STOCKHOLM UNIVERSITY
Department of Statistics
Econometrics II, Time Series Analysis, ST223G
Autumn semester 2019

Written Examination in Econometrics II, Time Series Analysis

Date	2020-01-16
Examiner:	Jörgen Säve-Söderbergh
Allowed tools:	1) Textbook: Wooldridge, J.M. *Introductory Econometrics: A Modern Approach*, Cengage, Boston. 2) Textbook: Montgomery, D.C., Jennings, C.L., and Kulachi, M., *Introduction to Time Series Analysis and Forecasting*, John Wiley & Sons, New Jersey. 3) Pocket calculator 4) Notes written in the text book are allowed.

- Note that no formula sheet is provided.
- Passing rate: 50% of overall total, which is 100 points. For detailed grading criteria, see the course description.
- The maximum number of points for each problem is stated after each question. If not indicated otherwise, to obtain the maximum number of points on each problem, detailed and clear solutions are required. Answers may be given in English or Swedish.

Good luck!

1. The following observations have been recorded.

6 8 7 6 9 8 6 7 10 3

We think of these observations as having been generated by an underlying random process. Your task is to estimate the sample autocorrelation function (sample ACF) r_k for the first two lags.

(a) Calculate r_1 och r_2. (8p)

(b) Although this is not done in our textbook *Montgomery et al*, you should also compute the sample partial autocorrelation function (sample PACF). This can be done by using a recursive formula. We use the notation r_{ii} for the sample PACF. From the recursive formula we have

$$r_{11} = r_1 \quad \text{and} \quad r_{22} = \frac{r_2 - r_{11}r_1}{1 - r_{11}r_1}$$

Use this to calculate r_{11} and r_{22}. (2p)

(c) The sample ACF r_1 is an estimator of the theoretical ACF ρ_1 defined on page 30 (in the first edition), also r_2 is an estimator of ρ_2.

Perform a hypothesis test of the null hypothesis $H_0 : \rho_k = 0$ for $k = 1, 2$. (10p)

2. The production of beer in the U.S.A. between the first quarter 1975 and the last quarter 1981 has been observed. There is seasonal variation which has been modelled by seasonal dymmy variables. There is also a positive trend that has been modelled by a linear trend function. The analysis was done with R:

```
lm(formula = Prod ~ time + Q1 + Q2 + Q3, data = dat1)

Residuals:
Min      1Q  Median      3Q     Max
-2.2782 -1.0202  0.2311  0.8592  1.8007

Coefficients:
            Estimate  Std. Error t value Pr(>|t|)
(Intercept) 32.98143   0.73020  45.167  < 2e-16 ***
time         0.38795   0.03227  12.022 2.13e-11 ***
Q1           3.18098   0.73659   4.319 0.000255 ***
Q2          11.12018   0.73305  15.170 1.81e-13 ***
Q3           9.51938   0.73092  13.024 4.24e-12 ***

Residual standard error: 1.366 on 23 degrees of freedom
Multiple R-squared:  0.9504,Adjusted R-squared:  0.9418
F-statistic: 110.2 on 4 and 23 DF,  p-value: 1.187e-14
```

The Durbin-Watson test statistic was also calculated to be equal to 1.822.

(a) Is the estimated regression model suitable for this data? Do statistical hypotheses tests to answer this question. (7p)

(b) Compute a crude estimate of r_1 (or $\hat{\rho}$ as Wooldridge writes) by utilizing an approximate relation between the Durbin-Watson test statistic DW and $\hat{\rho}$. (3p)

(c) Calculate forecasts for each quarter of 1982. Then compare the forecasts from the regression model with a possible competitor; a seasonal ARIMA model. Do this by calculating MSE for both models. Below you find the forecasts from the ARIMA model together with the actual observations from 1982. (10p)

1982	Forecasts from seasonal ARIMA	Actual observations
Quarter I	47.0800	47.84
Quarter II	55.4728	54.27
Quarter III	54.4764	52.31
Quarter IV	45.5577	42.03

3. Sheila is head of an expedition that investigates paleoanatomic finds on Gotland during the summer months. The expedition consumes buns during the coffee breaks. Sheila is worried about this consumption and has documented the number of buns that has been bought at a local bakery during the last five years.

