2020 Test

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, conducting homoscedasticity testing is possible only if the residuals are normally distributed.

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

   (d) ARIMAX modeling is an excellent choice for long term forecasting.

a) True
b) False. Homoscedasticity testing is possible no matter what distribution the residuals are
c) True
d) False. ARIMAX models are reasonably general, but they are not suitable for long term forecasting. Moreover, ARIMAX models are linear on the explanatory variables

2. Linear regression (3 p.)

   Consider the linear regression model

$$y_i = \beta_0 + \beta_1 x_i + \epsilon_i.$$

   You have a sample and you estimate the parameters β_0 and β_1 using traditional least squares estimators. In your sample, you have two separate subgroups that have sample sizes h and $n - h$, and you are worried about possible parameter instability.

   (a) Explain, step by step, how to apply a permutation test in testing parameter instability. (2 p.)

   (b) Give the null hypothesis of the test. (1/2 p.)

   (c) How would you proceed your modeling (what would you do), if the estimated p-value of the test statistic was very small? (1/2 p.)

(a) Steps of permutation test in testing parameter instability

Permutation tests rely on resampling the original data assuming the null hypothesis. Based on the resampled data it can be concluded how likely the original data is to occur under the null hypothesis. We will use the Chow test to know if there is a break in parameter estimators between the group. Chow test relies on permutation testing.

Step 1: Estimate the parameters of the linear model using the entire data and calculate the corresponding SSE.

Step 2: Estimate the parameters of the linear model separately for the subgroup 1 and for the subgroup 2, and calculate the corresponding SSE1 and SSE2.
Step 3: Calculate the statistic
Ch = (n − 2(k + 1))/(k + 1) * (SSE − (SSE1 + SSE2))/ (SSE1 + SSE2) ;
where k is the number of the explanatory variables. In the model, k = 1
=> Ch = (n − 1)/2 * (SSE − (SSE1 + SSE2))/ (SSE1 + SSE2) ;
Step 4: Divide the original entire sample randomly into two separate subgroups p1 and p2 that have sample sizes h and n − h.
Step 5: Estimate the parameters of the linear model separately for the subgroup p1 and for the subgroup p2, and calculate the corresponding SSEp1 and SSEp2.
Step 6 Calculate the value
Chp = (n − 1)/2 * (SSE − (SSEp1 + SSEp2))/(SSEp1 + SSEp2) :
Step 7: Repeat the steps 4, 5 and 6 10000 to 20000 times.
Step 8: Order the values Chp from the smallest to the largest and calculate the empirical (1 − α) · 100th percentile from the ordered sample. If the original statistic Ch is larger than the calculated percentile, then the null hypothesis is rejected (on significance level α).

(b) Null hypothesis of the test: the model parameters β_1 and β_2 are the same for both subgroups size n and n - h

(c) A result has statistical significance when it is very unlikely to have occurred given the null hypothesis. Since the p-value of the test statistic was very small, we are confident that the null hypothesis in (b) is wrong. In other words, model parameters β_1 and β_2 are not the same for the two different subgroups size n and n - h.

3. Interval bootstrapping/block bootstrapping (4 p.)

Assume that you have observed a time series $x_1, x_2, x_3, ..., x_{10263}$. Based on plotting the series and its estimated autocorrelation and partial autocorrelation -functions, you think that the observed series is an MA(1) process

$$x_t = \theta_1 \epsilon_{t-1} + \epsilon_t, \quad (\epsilon_t)_{t \in T} \sim WN(0, \sigma^2)$$

and you have estimated the parameter θ_1. Your estimate is 0.56.

(a) Explain, step by step, how to construct a 95% bootstrap confidence interval for the parameter θ_1. (3 p.)

(b) Explain how to use the confidence interval in testing (approximately) the significance of the parameter θ_1? (1 p.)

a) A 95% bootstrap confidence interval for the regression parameter θ_1 can be obtained as follows.
Step 1: Select exactly 10263 data points randomly with replacement from the original observations x_i, i = 1;...; 10263. Each data point x_i can be selected once, multiple times, or not at all.
Step 2: Calculate a new MA(1) process parameter estimate θ_1 from the new sample formed in step 1
Step 3: Repeat the previous steps (m-1) times. If m = 1000 then redo step 1-2 999 times

Step 4: Order all the obtained estimates θ_1 from the smallest to the largest. Include also the original estimate, which amounts to 1000 different numbers of θ_1

Step 5: Set the lower end of the bootstrap confidence interval to be smaller than or equal to the 25th ordered estimate and set the upper end of the bootstrap confidence interval to be larger than or equal to the 975th ordered estimate.

=> A 95% bootstrap confidence interval for the regression parameter θ_1 falls in the range of 25th and 975th θ_1 from the bootstrapping method

b) If we collect more data samples, for example, 1000 data sets, which are independent and identically distributed real-life observations, then we calculate the intervals for θ_1 from them. There is approximately 950 of the intervals obtained that will contain the true parameter θ_1.
Null hypothesis: $\theta_1 = 0.56$ is a good estimator for the MA(1) model. If $\theta_1 = 0.56$ doesn't lie in the interval found in b, the null hypothesis is rejected and we have to recalculate θ_1 or even consider a different time series model. If $\theta_1 = 0.56$ lies in the interval then the the null hypothesis is accepted: $\theta_1 = 0.56$ is a good estimator

4. Stationarity (3 p.)

Let x_t be a stochastic process such that, for all $t, \tau \in \mathbb{Z}$, we have that $E[x_t] = 0$, $E[x_t^2] = \sigma^2 < \infty$, $E[x_t x_{t-\tau}] = \gamma_\tau < \infty$, $E[x_t^4] = \kappa < \infty$, and $E[x_t^2 x_{t-\tau}^2] = \lambda_\tau < \infty$. Show that the process $y_t = x_t^2$ is weakly stationary.

Exercise 4: Show that $y_t = x_t^2$ is weakly stationary

x_t, for all $t, \tau \in \mathbb{Z}$, we have $E[x_t] = 0$, $E[x_t^2] = \sigma^2 < \infty$, $E[x_t x_{t-\tau}] = \gamma_\tau < \infty$, $E[x_t^4] = k < \infty$, $E[x_t^2 x_{t-\tau}^2] = \lambda_\tau < \infty$

$E[x_t^2] = \cancel{V[x_t]}\ \sigma^2$ independent of t

$V[x_t^2] = E(x_t^4) - E(x_t^2)^2 = k - \sigma^4 < \infty$

$Cov(x_t^2 x_{t-\tau}^2) = E[(x_t^2 - E[x_t^2])(x_{t-\tau}^2 - E[x_{t-\tau}^2])]$

$= E[(x_t^2 - \sigma^2)(x_{t-\tau}^2 - \sigma^2)]$

$= E[(x_t^2 x_{t-\tau}^2) - \sigma^2 x_t^2 - \sigma^2 x_{t-\tau}^2 + \sigma^4]$

$= \lambda_\tau - \sigma^4 - \sigma^4 + \sigma^4 = \lambda_\tau - \sigma^4$ independent of t

$\Rightarrow y_t = x_t^2$ is weakly stationary

$E(x_t) = \mu,\ \forall t \in T$

$Var(x_t) = \sigma^2 < \infty, \forall t \in T$

$Cov(x_t, x_{t-\tau}) = \gamma_\tau\ \forall t, \tau \in T$

for weakly stationary process ($cov(x_t, x_{t-\tau})$) is independent of t, depends only on τ

5. ARMA modeling (4 p.)

Assume that you observe a series $x_0, x_1, x_2, ..., x_{7305}$.

(a) Based on plotting the series, you observe a linear trend. You manage to stationarize the process by taking a difference. The obtained differenced series is $z_1, z_2, ..., z_{7305}$. Give the elements (z_t) of the stationarized process in terms of the elements of the original observed series. (1 p.)

(b) Based on plotting the stationarized series and its estimated autocorrelation and partial autocorrelation -functions, you think that the stationarized series $z_1, z_2, ..., z_{7305}$ is an autoregressive process of order 2. You estimate the parameters of the process and the estimated values are ϕ_1 and ϕ_2. Give predictions for z_{7306}, z_{7307} and z_{7308}. (2 p.)

(c) Using the predictions for z_{7306}, z_{7307} and z_{7308}, give predicted values for x_{7306}, x_{7307} and x_{7308}. (1 p.)

Exercise 5:

a) Since differenced series z is obtain by one time differencing of x_t

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

b) Time series $z_1, z_2, ..., z_{7305}$ is AR(2) with parameters ϕ_1 and ϕ_2

The 1-step prediction

$$\hat{z}_{7306|7305} = E(z_{7306} | z_{7305}, z_{7304}, ...)$$
$$= E(\phi_1 z_{7305} + \phi_2 z_{7304} + \varepsilon_{7306} | z_{7305}, z_{7304}, ...)$$
$$= \phi_1 E(z_{7305} | z_{7305}, ...) + \phi_2 E(z_{7304} | z_{7305} ...) + E(\varepsilon_{7306} | ...)$$
$$= \phi_1 z_{7305} + \phi_2 z_{7304} = z_{7306} \text{ (Prediction)}$$

The 2-step prediction

$$\hat{z}_{7307|7306} = E(z_{7307} | z_{7306}, z_{7305}, ...)$$
$$= E(\phi_1 z_{7306} + \phi_2 z_{7305} + \varepsilon_{7307} | z_{7306} ...)$$
$$= \phi_1 \hat{z}_{7306|7305} + \phi_2 z_{7305} = \phi_1(\phi_1 z_{7305} + \phi_2 z_{7304}) + \phi_2 z_{7305}$$

The 3 step prediction

$$\hat{z}_{7308|7307} = \phi_1 \hat{z}_{7307|7306} + \phi_2 \hat{z}_{7306|7305}$$
$$= \phi_1(\phi_1 z_{7305} + \phi_2 z_{7304}) + \phi_2 z_{7305} + \phi_1 z_{7305} + \phi_2 z_{7304}$$

c) We have $z_t = x_t - x_{t-1} \Rightarrow x_t = z_t + x_{t-1}$

- $x_{7306} = z_{7306} + x_{7305}$ (z_{7306} known from (b))
- $x_{7307} = z_{7307} + x_{7306} = z_{7307} + z_{7306} + x_{7305}$
- $x_{7308} = z_{7308} + x_{7307} = z_{7308} + z_{7307} + z_{7306} + x_{7305}$

6. 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 an MA(2)-process?

