Name Mr. Key
ID# 7777777

ECO 5375
Eco. & Bus. Forecasting

Prof. Tom Fomby
Fall 2013

MID-TERM EXAM

Instructions: Write in your name and student ID above. You have 1 hour and 20 minutes to complete this exam. This exam is worth a total of 94 points. The points for the separate question are broken out as follows:

Questions 1 – 10 are worth 2 points each.
Q11 = (2, 2, 2) = 6 points
Q12 = (2, 2) = 4 points
Q13 = 6 points
Q14 = (2, 2) =4 points
Q15 = (2, 6) =8 points
Q16 = (8, 2) = 10 points
Q17 = 3 points
Q18 = (2, 1, 1, 2, 2) = 8 points
Q19 = (2, 2) = 4 points
Q20 = (3, 4, 2, 2) = 11 points
Q21 = (2, 2, 4, 2) = 10 points

2 1. SAS programs have two basic steps. They are the Data step and the Procedure step.

2 2. The basic punctuation following each executable statement in SAS is
   a. Quotation mark (")
   b. Period (".")
   c. Semicolon (";")
   d. Dash ("-")

2 3. To put comments in a SAS program
   a. Enclose the comments between quotation marks as in: "content"
   b. Enclose the comments between /* and */ as in /*content*/
   c. Enclose the comments between (and) as in (content)
   d. Enclose the comments between # and # as in #content#

2 4. True or False. If a time series is slow-turning around a mean, it is probably a nonstationary time series and needs to be differenced before analyzing it.

2 5. If a time series y_t exhibits an exponential growth pattern, it can be transformed to stationarity by the transformation
   a. Δy_t
   b. $\Delta \log_e(y_t)$
   c. $\exp(y_t)$
   d. $\tan^{-1}(y_t)$

2 6. True or False. If some time series data, say x, data needs to be differenced to be made stationary, then the identify statement in the SAS Procedure ARIMA should read "identify var = x;"

2 7. Consider the following AR(1) Box-Jenkins model: $y_t = 18 + 0.6y_{t-1} + a_t$. This model implies that the mean of the y_t series is 45 = 18/(1 – 0.6). Suppose the last observation you have on the series is 43. The one-period ahead forecast for this model should be 43.8.

$\hat{y}_{T+h} = \mu + \phi^h(y_T - \mu) = 45 + 0.6(-2) = 43.8$

2 8. True or False. The "Damping" and "Cutting Off" patterns of the ACF and PACF of the stationary form of a time series provide a way to identify the orders of pure Box-Jenkins processes.

2 9. If the ACF has 3 spikes in it and then cuts off and if the PACF tails off, the ARMA model that is appropriate for the data is ARMA(0 , 3).

2 10. True or False. The two most popular Box-Jenkins transformations for handling nonstationarity in seasonal time series are using (i) a year-over-year differencing (Δ_s) or (ii) a year-over-year/period-to-period differencing ($\Delta_s\Delta_1$) of the data.

(2)

11. Consider the time series plotted in **Figure 1** in your handout.

(a) Does this time series look stationary to you? Explain your answer.

No. It is slow-turning around 10 and looks like a Random Walk without drift.

(2)

(b) Is it all right to apply the Box-Jenkins modeling approach to this data directly or should you transform the data first and, if so, how? Explain your answer.

This data needs to be differenced before we analyze the data using Box-Jenkins methods.