Year	2015	2016	2017	2018	2019
Number of buns	2043	2097	2245	2478	2673

(a) Help Sheila to do a forecast of the number of buns that the expedition is expected to consume this summer. Use the model equations as follows for exponential smoothing

$$
\begin{aligned}
l_t &= \alpha y_t + (1-\alpha)(l_{t-1} + b_{t-1}) \\
b_t &= \gamma(l_t - l_{t-1}) + (1-\gamma)b_{t-1}
\end{aligned}
$$

(This formulation is found on page 191 in the first edition with slightly different notation) Let $\alpha = 0.3$ and $\gamma = 0.1$.
To start the iteration we need starting values; l_0 and b_0. Fit a regression model to the entire data material and and use the estimated intercept as l_0 and the estimated regression coefficient as b_0. (10p)

(b) Since Sheila has not taken any statistics course, but has heard of so called naive forecasts, she wishes to do her own forecast and therefore googles "naive forecast". She finds *naive forecasting. Estimating technique in which the last period's actuals are used as this period's forecast, without adjusting them or attempting to establish causal factors. It is used only for comparison with the forecasts generated by the better (sophisticated) techniques.* Which forecast does Sheila do? (5p)

(c) Which forecast would you recommend? Do a comparison without any calculations. (5p)

4. The following model was found to fit a time series well.

$$y_t = 1.3y_{t-1} - 0.4y_{t-2} + \varepsilon_t$$

where ε_t is independent and normally distributed with expected value 0 and with known variance $\sigma_a^2 = 1$.

(a) What model is this? (2p)

(b) Compute ρ_1 for this model. (8p)

(c) Is the model stationary snd or invertible? (2p)

(d) Give a general comment on the autocorrelation function (ACF) $\rho_k, k = 1, 2, \ldots$ and the partial autocorrelation function (PACF) $\phi_{kk}, k = 1, 2, \ldots$ for this model. Explain using words or graphs. (4p)

(e) Rewrite the model using the backshift operator B. (4p)

5. Assume the model

$$y_t = a_t - \theta a_{t-1}$$

where $E(a_t) = 0$ and $\text{Var}(a_t) = \sigma^2$ and a_t are independent random variables.

(a) What kind of model is this? Is it invertible? (2p)

(b) Compute $E(y_t)$. (4p)

(c) Compute $\text{Var}(y_t)$. (4p)

(d) Compute the first and second order autocorrelation ρ_1 and ρ_2. (10p)

Appendix

Rules for the covariance

The *covariance* between X and Y is defined as

$$\mathrm{Cov}(X,Y) = E\{(X - E(X))(Y - E(Y))\}$$

The following formulas serves you well in this course.

(a) Relationship between the variance and the covariance

$$\mathrm{Cov}(X,X) = \mathrm{Var}(X).$$

(b) The order in which X and Y are mentioned in the covariance does not matter. We always have a symmetry:

$$\mathrm{Cov}(X,Y) = \mathrm{Cov}(Y,X).$$

(c) Multiplicative constants can be factored:

$$\mathrm{Cov}(aX,Y) = a\mathrm{Cov}(X,Y),$$

even

$$\mathrm{Cov}(aX,bY) = ab\mathrm{Cov}(X,Y).$$

(d) Additive constants has no effect on the covariance:

$$\mathrm{Cov}(X+a, Y+b) = \mathrm{Cov}(X,Y).$$

(e) The covariance between two sums

$$\mathrm{Cov}(X+Y, Z+W) = \mathrm{Cov}(X,Z)+\mathrm{Cov}(X,W)+\mathrm{Cov}(Y,Z)+\mathrm{Cov}(Y,W).$$

