STOCKHOLM UNIVERSITY
Department of Statistics
Johan Koskinen

WRITTEN RE-EXAMINATION, ECONOMETRICS II
2024-03-13

Time for examination: 08.00-13.00

Allowed tools: Pocket calculator, own formula sheet (1 double-sided A4 page), Course text-book (any edition): 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 6 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. (15 points)
Without accounting for your reasoning, indicate which **single** alternative is correct for each of the following part questions, a through e. Note that providing more than one alternative results in 0 marks for the part question.

(a) Which of the following statement is true if we regress an $I(1)$ (integrated of order 1) variable y_t against an $I(1)$ x_t variable:

(i) If $y_t - \beta x_t$ is $I(1)$ then OLS is consistent.

(ii) If $y_t - \beta x_t$ is $I(1)$ the model fulfills the Gauss-Markov assumptions for time-series.

(iii) $y_t - \beta x_t$ is always a stationary process.

(iv) If $y_t - \beta x_t$ is $I(0)$ then OLS is consistent.

(b) Why is cointegration useful:

(i) It makes it possible to model a stationary time series.

(ii) Ensures that the errors are serially correlated.

(iii) Describes an equilibrium relationship between two or more time series that are not stationary.

(iv) Allows for a test for Granger causality.

(c) Fixed effects for panel data:

(i) Cannot be correlated with explanatory variables.

(ii) Deal with unobserved variables that are time-varying.

(iii) Can model heteroscedasticity of the error term.

(iv) Are useful for dealing with unobserved variables that do not change over time.

(d) Suppose that the true model is

$$y_t = 0.35x_t + z_t + u_t,$$

where x_t and z_t are weakly dependent and stationary processes, and u_t is a white noise process with $E(u_t) = 0$. Assume that

$$z_t = 0.9z_{t-1} + v_t,$$

where v_t is a white noise process with $E(v_t) = 0$. Suppose further, that we regress y_t on x_t, and obtain the fitted values

$$\hat{y}_t = \hat{\beta}x_t, \quad (1)$$

where $\hat{\beta}$ is obtained as the OLS. Then:

(i) The proposed model underlying (1), i.e.

$$y_t = \beta x_t + \eta_t,$$

where η_t is a white noise process with $E(\eta_t) = 0$, s dynamically complete.

(ii) We can expect that the residuals obtained from (1) are independent.

(iii) We can expect that the residuals obtained from (1) have a unit root.

(iv) We can expect that the residuals obtained from (1) are serially correlated.

(e) In the time series regression

$$y_t = \alpha + \beta x_t + u_t,$$

the condition

$$E(u_t \mid x_1, \ldots, x_T) = 0,$$

(i) Is not necessary for unbiasedness of the OLS estimator for β.

(ii) Is necessary for unbiasedness of the OLS estimator for β.

(iii) Ensures that the errors u_t are homoscedastic.

(iv) Ensures that the errors are uncorrelated, i.e. $Corr(u_s, u_t) = 0$ for all $s \neq t$.

Problem 2. (10 points)

Write the following models using backshift operators. Clearly indicate the order of the different lag-polynomials, e.g. for an ARIMA(1,1,3),

$$(1 - \phi_1 B)(1 - B)y_t = \delta + (1 - \theta_1 B - \theta_2 B^2 - \theta_3 B^3)\epsilon_t.$$

(a) ARIMA$(0, 2, 2)$.

(b) ARIMA$(1, 2, 1)$

(c) ARIMA$(1, 1, 1) \times (1, 1, 2)$ with seasonal period $s = 4$

Problem 3. (17 points)
Consider three time series models of the form

$$y_t = \alpha + \beta y_{t-1} + \gamma_0 \epsilon_t + \gamma_1 \epsilon_{t-1} + \gamma_2 \epsilon_{t-2},$$

with parameters as in Table 1, and where $\epsilon_t \sim \mathcal{N}(0,1)$, indepdendently for all t.

