STOCKHOLMS UNIVERSITET
Statistiska institutionen
Oskar Gustafsson

WRITTEN EXAM, ECONOMETRICS II
2022-01-13

Time for examination: 8.00-13.00

Allowed tools: Pocket calculator, own formula sheet (1 double-sided A4 page), Course textbooks: Wooldridge, J.M. *Introductory Econometrics - a Modern Approach (any edition)* and Montgomery, Jennings and Kulachi - *Introduction to time series and forecasting.*
Note that no formula sheet will be provided.

The exam consists of 5 independent problems. Well motivated and clear solutions are required for full scoring on a problem. Don't forget to state any necessary assumptions or conditions where needed.
Passing rate: 50% of overall total, which is 100 points. For detailed grading criteria, see the course description. Answers may be given in English or Swedish.

Good luck!

Problem 1. (20 points)
Indicate which alternative that is correct. Answering more than one alternative result in 0 points on the sub-question. No motivation is required.

1. The following model: $(1 - 0.9B + 0.2B^3)y_t = (1 - 0.3B^2)\varepsilon_t$ is called:
   (a) an $ARMA(2, 1)$.
   (b) an $ARMA(3, 2)$.
   (c) an $ARIMA(2, 1, 1)$.
   (d) an $ARMA(2, 3)$.

2. The Granger causality test is used to:
   (a) determine the number of lags to be used in the model.
   (b) test if one time series has a causal (ceteris paribus) effect on another.
   (c) test if one time series can be used to predict future values of the other.
   (d) test for cointegration.

3. If a time series y_t is weakly stationary, which of the following statements is **not true** regarding the autocovariance function $cov(y_t, y_{t-k})$.
   (a) It equals $cov(y_t, y_{t+k})$.
   (b) It is independent of k.
   (c) It is independent of t.
   (d) It is equals zero for an $MA(k - 1)$-process.

4. Dynamical completeness:
   (a) is crucial for estimating causal effects.
   (b) together with weak stationarity and weak dependence implies strict exogeneity.
   (c) ensure that the OLS estimators are asymptotically normal (if we add TS.1'-TS.3')
   (d) implies that adding more lagged values of either $\mathbf{x}$ or y will not improve the forecasting performance.

5. What is true when using panel data methods?
   (a) Fixed effects estimation is more efficient than first difference when we have auto-correlated data.
   (b) Fixed effects = first difference when $T = 3$.
   (c) Random effects should be used when the unobserved variable is uncorrelated with the observed variables.
   (d) Fixed effects model takes into account omitted variables that vary both over time and entities.

Problem 2. (20 points)

A gaming studio that sell a certain type of computer games model the number of sold games according to:

$$y_t = \beta_0 + \beta_1 x_t + \beta_2 x_{t-1} + \beta_3 x_{t-2} + u_t, \tag{1}$$

where x_t denotes the number of visits to a certain gaming website in thousands. The following parameters has been estimated

$$\hat{y}_t = 10 + 5x_t + 10x_{t-1} + 2x_{t-2}, \tag{2}$$

Questions:

a) What type of model do we have in (1)?

b) In a lock-down during the recent pandemic, the number of visits to the website has increased by 1000 unique users. Assuming that the increase is only temporary, what is the transient effect on the number of sold games? Make a picture that show the dynamics over time.

c) Another idea is that the effect is "sticky", and that the number of visits has permanently increased. Calculate and interpret the long-run propensity in the number of sold games.

d) You are interested in testing if the long-run propensity is greater than 16. Rewrite the model in a way which let you test this hypothesis.

e) What assumptions need to be fulfilled for your test in (d) to be valid?

Problem 3. (15 points)

(a) Briefly explain the following concepts:

1. Weak stationarity.
2. Autocorrelation.
3. White noise.

(b) Write down the following models both with and without the back-shift operator

1. Random walk with drift.
2. ARMA(3,1).
3. ARIMA(1,1,1).

(c) Check which of the following models that are stationary and/or invertible. Also, determine p, d, and q in the ARIMA(p,d,q) notation.

1. $y_t = -4 + \varepsilon_t + 0.8\varepsilon_{t-1}$
2. $y_t = 5 + y_{t-1} + \varepsilon_t$
3. $y_t = 12 + 0.6y_{t-1} + 0.5y_{t-2}$

Problem 4. (20 points)

Consider the AR(2) model: $y_t = 4 + 1.2y_{t-1} - 0.4y_{t-2} + \varepsilon_t$, where $\varepsilon \sim N(0, 2)$

1. Check whether y_t is stationary.
2. Calculate $E[y_t]$.
3. Calculate $\rho_y(k)$ for $k = 1, 2$.
4. Assume we have observed some observations:

t	y
1	18
$\vdots$	$\vdots$
97	24
98	23
99	17

Use the model to compute the point forecasts for $t = 100, 101,$ and 102.

Problem 5. (25 points)

In the years 2003 to 2008 the number of goals scored by Zlatan Ibrahimovic in league games turned out as in the table below.

Year	Number of Goals
2003	3
2004	16
2005	7
2006	15
2007	17
2008	25

[**note:** both smoothed and forecasted number of goals does not have to be whole numbers]

a) Perform both simple exponential smoothing and double exponential smoothing on the data. Use 3 as your starting value and set the smoothing parameter to $\lambda = 0.2$.

b) Which method do you think is most suitable? Motivate.

c) Forecasts the number of goals scored by Zlatan in $2009 - 2012$ using both methods.

d) The number of scored goals turned out to be: $16, 14, 28, 30$. Calculate the *mean error* (ME) for the forecasts of both models. Was the result expected? Why?

e) Compute the *mean absolute error* (MAE) and *root mean squared error* RMSE for the forecasts, which method was more accurate?