Full Name: **Student Number:**

NAME (PRINT): ____
Last/Surname First /Given Name

STUDENT #: ____________________ **SIGNATURE:** ____________________

UNIVERSITY OF TORONTO MISSISSAUGA
APRIL 2014 FINAL EXAMINATION
STA457H5S
Applied Time Series Analysis
Ramya Thinniyam
Duration - 3 hours
Aids: Statistical Calculator (non-programmable without a text keypad)
1 page of double-sided Letter (8-1/2 x 11) sheet (hand-written)

The University of Toronto Mississauga and you, as a student, share a commitment to academic integrity. You are reminded that you may be charged with an academic offence for possessing any unauthorized aids during the writing of an exam. Clear, sealable, plastic bags have been provided for all electronic devices with storage, including but not limited to: cell phones, tablets, laptops, calculators, and MP3 players. Please turn off all devices, seal them in the bag provided, and place the bag under your desk for the duration of the examination. You will not be able to touch the bag or its contents until the exam is over.

If, during an exam, any of these items are found on your person or in the area of your desk other than in the clear, sealable, plastic bag; you may be charged with an academic offence. A typical penalty for an academic offence may cause you to fail the course.

*Please note, you **CANNOT** petition to **re-write** an examination once the exam has begun.*

INSTRUCTIONS:
-There are 7 questions - answer all questions
-There are 11 pages in total (including this page). There are 4 pages of output provided separately. Make sure you have all pages of questions and output before starting the exam.
- Do not write answers on output pages. Answer all questions on the question papers.
-For all true/false questions - no justification needed, just circle answer.
-For all other questions - show your work and justify answers to earn full marks. Correct answers with no justifications will not receive any marks.
-Round your answers to 4 decimal places where appropriate.
- You may copy numbers from output to answer questions unless the question specifically asks you to calculate from scratch.
-For all questions, consider p-values < 0.05 as statistically significant.
-Make sure to interpret and give practical conclusions to answer the questions of interest.

BEST WISHES! ☺

Question	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**TOTAL**
Value	10	10	10	18	30	15	7	**100**
Mark Earned								

Full Name: **Student Number:**

[10 marks - 1each]

1. TRUE/FALSE: ***If the statement is true under all conditions circle T, otherwise circle F.***

(a) The time domain approach is based on the idea that correlation between adjacent points in time is best explained by dependence of the current value on past values. T F

(b) All stationary models are causal. T F

(c) A series with a slow decay in autocorrelations suggests that the series has a linear trend. T F

(d) ARMA forecasts quickly reach the mean as the forecast horizon grows. T F

(e) Time series regression can capture and estimate trend and seasonal effects. T F

(f) The periodogram plot of a smooth time series would have large values for large frequencies. T F

(g) Trend should not be removed before fitting a time series regression model. T F

(h) For an MA(q) series, the PACF dies off and the ACF cuts off after lag q. The same is true for an ARIMA(p, 0, q) model. T F

(i) Time series regression can estimate stochastic trends. T F

(j) Overfitting leads to less efficient parameter estimates. T F

[10 marks]

2. Consider the following time series: $x_t = w_t + 2w_{t-1} - w_{t-2} + 0.6\, w_{t-3}$; where $w_t \sim wn(0, \sigma_w^2)$.

[2m]

(a) Is this process causal? Justify.

[1m]

(b) Find the mean function of x_t.

[7m]
(c) Find the auto-covariance function and then determine if x_t is weakly stationary. Justify.

[10 marks]
3. Given below is a SARIMA model:
$(1 + 0.1\,B - 0.06B^2)(1 + 0.8B^8)(1 - B^4)(1 - B)^2\ x_t = (1 - 0.8B^{12})(1 - 0.1B - 0.02B^2)\ w_t$.

[7m]
(a) Identify the model orders. Show your work in the space below and then write only the final answers in the chart.

s	p	d	q	P	D	Q

[3m]
(b) Suggest a transformation, y_t, of the original series x_t that would be stationary.
(Your answer y_t should be expressed in terms of x_t only.)