Model	α	β	γ_0	γ_1	γ_2
1	5	0	1	-0.7	0.3
2	5	0	1	0.9	0.5
3	.35	0.93	1	0	0

Table 1:

For each of the models 1, 2, and 3, with parameters as in Table 1 above, $T = 10,000$ samples are simulated and the autocorrelation function $\rho(k)$, $k = 1, \ldots, 12$, for each model is estimated. Figure 1 shows the estimated autocorrelation functions (B1, B2, B3) and snapshots of the first $t = 1, \ldots, 50$ observations for each time series (A1, A2, A3). Note that the figures in each panel are in random order, i.e. a time series in a given row does not necessarily correspond to the autocorrelation function in the same row. Pair each of the three models 1, 2, and 3, with one realisation (A1, A2, A3) and one autocorrelation function (B1, B2, B3). Clearly motivate your answers. Providing a correct pairing for a model without motivation results in 0 marks for the model.

Problem 4. (18 points)

years	2015	2016	2017	2018	2019	2020	2021	2022	2023
tries	62	71	66	78	84	74	86	73	91

Table 2: Number of tries scored in the Six Nations Champtionship, 2015-2023.

Scoring a try, grounding the ball behind the try-line, is the main form of scoring in rugby union. Table 2 provides the total number of tries scored in each of the nine Six Nations Championships prior to 2024. The Sic Nations is a tournament where each team play each other once. The teams are England, Scotland, Wales, Ireland, France, and Italy. The average number of tries scored is plotted in Figure 2.

Table 3 provides the averages y_T together with the smoothed values $\tilde{y}_T^{(1)}$ and $\tilde{y}_T^{(2)}$, obtained as

$$\tilde{y}_T^{(1)} = \lambda y_T + (1-\lambda)\tilde{y}_{T-1}^{(1)},$$
$$\tilde{y}_T^{(2)} = \lambda \tilde{y}_T^{(1)} + (1-\lambda)\tilde{y}_{T-1}^{(2)},$$

for $T = 1, \ldots, 9$, with $\tilde{y}_T^{(1)} = \tilde{y}_T^{(2)} = 7.03$, and λ chosen to minimise the sum of squares of the one-step ahead prediction errors (SSA) of a double exponential smoothing algorithm.

Year	2015	2016	2017	2018	2019	2020	2021	2022	2023
y_T	4.13	4.73	4.40	5.20	5.60	4.93	5.73	4.87	6.07
$\tilde{y}_T^{(1)}$	4.13	4.28	4.31	4.53	4.80	4.83	5.06	5.01	5.27
$\tilde{y}_T^{(2)}$	4.13	4.17	4.21	4.29	4.42	4.52	4.66	4.74	4.88

Table 3: Average number of tries scored in Six Nations Championships, 2015-2023, and their smoothed values. To be used for Problem 4.

(a) The values tried for the algorithm were $\lambda = .05, .15, .25, .35, .45, .55, .65, .75$. Based on Figure 3, which value of λ was used to compute the smoothed values in Table 3?

(b) Consider the first four Six Nations. Forecast the average number of tries scored per match for the 2019 Six Nations using DES and compute the forecast error. Note that you do not need to recompute the smoothed values in Table 3.

(c) Using the available data, forecast the average number of tries scored per game in the 2024 Six Nations with *simple exponential smoothing*.

(d) Using the available data, forecast the average number of tries scored per game in the 2024 Six Nations with *double exponential smoothing*.

(e) The 2024 Six Nations is still ongoing. Thus far 8 of the 15 games have been played and 42 tries have been scored. Based on this, assume that 37 tries are scored in the rest of the matches. Compute the forecast errors of the methods in (c) an (d). In addition, compute the relative forecasting errors in percentage form. Based on this measure, which one of the forecast methods in (c) and (d) performs the best?

Problem 5. (20 points)
Suppose that

$$y_t = -3 + 0.7y_{t-2} + \epsilon_t,$$

where $E(\epsilon_t) = 0$, $E(\epsilon_t^2) = \sigma^2$, and $Cov(\epsilon_t, \epsilon_s) = 0$, for all $t \neq s$. Assume further that all errors ϵ_t are normally distributed.

(a) In the process for y_t stationary? Motivate your answer.