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

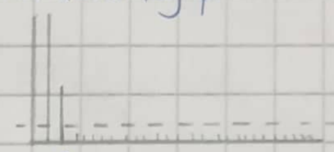
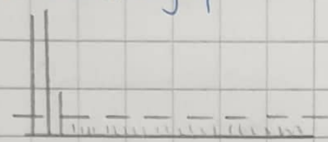
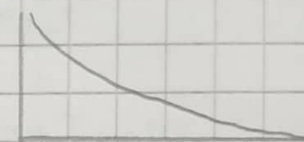
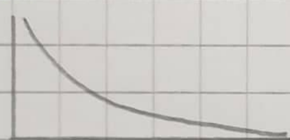
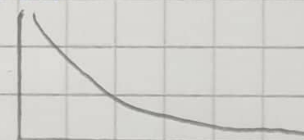
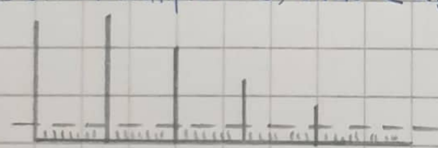
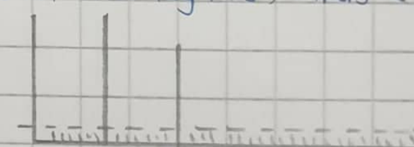
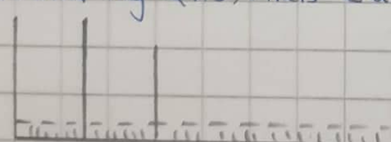
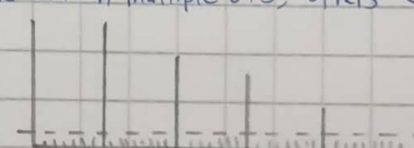
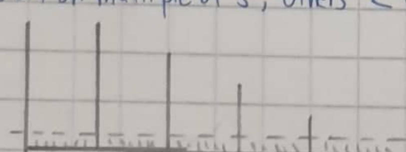
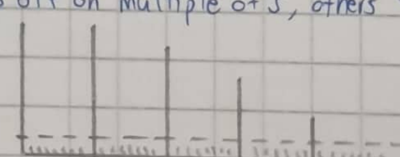
(c) Which one of the processes (Series 1, 2, 3, 4, 5 or 6) is an ARMA(2,2)-process?

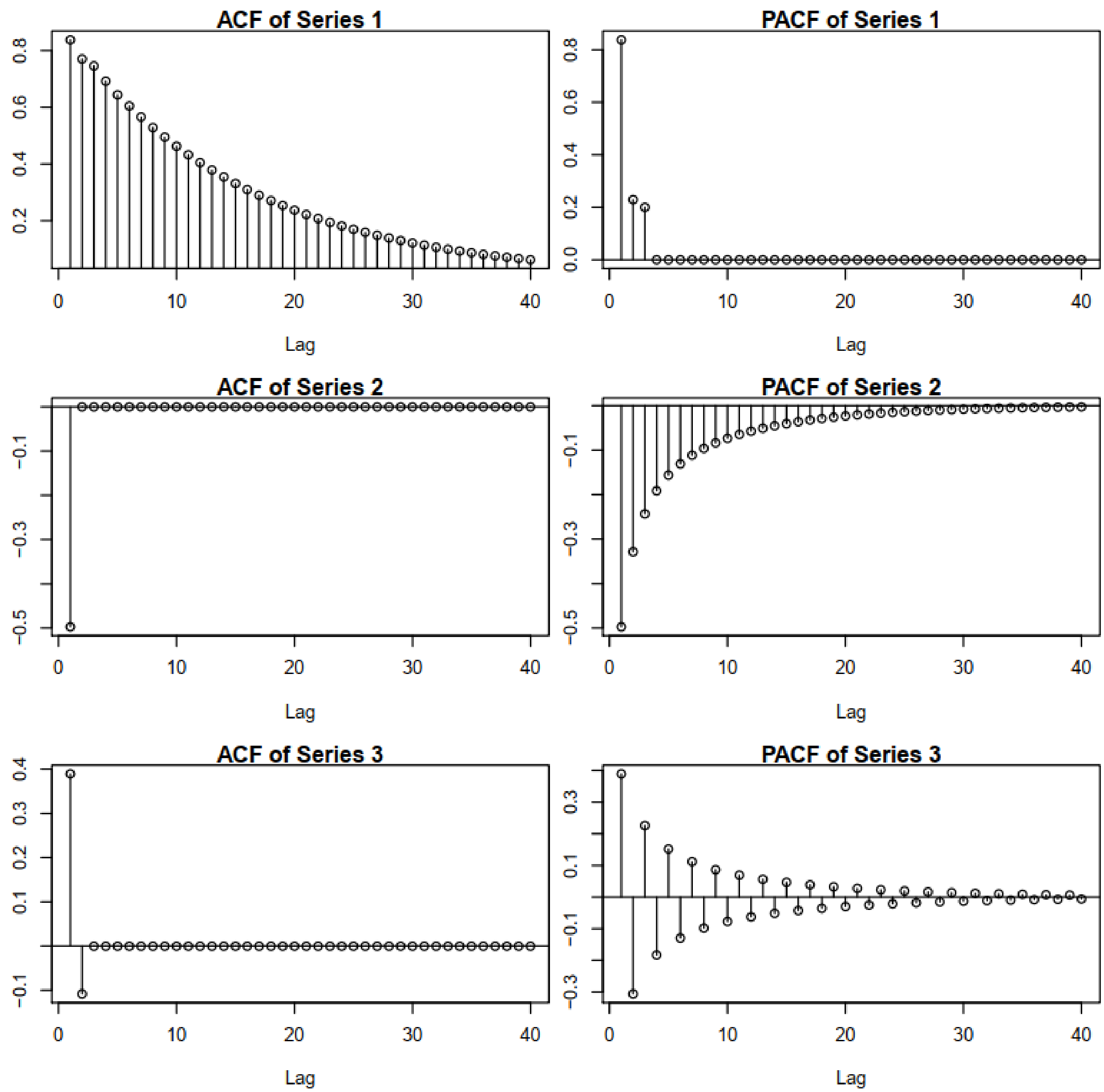
(d) Which one of the processes (Series 1, 2, 3, 4, 5 or 6) is a MA(1)-process?

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

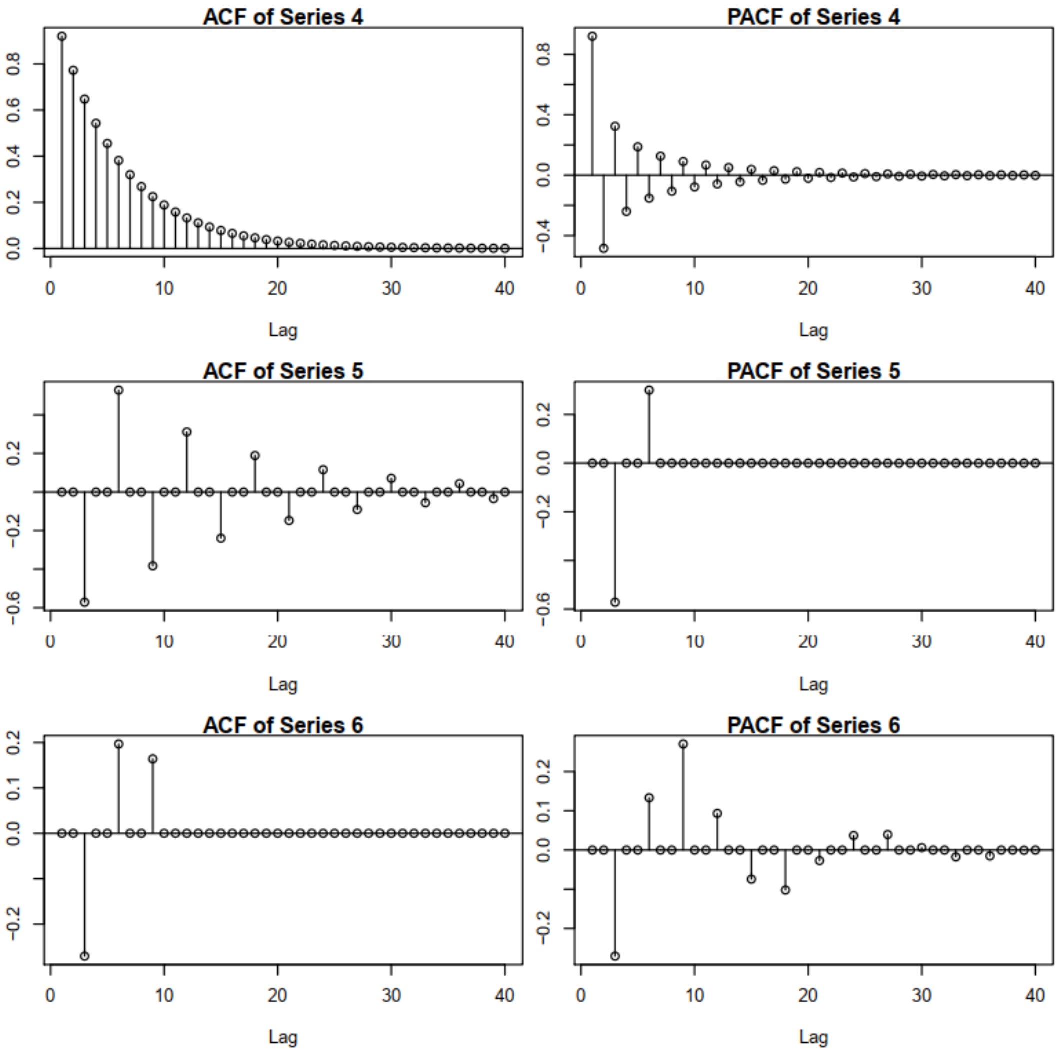
(f) Which one of the processes (Series 1, 2, 3, 4, 5 or 6) is a $SMA(3)_3$-process?