Full Name: **Student Number:**

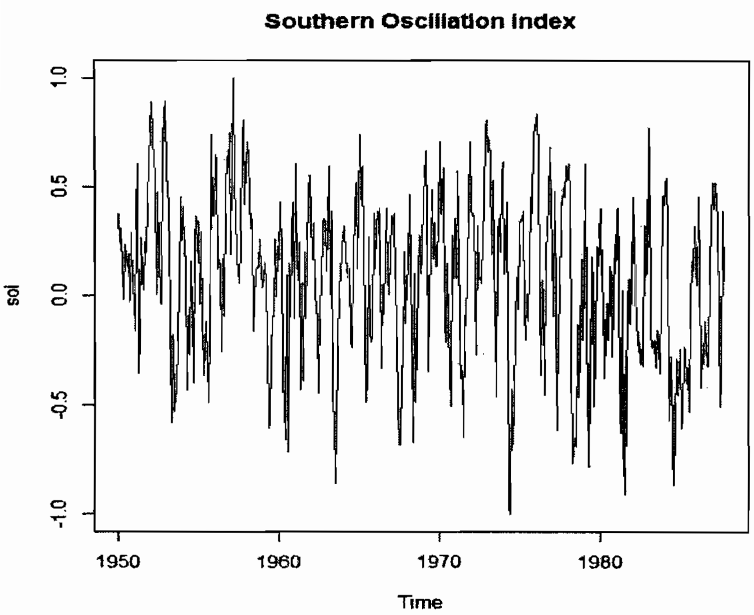
[18 marks]

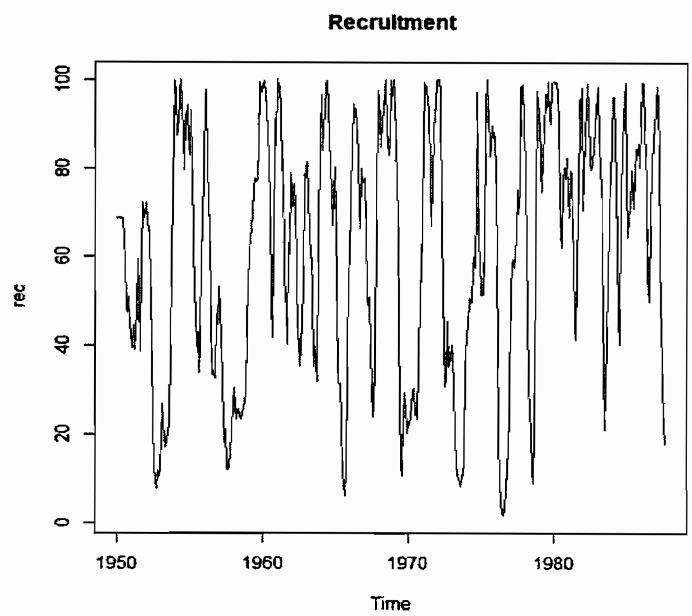
4. *El Nino and Fish Population:*

Consider the environmental series Southern Oscillation Index (soi) which measures changes in air pressure related to sea surface temperatures in the central Pacific Ocean. An associated time series is Recruitment (rec) which measures the number of new fish. Both series are monthly, measured over a total of 453 months between 1950-1987. The El Nino effect is the phenomenon of prolonged warming in the Pacific Ocean that occurs at irregular intervals every 3 to 7 years approximately.

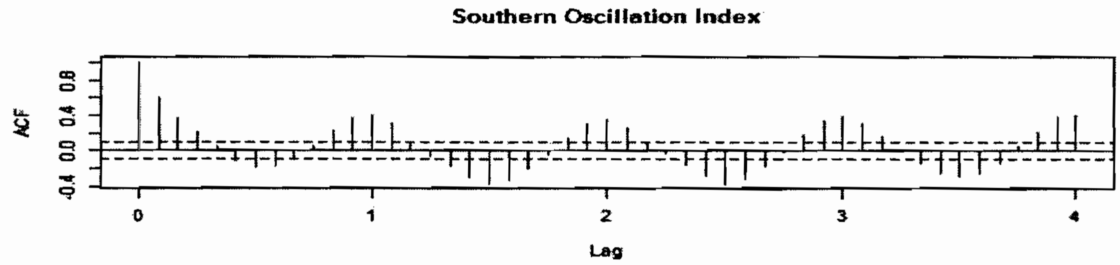
[6m]