(b) Compute the expected value of y_t

(c) Sketch the partial autocorrelation plot for the process y_t

(d) Assuming that $\sigma^2 = \frac{1}{2}$, suppose that we obtain observations on y_t

t	1	$\cdots$	97	98	99	100
y_t	-9.5	$\cdots$	-11.25	-10.21	-10.95	-11.1

Provide point forecasts for y_{101}, y_{102}, and y_{103}.

Problem 6. (20 points)
Suppose that the monthly sales y_t in a furniture store can be modelled as a function of the average price x_t of the sold items according to the following model

$$y_t = \alpha_0 + \beta_0 x_t + \beta_1 x_{t-1} + \beta_2 x_{t-2} + \epsilon_t,$$

where the error terms ϵ_t are independent, $E(\epsilon_t \mid x_t, x_{t-1}, x_{t-2}, \ldots) = 0$, and $V(\epsilon_t) = \sigma^2$.

(a) What kind of model is the model for y_t?

(b) Suppose that there is a momentary one-unit increase in the price at time t - what is the immediate change in sales due to this increase?

(c) Suppose that there is a momentary one-unit increase in the price at time t - what is the change in sales two periods *after* this increase?

(d) Suppose that there is a one-unit increase in price at time t that remains permanently - what is the long-term change in sales?

(e) Define the long-run propensity score

$$\theta_0 = \beta_0 + \beta_1 + \beta_2.$$

Explain, insofar as possible, how you can obtain an estimate $\hat{\theta}_0$ of θ_0 as well as the standard error of $\hat{\theta}_0$ (*You may find example 10.4 in JMW instructive*)