A special case of **(e)** is

$$\mathrm{Cov}(X, Z+W) = \mathrm{Cov}(X,Z) + \mathrm{Cov}(X,W).$$

(f) The covariance between two (different) linear combinations of X and Y; $aX+bY$ and $cX+dY$

$$\mathrm{Cov}(aX+bY, cX+dY) = ac\mathrm{Var}(X) + bd\mathrm{Var}(Y) + (ad+bc)\,\mathrm{Cov}(X,Y).$$

(g) Important special case of **(f)**: $V(aX+bY) = a^2V(X) + b^2V(Y) + 2ab\mathrm{Cov}(X,Y)$.

Department of Statistics

Stockholms universitet

Correction sheet

Date: 200116

Room: Uggleviksssalen

Exam: Econometrics II

Course: Econometrics

Anonymous code: 0032-TTM

☒ I authorise the anonymous posting of my exam, in whole or in part, on the department homepage as a sample student answer.

NOTE! ALSO WRITE ON THE BACK OF THE ANSWER SHEET

Mark answered questions

1	2	3	4	5	6	7	8	9	Total number of pages
X	X	X	X	X					5 &
Teacher's notes 18	20	15	20	20					

Points	Grade	Teacher's sign.
93	A	JSS

SU, DEPARTMENT OF STATISTICS

Room: Ugglevikssalen **Anonymous code:** 0032-TTM **Sheet number:** 1

1)

Nr	1	2	3	4	5	6	7	8	9	10
y	6	8	7	6	9	8	6	7	10	3

Slumpvis genererade

Antal $n=10$

$$\bar{y} = \frac{6+8+7+6+9+8+6+7+10+3}{10} = 7$$ R

a) $r_1 = ?$ $r_2 = ?$ (ACF)

$$r_k = \frac{c_k}{c_0}, \quad k = 0, 1, \ldots K$$

$$c_k = \frac{1}{n}\sum_{t=1}^{n-k}(y_t-\bar{y})(y_{t+k}-\bar{y}) \quad k = 0, 1, 2, \ldots K$$

$$c_0 = \frac{1}{10}\sum_{t=1}^{10}(y_t-\bar{y})(y_t-\bar{y}) = \frac{1}{10}\left((6-7)(6-7)+(8-7)(8-7)+\ldots+(3-7)(3-7)\right) =$$

$$= 3{,}4$$ R

$k=1$ $$c_1 = \frac{1}{10}\sum_{t=1}^{9}(y_t-\bar{y})(y_{t+1}-\bar{y}) = \frac{1}{10}\left((6-7)(8-7)+(8-7)(7-7)+\ldots+(10-7)(3-7)\right) =$$

$$= -1{,}4$$

$$r_1 = \frac{-1{,}4}{3{,}4} \approx -0{,}4118$$ R

$k=2$ $$c_2 = \frac{1}{n}\sum_{t=1}^{8}(y_t-\bar{y})(y_{t+2}-\bar{y}) = \frac{1}{10}\left((6-7)(7-7)+(8-7)(6-7)+\ldots+(7-7)(3-7)\right) =$$

$$= -0{,}\cancel{07}$$ -0.7 $+2p$

$$r_2 = \frac{-0{,}07}{3{,}4} \approx -0{,}0206$$ ✓

Svar: $r_1 = -0{,}4118$, $r_2 = -0{,}0206$ $\cdot 10$

b) $r_{11} = ?$, $r_{22} = ?$ (PACF)

$r_{11} = r_1$

$r_{22} = \dfrac{r_2 - r_{11} \cdot r_1}{1 - r_{11} r_1}$

$r_{11} = r_1 = -0{,}4118$ R

$r_{22} = \dfrac{-0{,}0206 - (-0{,}4118) \cdot (-0{,}4118)}{1 - (-0{,}4118)(-0{,}4118)} = -0{,}1626$ $\surd$