(2)

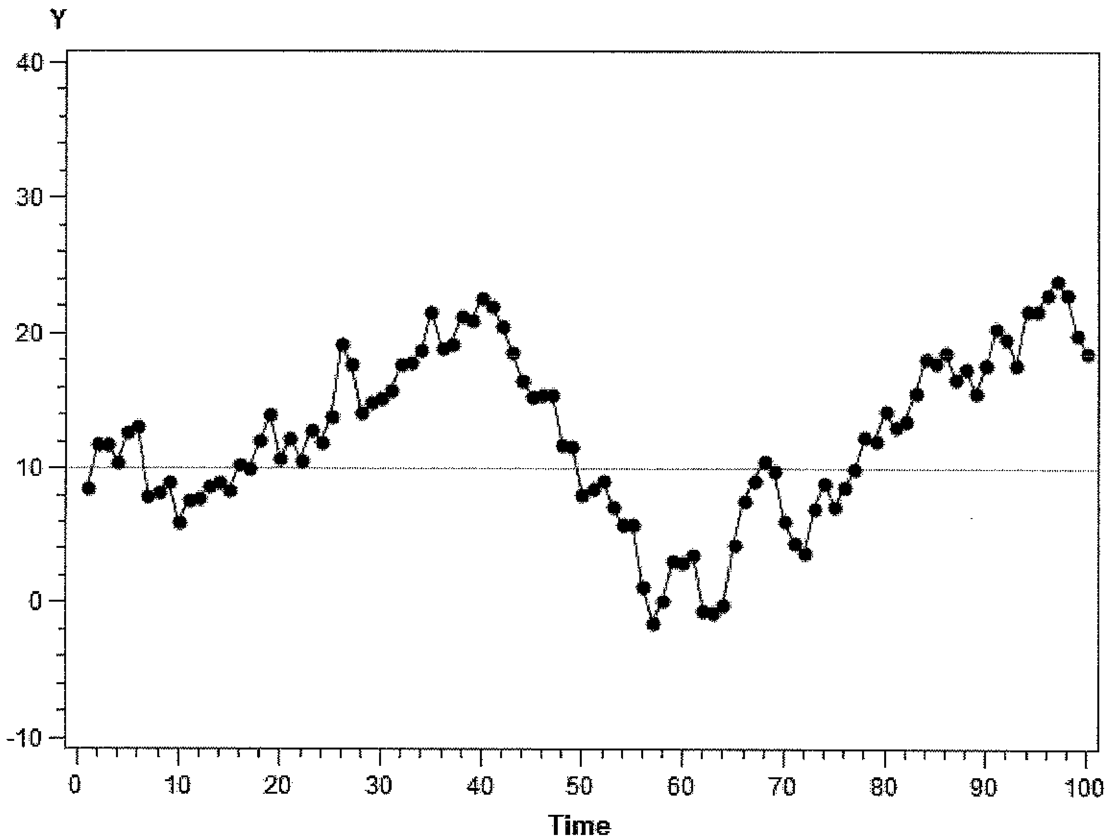
(c) In the below space, formally write out the requirements for a time series to be stationary: Let y_t be the stationary time series. Then for y_t ~~be~~ to be stationary it is required that

(i) $E(y_t) = \mu \quad \forall t$ (Constant mean)

(ii) $Var(y_t) = \sigma_y^2 \quad \forall t$ (Constant variance)

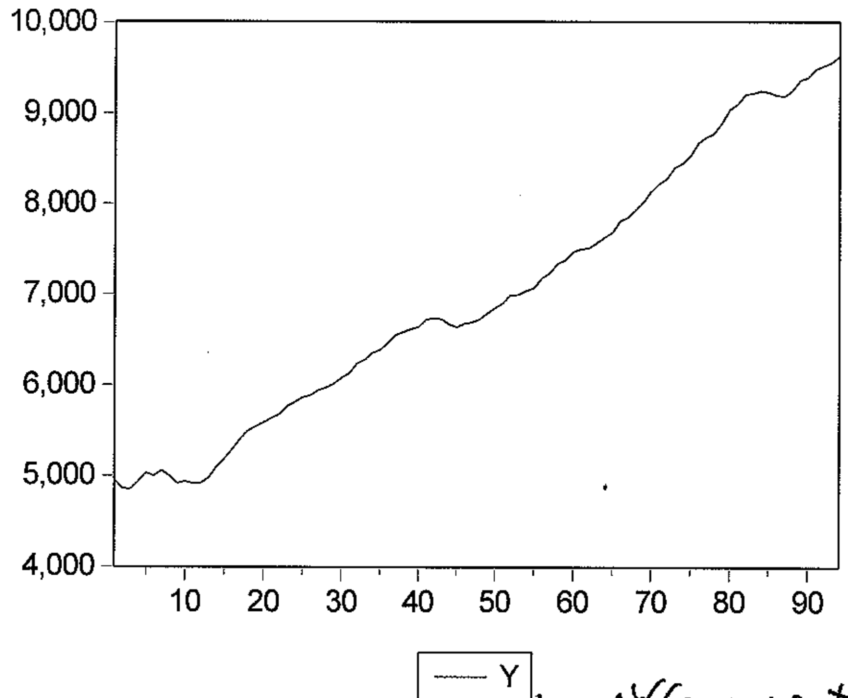
(iii) $Cov(y_t, y_{t-j}) = \gamma_j \quad \forall t$ (Constant covariance)

