ISyE4031 Regression and Forecasting
Practice Problems 3
Spring 2016

1. An annual time-series regression model was fit to data collected for 30 years. After studying and analyzing the results, the error terms were found to be first-order autocorrelated, and the independent variable is lagged, i.e., the value of x in year $t-1$ also affects the value of y in year t. So, the following pair of models was studied:
$y_t = \beta_0 + \beta_1 x_t + \beta_2 x_{t-1} + \beta_3 t + \varepsilon_t$ and $\varepsilon_t = \phi\, \varepsilon_{t-1} + a_t$ where the a_t are *i.i.d.* $N(0,\sigma^2)$ for each t, $t = 1,\dots,30$. By using the 30 data points and applying the *approximate (Cochrane-Orcutt) method*, the following estimates of the coefficients are obtained as a result of the transformed regression model: $\hat{\beta}_0 = 10$, $\hat{\beta}_1 = 3.2$, $\hat{\beta}_2 = 1.1$, $\hat{\beta}_3 = 1.5$, $\hat{\phi} = 0.74$. Use the fitted model to forecast the value of y in year $t = 31$, given that $y_{30} = 85$, $x_{29} = 8.5$, $x_{30} = 4.75$, and $\hat{x}_{31} = 6.9$.

2. Seven years of monthly data on the number of airline miles flown (in millions) in the UK were recorded and studied.

a. Suppose that the Holt's Trend Corrected (Double) Exponential Smoothing method was applied to the data and the following estimates were obtained for the last few months. Determine the point forecast for the y_{87} forecast made in time period 84, i.e., $\hat{y}_{87}(84)$.

Year	t	y_t	l_t	b_t
⋮	⋮	⋮	⋮	⋮
	82	12.38	12.77	0.036
7	83	11.59	11.71	−0.019
	84	12.77	12.66	0.029

b. Suppose that the additive Holt-Winters method was applied to the monthly airline mileage data with $\alpha = 0.2$, $\gamma = 0.3$, $\delta = 0.4$, and the following level, growth, and seasonality estimates were obtained for the last year. Calculate the point estimate of next February's airline miles flown y_{86} forecast made in month 84, i.e., $\hat{y}_{86}(84)$.

Month	t	y_t	l_t	b_t	sn_t
Jan	73	10.84	12.92	0.139	−1.93
Feb	74	10.43	12.99	0.117	−2.38
Mar	75	13.58	13.25	0.161	−0.02
Apr	76	13.40	13.63	0.227	−0.76
May	77	13.10	13.57	0.139	0.23
June	78	14.93	13.56	0.095	1.72

Month	t	y_t	l_t	b_t	sn_t
July	79	14.14	13.48	0.041	1.10
Aug	80	14.05	13.26	−0.035	1.41
Sep	81	16.23	13.29	−0.014	2.76
Oct	82	12.38	13.15	−0.054	−0.44
Nov	83	11.59	13.17	−0.031	−1.76
Dec	84	12.77	13.22	−0.006	−0.65

c. Suppose that airline miles flown was observed as 13.05 in January of year 8, i.e., $y_{85} = 13.05$. Update the level, the growth rate, and the seasonal factor estimates by using the smoothing constants given in part (b).

3. More Exponential Smoothing methods questions.

a. The growth rate smoothing constant, γ, can be negative, if the time series is decreasing over time. True or False? (Circle the correct answer.)

b. Suppose that you applied the simple exponential smoothing method with $\alpha = 0.1$ to the weekly demand data of a certain product by using 20 weeks of data. If $l_{19} = 10$ units, $y_{20} = 15$, and $s = 1.5$, a 95% prediction interval computed in week 20 for y_{23} is (circle the correct answer):

i. $[10 \pm (1.96)(1.5)(\sqrt{1.02})]$

ii. $[10.5 \pm (1.96)(1.5)]$

iii. $[10.5 \pm (1.96)(1.5)(\sqrt{1.02})]$

iv. $[10.5 \pm (1.96)(1.5)(\sqrt{1.1})]$

v. $[10 \pm (1.96)(1.5)(\sqrt{1.2})]$

c. Suppose that four different exponential smoothing methods were applied to a certain time series $\{y_t\}$. Based on the results shown below, which method should be selected, and what type of time series could $\{y_t\}$ be?

Exponential Smoothing Methods

	Single	Double (Holt's Trend Corrected)	Holt-Winters Additive	Holt-Winters Multiplicative
MAPE	8.3	8.1	8.3	8.8
MAD	0.17	0.16	0.18	0.21
MSD	0.45	0.41	0.52	0.81

d. The additive Holt-Winters method was applied to the quarterly power loads for a utility company, and to find the initial estimates for the level, trend, and four seasonal factors, a least squares regression line was fitted, i.e., $\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 t$. By using the following fitted values, find the initial seasonal factor for Quarter 4, i.e. sn_0.

