Stockholm University

STOCKHOLM UNIVERSITY
Department of Statistics
Fall 2019, period A-B

Andriy Andreev (examiner)
Ulf Högnäs

FINANCIAL STATISTICS

2019-10-29

Time:	15.00 - 20.00
Place:	*Värtasalen*
Approved aid:	Hand-held calculator with no stored text, data or formulas
Provided aid:	*Formula Sheet and Probability Distribution Tables*, returned after the exam

- **Problems 1 – 4: MULTIPLE CHOICE QUESTIONS – max 38 points**
  - A total of four multiple choice questions with five alternative answers per question one of which is the correct answer. Mark your answers on the attached **answer form**.
  - Marking more than one alternative will result in zero points for that question.

- **Problems 5 – 6: COMPLETE WRITTEN SOLUTIONS – max 22 points**
  - Use only the provided **answer sheets** when submitting your solutions and answers.
  - For full marks, clear, comprehensive and well-motivated solutions are required. Unclear and un-explained solutions may result in point deductions even if the final answer is correct.
  - Check your calculations and solutions before submitting. Careless mistakes may result in unnecessary point deductions.

- The maximum number of points is stated for each question. The maximum total number of points is 38 + 22 = 60. At least 30 points is required to pass (grades A-E). The grading scale is as follows:

  A: 54 – 60 points
  B: 48 – 53 points
  C: 42 – 47 points
  D: 36 – 41 points
  E: 30 – 35 points
  Fx: 24 – 29 points
  F: 0 – 23 points

- NOTE! Fx and F are failing grades that require re-examination. Students who receive the grade Fx or F cannot supplement for a higher grade.

- Outlines of solutions will be posted on Mondo within several days after the exam.

1. (Multiple choice, 2 points + 4 points = 6 points, Multiple linear regression)

A real estate analyst developed a linear regression model for home values in King County, Washington (Seattle). Using a random sample of 1000 single-family homes, she estimated the model

$$Y = \beta_0 + \beta_1 \cdot x_1 + \beta_2 \cdot x_2 + \beta_3 \cdot x_3 + \beta_4 \cdot x_4 + \beta_5 \cdot x_5 + \beta_6 \cdot x_6 + \varepsilon$$

where Y is the value of the home in thousands of USD, x_1 is the number of bedrooms, x_2 is the number of bathrooms, x_3 is the living area measured in square feet, and x_4 is a dummy variable which takes the value 1 if the house borders a waterfront (lake or ocean front).

The variables x_5 and x_6 are dummy variables for the condition of the house. The variable x_5 takes the value 1 if the house has condition category "2" (good condition) and zero otherwise; x_6 takes the value 1 if the house has condition category "3" (excellent condition) and zero otherwise. Houses that do not belong to category "2" or "3" belong to category "1," which is the base category (poor or fair condition).

You can find the output here:

```
. . regress price bedrooms bathrooms sqft_living waterfront i.condition
```

Source	SS	df	MS			
				Number of obs	=	1,000
				F(6, 993)	=	225.87
Model	67510546.9	6	11251757.8	Prob > F	=	0.0000
Residual	49466168.9	993	49814.873	R-squared	=	0.5771
				Adj R-squared	=	0.5746
Total	116976716	999	117093.81	Root MSE	=	223.19

price	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
bedrooms	-67.66534	10.05724	-6.73	0.000	-87.40122	-47.92946
bathrooms	65.70212	14.41065	4.56	0.000	37.4233	93.98095
sqft_living	.265694	.0131683	20.18	0.000	.2398531	.2915348
waterfront	987.5802	113.5808	8.69	0.000	764.6943	1210.466
condition						
2	63.56032	16.24994	3.91	0.000	31.67217	95.44848
3	172.6911	31.87982	5.42	0.000	110.1315	235.2506
_cons	29.45056	29.91311	0.98	0.325	-29.24961	88.15072

a. (2 points) **Find the estimated expected value in thousands of dollars of a randomly selected home that has three bedrooms, two bathrooms, 1500 square feet living area, does not border a waterfront, and belongs to category "1" (poor or fair condition), according to the model.**