Svar $r_{11} = -0{,}4118$, $r_{22} = -0{,}1626$

c) r_1 är en estimator av ρ_1

r_2 — " — ρ_2

Testa: $\begin{cases} H_0: \rho_k = 0 \\ H_A: \rho_k \neq 0 \end{cases}$ $k = 1, 2$

Antag Gaussian vitt brus $\Rightarrow$ $r_k \sim N(0, \frac{1}{n})$

Teststatistika $Z = \dfrac{r_k}{\sqrt{\frac{1}{n}}} = r_k \sqrt{n}$ R

Förkasta H_0 för $|Z| > Z_{\alpha/2}$

Antag $\alpha = 0{,}05 \Rightarrow Z_{\alpha/2} = 1{,}96$

$k=1$ $|Z_1| = |-0{,}4118 \cdot \sqrt{10}| \approx 1{,}302$ $\Rightarrow$ ej förkasta H_0

$k=2$ $|Z_2| = |-0{,}0206 \cdot \sqrt{10}| \approx 0{,}0651$ $\Rightarrow$ ej förkasta H_0

Svar: Vi kan ej utesluta att ρ_1 och ρ_2 är lika med 0. R

SU, DEPARTMENT OF STATISTICS

Room: Uggleviksalen **Anonymous code:** 0032-TTM **Sheet number:** 2

2) Ölproduktion mellan Q1:1975 – Q4:1981

$\Rightarrow n = 28$

Durbin-Watson, $d = 1{,}822$

a) Hypotestesta om modellen är lämplig.

Durbin-Watson

$$\begin{cases} H_0: \phi = 0 \\ H_1: \phi > 0 \end{cases}$$

Från tabell A5 i Montgomery: $d_L = 1{,}14$ (vid $\alpha = 0{,}05$, $n = 30$) R

$d_U = 1{,}74$

$\Rightarrow d > d_U \Rightarrow$ kan inte förkasta H_0 R

Svar: Hypotestestet tyder på att modellen fungerar. Vi kan inte påvisa positiv autokorrelation i modellens feltermer.

b) Estimera r_1 med hjälp av d.

$d \approx 2(1 - r_1)$ (enligt ekv. 3,97 i Montgomery)

$\Rightarrow 1{,}822 = 2(1 - r_1) \Leftrightarrow r_1 = 1 - \frac{1{,}822}{2} = 0{,}089$

Svar: $r_1 = 0{,}089$ R

c)

1982	ARIMA	observation
Q1	47,0800	47,84
Q2	55,4728	54,27
Q3	54,4764	52,31
Q4	45,5577	42,03

Forecast för varje kvartal under 1982 och jämför. Antat time = 1 vid 1975:I

$prod = 32,98143 + 0,38795 \cdot time + 3,18098\, Q1 + 11,12018\, Q2 + 9,51938 \cdot Q3$

$prod_{1982I} = 32,98143 + 0,38795 \cdot 29 + 3,18098 \cdot 1 \approx 47,413$

$prod_{1982II} = 32,98143 + 0,38795 \cdot 30 + 11,12018 \cdot 1 \approx 55,7401$

$prod_{1982III} = 32,98143 + 0,38795 \cdot 31 + 9,51938 \cdot 1 \approx 54,5273$

$prod_{1982IV} = 32,98143 + 0,38795 \cdot 32 \approx 45,3958$

R

$MSE = \frac{1}{n} \sum_{t=1}^{n} (y_t - \hat{y}_t(t-1))^2$

$MSE_{REG} = \frac{1}{4} \sum_{t=1}^{4} \left[(47,84 - 47,413)^2 + (54,27 - 55,7401)^2 + (52,31 - 54,5273)^2 + (42,03 - 45,3958)^2 \right] \approx$

$\approx \underline{4,647}$ R

$MSE_{ARIMA} = \frac{1}{4} \sum_{t=1}^{4} \left[(47,84 - 47,08)^2 + (54,25 - 55,4728)^2 + (52,31 - 54,4764)^2 + (42,03 - 45,5573)^2 \right] =$

$\approx \underline{4,803}$ Ok

Svar: Regressionsmodellen fick något lägre MSE än ARIMA-modellen $MSE_{REG} = 4,647$, $MSE_{ARIMA} = 4,803$, så om vi skulle välja modell utefter MSE, så är regressionsmodellen att föredra.