Year	Quarter	$y_t - \hat{y}_t$
1	1	3.4
	2	-10.5
	3	8.6
	4	-6.1

Year	Quarter	$y_t - \hat{y}_t$
2	1	5.8
	2	-9.1
	3	11
	4	-2.9

Year	Quarter	$y_t - \hat{y}_t$
3	1	5.2
	2	-7.7
	3	9.4
	4	-4.5

4. 150 daily viscosity readings for a chemical product were recorded, and a nonseasonal MA(2) model, $y_t = \delta + a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2}$, was studied. The estimated model was found as:

$$\hat{y}_t = 35 + \hat{a}_t - 0.52\hat{a}_{t-1} + 0.65\hat{a}_{t-2} \ .$$

a. (9) Given $y_1 = 39.9$, $y_2 = 31.9$, $y_3 = 37.5$, calculate the first three point estimates of $\{y_t\}$, i.e., $\hat{y}_1$, $\hat{y}_2$, and $\hat{y}_3$.

b. (3) Given $\hat{a}_{149} = -3.32$ and $\hat{a}_{150} = -2.28$, calculate the forecasts for future period 151 from $t = 150$, i.e., $\hat{y}_{151}$.

5. More ARIMA questions.

a. Consider the tentative ARIMA(1,1) model, $\hat{y}_t = 25 + 0.8\, y_{t-1} + a_t - 0.2a_{t-1}$ where a_t is a random shock, distributed *i.i.d.* $N(0,\sigma^2)$. Calculate r_1 and r_2 by using ρ_1 and ρ_2 estimates.

b. Estimate the mean value of the time series, $\hat{\mu}$ of the ARIMA(1,1) model in part (a).

c. Consider the tentative AR(2) model $y_t = 2 + 0.6\, y_{t-1} + 0.3\, y_{t-2} + a_t$ where a_t is a random shock, each distributed *i.i.d.* $N(0,\sigma^2)$. Is the time series stationary and/or invertible? Explain by stating all conditions.

6. Consider the Sample Autocorrelation (SAC) and Sample Partial Autocorrelation (SPAC) functions for the following time series. Circle the most appropriate tentative model.

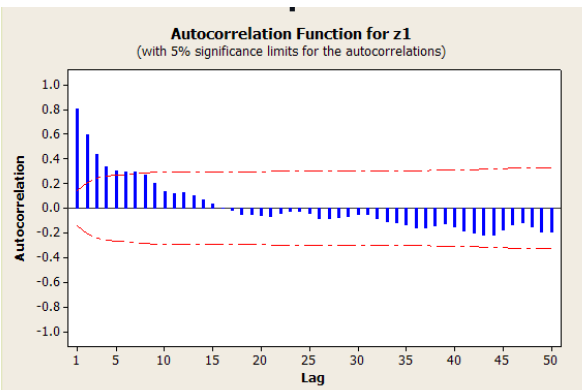
I. a. Nonseasonal AR(1): $z_t = \delta + \phi_1 z_{t-1} + a_t$

b. Nonseasonal MA(1): $z_t = \delta + a_t - \theta_1 a_{t-1}$

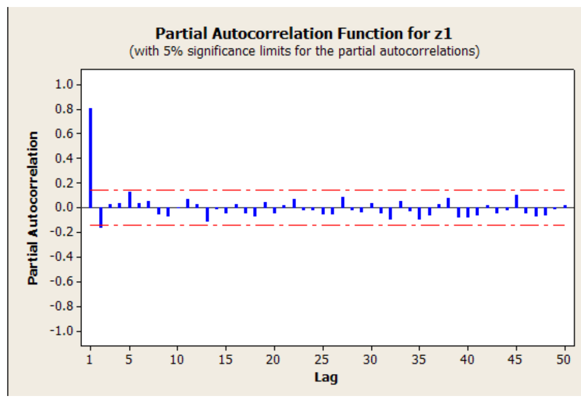
c. Nonseasonal MA(2): $z_t = \delta + a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2}$

d. Nonseasonal ARIMA(1,1): $z_t = \delta + \phi_1 z_{t-1} + a_t - \theta_1 a_{t-1}$

e. The time series is nonstationary.