(A) 323

(B) 356

(C) 391

(D) 460

(E) 524

b. (4 points) The analyst wanted to find the Variance Inflation Factor of the variable x_4 (waterfront), so in addition to the first model, she ran a second model, the output of which can be found below.

```
. . regress waterfront bedrooms bathrooms sqft_living i.condition

      Source |       SS           df       MS      Number of obs   =     1,000
-------------+----------------------------------   F(5, 994)       =      6.31
       Model |   .1225576         5  .02451152   Prob > F        =    0.0000
    Residual |  3.8614424       994  .003884751   R-squared       =    0.0308
-------------+----------------------------------   Adj R-squared   =    0.0259
       Total |      3.984       999  .003987988   Root MSE        =    .06233

------------------------------------------------------------------------------
  waterfront |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    bedrooms |  -.0098864    .002791    -3.54   0.000    -.0153633   -.0044095
   bathrooms |   .0060017   .0040198     1.49   0.136    -.0018865    .0138899
 sqft_living |   .0000115   3.66e-06     3.13   0.002     4.29e-06    .0000187
             |
   condition |
          2  |   .0024301   .0045372     0.54   0.592    -.0064735    .0113338
          3  |  -.0041373   .0089017    -0.46   0.642    -.0216055    .0133309
             |
       _cons |   .0000435   .0083534     0.01   0.996    -.0163488    .0164358
------------------------------------------------------------------------------
```

What is the VIF of the variable waterfront in the original model? Choose the value closest to your answer.

(A) 0.0308

(B) 1.00

(C) 1.03

(D) 1.50

(E) 2.36

2. (Multiple choice, 2 points + 6 points = 8 points, Normal distribution)

An casual investor owns a portfolio consisting of three stocks. The weights are as follows: 50% of stock A, 30% of stock B, and 20% of stock C. Let U be the annual return of stock A, let V be the annual return of stock B, and let W be the annual return of stock C. The investor makes the following assumptions about the annual returns:

$$\begin{aligned}
\mathrm{E}[U] &= 0.10;\ \sigma_U = 0.20 \\
\mathrm{E}[V] &= 0.15;\ \sigma_V = 0.30 \\
\mathrm{E}[W] &= 0.20;\ \sigma_W = 0.50 \\
\rho_{U,V} &= 0.4 \\
\rho_{U,W} &= \rho_{V,W} = 0.
\end{aligned}$$

a. (2 points) Assume that all the returns are normally distributed. **Find the expected value of the portfolio.**

(A) 0.125
(B) 0.130
(C) 0.135
(D) 0.140
(E) 0.150

b. (6 points) Again, assume that all the returns are normally distributed. **Find the probability that that return of the portfolio is higher than** 0.2.

(A) 0.34
(B) 0.36
(C) 0.38
(D) 0.40
(E) 0.42

3. (Multiple choice, 6 points + 4 points = 10 points, logistic regression)

As part of their business research, a dating app offers "premium membership" for the price of 149 kr per year, to a randomly selected sample of 2800 of their active members. The business research team collects data and decide on the following logistic regression model:

$$y = \beta_0 + \beta_1 \cdot x_1 + \beta_2 \cdot x_2 + \beta_3 \cdot x_1 \cdot x_2.$$

y = the log-odds of buying premium membership.

x_1 = male (1 = yes, 0 = no).

x_2 = average hours spent per week on the app.

The third term is an interaction term between the variables "male" and "hours." Their output can be found here:

```
Call:
glm(formula = premium ~ male + hours + male * hours,
    family = binomial(link = "logit"), data = website)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-2.2910  -0.2513  -0.1813  -0.1499   3.0588

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept) -4.68826    0.22275 -21.047  < 2e-16 ***
male         0.85741    0.35641   2.406   0.0161 *
hours        0.33992    0.04240   8.017 1.09e-15 ***
male:hours   0.37752    0.08561   4.410 1.03e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 1076.08  on 2799  degrees of freedom
Residual deviance:  781.62  on 2796  degrees of freedom
AIC: 789.62

Number of Fisher Scoring iterations: 7
```

a. (6 points) **Find the estimated probability that a <u>woman</u> who spends 4 hours a week on the app will accept the offer to pay for premium membership.**

(A) 0.0346
(B) 0.0475
(C) 0.0583
(D) 0.0612
(E) 0.0701

b. (4 points) **Find the approximate number of hours spent on the app that would give a <u>male</u> member a 50% chance of buying premium membership, according to the model.** Round to the nearest hour.

