STOCKHOLMS UNIVERSITET
Statistiska institutionen
Oskar Gustafsson

REXAM, ECONOMETRICS II
2022-02-11

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 4 (mostly) 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. (25 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: $y_t = 10 + 0.8y_{t-1} - 0.3y_{t-2} + \varepsilon_t - 0.6\varepsilon_{t-2}$ is called:
   (a) an $ARMA(1,1)$.
   (b) an $ARMA(1,2)$.
   (c) an $ARIMA(2,1,2)$.
   (d) an $ARMA(2,2)$.

2. What is the main issue of regressing two variables with unit roots ($I(1)$ variables) on each other?
   (a) We may get problems if they are cointegrated.
   (b) There are no problems with this.
   (c) We will get a very low R^2.
   (d) We may get strong spurious results that makes no sense.

3. The third first difference ($\Delta^3 y_t = (1-B)^3 y_t$) is given by:
   (a) $y_t - y_{t-3}$.
   (b) $y_t - 3y_{t-1} - y_{t-2} + y_{t-3}$.
   (c) $y_t - 3y_{t-1} + 3y_{t-2} - y_{t-3}$.
   (d) none of the above.

4. What is **not** true regarding pooling of cross-sections over time:
   (a) We can estimate the regression by weighted least squares.
   (b) The resulting coefficients can be interpreted in the same way as for multiple linear regression.
   (c) We can take into account that the relation between the dependent variable and the independent variables my change over time.
   (d) We can control for entity specific unobserved factors that are constant over time.

5. If you estimate an ARMA model to a quarterly time series, y_t, and find that your residuals are stationary but that $\hat{\rho}_y(4)$, and $\hat{\rho}_y(8)$ are high and statistically significant you should probably:
   (a) Take the first difference of your data.
   (b) Try with an ARCH or a GARCH model instead.
   (c) Add more lags to the model.
   (d) Try with a seasonal ARIMA model instead.

Problem 2. (25 points)
Suppose you are interested in how a particular policy change that limit the number of rental kick-scooters ("elspakcyklar") affect the number of injuries per 100 000 people caused by dangerously parked scooters on the streets. Suppose that we have data on the number of injuries over several time periods both in Stockholm and in Gothenburg before and after the time point T_{change} at which Stockholm placed a restriction on the number of allowed bikes, while Gothenburg did not. Your goal is to estimate the effect that the policy change has on the number of injuries using the difference in difference model. This may be used as decision basis for whether the policy change should be kept or not.

Questions:

a) Do you think this is a good idea? Give a short motivation to your answer.

b) What is the key assumption that has to hold in order for the approach to successfully estimate the causal effect of the policy change? Do you think it is a reasonable assumption?

c) Write up the model in terms of a regression model.

d) Give a short interpretation on each coefficient in c).

e) Given that all necessary assumptions are fulfilled, how would you test if the policy change significantly reduce on the number of injuries? Write up H_0, H_a, and the test statistic.

Problem 3. (25 points)

Consider the data set

Time	y_t
1	2
2	7
3	14
4	16
5	13
6	17
7	17
8	30
9	36
10	38
11	42
12	44
13	38
14	36

Table 1: Simulated Data

a) Calculate the first difference of the data in Table 1 for the first 10 periods. Call the new differenced data x_t.

b) Use the first differenced data to calculate the fitted values using the ARMA-model: $x_t = 2 + 0.5x_{t-1} + \varepsilon_t + 0.5\varepsilon_{t-1}$.

c) Reverse the fist difference by calculating $\hat{y}_t = y_2 + \sum_{k=0}^{t-3} \hat{x}_{t-k}$ for time periods $t = 3, \ldots 10$. This gives you the fitted values for the time series in levels.
[*In case you are not able to do this, continue with the differenced data, x_t, from here.*]

d) Calculate the model residuals $\hat{e}_t = y_t - \hat{y}_t$ and calculate the residual variance.

e) Calculate the autocorrelation on lag length 1 of the residuals. What does your result suggest regarding the model choice?

f) Calculate forecasts for $\hat{y}_t$ for $t = 11, 12, 13$, and 14. Also compute the $RMSE$.
[*Hint: this can be done by first making forecasts on x_t and then do the same thing as in (c) but using y_{10} instead of y_2.*]

Problem 4. (25 points)

Consider the same data as in 3.

a) Use Holt's method to smooth first 10 observations of the (non-differenced) y-data in Table 1 (from Problem 3) with starting values $L_0 = T_0 = 2$ and smoothing parameters $\alpha = 0.2$, and $\beta = 0.3$.

b) Use your smoothed values as fitted values and calculate the model residuals.

c) Calculate the autocorrelation on lag length 1 of the residuals. What does your result suggest regarding the model choice?

d) How do the autocorrelation differ from Problem 3? What is a likely reason for this?

[*In case you were not able to calculate the autocorrelations, explain what you think should happen.*]

e) Calculate forecasts for $\hat{y}_t$ for $t = 11, 12, 13$, and 14.

f) Compute RMSE and compare the results with the ARIMA model in Problem 3.