SU, DEPARTMENT OF STATISTICS

Room: Uggleviksalen **Anonymous code:** 0032-TTM **Sheet number:** 3

3)

	0	1	2	3	4	5	6
	År	2015	2016	2017	2018	2019	
	Bullar	2043	2097	2245	2478	2673	

$$\begin{cases} L_t = \alpha y_t + (1-\alpha)(L_{t-1} + b_{t-1}) \\ b_t = \gamma (L_t - L_{t-1}) + (1-\gamma) b_{t-1} \end{cases}$$

$\alpha = 0,3$ $\gamma = 0,1$

a) Modell för att få fram startvärdena:

$\text{bullar} = \beta_0 + \text{tid}\,\beta_1$

OLS $\Rightarrow$ $\beta_1 = 164,1 = b_0$

$\beta_0 = 1814,9 = L_0$ R

$L_1 = 0,3 \cdot 2043 + 0,7 \cdot (1814,9 + 164,1) \approx 1998,2$

$b_1 = 0,1 \cdot (1998,2 - 1814,9) + 0,9 \cdot 164,1 \approx 166,02$

$L_2 = 0,3 \cdot 2097 + 0,7 (1998,2 + 166,02) \approx 2144,05$

$b_2 = 0,1 \; (2144,05 - 1998,2) + 0,9 \; 166,02 \approx 164,003$

$L_3 = 0,3 \cdot 2245 + 0,7 (2144,05 + 164,003) \approx 2289,14$

$b_3 = 0,1 (2289,14 - 2144,05) + 0,9 \cdot 164,003 \approx 162,112$

$L_4 = 0,3 \cdot 2478 + 0,7 \; (2289,14 + 162,112) \approx 2459,28$

$b_4 = 0,1 (2459,28 - 2289,14) + 0,9 \cdot 162,112 \approx 162,915$

$L_5 = 0,3 \cdot 2673 + 0,7 \; (2459,28 + 162,915) \approx 2637,44$

$b_5 = 0,1 \; (2637,44 - 2459,28) + 0,9 \cdot 162,915 \approx 164,44$

R

$\hat{y}_6 = \hat{y}_{5+1} = L_5 + 1 \cdot b_5 = 2637,44 + 164,44 \approx 2801,88$

Svar: Antal bullar uppskattas till 28 01|88 st

c) Svar: Jag skulle rekomendera den uträknade forcasten. Ok

SU, DEPARTMENT OF STATISTICS

Room: Uggleviksalen **Anonymous code:** 0032-TTM **Sheet number:** 4

4) $Y_t = 1{,}3\, Y_{t-1} - 0{,}4\, Y_{t-2} + \varepsilon_t$, ε_t oberoende

$\varepsilon_t \sim N(0,1)$, $E(\varepsilon_t) = 0$, $\sigma_a^2 = 1$

a) Det är en AR(2)-modell:

$Y_t = \delta + \phi_1 Y_{t-1} + \phi Y_{t-2} + \varepsilon_t$

$\delta = 0$

$\phi_1 = 1{,}3$

$\phi_2 = -0{,}4$

Svar: A(2)-modell R

b) Sökt ρ_1

Yule-Walker: $\rho(k) = \phi \rho(k-1) + \phi_2 \rho(k-2)$, $k = 1, 2, \ldots$

AR(2) => k=2

För $\rho_1 = \frac{\phi_1}{1-\phi_2} = \frac{1{,}3}{1-(-0{,}4)} \approx 0{,}9286$ R