(A) 1 hour
(B) 2 hours
(C) 3 hours
(D) 5 hours
(E) 8 hours

4. (Multiple choice. 2 points + 6 points + 6 points = 14 points. ARMA, ACF and PACF)

a. (2 points) Consider the following time series model.

$$X_t = 0.1 + \varepsilon_t$$

Assume that $E[\varepsilon_t] = 0$ and that $\mathrm{Var}(\varepsilon_t)$ is constant for all t, and that $\varepsilon_s, \varepsilon_t$ are independent for all s, t where $s \neq t$. **Find the correlation between X_{100} and X_{101}.**

(A) -1
(B) -0.1
(C) 0
(D) 0.1
(E) 1

b. (6 points) Consider the following time series model

$$Y_t = \varepsilon_t + 0.1 \cdot \varepsilon_{t-1} + 0.1 \cdot \varepsilon_{t-2}.$$

Assume that $E[\varepsilon_t] = 0$ and that $\mathrm{Var}(\varepsilon_t)$ is constant for all t, and that $\varepsilon_s, \varepsilon_t$ are independent for all s, t where $s \neq t$. **Find the correlation between Y_{12} and Y_{13}.** Choose the alternative closest to your answer. **You may turn in your calculations on a sheet for partial credit for this multiple choice problem.**

(A) 0.108
(B) 0.155
(C) 0.178
(D) 0.211
(E) 0.305

c. (6 points) Figure 1 shows the ACF and PACF plot of an ARMA(p, q) time series. Use what you have learned from this course to interpret the plots. **Which of these models best describes the time series that generated these plots?**

(A) $X_t = 0.4 \cdot X_{t-1} + 0.3 \cdot X_{t-2} + \varepsilon_t$
(B) $X_t = 0.4 \cdot X_{t-1} + \varepsilon_t$
(C) $X_t = \varepsilon_t + 0.4 \cdot \varepsilon_{t-1}$
(D) $X_t = \varepsilon_t + 0.4 \cdot \varepsilon_{t-1} + 0.3 \cdot \varepsilon_{t-2}$
(E) $X_t = 1 \cdot X_{t-1} + \varepsilon_t + 0.3 \cdot \varepsilon_{t-1}$.

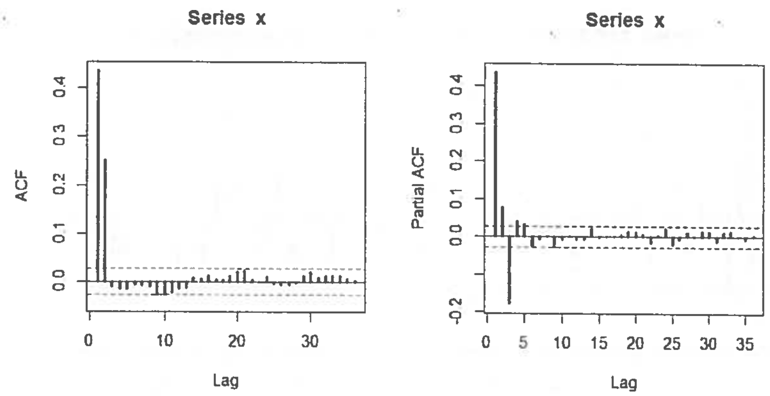

Figure 1: The ACF and PACF plots from the time series in question 4c.