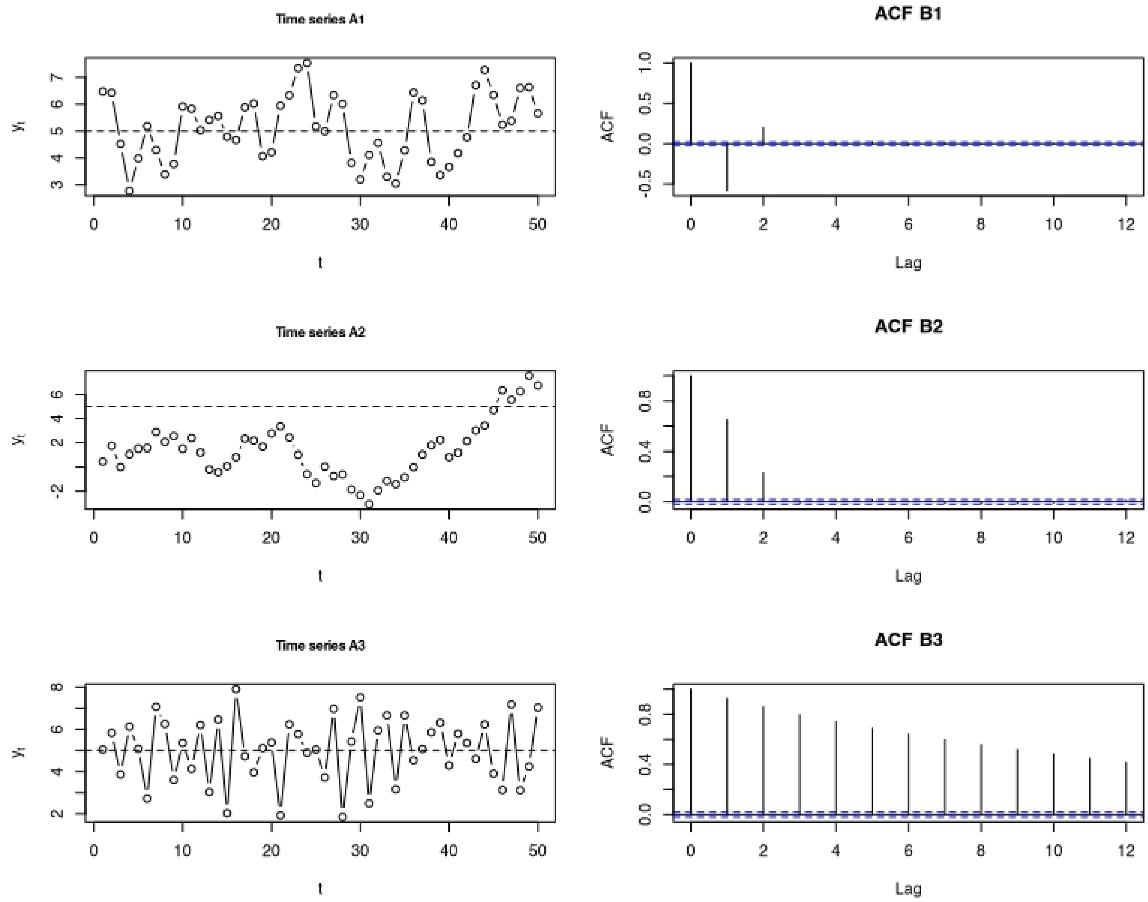

Figure 1: Three time series y_t, $t = 1, \ldots, 50$, (left panel) and three estimated autocorrelation functions (right panel) in Problem 3. The dashed horisontal lines in the figures on the left panel correspond to $E(y_t)$. The two dashed horisontal lines in the figures on the right panel correspond to the 95% confidence bands for the sample autocorrelation. Note that the figures in each panel are in random order, i.e. a time series in a given row does not necessarily correspond to the autocorrelation function in the same row.

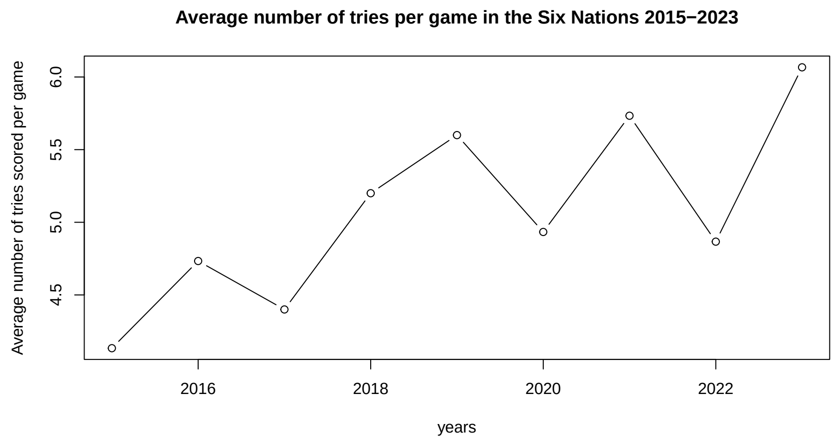

Figure 2: Average number of tries For problem 4

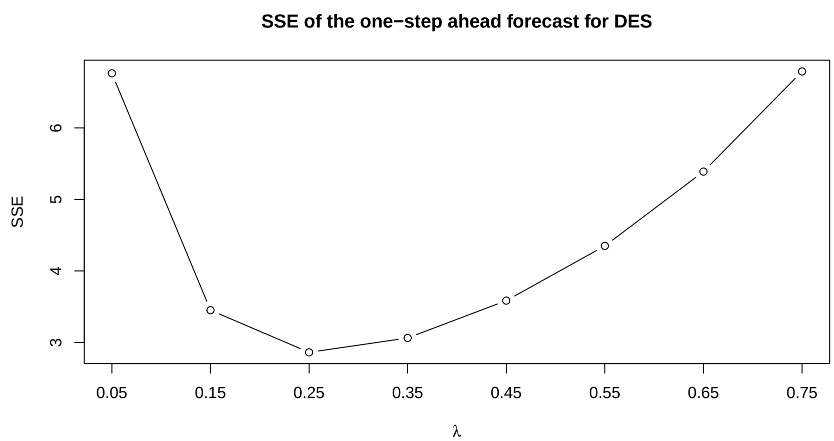

Figure 3: Sum of squares of the one-step ahead prediction errors (SSE) for double exponential smoothing (DES) as a function of λ. To be used for problem 4