Svar: $\rho_1 = 0{,}9286$

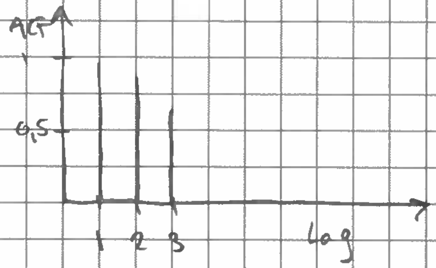
c) För att en AR(2)-modell ska vara stationär så krävs:

$$\begin{cases} \phi_1 + \phi_2 < 1 & \text{Test: } 1{,}3 + (-0{,}4) = 0{,}9 < 1 \quad (OK) \\ \phi_2 - \phi_1 < 1 & -0{,}4 - 1{,}3 = -1{,}7 < 1 \quad (OK) \\ |\phi_2| < 1 & |-0{,}4| = 0{,}4 < 1 \quad (OK) \end{cases}$$

=> Den är stationär. R

d) $\rho_2 = \phi_1 \rho_1 + \phi_2 = 1{,}3 \cdot 0{,}9286 + (-0{,}4) \approx 0{,}8071$

$\rho_3 = \phi_1 \rho_2 + \phi_2 \rho_1 = 1{,}3 \cdot 0{,}8071 + (-0{,}4) \cdot 0{,}9286 \approx 0{,}6779$



ACF är avtagande

e) Backshift operatorn B

$$\left.\begin{aligned} \phi(B)\, Y_t &= \delta + \varepsilon_t \\ \phi(B) &= 1 - \phi_1 B - \phi_2 B^2 \end{aligned}\right\} \Rightarrow (1 - \phi_1 B - \phi_2 B^2) Y_t = \delta + \varepsilon_t$$

$$\Rightarrow (1 - 1{,}3 B + 0{,}4 B^2)\, Y_t = \varepsilon_t$$

<u>Svar:</u> $(1 - 1{,}3 B + 0{,}4 B^2)\, Y_t = \varepsilon_t$

R

SU, DEPARTMENT OF STATISTICS

Room: Uggleviksalen Anonymous code: 0032-TTM Sheet number: 5

5) $Y_t = a_t - \theta a_{t-1}$

$E(a_t)=0$, $Var(a_t)=\sigma^2$, a_t - oberoende slumpvar.

a) MA(1) -modell : Invertibel om $|\theta|<1$

$Y_t = \mu + \varepsilon_t - \theta_1 \varepsilon_{t-1}$ θ är inte känd.

Men om θ uppfyller $-1<\theta<1$, då invertibel

Här: $\mu = 0$, $\theta_1 = \theta$

Svar: MA(1)-modell, invertibel om $|\theta|<1$

b) $E(y_t) = ?$

$E(y_t) = E(a_t - \theta a_{t-1}) = E(a_t) - \theta E(a_{t-1}) = 0 - \theta \cdot 0 = 0$

Svar: $E(y_t) = 0$ R

c) $Var(y_t) = ?$

$Var(y_t) = Var(a_t - \theta \cdot a_{t-1}) = Var(a_t) + \theta^2 Var(a_{t-1}) = \sigma^2 + \theta^2\sigma^2 =$
$= \sigma^2(1+\theta^2)$

Svar: $Var(y_t) = \sigma^2(1+\theta^2)$ R

d) $\rho_1 = ?$, $\rho_2 = ?$

För en MA(1)-modell så kommer autokorrelationen för $k>1$ (ρ_k) vara 0. $\Rightarrow \rho_2 = 0$

ρ_1 för en MA(1)-modell,

$\rho_1 = \frac{-\theta_1}{1+\theta_1^2}$, $\theta_1 = \theta \Rightarrow \rho_1 = \frac{-\theta}{1+\theta^2}$ R

Svar: $\rho_1 = \frac{-\theta}{1+\theta^2}$, $\rho_2 = 0$ R