Time series model ACF and PACF

	ACF	PACF
AR(p) $x_t = \varphi_1 x_{t-1} + \dots + \varphi_p x_{t-p} + \varepsilon_t$	Tails off 	Cuts off after lag p 
MA(q) $x_t = \theta_1 \varepsilon_{t-1} + \dots + \theta_q \varepsilon_{t-q} + \varepsilon_t$	Cuts off after lag q 	Tails off 
ARMA(p,q) $x_t = \varphi_1 x_{t-1} + \dots + \varphi_p x_{t-p} + \theta_1 \varepsilon_{t-1} + \dots + \theta_q \varepsilon_{t-q} + \varepsilon_t$	Tails off 	Tails off 
SAR(P)$_s$ $x_t = \phi_1 x_{t-s} + \dots + \phi_P x_{t-Ps} + \varepsilon_t$	Tails off on multiple of s, others $< \alpha_{sig}$ 	Cuts off after lag P*s, others $< \alpha_{sig}$ 
SMA(Q)$_s$ $x_t = \Theta_1 \varepsilon_{t-s} + \dots + \Theta_Q x_{t-Qs} + \varepsilon_t$	Cuts off after lag Q×s, others $< \alpha_{sig}$ 	Tails off on multiple of s, others $< \alpha_{sig}$ 
SARMA(P,Q)$_s$ $x_t = \Phi_1 x_{t-s} + \dots \Phi_P x_{t-Ps} + \dots + \Theta_1 \varepsilon_{t-s} + \dots + \Theta_Q \varepsilon_{t-Qs} + \varepsilon_t$	Tails off on multiple of s, others $< \alpha_{sig}$ 	Tails off on multiple of s, others $< \alpha_{sig}$ 

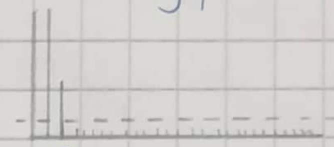
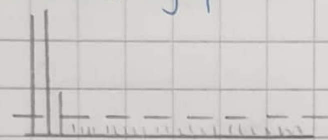
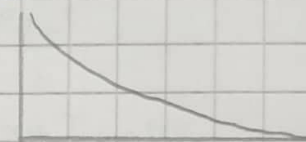
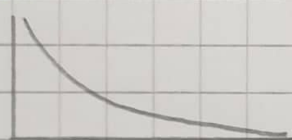
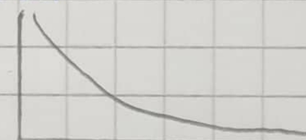
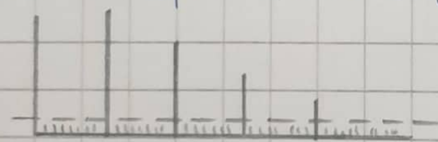
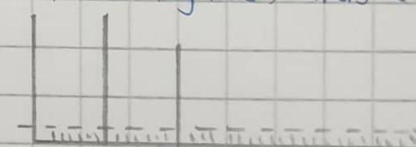
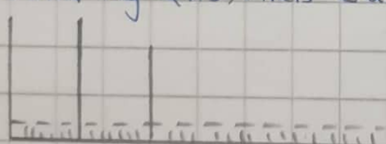
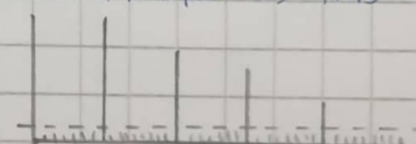
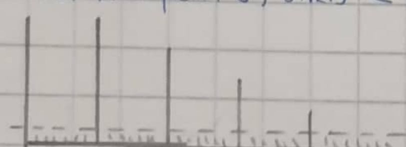
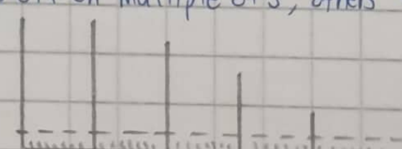


Series 1: AR(3)
Series 2: MA(1)
Series 3: MA(2)



Series 4: ARMA(2,2)
Series 5: SAR(2)3
Series 6: SMA(3)3

Time series model ACF and PACF

	ACF	PACF
AR(p) $x_t = \varphi_1 x_{t-1} + \dots + \varphi_p x_{t-p} + \varepsilon_t$	Tails off 	Cuts off after lag p 
MA(q) $x_t = \theta_1 \varepsilon_{t-1} + \dots + \theta_q \varepsilon_{t-q} + \varepsilon_t$	Cuts off after lag q  	Tails off 
ARMA(p,q) $x_t = \varphi_1 x_{t-1} + \dots + \varphi_p x_{t-p} + \theta_1 \varepsilon_{t-1} + \dots + \theta_q \varepsilon_{t-q} + \varepsilon_t$	Tails off 	Tails off 
SAR(P)$_s$ $x_t = \phi_1 x_{t-s} + \dots + \phi_P x_{t-Ps} + \varepsilon_t$	Tails off on multiple of s, others $< \alpha_{sig}$  	Cuts off after lag P*s, others $< \alpha_{sig}$ 
SMA(Q)$_s$ $x_t = \Theta_1 \varepsilon_{t-s} + \dots + \Theta_Q x_{t-Qs} + \varepsilon_t$	Cuts off after lag Q×s, others $< \alpha_{sig}$  	Tails off on multiple of s, others $< \alpha_{sig}$ 
SARMA(P,Q)$_s$ $x_t = \Phi_1 x_{t-s} + \dots \Phi_P x_{t-Ps} + \dots + \Theta_1 \varepsilon_{t-s} + \dots + \Theta_Q \varepsilon_{t-Qs} + \varepsilon_t$	Tails off on multiple of s, others $< \alpha_{sig}$  	Tails off on multiple of s, others $< \alpha_{sig}$ 

2019 Test

1. True or False (6 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, traditional least-squares estimators can be applied only if the residuals are normally distributed.

(b) In the context of linear regression, the coefficient of determination is a measure of heteroscedasticity.

(c) An autoregressive process of order 1 is always stationary.

(d) The theoretical autocorrelation function of a pure autoregressive process of order 3 is equal to 0 after 3.

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

(f) Autoprojective time series models are models that involve only the time series to be forecasted.

a) False. Least square estimator does not require that the error term follows a normal distribution to produce unbiased estimates with the minimum variance. However, satisfying this assumption allows you to perform statistical hypothesis testing and generate reliable confidence intervals and prediction intervals.
b) False. Residual variance is a measure of heteroscedasticity
c) False. It is stationary only if its roots of the lag polynomials lie inside the unit circle.
d) False. This is the attribute of moving average MA model, not AR model
e) True
f) True

2. Linear regression (6 p.)

Consider a random sample $(x_1, y_1), (x_2, y_2), (x_3, y_3) \dots, (x_{5117}, y_{5117})$ from the linear regression model

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \epsilon_i.$$

You have estimated the parameters β_0, β_1 and β_2 from the sample using traditional least squares estimators.

(a) Explain, step by step, how to construct a 95% bootstrap confidence intervals for the parameters β_1 and β_2. (4 p.)

