x_t - 2z_s + 2] = 0$

$\Rightarrow E[x_t z_s] - 2 - 2\times 1 + 2 = 0 \Rightarrow E[x_t z_s] = 2 \quad \forall t,s \in Z$

$\Rightarrow Var(y_t) = 2\times 2 + 1 = 5 < \infty \quad \forall t \in Z$

* Covariance of y_t

- $E[(x_t - 2)(x_s - 2)] = E[(x_t x_s - 2x_t - 2x_s + 4)]$

$= E[x_t x_s] - 2\times 2 - 2\times 2 + 4 = E[x_t x_s] - 4 = \frac{2}{1+|t-s|}$

$\Rightarrow E[x_t x_s] = \frac{2}{1+|t-s|} + 4 \quad \forall t,s \in Z$

- $E[(z_t - 1)(z_s - 1)] = E[z_t z_s - z_t - z_s + 1]$

$= E[z_t z_s] - 1 - 1 + 1 = E[z_t z_s] - 1 = \frac{3}{1+(t-s)^2}$

$\Rightarrow E[z_t z_s] = \frac{3}{1+(t-s)^2} + 1 \quad \forall t,s \in Z$

Nguyen Xuan Binh 887799

$$\text{Cov}(y_t, y_{t-\tau}) = E[(y_t - E[y_t])(y_{t-\tau} - E[y_{t-\tau}])]$$
$$= E[(y_t - 3)(y_{t-\tau} - 3)]$$
$$= E[y_t y_{t-\tau} - 3y_t - 3y_{t-\tau} + 9]$$
$$= E[y_t y_{t-\tau}] - 3\times 3 - 3\times 3 + 9$$
$$= E[y_t y_{t-\tau}] - 9$$

We have: $E[y_t y_{t-\tau}] = E[(x_t + z_t)(x_{t-\tau} + z_{t-\tau})]$

$$= E[x_t x_{t-\tau} + x_t z_{t-\tau} + z_t x_{t-\tau} + z_t z_{t-\tau}]$$
$$= \left(\frac{2}{1+|t-(t-\tau)|} + 4\right) + 2 + \left(\frac{3}{1+(t-(t-\tau))^2} + 1\right) + 2$$
$$= \frac{2}{1+|\tau|} + \frac{3}{1+\tau^2} + 9$$

$$\Rightarrow \text{Cov}(y_t, y_{t-\tau}) = \frac{2}{1+|\tau|} + \frac{3}{1+\tau^2} + 9 - 9 = \frac{2}{1+|\tau|} + \frac{3}{1+\tau^2} = \gamma_\tau \ \forall t, \tau \in \mathbb{Z}$$

Since $E[y_t] = 3$, $\text{Var}(y_t) = 5 < \infty$ and $\text{Cov}(y_t, y_{t-\tau}) = \gamma_\tau \ \forall t, \tau \in \mathbb{Z}$

$\Rightarrow$ Process y_t is weakly stationary

Exercise 4

4. ARMA modeling (6 p.)

   Assume that you observe a series

$$x_0, x_1, x_2, ..., x_{7300}, x_{7301}, x_{7302}, x_{7303}, x_{7304}, x_{7305},$$

   where $x_{7300} = 201.3$, $x_{7301} = 219.8$, $x_{7302} = 241.4$, $x_{7303} = 262.7$, $x_{7304} = 281.5$ and $x_{7305} = 300.8$.

   (a) Based on plotting the series, you observe a linear trend. You manage to stationarize the series by taking a difference. The obtained differenced series is $z_1, z_2, ..., z_{7305}$. Give the values of the elements $z_{7301}, z_{7302}, z_{7303}, z_{7304}$ and z_{7305} of the stationarized series. (1 p.)

   (b) After stationarization, you center the series by subtracting the sample mean $\bar{z} = 20.1$ from the observations $z_1, z_2, ..., z_{7305}$. The obtained (stationarized and) centered series is $y_1, y_2, ..., y_{7305}$. Give the values of the elements $y_{7301}, y_{7302}, y_{7303}, y_{7304}$ and y_{7305}. (1 p.)

   (c) Based on plotting the stationarized and centered series and its estimated autocorrelation and partial autocorrelation -functions, you think that $y_1, y_2, ..., y_{7305}$ is an autoregressive process of order 2. You estimate the parameters of the process and the estimated values are $\phi_1 = 0.5$ and $\phi_2 = -0.2$. Give predictions for y_{7306}, y_{7307} and y_{7308}. (2 p.)

   (d) Using the predictions for y_{7306}, y_{7307} and y_{7308}, give predicted values for x_{7306}, x_{7307} and x_{7308}. (2 p.)

