Econometrics II

Written examination

Thursday August 15, 2019, 10:00 - 15:00

Examiner: Andreas Rosenblad, Department of Statistics, Stockholm University

Instructions

Allowed tools:

- Pocket calculator
- Text book: Wooldridge, J.M. *Introductory Econometrics: A Modern Approach.* Cengage Learning, Boston.
- Text book: Montgomery, D.C., Jennings, C.L., and Kulachi, M. *Introduction to Time Series Analysis and Forecasting.* John Wiley, New Jersey.
- Notes written in the text books 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 subproblem is given in the right margin. 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.

Solutions to the exam questions will be uploaded to Athena after the exam is finished. The corrected exams will be available at the student office of the Department of Statistics within three weeks after the date of the exam.

Question 1 (24 points)

The R package **wooldridge** contains the data set PHILLIPS, which gives data on unemployment and inflation rates in the US during the years 1948-2003. We are interested in estimating the regression model

$$unem_t = \beta_0 + \beta_1 unem_{t-1} + \beta_2 inf_{t-1} + u_t, \quad t = 1948, 1949, ..., 1999$$

where $unem$ gives the unemployment rate in percent and inf gives the inflation rate, defined as the percentage change in consumer price index (CPI). The error terms u_t are assumed to fulfill the usual requirements for OLS in time series regression. The R code and parts of the output for estimating this model are given below, together with the values of $unem$ and inf for the years 1948-1952 and 1997-2001.

```
> library(wooldridge)
> library(dynlm)
> phillips.ts <- ts(phillips, start = 1948)
> out.phillips <- dynlm(unem ~ L(unem) + L(inf), data = phillips.ts,
end = 1999)
> summary(out.phillips)

Time series regression with "ts" data:
Start = 1949, End = 1999

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  1.23810    0.46693   2.652   0.0108 *
L(unem)      0.65234    0.08115   8.038 1.93e-10 ***
L(inf)       0.18676    0.03984   4.688 2.31e-05 ***

> phillips.ts[,1:3]
     year unem  inf
1948 1948  3.8  8.1
1949 1949  5.9 -1.2
1950 1950  5.3  1.3
1951 1951  3.3  7.9
1952 1952  3.0  1.9
[...]
1997 1997  4.9  2.3
1998 1998  4.5  1.6
1999 1999  4.2  2.2
2000 2000  4.0  3.4
2001 2001  4.8  2.8
```

(a) Calculate the one-step-ahead forecast of the US unemployment rate made in 1999. What is the forecast error of this one-step-ahead forecast? (8)

(b) Calculate the mean absolute error (MAE) and the mean absolute percent forecast error (MAPE) of the one-step-ahead forecast from the previuos question. (8)

(c) We are interested in estimating the unemployment time series $unem_t$ using simple exponential smoothing (SES) with discount factor $\lambda = 0.5$. With $y_t = unem_t$, calculate the smoothed (i.e. local mean) values $\tilde{y}_1$, $\tilde{y}_2$, and $\tilde{y}_3$ using the initial value $\tilde{y}_0 = y_1$. (8)

Question 2 (18 points)
A stationary times series process y_t is given by the equation

$$y_t = 3 + \varepsilon_t + \varepsilon_{t-1} + 0.8\varepsilon_{t-2} + 0.4\varepsilon_{t-2} + 2 + 1.4\varepsilon_{t-3}$$

where ε_t is a white noise process.

(a) What is this time series process called? (4)

(b) Calculate $E(y_t)$. (4)

(c) Assuming that $\sigma_\varepsilon^2 = \sigma^2 = \mu$, calculate $Var(y_t)$. (6)

(d) Assuming that $\sigma_\varepsilon^2 = \sigma^2 = \mu$, calculate $\gamma_y(4)$. (4)

Question 3 (16 points)
A stationary times series process y_t is given by the equation

$$y_t = 4 + 0.8y_{t-1} - 0.5y_{t-2} + \varepsilon_t$$

where ε_t is a white noise process.

(a) What is this time series process called? (4)

(b) Show that y_t is a stationary process. (6)

(c) Calculate $E(y_t)$. (6)

Question 4 (6 points)
Express the time series process

$$y_t = \varepsilon_t - \theta_2\varepsilon_{t-2} - \theta_4\varepsilon_{t-4}$$

in terms of the backshift operator B.

Question 5 (36 points)
Give the correct answer for the following multiple-choice questions. No motivation is needed.

(a) Which of the following four conditions is necessary for calculating the autocorrelation function (ACF) for a time series? (3)

A. The time series has to be stationary.
B. The time series has to be normally distributed.
C. The time series has to be persistent
D. The time series has to be symmetric.