(b) Assume that 0 is in the confidence interval that corresponds to β_1, but it is not in the confidence interval that corresponds to β_2. How would you interpret that? (2 p.)

a) A 95% bootstrap confidence interval for the regression parameter β_1 and β_2 can be obtained as follows.
Step 1: Select exactly 5117 data points randomly with replacement from the original observations (x_i, y_i), i = 1;...; 5117. Each data point (x_i, y_i) can be selected once, multiple times, or not at all.
Step 2: Calculate a new regression model parameter estimate β_1 and β_2 from the new sample formed in step 1
Step 3: Repeat the previous steps (m-1) times. If m = 1000 then redo step 1-2 999 times
Step 4: Order all the obtained estimates β_1 and β_2 from the smallest to the largest. Include also the original estimate, which amounts to 1000 different numbers of β_1 and β_2
Step 5: Set the lower end of the bootstrap confidence interval to be smaller than or equal to the 25th ordered estimate and set the upper end of the bootstrap confidence interval to be larger than or equal to the 975th ordered estimate for both β_1 and β_2
=> A 95% bootstrap confidence interval for the regression parameter β_1 and β_2 falls in the range of 25th and 975th β_1 and β_2 from the bootstrapping method
b) Since 0 is in the confidence interval for β_1, there is likelihood that β_1 true value lies around 0. In other words, β_1 will contribute very little or none to the regression model. On the other hand, there are no 0 value in the confidence interval for β_2, which means that the regression model definitely depends on the parameter β_2 to some extent.
Possible interpretations could be:
+ There is little correlation between x_{i1} and y_i
+ The explanatory variables x_{i1} and x_{i2} are linearly dependent, and thus we can omit the variable x_{i1} from the model to achieve better regression accuracy

3. Stationarity (6 p.)

Let x_t and z_t be weakly stationary stochastic processes such that, for all $t, s \in \mathbb{Z}$, we have that $E[x_t z_s] = 0$. Let $y_t = x_t + z_t$. Show that the process y_t is weakly stationary.

Exercise 3:

x_t and z_t are weakly stationary stochastic processes, $E[x_t z_s] = 0 \ \forall t, s \in \mathbb{Z}$

$\Rightarrow E(x_t) = \mu_x$, $E(z_t) = \mu_z$, μ_x, μ_z independent of t

$Var(x_t) = \sigma_x^2 < \infty$, $Var(z_t) = \sigma_z^2 < \infty \ \forall t \in T$

$Cov(x_t, x_{t-\tau}) = \gamma_\tau \ \forall t, \tau \in T$, $Cov(z_t, z_{t-\tau}) = \kappa_\tau \ \forall t, \tau \in T$

- We have: $y_t = x_t + z_t$

  * $E[y_t] = E[x_t + z_t] = E[x_t] + E[z_t] = \mu_x + \mu_z$ independent of t
  * $Var[y_t] = E[y_t^2] - E[y_t]^2 = E[x_t^2 + 2x_t z_t + z_t^2] + (\mu_x + \mu_z)^2$

    $(E[x_t^2] = Var[x_t] + E[x_t]^2 = \sigma_x^2 + \mu_x^2)$

    $\Rightarrow Var[y_t] = \sigma_x^2 + \mu_x^2 + \sigma_z^2 + \mu_z^2 + \mu_x^2 + 2\mu_x\mu_z + \mu_z^2$

    $= \sigma_x^2 + \sigma_z^2 + 2(\mu_x^2 + \mu_z^2 + \mu_x\mu_z) < \infty$

  * $Cov(y_t, y_{t-\tau}) = E[(y_t - E[y_t])(y_{t-\tau} - E[y_{t-\tau}])]$ (1)

    $= E[(y_t - \mu_x - \mu_z)(y_{t-\tau} - \mu_x - \mu_z)]$

We have $Cov(x_t, x_{t-\tau}) = \gamma_\tau \Rightarrow E[(x_t - E[x_t])(x_{t-\tau} - E[x_{t-\tau}])] = \gamma_t$

$\Rightarrow E[x_t x_{t-\tau}] = \gamma_\tau + \mu_x^2$

Similarly $E[z_t z_{t-\tau}] = \kappa_\tau + \mu_z^2$

From (1): $Cov(y_t, y_{t-\tau}) = E[y_t y_{t-\tau}] - (\mu_x + \mu_z)^2$

$= E[(x_t + z_t)(x_{t-\tau} + z_{t-\tau})] - (\mu_x + \mu_z)^2$

$= E[x_t x_{t-\tau} + x_t z_{t-\tau} + z_t x_{t-\tau} + z_t z_{t-\tau}] - (\mu_x + \mu_z)^2$

$= \gamma_\tau + \mu_x^2 + \kappa_\tau + \mu_z^2 - (\mu_x + \mu_z)^2$ independent of t

4. ARMA modeling (6 p.)

Assume that you have observed a series $x_0, x_1, x_2, \ldots, x_{7305}$.

(a) Based on plotting the series, you observe a linear trend. You manage to stationarize the process by taking a difference. Give the elements of the obtained stationary process in terms of the elements of the original observed series. (1 p.)

(b) Based on plotting the stationarized series and its estimated auto-correlation and partial autocorrelation -functions, you think that the stationarized series is a pure autoregressive process of order 2. Give the definition of an autoregressive process of order 2. (2 p.)

(c) You decide to apply traditional ARMA-modeling based prediction to calculate the 1, 2, 3 and 4 step predictions for the stationarized series. What are the predicted values of x_{7306}, x_{7307}, x_{7308} and x_{7309} of the original observed series? (3 p.)

Exercise 4 Done in 2020 test

5. 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 an AR(2)-process?

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

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

(d) Which one of the processes (Series 1, 2, 3, 4, 5 or 6) is a $SAR(2)_6$-process?

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

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

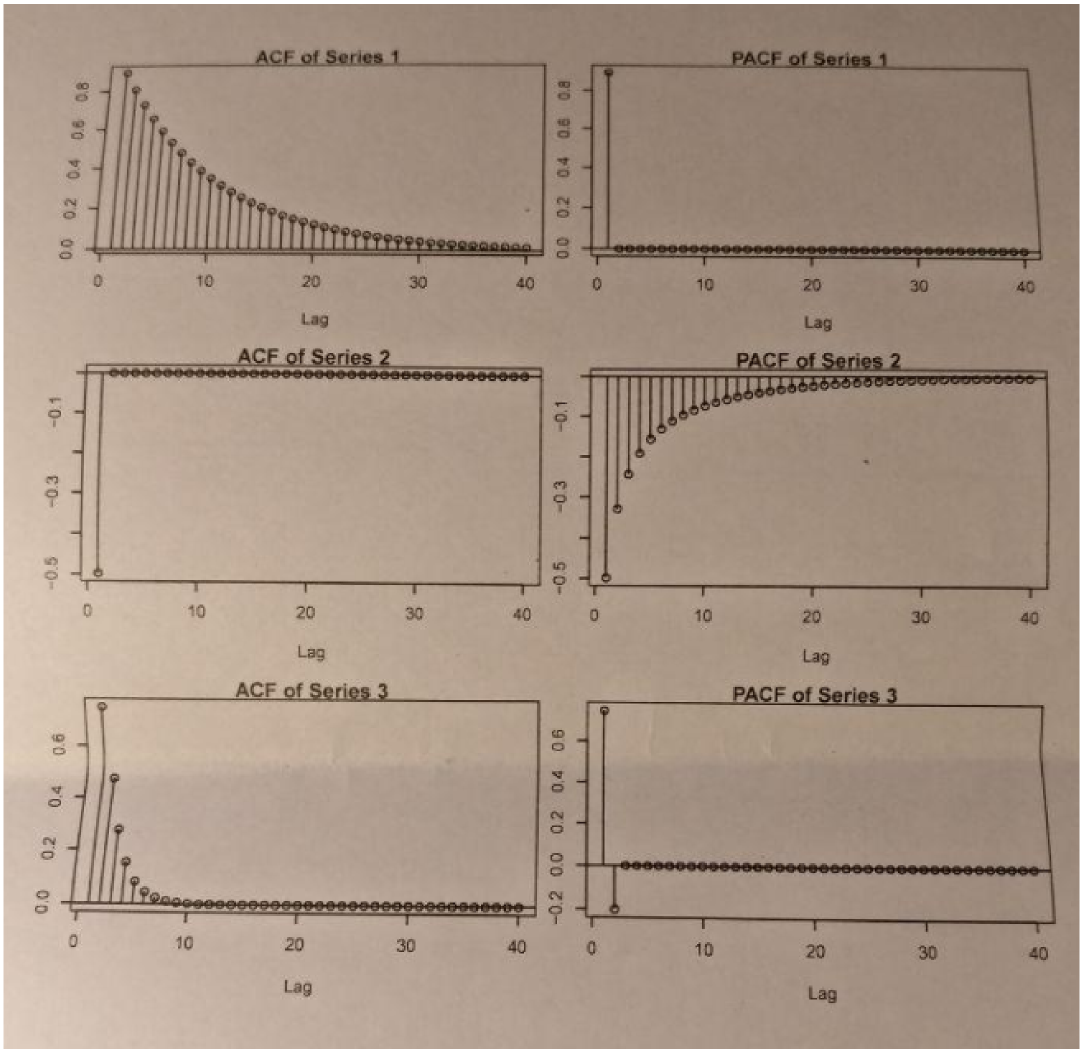


Figure 1

Series 1: AR(1)
Series 2: MA(1)
Series 3: AR(2)