Nguyen Xuan Binh 887799

Time series: $x_0, x_1, \dots x_{7304}, x_{7305}$

a) The time series is stationarized by taking a difference

$\Rightarrow z_t = Dx_t = x_t - x_{t-1}$

Values of the elements z_{7301} to z_{7305} of the stationarized series are

- $z_{7301} = x_{7301} - x_{7300} = 219.8 - 201.3 = 18.5$
- $z_{7302} = x_{7302} - x_{7301} = 241.4 - 219.8 = 21.6$
- $z_{7303} = x_{7303} - x_{7302} = 262.7 - 241.4 = 21.3$
- $z_{7304} = x_{7304} - x_{7303} = 281.5 - 262.7 = 18.8$
- $z_{7305} = x_{7305} - x_{7304} = 300.8 - 281.5 = 19.3$

b) The series is centered by subtracting the sample mean $\bar{z} = 20.1$ from z_1 to z_{7305} and obtained center series is y_1 to y_{7305}

$\Rightarrow y_t = z_t - \bar{z} = z_t - 20.1$. Values of y_{7301} to y_{7305} are

- $y_{7301} = z_{7301} - 20.1 = 18.5 - 20.1 = -1.6$
- $y_{7302} = z_{7302} - 20.1 = 21.6 - 20.1 = 1.5$
- $y_{7303} = z_{7303} - 20.1 = 21.3 - 20.1 = 1.2$
- $y_{7304} = z_{7304} - 20.1 = 18.8 - 20.1 = -1.3$
- $y_{7305} = z_{7305} - 20.1 = 19.3 - 20.1 = -0.8$

c) Series y_t is AR(2), where $\phi_1 = 0.5$ and $\phi_2 = -0.2$

$\Rightarrow y_t = 0.5y_{t-1} - 0.2y_{t-2} + \epsilon_t$, $\epsilon_t \in$ i.i.d WN, $E[\epsilon_t] = 0$

Give predictions for $y_{7306}, y_{7307}, y_{7308}$

- One step prediction: $\hat{y}_{7306} = E(y_{7306} \mid y_{7305}, \dots)$
  $= E(\phi_1 y_{7305} + \phi_2 y_{7304} + \epsilon_{7306} \mid y_{7305}, \dots)$
  $\Rightarrow \hat{y}_{7306} = \phi_1 y_{7305} + \phi_2 y_{7304} = 0.5 \times (-0.8) - 0.2 \times (-1.3) = -0.14$ (answer)
- Two step prediction: $\hat{y}_{7307} = E(y_{7307} \mid y_{7306}, \dots)$
  $= E(\phi_1 y_{7306} + \phi_2 y_{7305} + \varepsilon_{7307} \mid y_{7306}, \dots)$
  $\Rightarrow \hat{y}_{7307} = \phi_1 y_{7306} + \phi_2 y_{7305} = 0.5 \times (-0.14) - 0.2 \times (-0.8) = 0.09$ (answer)
- Three step prediction: $\hat{y}_{7308} = E(y_{7308} \mid y_{7307}, \dots)$
  $= E(\phi_1 y_{7307} + \phi_2 y_{7306} + \varepsilon_{7308} \mid y_{7307}, \dots)$
  $\Rightarrow \hat{y}_{7308} = \phi_1 y_{7307} + \phi_2 y_{7306} = 0.5 \times 0.09 - 0.2 \times (-0.14) = 0.073$ (answer)

d) Use predictions for y_{7306} to y_{7308} to predict x_{7306} to x_{7308}

$$\left.\begin{aligned} z_t &= Dx_t = x_t - x_{t-1} \\ y_t &= z_t - 20.1 \end{aligned}\right\} \Rightarrow \begin{aligned} y_t &= x_t - x_{t-1} - 20.1 \\ \Rightarrow x_t &= y_t + x_{t-1} + 20.1 \end{aligned}$$

Predicted values x_{7306} to x_{7308} are:

- $x_{7306} = y_{7306} + x_{7305} + 20.1 = -0.14 + 300.8 + 20.1 = 320.76$
- $x_{7307} = y_{7307} + x_{7306} + 20.1 = 0.09 + 320.76 + 20.1 = 340.95$
- $x_{7308} = y_{7308} + x_{7307} + 20.1 = 0.073 + 340.95 + 20.1 = 361.123$

Exercise 5

5. Autocorrelations (6 p.)

Figures 1 and 2 display the theoretical autocorrelation and partial autocorrelation -functions of six different processes. Answer to the following questions. You do not have to justify your answers. (Every correct answer +1 p., every wrong answer 0 p., no answer 0 p.)

(a) Which one of the processes (Series 1, 2, 3, 4, 5 or 6) is a MA(1)-process?

(b) Which one of the processes (Series 1, 2, 3, 4, 5 or 6) is an MA(2)-process?

(c) Which one of the processes (Series 1, 2, 3, 4, 5 or 6) is an AR(3)-process?