Figure 1



Now let us conduct a unit root test on the data in **Figure 2.** In the **EVIEWS Computer Output # 1** there is some information that should allow you to answer the following 3 questions.

Figure 2



needs to be difference to achieve stationarity.

2 12. The null hypothesis of the Dickey-Fuller test for the data in Figure 2 is
2 the data has a unit root and. The alternative hypothesis is the data follows a deterministic trend.

13. Using **EVIEWS Output # 1** and the correct Dickey-Fuller case for this data (Y), report the following information:

1 (a) The appropriate case for the Dickey-Fuller test is Zero Mean / Single Mean / Trend (circle a choice)

1 (b) The number of augmenting terms chosen for the test is = 2

1 (c) Dickey-Fuller t-statistic (tau) = -1.983249

1 (d) Probability Value of DF t-statistic = 0.6023

2 (e) This test result indicates that the time series Y (is/is not) stationary and (does/does not) need to be differenced to make the series stationary.

Now consider the **SAS Computer Program and Output # 2.** Use them to answer the following 4 questions. We wish to determine the best Box-Jenkins model for this data. The series is plotted in **Figure 2. Assume that the time series is observed <u>monthly</u>.**

14. Using the sample ACF and sample PACF that is provided by **Computer Output # 2**, give me a tentative identification of the p and q values for the Y series. Explain your answer.

(2)

p = 1, q = 0.

Explanation: The ACF is damping out (i.e. tailing off) while the PACF has one spike in it (note: don't count the zero lag) and then cuts off. This is the signature of an AR(1) model.

(2)

15. Using **Computer Output # 2**, fill in the following P-Q box. Be sure to tell me what the entries of the cells of your box are. Which model is indicated to be the best model in the P-Q box? Explain your reasoning.

Reasoning: The AR(1) model has the lowest SBC while the AR(2) model has the lowest AIC. There is a split decision. However, in the AR(2) model the overfitting coefficient AR1,2 is statistically insignificant thus we prefer the AR(1) model. The residuals of the AR(1) model are white noise!

(2)

Q

(6)

P	0	1	2
0	985.1017 987.6343 30.34 (0.1723)	977.6966 982.7818 18.01 (0.7568)	976.0719 983.6697 15.89 (0.8214)
1	974.0883 979.1525 16.23 (0.8452)	974.1618 981.7596 14.96 (0.8641)	
2	973.9127 987.5705 15.05 (0.8602)		

Legend: AIC
SBC
Q_{24}
(p-value)

16. Use **Computer Output #2** to conduct an overfitting exercise on the model you chose from the P-Q box. Below, report the overfitting coefficient of each overfitting model and its corresponding t-statistic. What conclusion do you draw from the overfitting exercise? Explain your answer.

4

Overfitting Model 1 is ARMA(2 , 0).
The overfitting coefficient is 0.15330 .
The T-statistic of the overfitting coefficient is 1.47 .
Therefore the overfitting coefficient from this model is statistically (significant/insignificant). Circle one alternative.