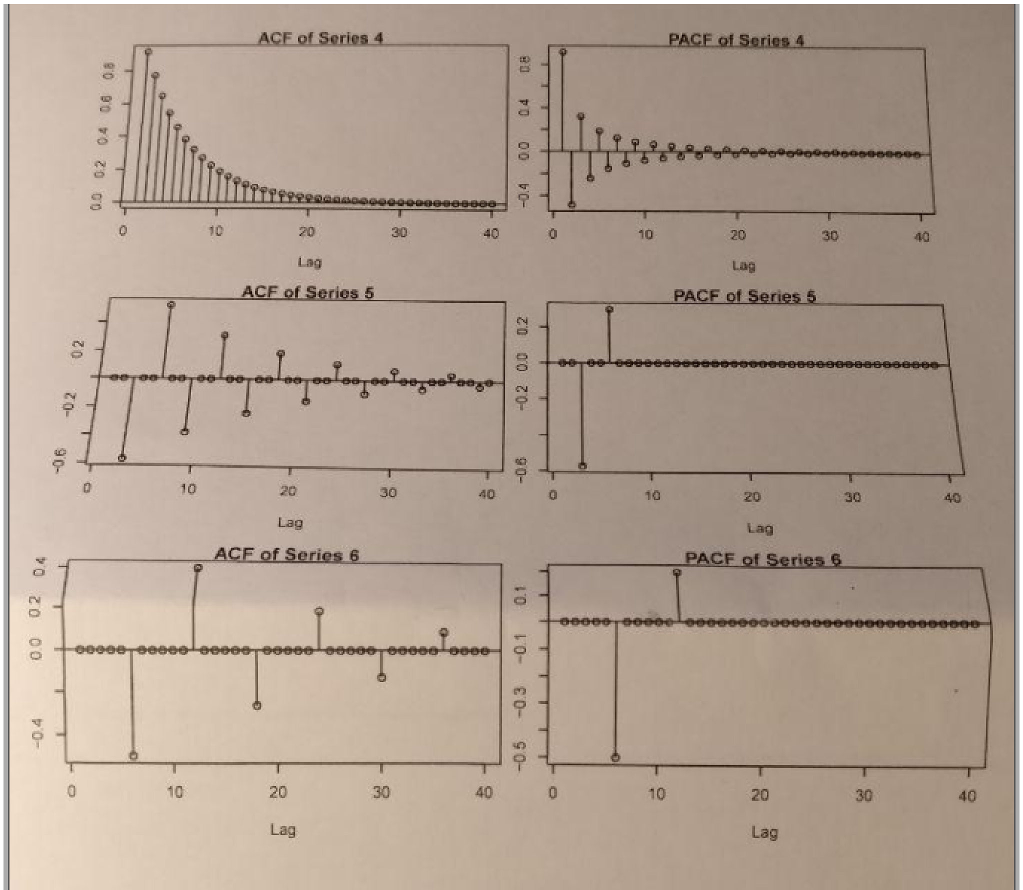


Figure 2

Series 4: ARMA(2,2)
Series 5: SAR(2)3
Series 6: SAR(2)6

2018 Test Autumn

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, 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 heteroscedasticity.

(c) In time series analysis, differencing and seasonal differencing can be applied in order to stationarize the process.

(d) The theoretical partial autocorrelation function of a moving average process of order 3 is equal to 0 after 3.

(e) Autoregressive processes can be applied in predicting only if the residual terms of the process are normally distributed.

(f) ARIMAX models have been designed for long term forecasting.

a) True. In particular, least squares estimates for regression models are highly sensitive to outliers.
b) False. VIF are calculated in order to detect multicollinearity
c) True
d) True
e) False. Residual terms of AR model is i.i.d white noise. White noise distribution is continuous and there are no assumptions of normal distribution
f) False. ARIMAX models are reasonably general, but they are not suitable for long term forecasting. Moreover, ARIMAX models are linear on the explanatory variables

2. Linear regression (6 p.)

Consider the linear regression model

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

You have a sample and you estimate the parameters β_0 and β_1 using traditional least squares estimators. In your sample, you have two separate subgroups that have sample sizes h and $n-h$, and you are worried about possible parameter instability.

(a) Explain, step by step, how to apply a permutation test in testing parameter instability. (4 p.)

(b) Give the null hypothesis of the test. (1 p.)

(c) How would you proceed your modeling (what would you do), if the estimated p-value of the test statistic was very small? (1 p.)

Done in 2020 test

3. Autocorrelation (6 p.)

Let ρ_k denote the kth autocorrelation coefficient of a stationary stochastic process $(x_t)_{t \in T}$.

(a) Prove that $\rho_0 = 1$. (1 p.)

(b) Prove that $\rho_{-k} = \rho_k$ for all $k \in \mathbb{Z}$. (2 p.)

(c) Prove that $|\rho_k| \leq 1$ for all $k \in \mathbb{Z}$. (3 p.)

Week 3, page 14:

(i) $\rho_0 = \frac{\gamma_0}{\gamma_0} = 1$

(ii) $\rho_{-k} = \frac{\gamma_{-k}}{\gamma_0} = \frac{cov(X_t, X_{t-(-k)})}{\gamma_0}$

$= \frac{cov(X_t, X_{t+k})}{\gamma_0} = \frac{cov(X_{t+k}, X_t)}{\gamma_0}$

$\overset{\text{stationarity}}{=} \frac{cov(X_t, X_{t-k})}{\gamma_0} = \frac{\gamma_k}{\gamma_0} = \rho_k$

(iii) $|\rho_k| = \left| \frac{\gamma_k}{\gamma_0} \right| = \left| \frac{cov(X_t, X_{t-k})}{var(X_t)} \right|$

$= \frac{|E[(X_t - \mu)(X_{t-k} - \mu)]|}{var(X_t)} \overset{*}{\leq} \frac{\sqrt{E[(X_t-\mu)^2]E[(X_{t-k}-\mu)^2]}}{var(X_t)}$

$\overset{\text{stationarity}}{=} \frac{\sqrt{E[(X_t-\mu)^2]E[(X_t-\mu)^2]}}{var(X_t)} = \frac{\sqrt{(var(X_t))^2}}{var(X_t)} = \frac{var(X_t)}{var(X_t)} = 1$

* Cauchy-Schwarz inequality for random variables:

$|E[XY]|^2 \leq E[X^2]E[Y^2]$

4. ARMA modeling (6 p.)

Assume that you have observed a series $x_1, x_2, x_3, ..., x_{5012}$.

(a) Based on plotting the series, you observe a linear trend. You manage to stationarize the process by taking a difference. Give the elements of the obtained stationary process in terms of the elements of the original observed series. (1 p.)

(b) Based on plotting the stationarized series and its estimated autocorrelation and partial autocorrelation -functions, you think that the observed series is a pure invertible moving average process of order 2. Give the definition of a moving average process of order 2. (2 p.)

(c) You decide apply traditional ARMA-modeling based prediction to calculate the 1, 2, 3 and 4 step predictions for the stationarized series. What are the predicted values of x_{5013}, x_{5014}, x_{5014} and x_{5015}? (3 p.)

< Not done yet>

5. 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(3)-process?

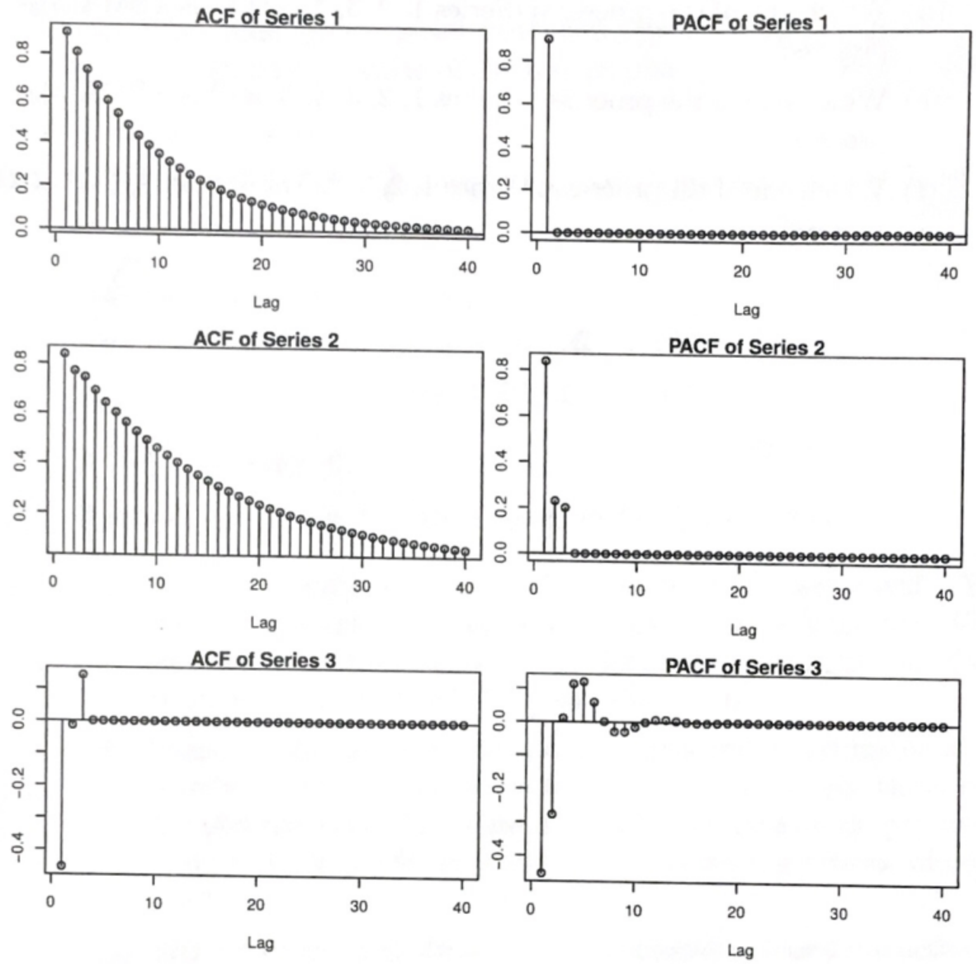
(b) Which one of the processes (Series 1, 2, 3, 4, 5 or 6) is an AR(1)-process?