SAC



SPAC

II. Seasonal length, $L = 12$:

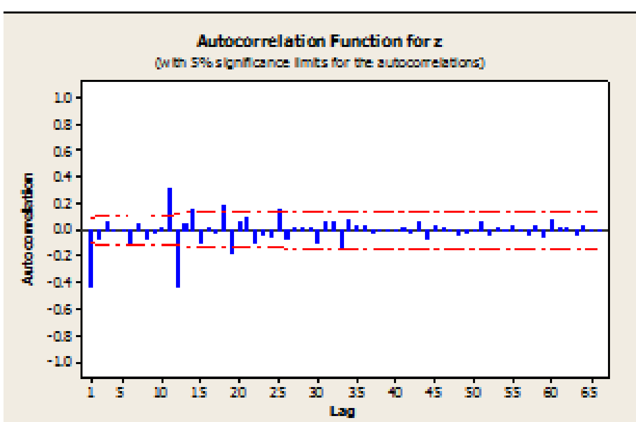
a. Nonseasonal MA(1): $z_t = \delta + a_t - \theta_1 a_{t-1}$, Seasonal AR(1): $z_t = \delta + \phi_{1,L}\, y_{t-12} + a_t$

b. Nonseasonal MA(1): $z_t = \delta + a_t - \theta_1 a_{t-1}$, Seasonal ARIMA(1,1): $z_t = \delta + \phi_{1,L}\, y_{t-12} + a_t - \theta_{1,L}\, a_{t-12}$

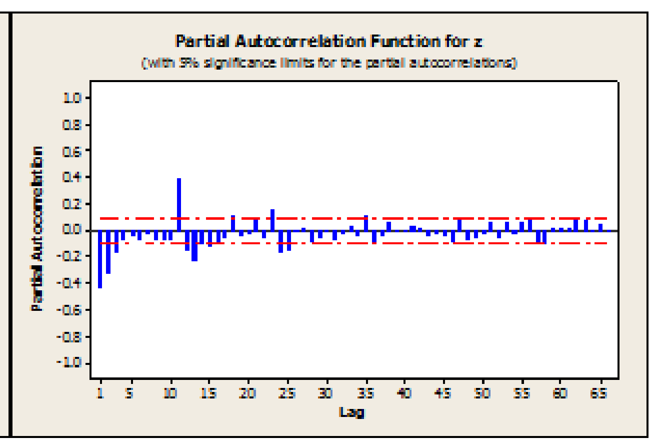
c. Nonseasonal AR(2): $z_t = \delta + \phi_1 z_{t-1} + \phi_2 z_{t-2} + a_t$, Seasonal MA(1): $z_t = \delta + a_t - \theta_{1,L}\, a_{t-12}$

d. Nonseasonal AR(2): $z_t = \delta + \phi_1 z_{t-1} + \phi_2 z_{t-2} + a_t$, Seasonal ARIMA(1,1): $z_t = \delta + \phi_{1,L}\, y_{t-12} + a_t - \theta_{1,L}\, a_{t-12}$

e. The time series is nonstationary.



SAC



SPAC

7. The monthly values for Midwestern housing starts $y_1, y_{2, \ldots,} y_{232}$ were recorded, and a time series analysis was conducted. According to the preliminary analysis the time series was found to be non-stationary. The first regular differenced and first seasonal differenced transformations were applied to $\{y_t\}$, and the working time series $\{z_t\}$, such that $z_t = y_t - y_{t-1} - y_{t-L} - y_{t-L-1}$, where L = 12, was obtained. Based on the SAC and SPAC of $\{z_t\}$, a combination of nonseasonal AR(3) and seasonal MA(1) models was identified as the tentative model, i.e., $z_t = \delta + \phi_1 z_{t-1} + \phi_2 z_{t-2} + \phi_3 z_{t-3} + a_t - \theta_{1,12} a_{t-12}$, and the following results were produced.

Final Estimates of Parameters

Type	Coef	SE Coef	T	P
AR 1	-0.4931	0.0683	-7.22	0.000
AR 2	-0.2145	0.0758	-2.83	0.005
AR 3	-0.0346	0.0682	-0.51	0.612
SMA 12	0.9290	0.0373	24.90	0.000
Constant	0.00507	0.02384	0.21	0.832

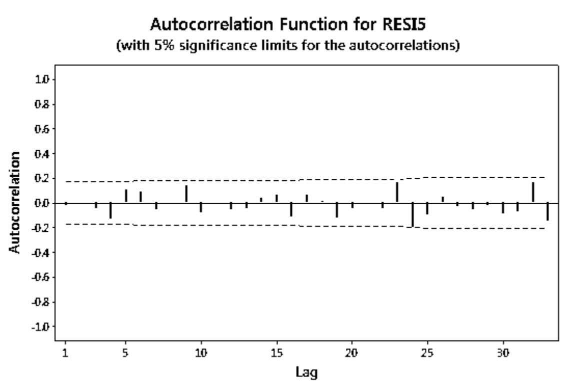
Modified Box-Pierce (Ljung-Box) Chi-Square statistic