4

Overfitting Model 2 is ARMA(1 , 1).
The overfitting coefficient is 0.32538 .
The T-statistic of the overfitting coefficient is 1.32 .
Therefore the overfitting coefficient from this model is statistically (significant/insignificant). Circle one alternative.

2

My conclusion is The AR(1) model is the best model. The overfitting coefficients of the overfitting models are both statistically insignificant.

17. In the below space write out the final model that you have chosen for the Y time series in **Computer Output # 2** with accompanying t-statistics, standard errors, goodness-of-fit measures, and a test statistic for white noise residuals with accompanying p-value. (You can report your estimated model either in the intercept-form or the deviation-from-mean form.) $(1-0.36274B)(y_t-47.76943)=a_t$

3

$y_t - 47.76943 = 0.36274(y_{t-1} - 47.76943) + a_t$: Deviation-from-mean form.

(7.27282) (0.09787) (7.27282)

$[6.72]$ $[3.71]$ $[6.72]$

(·) = se
[·] = t-stat

Intercept Form:
$y_t = 31.15467 + 0.36274 y_{t-1} + a_t$
$(0.09789)[3.71]$

AIC = 974.0863
SBC = 979.1535
Q_{24} = 16.23
p-value = 0.8452

18. Suppose that I told you that the correct model for a time series is

$$y_t = 30 + 0.7 y_{t-1} + a_t \quad .$$

2 (a) The above model is a ARMA(1 , 0). (Fill in the blanks.)

1 (b) The above model's intercept is 30 . (Fill in the blank.)

1 (c) The above model implies that the population mean of y is $\mu =$ 100 (Fill in the blank.)

$$\mu = \frac{30}{1-0.7} = \frac{30}{0.3} = 100$$

(d) The minimum mean square forecast equation for the above model for an h-step ahead forecast is

2

$$\hat{y}_{T+h} = \mu + \phi_1^h (y_T - \mu)$$

with the standard error of the h-step ahead forecast given by

$$se(\hat{y}_{T+h}) = \hat{\sigma}_a (1 + \hat{\phi}_1^2 + \hat{\phi}_1^4 + \cdots + \hat{\phi}_1^{2(h-1)})^{1/2} .$$

Suppose that the last available observation that you have on y is $y_T = 105$ and that the standard error of your model is $\hat{\sigma}_a = 3$. In the space below derive the **two-step-ahead forecast** for y $(\hat{y}_{T+2})$ and a 95% confidence interval for your forecast. Show your work if you expect full credit.

$$\hat{y}_{T+2} = 100 + (0.7)^2 (105 - 100)$$
$$= 100 + .49(5) = 102.45$$

$$se(\hat{y}_{T+2}) = 3(1 + 0.7^2)^{1/2} = 3(1.49)^{1/2} = 3(1.22) = 3.66$$

$$95\% \text{ C.I.} \Rightarrow \hat{y}_{T+2} \pm 1.96\, se(\hat{y}_{T+2}) = 102.45 \pm 1.96(3.66)$$
$$\Rightarrow (95.28, 109.62)$$

(e) Suppose that the forecast horizon of interest goes to infinity, $h \to \infty$. Given the above model, what will the ∞-step-ahead forecast be? $\hat{y}_{T+\infty}$ = 100.

2

Explain how you got your answer: The infinite horizon forecast for any stationary Box-Jenkins model is the sample mean.

19. Consider **Computer Output #3**. This output examines an Electricity Production data set (1972 – 1989). .

(a) Explain to me how you would informally inspect this data vis-à-vis an autocorrelation function to determine whether or not there is substantial seasonality in the data or not. Look for the spikes at the seasonal lags of 12, 24, 36, 48 .... of the first differenced data. The data has trend in it hence the need to look at the ACF of the first differenced data.

2

(b) What are the results of Computer Output #3 suggesting? Explain your answer thoroughly.

2

The ACF of the differenced data does have spikes at the seasonal lags thus the indication that the data does have seasonality in it.