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

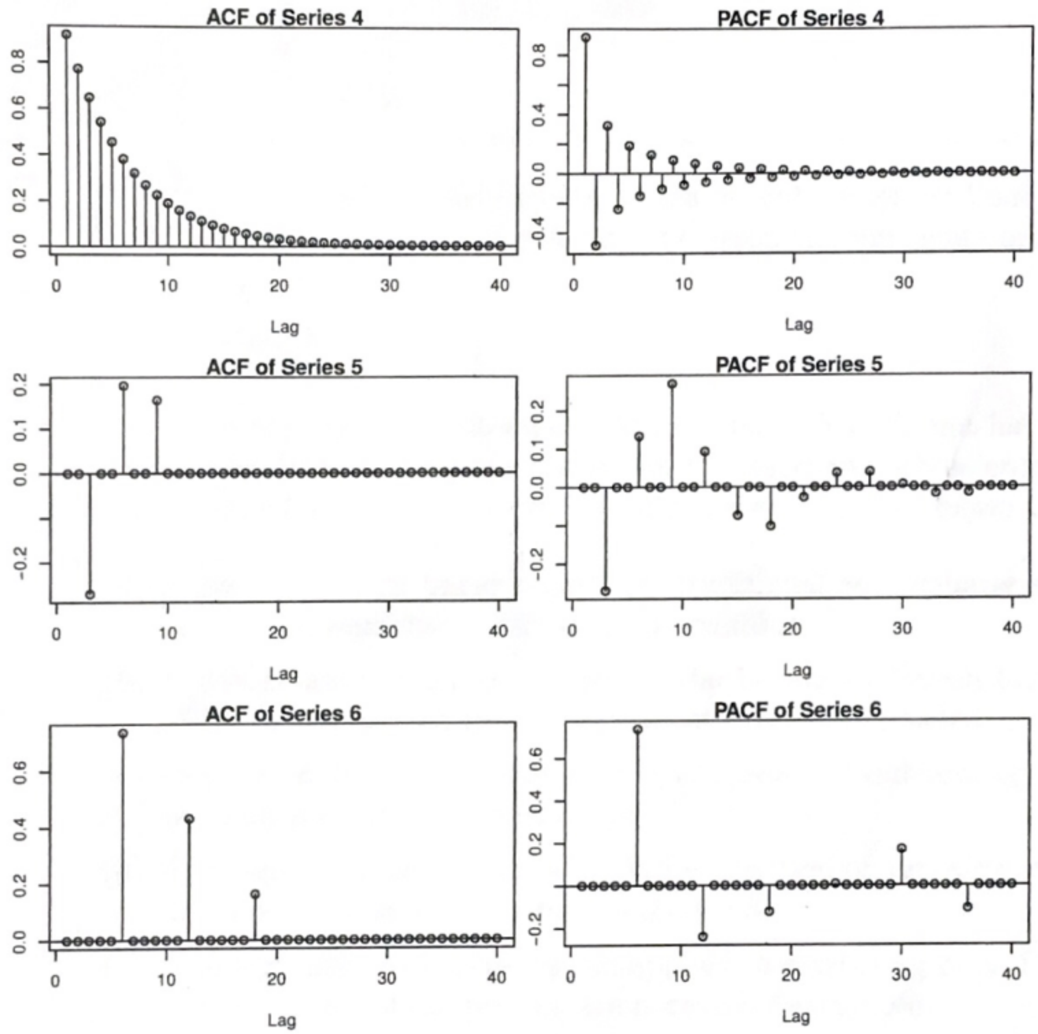
(d) Which one of the processes (Series 1, 2, 3, 4, 5 or 6) is a $\mathrm{SMA}(3)_3$-process?

(e) Which one of the processes (Series 1, 2, 3, 4, 5 or 6) is a $\mathrm{SMA}(3)_6$-process?

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



Series 1: AR(1)
Series 2: AR(3)
Series 3: MA(3)



Series 4: ARMA(2,2)
Series 5: SMA(3)3
Series 6: SAR(3)6

2018 Test Spring

1. True or False (6 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, traditional least-squares estimators can be applied only if the residuals are normally distributed.

(b) In the context of linear regression, the coefficient of determination is a measure of heteroscedasticity.

(c) In time series analysis, differencing can be applied in order to remove a linear trend.

(d) The theoretical partial autocorrelation function of an autoregressive process of order 3 is equal to 0 after 3.

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

(f) Autoprojective time series models are models that involve only the time series to be forecasted.

a) False
b) False
c) True
d) False
e) True
f) True

2. Linear regression (6 p.)

Consider the linear regression model

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \epsilon_i.$$

You have a sample and you estimate the parameters β_0, β_1 and β_2 using traditional least squares estimators. You are worried about possible heteroscedasticity and you decide to apply the White homoscedasticity test.

(a) Give the corresponding White test model and the White homoscedasticity test statistic. (2 p.)

(b) Give the null hypothesis of the test. (1 p.)

(c) Assume that the ϵ_i are normally distributed. What is the distribution of the test statistic under the null hypothesis? (1 p.)

(d) Assume that the ϵ_i are not normally distributed. Explain how, in this case, you can estimate the p-value of the test statistic. (2 p.)

<Not Done Yet>

a) The White test model is

ei2 = γ0 + γ1xi1 + γ2xi2 + γ3xi1^2 + γ4xi2^2 + γ5xi1xi2 + δi

And White homoscedasticity test statistic

The test statistic is

$$\chi^2 = \sum_{i=n+1}^{n+h} \frac{u_i^2}{s^2}.$$

b) The null hypothesis of the test

The null hypothesis H_0: $\beta_1 = \beta_2$, $\sigma_1^2 = \sigma_2^2$.

- The parameter β_1 is related to the observations $1, ..., n$.
- The parameter β_2 is related to the observations $n+1, ..., n+h$.

c) If the residuals are normally distributed, the the test statistic follows, under the null, the χ2(h) distribution.

d) The p-value can be estimated using permutations

3. Stationarity (6 p.)

Let ϵ_t be iid and assume that $E[\epsilon_t] = 0$ and $E[(\epsilon_t)^2] = \sigma^2, \sigma^2 < \infty$. Let $x_0 = \epsilon_0$ and let $x_t = x_{t-1} + \epsilon_t, t > 0$. Let D denote the difference operator. Show that the process $y_t = D^2 x_t$ is weakly stationary.

Exercise 3

ϵ_t is i.i.d, $E[\epsilon_t] = 0$, $E[(\epsilon_t)^2] = \sigma^2$, $\sigma^2 < \infty$. Let $x_0 = \epsilon_0$ and let $x_t = x_{t-1} + \epsilon_t$, $t > 0$. D is difference operator. Show $y_t = D^2 x_t$ is stationary

We have: $E[y_t] = E[D^2 x_t] = E[D(x_t - x_{t-1})]$

$= E[x_t - 2x_{t-1} + x_{t-2}]$

We have: $\left.\begin{array}{l} E[\epsilon_0] = 0,\ E[x_0] = E[\epsilon_0] = 0 \\ E[\epsilon_1] = E[x_0 + \epsilon_1] = 0 + 0 = 0 \\ E[\epsilon_2] = E[x_1 + \epsilon_2] = 0 + 0 = 0 \end{array}\right\} \Rightarrow$ This pattern go on recursively $\Rightarrow E[x_t] = 0$

$\Rightarrow E[y_t] = 0 - 2 \times 0 + 0 = 0$ independent of t

□ Variance: $Var(y_t) = E[(y_t - E[y_t])^2] = E[y_t^2]$

$= E[D^4 x_t^2] = E[(1-L)^4 x_t^2] = E[(1 - 4L + 6L^2 - 4L^3 + L^4) x_t^2]$

$= E[x_t^2 - 4x_t x_{t-1} + 6x_t x_{t-2} - 4x_t x_{t-3} + x_t x_{t-4}]$

Analysis: $x_t = x_{t-1} + \epsilon_t = x_{t-2} + \epsilon_t + \epsilon_t = x_{t-3} + \epsilon_t + \epsilon_t + \epsilon_t$

$\Rightarrow x_t = x_{t-k} + k\epsilon_t$, $k \leq t$

If $k = t \Rightarrow x_t = x_0 + t\epsilon_t = \epsilon_t + t\epsilon_t = (t+1)\epsilon_t$

$\cancel{\Rightarrow E[y_t^2] = E[(t+1)^2 \epsilon_t^2 - 4t^2\epsilon_t^2 + 6(t-1)^2\epsilon_t^2 - 4}$

$\Rightarrow E[y^2_t] = E[(t+1)^2\epsilon_t^2 - 4(t+1)t\epsilon_t^2 + 6(t+1)(t-1)\epsilon_t^2 - 4(t+1)(t-2)\epsilon_t^2$
$+ (t+1)(t-3)\epsilon_t^2]$, where $E[(\epsilon_t)^2] = \sigma^2$

$\Rightarrow E[y^2_t] = [(t+1)^2 - 4(t+1)t \cdots + (t+1)(t-3)]\sigma^2 < \infty$

□ Covariance: $Cov(y_t, y_{t-\tau}) = E[(y_t - E[y_t])(y_{t-\tau} - E[y_{t-\tau}])]$

$= E[y_t y_{t-\tau}] = E[D^4 x_t x_{t-\tau}]$

We know that $x_t x_s = f(t) * \sigma^2 \Rightarrow Cov(y_t, y_{t-\tau}) = E[D^4 f(t)\sigma^2]$

$= D^4 f(t)\sigma^2 < \infty$

4. Interval bootstrapping/block bootstrapping (6 p.)

Assume that you have observed a time series $x_1, x_2, x_3, \ldots, x_{10263}$. Based on plotting the series and its estimated autocorrelation and partial autocorrelation -functions, you think that the observed series is a stationary AR(2) process

$$x_t = \phi_1 x_{t-1} + \phi_2 x_{t-2} + \epsilon_t, \quad (\epsilon_t)_{t\in T} \sim WN(0, \sigma^2)$$

and you have estimated the parameters ϕ_1 and ϕ_2.