(b) Which of the following four statements is true for a time series that is stationary, weakly dependent, and has serial correlation? (3)

A. $\bar{R}^2$ is an inconsistent while R^2 is a consistent estimator of the population parameter.

B. $\bar{R}^2$ is a consistent while R^2 is an inconsistent estimator of the population parameter.

C. Both $\bar{R}^2$ and R^2 are inconsistent estimators of the population parameter.

D. Both $\bar{R}^2$ and R^2 are consistent estimators of the population parameter.

(c) For time series data on two variables y and z, where y_t and z_t are dated contemporaneously, what is the model (3)

$$y_t = \alpha_0 + \beta_0 z_t + \beta_1 z_{t-1} + \beta_2 z_{t-2} + \beta_3 z_{t-3} + u_t, \quad t = 1, 2, ..., n$$

called?

A. An infinite distributed lag model of order 3.

B. An infinite distributed lag model of order 4.

C. A finite distributed lag model of order 3.

D. A finite distributed lag model of order 4.

(d) What is the process represented by the model (3)

$$y_t = y_{t-1} + u_t, \quad t = 1, 2, ..., n$$

called?

A. An MA(1) process.

B. An AR(2) process.

C. A random walk process.

D. A random walk with drift process.

(e) Which of the following is an advantage of using panel data? (3)

A. The dependent variable y can be differenced across time for different cross-sectional units.

B. The dependent variable y can be added across time for different cross-sectional units.

C. The dependent variable y can be differenced across time for the same cross-sectional units.

D. The dependent variable y can be added across time for the same cross-sectional units.

(f) Which of the following four methods is not a variable selection method? (3)

A. Forward selection.

B. Weighted least squares regression.

C. Backward elimination.

D. All possible regressions.

(g) What will happen with the smoothed values obtained from a first-order exponential smoother when the discount factor $\lambda \to 1$? (3)

A. They will be more smoothed.

B. They will be less smoothed.

C. They will approach a standard normal distribution.

D. They will approach a Student's t distribution.

(h) Which of the following four statements about spurious regressions is true? (3)

A. The OLS estimates of the population parameters are efficient and unbiased and the t statistic is valid.

B. R^2 may be large even if the explanatory variables and the dependent variable are independent times series processes.

C. Spurious regressions are limited to $I(0)$ processes, and are not possible for $I(1)$ processes.

D. Spurious regressions are limited to $I(1)$ processes, and are not possible for $I(0)$ processes.

(i) If the forecast lead time is always of the same length and the forecast is revised each time period, what is the employed method called? (3)

A. Prediction interval.

B. Forecast horizon.

C. Forecast interval.

D. Moving horizon.

(j) Which of the following is not a problem with judgment forecasts? (3)

A. They are often consistent.

B. They are often too optimistic.

C. The are known for underestimating the variability.

D. They give too much attention to recent events.

(k) For a white noise error term ε_t, what is the time series process y_t given by the equation (3)

$$y_t = 0.50y_{t-1} - 0.15y_{t-2} + 0.25y_{t-1} + \varepsilon_t$$

called?

A. An $ARIMA(2,0,0)$ process.

B. An $AR(3)$ process.

C. An $ARIMA(0,0,3)$ process.

D. An $ARMA(1,1)$ process.

(l) Which of the following types of variables cannot be included in fixed effects models? (3)

A. Dummy variables.

B. Discrete dependent variables.

C. Time-varying independent variables.

D. Time-constant independent variables.

Department of Statistics

Stockholms universitet

Correction sheet

Date: 2019-08-15

Room: Uggleviksalen

Exam: Econometrics II

Course: Econometrics

Anonymous code: 0021-XZA

☒ 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					3 TF
Teacher's notes 24	18	16	6	33					

Points	Grade	Teacher's sign.
97	A	AR

SU, DEPARTMENT OF STATISTICS

Room: Ugglevikssalen Anonymous code: 0021-XZA Sheet number: 113.