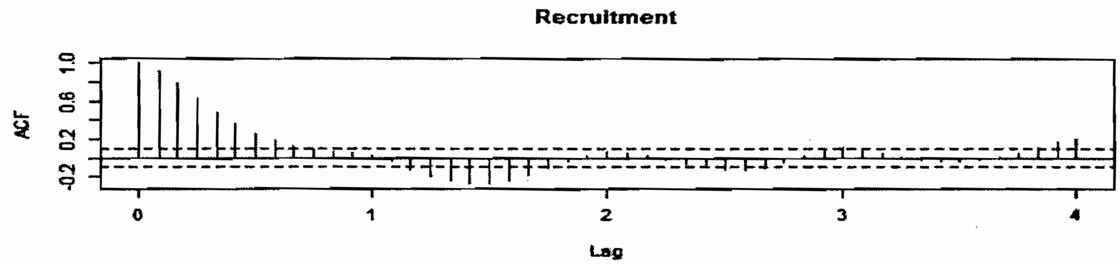
(a) Below are the two time series plots. Describe the properties of each series.

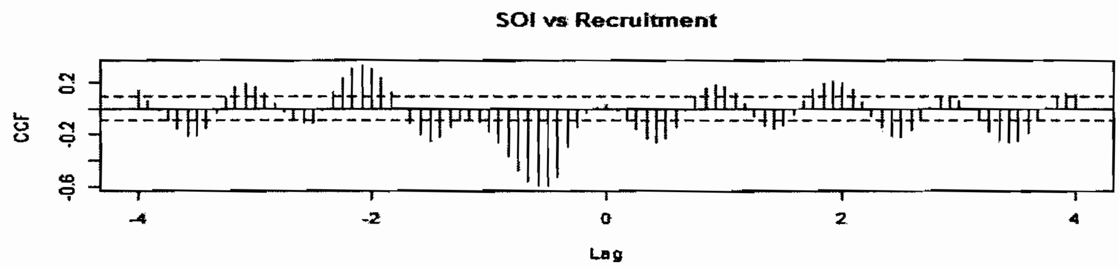




(b) Displayed here are the auto-correlation (ACF) plots for each series and the cross-correlation (CCF) of the two series. The lag axes are labelled in years. A negative lag in the CCF means that SOI leads Recruitment.







[2m]
i) For each series, two prominent observations can be made about autocorrelations. Name these two observations by referencing the ACF plots.

[2m]
ii) Look at the CCF plot. At which lag is there a large peak? Interpret this in terms of the practical problem. (Your answer should include the lag number and then a practical interpretation involving Recruitment and SOI).

[2m]
(c) Suggest an appropriate linear regression model for these data by using your answer from part ii) of (b) above. Write out the model using proper notation and define any variables you introduce.

[2m]
(d) If you were to fit the regression model suggested in (c), would you expect the estimate of the slope parameter to be positive or negative? Justify.

[2m]
(e) Define "spurious correlation" in the time series regression context. (Give a general definition.)

[2m]
(f) If you were to fit the regression model that you suggested in (c) for the El Nino & Fish Population example, do you need to worry about spurious correlation? Why or why not?

[30 marks]

5. This question pertains to the El Nino and Fish Population example from Question 4. Three SARIMA models (*model1,model2,model3*) were fit to the Recruitment time series. Refer to the output.

[2m]
(a) What type of SARIMA model was fit in *model1*?
Use proper notation ARIMA $(p, d, q) \times (P, D, Q)_S$ and specify all the model orders.

[2m]
(b) What type of SARIMA model was fit in *model2*?
Use proper notation ARIMA $(p, d, q) \times (P, D, Q)_S$ and specify all the model orders.

[3m]
(c) Referring to the ACF and PACF plots of Recruitment, explain why *model1* was fit to the data.

[3m]
(d) Referring to the ACF and PACF plots of Recruitment, explain why *model2* was proposed.

[3m]
(e) For *model1*, write out the fitted model. Define any notation/variables you use in your equation.

[3m]
(f) For *model2*, write out the fitted model without using the backward shift notation.

[3m]
(g) What type of SARIMA model was fit in *model3*?
Use proper notation ARIMA $(p, d, q) \times (P, D, Q)_s$ and specify all the model orders.
Explain why there is no constant or mean term estimated in *model3*.

[3m]
(h) Refer to the p-values for the Ljung-Box statistic for *model3.* What do you conclude?
Give an intuitive reason for why this occurred.