(a) Explain, step by step, how to construct a 95% bootstrap confidence intervals for the parameters ϕ_1 and ϕ_2. (4 p.)

(b) Assume that 0 is in the confidence interval that corresponds to ϕ_1, but it is not in the confidence interval that corresponds to ϕ_2. How would you interpret that? (2 p.)

Done in 2019 test

5. 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 an AR(2)-process?

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

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

   (d) Which one of the processes (Series 1, 2, 3, 4, 5 or 6) is a $SAR(2)_6$-process?

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

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

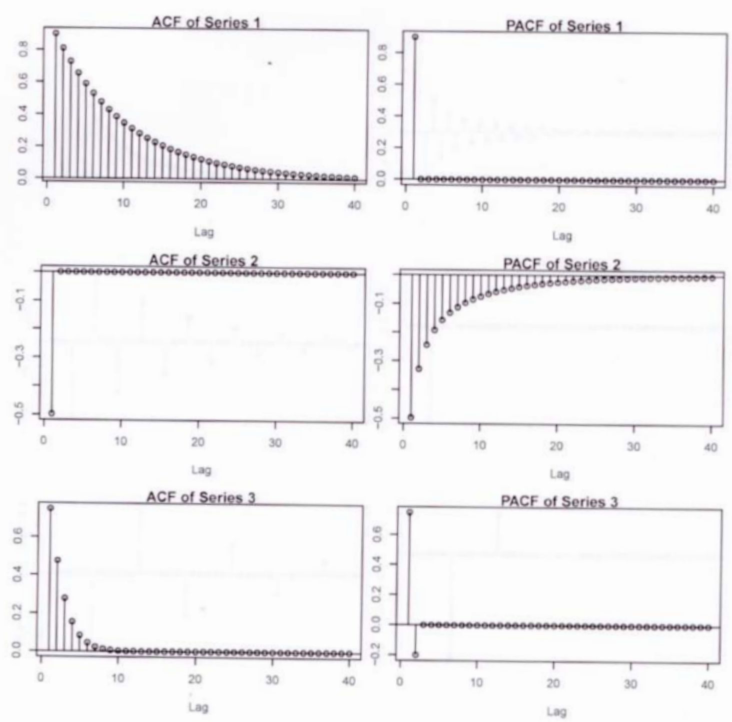


Figure 1

Series 1: AR(1)
Series 2: MA(1)
Series 3: AR(2)

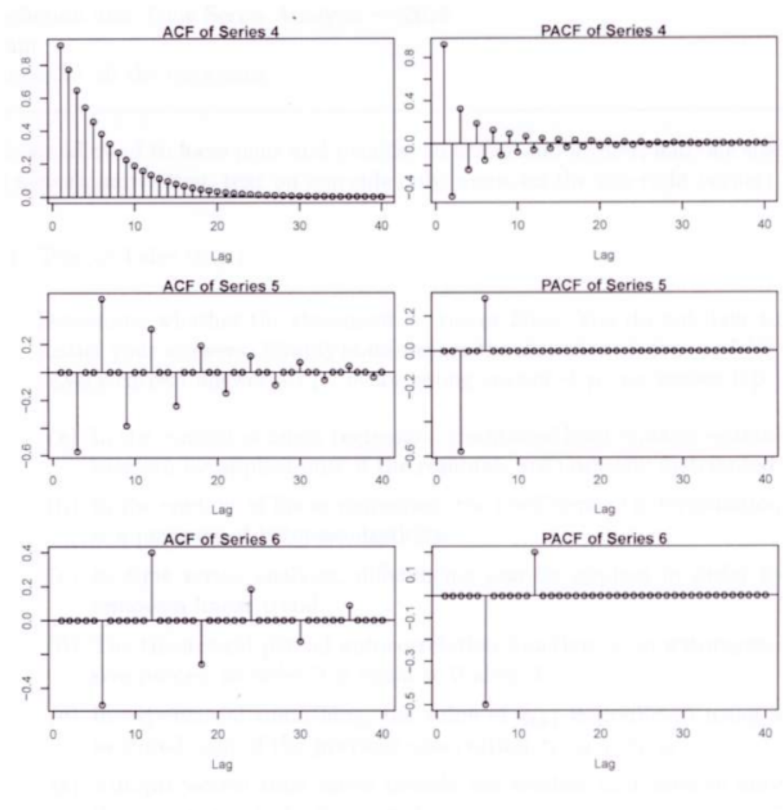


Figure 2

Series 4: ARMA(2,2)
Series 5: SAR(2)3
Series 6: SAR(2)6

2017 Test Autumn

1. Ovatko seuraavat väittämät aina totta? Vastaa **1** = Kyllä tai **2** = Ei. Oikeasta vastauksesta 1 piste, väärästä -1 piste ja tyhjästä 0. Jos tehtävän kokonaispisteet on alle 0, niin tehtävän pisteitä ei huomioida arvostelussa.

   (a) Varianssin inflaatiotekijää voidaan käyttää poikkeavien havaintojen tunnistamiseen.

   (b) Stokastinen prosessi $x_t - 0.4x_{t-1} = \epsilon_t + \epsilon_{t-1}$, missä $(\epsilon_t)_{t \in T}$ on valkoinen kohina, on kääntyvä.

   (c) Stokastinen prosessi $x_t - 0.4x_{t-1} = \epsilon_t + \epsilon_{t-1}$, missä $(\epsilon_t)_{t \in T}$ on valkoinen kohina, on stationaarinen.

   (d) Lineaarisessa regressiossa Cookin etäisyyttä voidaan käyttää poikkeavien havaintojen tunnistamiseen.

   (e) Jakautuneen viiveen mallissa aikasarjaa y_t selitetään lineaarisesti jonkin toisen aikasarjan x_t avulla.

   (f) Lineaarisessa regressiossa ei-merkitsevän lineaarisesti riippumattoman selittäjän lisääminen kasvattaa mallin selitysastetta.

a) VIF can be used to detect outliers? False. VIF detects multicollinearity. Cook's distance method can
b) The stochastic process is invertible? False. MA polynomial: 1 - L = 0 => |L| = 1 => the stochastic process is not invertible

c) The stochastic process is stationary? True. AR polynomial: 1 - 0.4L = 0 => |L| = 2.5 => The stochastic process is stationary
d) In linear regression, Cook's distance is used to identify outliers. True
e) In the distributed lag model, the time series yt is explained linearly by other time series xt. True
f) In linear regression, the addition of a nonsignificant linearly independent exponent increases the degree of explanation of the model

2. Vastaa seuraaviin kysymyksiin lyhyesti (noin 2-5 riviä/kysymys):

   (a) Selitä mitä tarkoittaa homoskedastisuus lineaarisessa regressiossa ja kuinka testaat sitä apuregression avulla. (3p)

   (b) Selitä kuinka testaat regression merkitsevyyttä permutaatiotestillä. (3p)

3. (a) Mitä tarkoittaa prosessin (heikko) stationaarisuus? (2p)

   (b) Määrittele stationaarisen prosessin autokorrelaatiofunktio. (1p)

   (c) Osoita, että MA(1)-prosessi $x_t = \epsilon_t + \theta_1 \epsilon_{t-1}$, missä $\epsilon_t \sim WN(0, \sigma^2)$ on valkoinen kohina varianssilla σ^2, on stationaarinen ja laske sen autokorrelaatiofunktio. (3p)

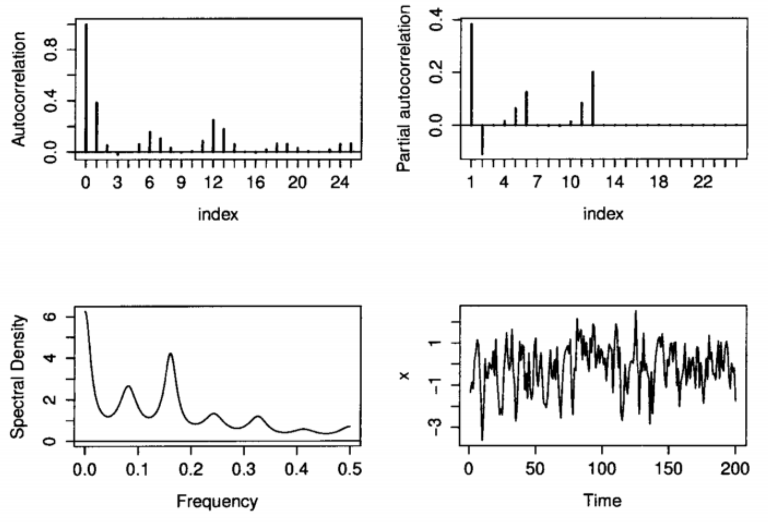
4. Johda stationaarisen AR(2)-prosessin

$$x_t = \phi_1 x_{t-1} + \phi_2 x_{t-2} + \epsilon_t, \quad \epsilon_t \sim i.i.d.(0, \sigma^2),$$

kahden aika-askeleen optimaalinen ennuste keskineliövirheen mielessä, kun x_t on havaittu ajanhetkeen t asti. (6p)