5. (Essay type question, 2 points + 2 points + 6 points = 10 points, Box-Jenkins)

   a. (2 points) A statistician studied the weekly returns of the the spot price of Silver, in USD per Troy Ounce (31.1 grams). Let X_t be the Price at the end of week t and let Y_t be the return, then

   $$Y_t = \frac{X_t}{X_{t-1}} - 1.$$

   The statistician ran an Augmented Dickey Fuller test (with 10 lags); she chose a significance level $\alpha = 0.05$. The p-value of the test was 0.01. **State the hypotheses and interpret the outcome of the test.**

   b. (2 points) Use the ACF and PACF plots in figure 2 to form a first impression of the series. **Based on these plots and no other information, what kind of ARMA(p, q) do you think would match these plots and why? Indicate your choice of p and q.** Two sentences should be enough to motivate your choice.

   c. (6 points) Using STATA and three years of weekly data, the statistician estimated four different models. On the following pages, you can find the output (labeled MODEL 1 and so forth). **Based on lecture, explain in two to three sentences how you could use the information on these outputs to choose model. Mention at least two different statistics that you would take into account. What would your choice be and why?** If you are unsure, pick your two best candidates and discuss why you picked those.

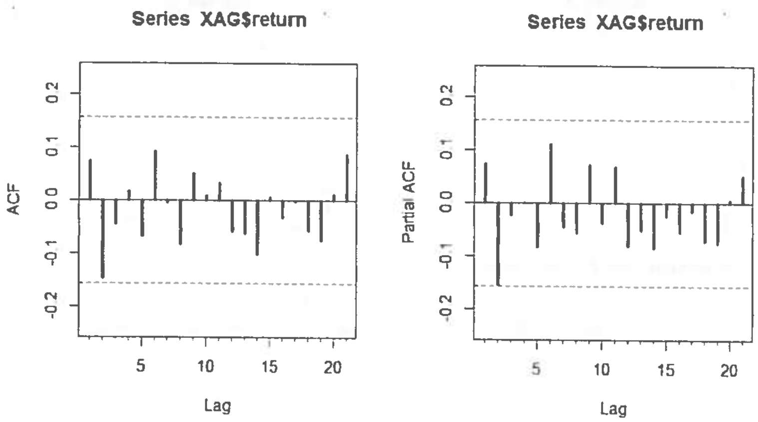

Figure 2: The ACF and PACF plots from the time series in question 5.

MODEL 1

ARIMA regression

Sample: 2 - 157

Number of obs = 156
Wald chi2(.) = .
Prob > chi2 = .

Log likelihood = 316.3159

D.return	Coef.	OPG Std. Err.	z	P>|z|	[95% Conf.	Interval]
return						
_cons	7.29e-06	.0025517	0.00	0.998	-.004994	.0050086
/sigma	.0318536	.0017707	17.99	0.000	.028383	.0353242

Note: The test of the variance against zero is one sided, and the two-sided confidence interval is truncated at zero.

. . estat ic

Akaike's information criterion and Bayesian information criterion

Model	Obs	ll(null)	ll(model)	df	AIC	BIC
.	156	.	316.3159	2	-628.6317	-622.532

Note: N=Obs used in calculating BIC; see [R] BIC note.

MODEL 2

ARIMA regression

Sample: 2 - 157	Number of obs	=	156
	Wald chi2(2)	=	0.88
Log likelihood = 362.7629	Prob > chi2	=	0.6445

D.return	Coef.	OPG Std. Err.	z	P>|z|	[95% Conf.	Interval]
return						
_cons	.0000359	.0000511	0.70	0.482	-.0000643	.0001361
ARMA						
ar						
L1.	.0747524	.0804167	0.93	0.353	-.0828614	.2323663
ma						
L1.	-1.000001	1036.655	-0.00	0.999	-2032.806	2030.806
/sigma	.0232822	12.06791	0.00	0.499	0	23.67595

Note: The test of the variance against zero is one sided, and the two-sided confidence interval is truncated at zero.

. . estat ic

Akaike's information criterion and Bayesian information criterion

Model	Obs	ll(null)	ll(model)	df	AIC	BIC
.	156	.	362.7629	4	-717.5258	-705.3263

Note: N=Obs used in calculating BIC; see [R] BIC note.