[3m]
(i) For *model1*, find the 3-step ahead prediction of recruitment, x_{456}^{453}, and its corresponding prediction error P_{456}^{453}?

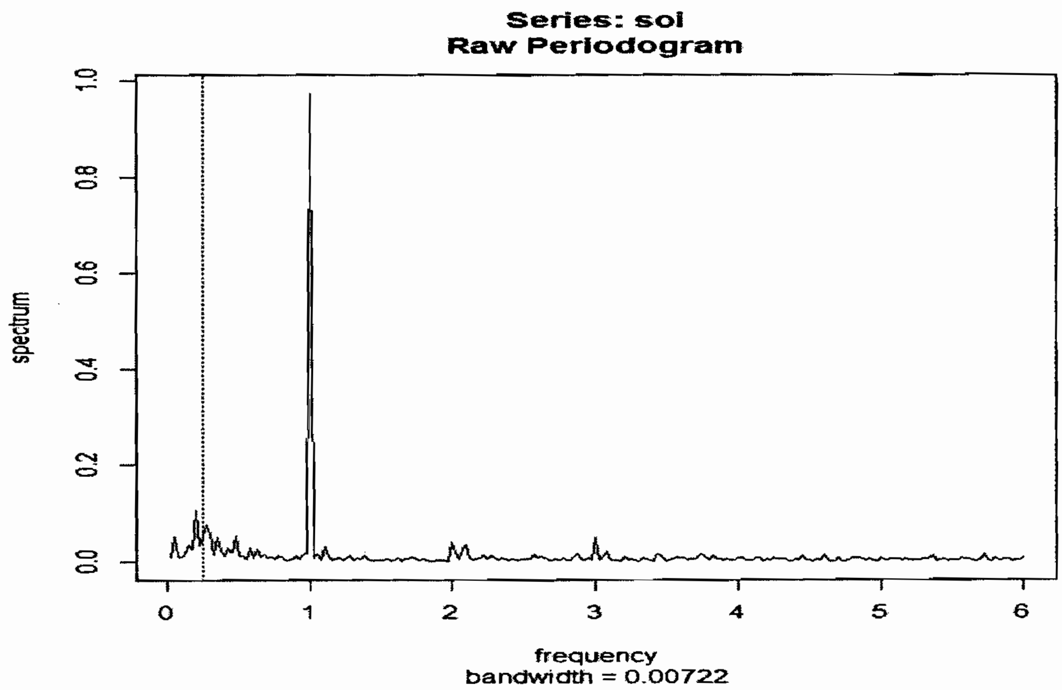
[5m]
(j) Amongst the three models that were fit, which would you select as the "best" model?
Name the model and fully justify your choice. Also describe if there are concerns with your selected model.

[15 marks]

6. Once again, we will work with the El Nino and Fish Population example from Question 4. Since both series exhibit periodic behaviour, we will perform Spectral Analysis to measure the strength of these cycles.
In this question, we will focus on the SOI series only. Refer to the output.

[5m]

(a) Below is the periodogram of the SOI series, where the frequency axis is labelled in multiples of $\Delta = 1/12$.



Two peaks are prominent in the periodogram. Identify the frequencies/periods of each of these cycles and interpret them in terms of the practical problem. Your answer should include the frequency/period values and what each represents. (It will help to reread the description of the data in Question 4).

[7m]

(b) The original length of the SOI series is $n=453$ but the series has been centered and padded to a length of $n' = 480$ (480 is the next highly composite number that allows fast computation of the Fourier transform).

You identified two cycles in part (a). For each of these cycles, find an approximate 95% confidence interval for the spectrum for the SOI series. Refer to the output and show your work.
Your answer should include 2 confidence intervals - each CI should be displayed with a lower and upper bound and then circled.

[3m]

(c) What problem do you notice about the two confidence intervals calculated in (b)?
Suggest a potential solution to this problem.

[7 marks]

7. This question asks you to define and describe the importance of some of the properties of ARMA models.

[3m]

(a) In words (without formulas), describe what it means for a time series to be weakly stationary? Why is stationarity important for time series analysis?

[2m]

(b) In plain English, explain what a causal model is and why causality is important for ARMA processes.

[2m]

(c) Why is parameter redundancy problematic?

---END OF EXAM---