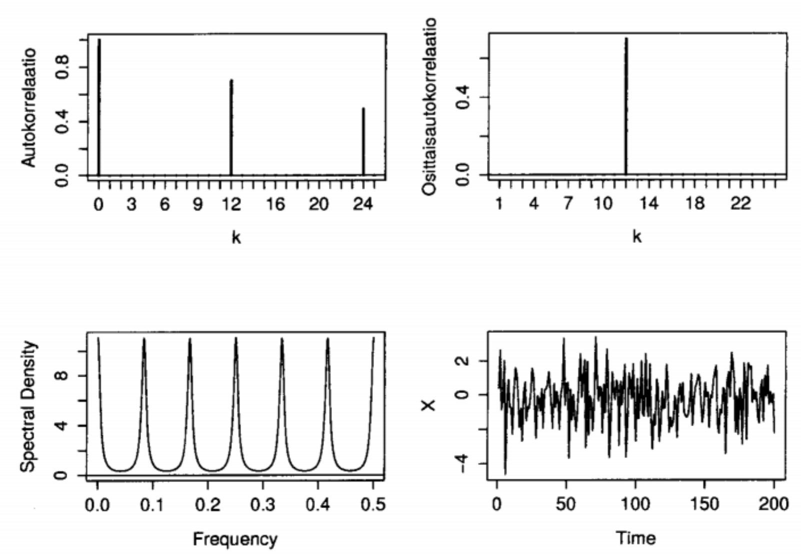
5. Seuraavalla sivulla on kuvattuna eräiden ARIMA-prosessien autokorrelaatiot, osittaisautokorrelaatiot, spektrit ja simuloidut polut. Valitse seuraavista vaihtoehdoista kuviin sopivat prosessit: MA(1), AR(1), ARMA(1,1), AR(12), SMA$(1)_{12}$, SAR$(1)_{12}$. Perustele valintasi huolellisesti.

Kuva 1: Aikasarja 1



Time series 1: AR(12)

Kuva 2: Aikasarja 2



Time series 2:
SAR(1)12

2017 Test Spring

1. Ovatko seuraavat väittämät aina totta? Vastaa **1** = Kyllä tai **2** = Ei. Oikeasta vastauksesta 1 piste, väärästä -1 piste ja tyhjästä 0. Jos tehtävän kokonaispisteet on alle 0, niin tehtävän pisteitä ei huomioida arvostelussa.

   (a) Varianssin inflaatiotekijää voidaan käyttää poikkeavien havaintojen tunnistamiseen.

   (b) Stokastinen prosessi $x_t - 0.4x_{t-1} = \epsilon_t + \epsilon_{t-1}$, missä $(\epsilon_t)_{t \in T}$ on valkoinen kohina, on kääntyvä.

   (c) Stokastinen prosessi $x_t - 0.4x_{t-1} = \epsilon_t + \epsilon_{t-1}$, missä $(\epsilon_t)_{t \in T}$ on valkoinen kohina, on stationaarinen.

   (d) Bootstrapilla voidaan laskea regressiokertoimien luottamusvälejä ilman jakaumaoletuksia.

   (e) Dynaamiset regressiomallit voivat huomioida riippuvuutta myös muista aikasarjoista.

   (f) Stationaarisen prosessin odotusarvo on nolla.

2. Vastaa seuraaviin kysymyksiin lyhyesti:

   (a) Selitä mitä tarkoittaa homoskedastisuus lineaarisessa regressiossa ja kuinka testaat sitä apuregression avulla. (3p)

   (b) Selitä kuinka testaat regression vakioparametrisuutta permutaatiotestillä. (3p)

3. (a) Mitä tarkoittaa prosessin (heikko) stationaarisuus? (2p)

   (b) Osoita, että MA(q)-prosessi $x_t = \epsilon_t + \theta_1\epsilon_{t-1} + \theta_2\epsilon_{t-2} + \cdots + \theta_q\epsilon_{t-q}$, missä $\epsilon_t \sim WN(0, \sigma^2)$ on valkoinen kohina varianssilla σ^2, on stationaarinen ja totea, että sen autokorrelaatiofunktion katkeaa viiveellä q. (4p)

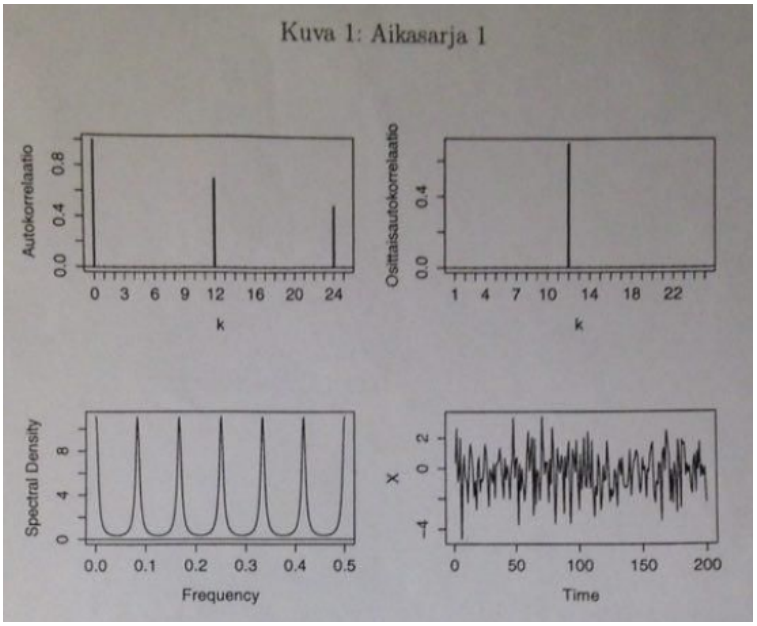
4. (a) Määrittele ARIMA(0,1,1)-prosessi. (2p)

   (b) Osoita, että ARIMA(0,1,1)-prosessi x_t toteuttaa eksponentiaalisen tasoituksen ehdon

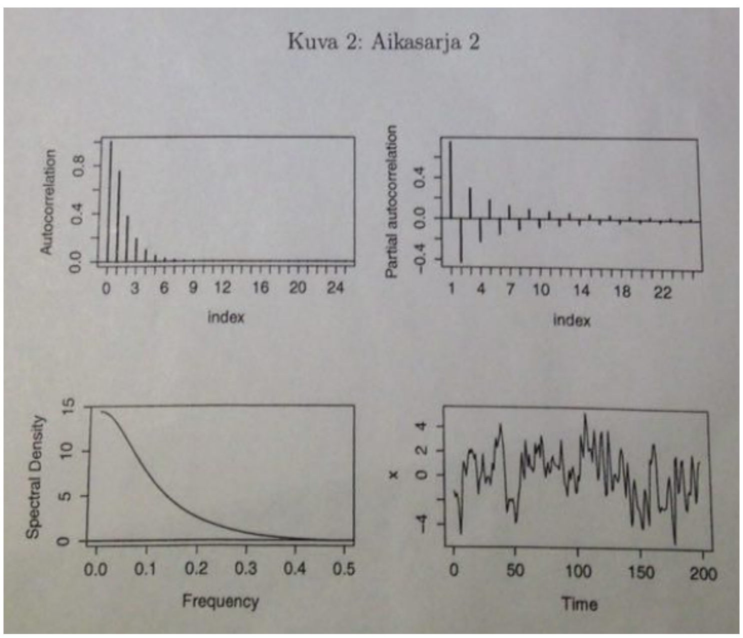
$$\hat{x}_{t+1|t} = \alpha x_t + (1-\alpha)\hat{x}_{t|t-1},$$

   sopivalla parametrin α arvolla. (4p)

5. Seuraavalla sivulla on kuvattuna eräiden ARIMA-prosessien autokorrelaatiot, osittaisautokorrelaatiot, spektrit ja simuloidut polut. Valitse seuraavista vaihtoehdoista kuviin sopivat prosessit: MA(1), AR(1), ARMA(1,1), AR(12), $SMA(1)_{12}$, $SAR(1)_{12}$. Perustele valintasi huolellisesti.



Time series: SAR(1)12



Time series: ARMA(1,1)

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.)

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;...; 862. Each observation (y_i, x_{i1}) can be selected once, multiple 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 estimate, from the smallest to the 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 25th ordered estimate 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 975th 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 25th and 975th values of β_1 from the bootstrapping method, because (975 - 25)/1000 = 95%. This is how the bootstrap samples is used in constructing the confidence interval for parameter β_1
How many bootstrap samples do you take: A total of 999 bootstrap samples, together with the original sample, which amounts to 1000 samples
Sample sizes of your bootstrap samples: each sample of my bootstrap samples has exactly 862 observations
b) Based on the information, it seems that the parameter β_1 is truly significant
Null hypothesis: β_1 parameter is significant. If values near 0 are included in the 95% interval, there are chances that value β_1 can be a null value. In other words, β_1 is insignificant. But the obtained 95% interval (-17.3, -7.2) does not contain 0 nor values near 0 so parameter β_1 **is significant at a 0.05 level**
=> Null hypothesis is accepted: parameter β_1 is significant and should not be rejected from the linear regression model