MODEL 3

```
ARIMA regression

Sample:  2 - 157                                Number of obs      =        156
                                                Wald chi2(1)       =       0.00
Log likelihood =  362.3265                      Prob > chi2        =     0.9991

------------------------------------------------------------------------------
             |                 OPG
    D.return |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
return       |
       _cons |   .0000359   .0000481     0.75   0.455    -.0000583    .0001302
-------------+----------------------------------------------------------------
ARMA         |
          ma |
         L1. |  -1.000001   869.4639    -0.00   0.999    -1705.118    1703.118
-------------+----------------------------------------------------------------
      /sigma |   .0233368   10.14534     0.00   0.499            0    19.90784
------------------------------------------------------------------------------
Note: The test of the variance against zero is one sided, and the two-sided
      confidence interval is truncated at zero.

. . estat ic

Akaike's information criterion and Bayesian information criterion

-----------------------------------------------------------------------------
       Model |        Obs  ll(null)  ll(model)      df         AIC        BIC
-------------+---------------------------------------------------------------
           . |        156         .   362.3265       3    -718.653  -709.5034
-----------------------------------------------------------------------------
               Note: N=Obs used in calculating BIC; see [R] BIC note.
```

MODEL 4

```
ARIMA regression

Sample:  2 - 157                                Number of obs      =       156
                                                Wald chi2(2)       =      0.88
Log likelihood =  362.7629                      Prob > chi2        =    0.6445

------------------------------------------------------------------------------
             |                 OPG
    D.return |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
return       |
       _cons |   .0000359   .0000511     0.70   0.482    -.0000643    .0001361
-------------+----------------------------------------------------------------
ARMA         |
          ar |
         L1. |   .0747524   .0804167     0.93   0.353    -.0828614    .2323663
             |
          ma |
         L1. |  -1.000001   1036.655    -0.00   0.999    -2032.806    2030.806
-------------+----------------------------------------------------------------
      /sigma |   .0232822   12.06791     0.00   0.499            0    23.67595
------------------------------------------------------------------------------
Note: The test of the variance against zero is one sided, and the two-sided
      confidence interval is truncated at zero.

. . estat ic

Akaike's information criterion and Bayesian information criterion

-----------------------------------------------------------------------------
       Model |    Obs    ll(null)   ll(model)     df          AIC         BIC
-------------+---------------------------------------------------------------
           . |    156           .    362.7629      4    -717.5258   -705.3263
-----------------------------------------------------------------------------
               Note:  N=Obs used in calculating BIC; see [R] BIC note.
```

6. (Essay type question. 2 points + 4 points + 6 points = 12 points, GARCH)

A financial statistics student analyzed the weekly price development of the cryptocurrency *Mendacium*, measured in USD, over the two-year year period ending on October 25, 2019 (see Figure 4). To this end, she modelled the *logarithmic returns* of the price. If X_t is the closing price of one Mendacium at the end of week t and Y_t is the logarithmic return then,

$$Y_t = \log\left(\frac{X_t}{X_{t-1}}\right)$$

by definition. After some testing and analysis, she decided on the following GARCH(1, 1) model

$$\begin{aligned} Y_t &= \beta_0 + \varepsilon_t \\ h_t &= a_0 + a_1\varepsilon_{t-1}^2 + b_1 h_{t-1} \end{aligned}$$

where $\mathrm{Var}(\varepsilon_t) = h_t$ and $\mathrm{E}(\varepsilon_t) = 0$. You can find the rounded estimates of the GARCH coefficients below, along with the last five weeks of data.

a0	a1	b1
9.2e-05	0.095	0.90

week	Close	logreturn	h.hat
99	3.4882	-0.0825	0.0318
100	4.1154	0.1653	0.0288
101	3.8037	-0.0788	0.0318
102	3.7880	-0.0041	0.0344
103	3.0219	-0.2259	0.0316
104	3.0177	-0.0014	0.0332

a. (2 points) The analyst estimated the average logarithmic return and got $\hat{\beta} = 0.0106$. **Find a forecast of the price of Mendacium at week 105.**

b. (4 points) **Find the forecast of the variance of the logarithmic return Y_t for week 105.** Tip: you need to use the information given in 6a to solve this.

c. (6 points) **Find the forecast of the variance of the logarithmic return Y_t for week 106.** Tip: you need to use the information given in 6a and your results from 6b to solve this.