Question 1.

a) The one-step-ahead forecast of $unem_{2000}$ is given by:

$$\hat{f}_{99} = \hat{\beta}_0 + \hat{\beta}_1 \, unem_{99} + \hat{\beta}_2 \, inf_{99}$$
$$= 1{,}23810 + 0{,}65234 \times 4{,}2 + 0{,}18676 \times 2{,}2$$
$$= \underline{4{,}3888}$$

The forecast error is given by:

$$\hat{e}_{2000} = unem_{2000} - \hat{f}_{99}$$
$$= 4{,}0 - 4{,}3888$$
$$= \underline{-0{,}3888}.$$

/ 8

b) The mean absolute error is given by:

$$MAE = \frac{1}{n} \sum_{t=1}^{n} |e_t(1)|$$

where n is the number of observations for which a forecast has been made and $\sum |e_t(1)|$ is the sum of the forecast errors given in absolute value. Since we have only one forecast,

$$MAE = |\hat{e}_{2000}|$$
$$= \underline{0{,}3888}$$

The mean absolute percent forecast error is given by:

$$MAPE = \frac{1}{n} \sum_{t=1}^{n} |re_t(1)| \quad \text{where} \quad re_t(1) = \frac{e_t(1)}{y_t} \times 100.$$

⟹

For the forecast error we have calculated above, we get:

$$re_t(1) = \frac{-0,3888}{4,0} \times 100$$

$$= -9,72.$$

Since we have only one forecast, $n=1$ and:

MAPE = 9,72 %.

/8

c. $\tilde{y}_0 = y_1 \Leftrightarrow \tilde{y}_0 = 3,8$. We also know that $\lambda = 0,5$.

The simple exponential smoother is given by:

$$\tilde{y}_t = \lambda y_t + (1-\lambda)\tilde{y}_{t-1}$$

Hence, for $\tilde{y}_1$, $\tilde{y}_2$ and $\tilde{y}_3$, we get respectively:

$$\tilde{y}_1 = 0,5 \times y_1 + 0,5 \times \tilde{y}_0 = 0,5 \times 3,8 + 0,5 \times 3,8 = 3,8.$$

$$\tilde{y}_2 = 0,5 \times y_2 + 0,5 \times \tilde{y}_1 = 0,5 \times 5,9 + 0,5 \times 3,8 = 4,85.$$

$$\tilde{y}_3 = 0,5 \times y_3 + 0,5 \times \tilde{y}_2 = 0,5 \times 5,3 + 0,5 \times 4,85 = 5,075.$$

/8

Question 2.

a. This is a third-order moving average process, i.e an MA(3) process.

/4

b. $E(y_t) = \mu$

$= 3+2.$

$= 5.$

/4

c. The variance of y_t is given by

$$Var(y_t) = \sigma^2(1 + \theta_1^2 + \theta_2^2 + \theta_3^2).$$

SU, DEPARTMENT OF STATISTICS

Room: Ugglevikssalen Anonymous code: 0021-XZA Sheet number: 2/3

(Q2 ctd).

Hence, assuming $\sigma^2 = \mu = 5$, we get:

$Var(y_t) = 5 \times (1 + 1^2 + (0,8+0,4)^2 + 1,4^2)$.

$= 27.$

6

d. $\gamma_y(4) = 0$ because $q = 3$ and $\gamma_y(k) = 0$ for all $k > q$.

4

Question 3.

a. This is a second-order autoregressive process, i.e an AR(2) process.

4

b. An AR(2) process is stationary if the following condition holds:

$\varphi_1 + \varphi_2 < 1$.

$\varphi_2 - \varphi_1 < 1$

$|\varphi_2| < 1$

Applying this condition to the model, we get:

$0,8 - 0,5 = 0,3 < 1$.

$-0,5 - 0,8 = -1,3 < 1$.

$|-0,5| = 0,5 < 1$.

Conclusion: the process is stationary.

6

c. For an AR(2) process, E(y) is given by:

$$E(y_t) = \frac{\delta}{1 - \varphi_1 - \varphi_2}$$

→

Hence, we get:

$$E(y_t) = \frac{4}{1-0{,}8+0{,}5}$$

$$= \frac{4}{0{,}7}$$

$$\approx 5{,}71429.$$

/6

Question 4.

The MA(4) process $y_t = \varepsilon_t - \theta_2 \varepsilon_{t-2} - \theta_4 \varepsilon_{t-4}$ can be expressed in terms of the backshift operator as:

$$y_t = \mu + (1 - \theta_2 B^2 - \theta_4 B^4)\varepsilon_t$$

where $\mu = 0$ and $\{\varepsilon_t\}$ is white noise.

/6

(Q5 next page).

SU, DEPARTMENT OF STATISTICS

Room: Ugglevikssalen Anonymous code: 0021-XZA Sheet number: 3/3.

Question 5

a. A. /3
b. D. /3
c. C. /3
d. C. /3
e. C. /3
f. B. /3
g. B. /3
h. B. /3
i. D /3
j. A /3
k. A /3
l. B. /0

/33