(d) Which one of the processes (Series 1, 2, 3, 4, 5 or 6) is an ARMA(2,2)-process?

(e) Which one of the processes (Series 1, 2, 3, 4, 5 or 6) is a $SMA(3)_3$-process?

(f) Which one of the processes (Series 1, 2, 3, 4, 5 or 6) is a $SAR(2)_3$-process?

Nguyen Xuan Binh 887799

Exercise 5

a) Series 2

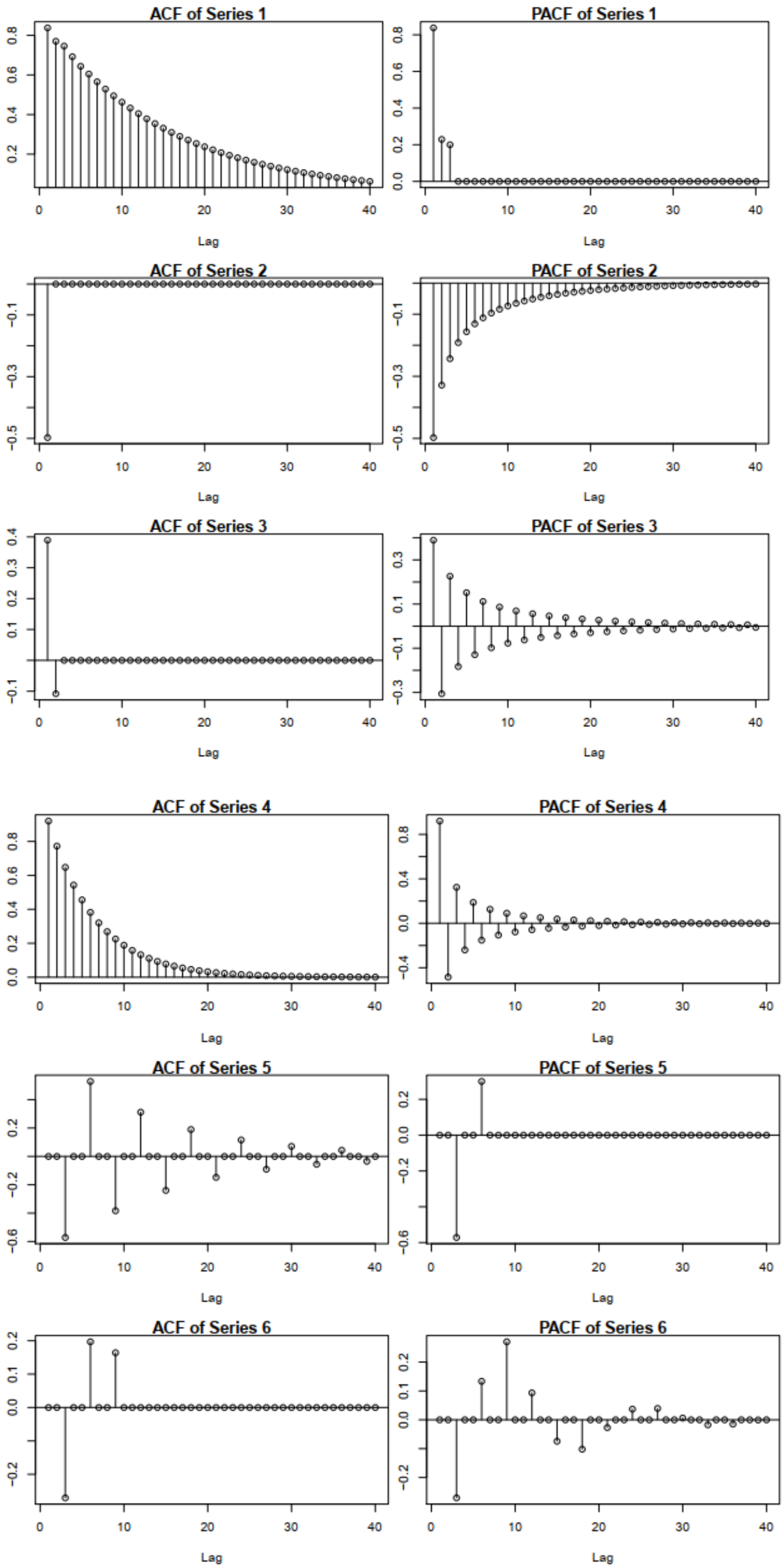
b) Series 3

c) Series 1

d) Series 4

e) Series 6

f) Series 5


ACF of Series 1
PACF of Series 1
ACF of Series 2
PACF of Series 2
ACF of Series 3
PACF of Series 3
ACF of Series 4
PACF of Series 4
ACF of Series 5
PACF of Series 5
ACF of Series 6
PACF of Series 6
Lag