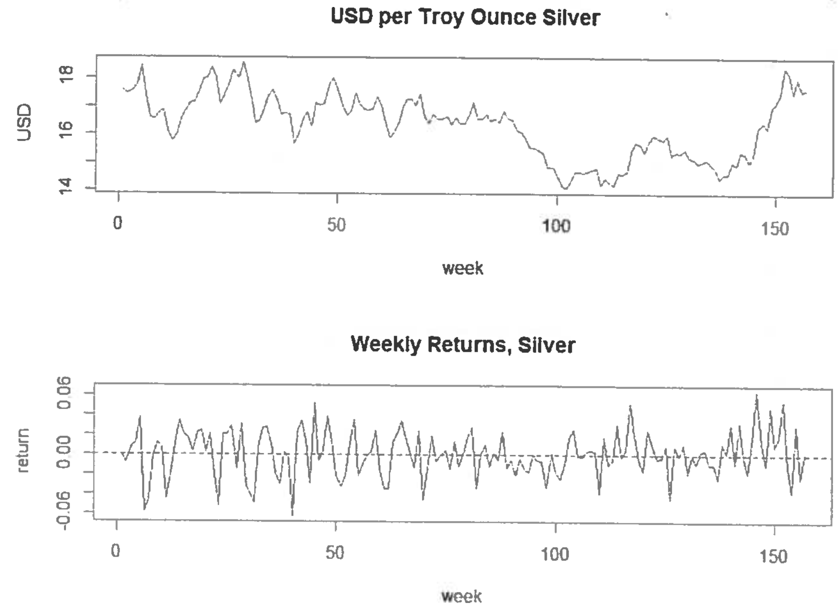

Figure 3: The spot price and weekly returns of Silver from Question 5.

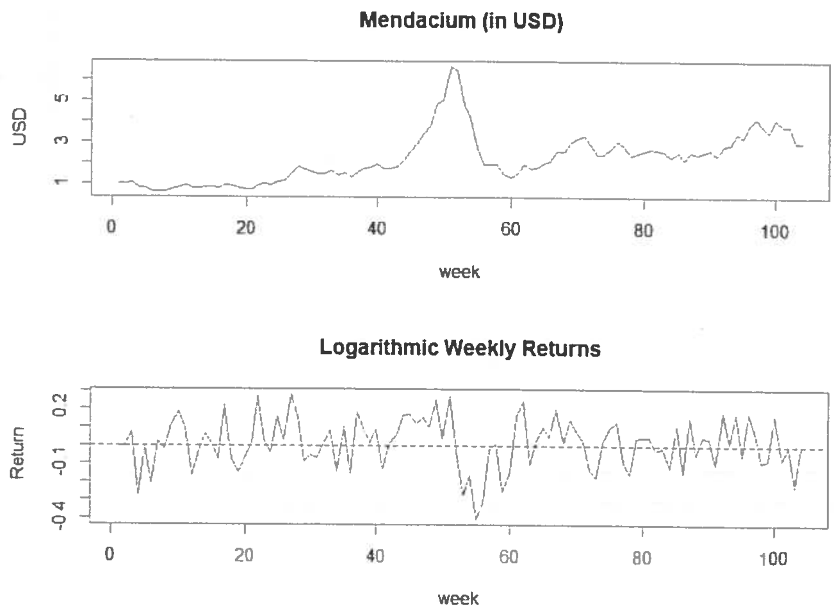

Figure 4: The price and weekly logarithmic returns of the (totally made-up) cryptocurrency Mendacium.