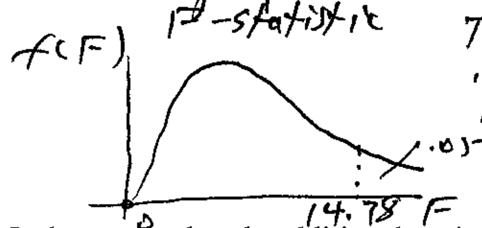
20. Consider **Computer Output #4**. It is the case that the data needs to be logged to make the variance of the data around trend more uniform. This output is testing for seasonal differencing for the Electricity Production data set that we examined in Computer Output # 3 above.

(a) Which test is being performed here? What is the null hypothesis of the test? What is the alternative hypothesis of the test? The Hasza-Fuller Test is being conducted here. The null hypothesis is that the transformation $\Delta_s\Delta_1$ is appropriate whereas the alternative hypothesis is that the $\Delta_s\Delta_1$ transformation is not appropriate.

3

(b) Draw me a sampling distribution of the test statistic under the assumed truth of the null hypothesis and in the drawing indicate the appropriate acceptance and rejection regions at the 5% level of significance. What do you conclude from the test? Sampling distribution of Hasza-Fuller F-statistic

4



This distribution is skewed to the right.

critical value = 14.78

212 observations ⇒ $17\frac{2}{3}$ years. use m = 20

Since the Hasza-Fuller F-statistic = 5.49 < 14.78 we accept H_0 and $\Delta_s\Delta_1$ is appropriate filter of the data.

(c) Is there a need to do additional testing for differencing on this data? Explain your answer. No. Only if the null hypothesis of the Hasza-Fuller test is rejected do we go to the Dickey-Hasza-Fuller test to test the appropriateness of the Δ_s filter.

2

(d) In the below space, write out a mathematical expression (possibly using backshift operators) for the stationary form of the Electricity production series.

2

$$(1-B)(1-B^{12})y_t = (y_t - y_{t-1}) - (y_{t-12} - y_{t-13})$$
$$= y_t - y_{t-1} - y_{t-12} + y_{t-13}$$

21. Consider **Computer Output #5**. It analyzes the previous Electricity Production data. Use it to answer the following questions.

(a) Look at the SAS program file. Which model is the so-called AIRLINE model? Model 1, Model 2, Model 3, or Model 4? The AIRLINE model is Model __1__. (Fill in the blank.)

2

(b) Examine the sample ACF and sample PACF. Make a tentative identification of an appropriate multiplicative, seasonal Box-Jenkins model for the Electricity Production data. Be sure and explain your reasoning. Reasoning: ACF cuts off after lag $q + sQ = 1 + 12 \cdot 1 = 13$. PACF damps out.

2

My tentative identification for the lelec data is

d=_1_, D=_1_, p=_0_, P=_0_, q=_1_, Q=_1_.

The Airline Model.

Only have to report p-value.

Q_{24}

(c) Fill in the following blanks (assuming the appropriate differencing):

p-value in parentheses

Model	p	P	q	Q	AIC	SBC	Chi p-value (lag=24)
1	0	0	1	1	-863.323	-853.443	37.10 (0.0231)
2	0	0	2	1	-869.886	-856.713	19.84 (0.5314)
3	1	0	2	1	-867.975	-851.508	19.63 (0.4815)
4	0	1	2	1	-868.538	-852.071	18.14 (0.5779)

4

(d) Given the results of part (c), which model do you prefer? Explain your answer.

2

I prefer Model 2 which is a slight modification of the Airline Model. It has the lowest AIC and SBC Goodness-of-Fit measures and the residuals of the model are white noise as the p-value of the Q_{24} statistic is $0.5314 > 0.05$. (Also you will note that all of the coefficients of the model (apart from the mean) are highly statistically significant. Not shown here in total, all overfitting models of Model 2 have statistically insignificant overfitting coefficients and their Goodness-of-fit measures are worse.)