Department of Statistics

Stockholms universitet

Correction sheet

Date: 29/10 - 2019

Room: Värtasalen

Exam: Financial Statistics

Course: Financial Statistics

Anonymous code: 0025-KOA

☒ 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	X				3
Teacher's notes 6	2	10	12	9	12				

Points	Grade	Teacher's sign.
51	B	AA

STOCKHOLM UNIVERSITY
Department of Statistics

Fall 2019 A-B

ANSWER FORM Exam – Financial Statistics
2019-10-29

Anonymous code: 0025 - KOA (*write clearly!*)

Mark your answers with a clear cross (X) in the corresponding boxes below.

NOTE! Only one cross per question. If more than one alternative has been marked, zero points will be awarded for that question.

NOTE! If, after checking your calculations properly, you are convinced that the correct answer is not included among the given alternatives, write your answer in the margin to the right and explain you reasoning on the back.

	A	B	C	D	E		
1a.		X				(2)	
1b.			X			(4)	
2a.			X			(2)	
2b.	X					(6)	-6
3a.	X					(6)	
3b.				X		(4)	
4a.					X	(2)	-2
4b.	X					(6)	
4c.				X		(6)	

30/38

SU, DEPARTMENT OF STATISTICS

Room: Värtasalen Anonymous code: 0025-KOA Sheet number: 1

5.

A. $Y_t = \frac{X_t}{X_{t-1}} - 1$

X_t = Price at end of week

Y_t - return

Dickey fuller

$H_0: a_1 = 1$ non-stationary

$H_1: a_1 < 1$ stationary

Since our p-value is less than our crit value
p-value 0.01 < crit value 0.05 we reject the
null. This is stationary.

2

B. What kind of ARMA?

This is an ARMA(0,0) since none of the lags go outside the confidence interval. none of the lags are large enough so it is only noise.

C. In choosing the best model you can use AIC/BIC and the coefficients p-values, and the 95% confidence interval

AIC

M1 -628.63
M2 -717.5258
M3 -718.653 ← best AIC
M4 -717.5258

2

Model 3 has best AIC but bad p-values on its coefficent but when looking at the other 3 models they also have bad p-values. So in conclusion I choose model 3.

5

SU, DEPARTMENT OF STATISTICS

Room: Värtasalen Anonymous code: 0025-KOA Sheet number: 2

6.

$$Y_t = \log\left(\frac{X_t}{X_{t-1}}\right)$$

a) $\hat{\beta} = 0.0106 \Rightarrow Y_t = 0.0106$

$$0.0106 = \log\left(\frac{X}{3.0177}\right)$$

$$\log(x) = 0.0106$$

$$\log(1.0106) \approx 0.0106$$

$$\frac{X}{3.0177} = 1.0106$$

$$X = 3.04968762$$

answer: $3.04968762 = X_{105}$

2

b) $h_{105} = a_0 + a_1 \varepsilon_{104}^2 + b_1 h_{104}$

$h_{104} = 0.0332$

$a_0 = 9.2 \cdot 10^{-5}$

$a_1 = 0.095$

$b_1 = 0.90$

$\varepsilon_{104} = -0.012$

$$\hat{h}_{105} = 9.2 \cdot 10^{-5} + 0.095 \cdot -0.012^2 + 0.90 \cdot 0.0332$$

$$\hat{h}_{105} = 0.02998568$$

4

~~$0.0332 = 9.2 \cdot 10^{-5} + 0.095 \cdot \varepsilon_{103}^2 + 0.9 \cdot 0.0316$~~

~~$\varepsilon_{103} = 0.22166631552$~~

$$-0.0014 = 0.0106 + \varepsilon_{104} \rightarrow Y_{104} = \hat{\beta} + \varepsilon_{104}$$

$$-0.012 = \varepsilon_{104}$$

Turn →

c)

$\hat{h}_{106} = a_0 + a_1 \varepsilon_{105}^2 + b_1 h_{105}$

$\hat{h}_{106} = 9{,}2 \cdot 10^{-5} + 0.095 \cdot 0^2 + 0{,}90 \cdot 0{,}02998568 = 0.027079112$

$Y_{105} = 0.0106 + \varepsilon_{105}$

$0 = \varepsilon_{105}$

$\hat{h}_{106} = 0.027079112$

(|)