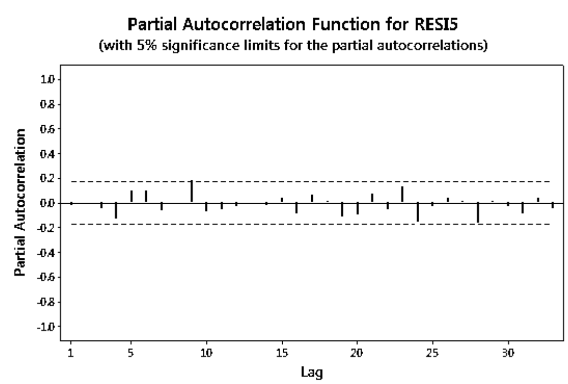
Lag	12	24	36	48
Chi-Square	8.7	18.6	32.2	41.0
DF	7	19	31	43
P-Value	0.276	0.481	0.407	0.558

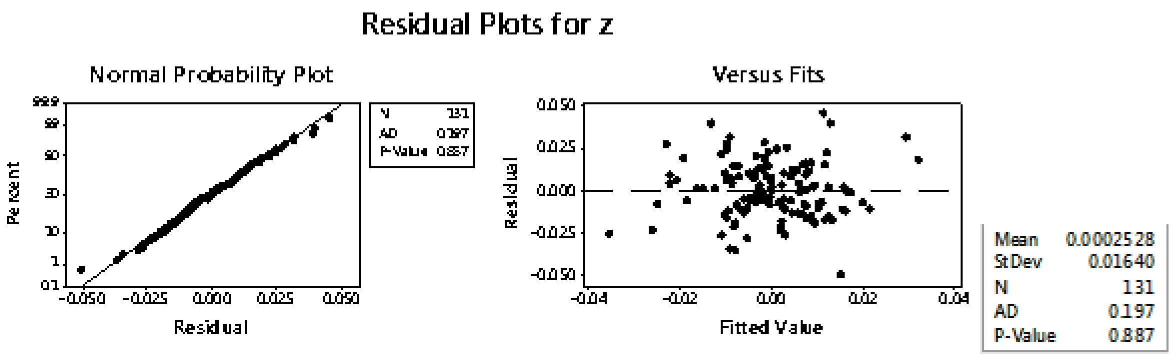
a. Determine which parameters should remain in the model and explain the reasons. Use the *p*-values associated with the *t*-statistics. Assume a 5% significance level.

b. Use Q^*-statistics (Ljung-Box) to test model adequacy, and explain the reasons. Use the appropriate *p* values and a 5% significance level for testing (state explicitly by writing the *p* values). If the significance level were 10%, would your answer be different? What do you think about the strength of the model adequacy?

c. Consider the autocorrelation and partial autocorrelation functions for the residuals (RSAC and RSPAC, respectively), the normality plot, and the residuals vs. fits plot below. Are the random shock assumptions ($a_t \sim i.i.d.\ N(0,\sigma^2)$) justified? State the assumptions and the reasons explicitly by referring to the plots and statistical tests whenever possible.







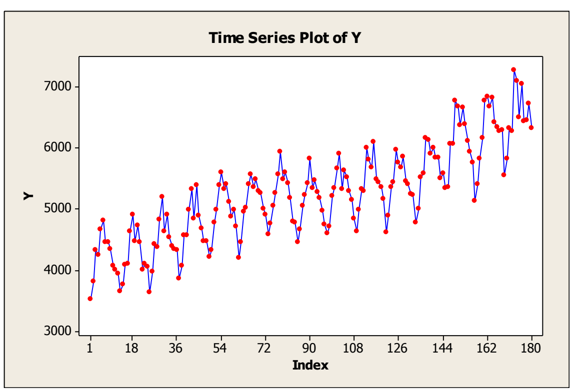
8. Miscellaneous Short-Answer Questions. Please answer the following questions.

a. Suppose that the true relationship between y and t is given by the nonlinear model: $y_t = \beta_0 \beta_1^t e^{\varepsilon}$. A natural log transformation was applied so that a simple linear regression solution can be found. The solution to the transformed model was obtained as $\hat{y}_t^* = 10 - 0.05\,t$. What is the estimated *percentage* increase/decrease from y_{t-1} to y_t?

b. The Holt-Winters method was applied to the quarterly sales data, and the initial seasonal factors for the first three quarters were estimated as $sn_{-3} = 105$, $sn_{-2} = 25$, $sn_{-1} = -45$. Determine the initial estimate for the seasonal factor of Quarter 4 (sn_0).

c. To obtain the initial estimates for the Holt's Trend Corrected method, we can fit a least squares trend line to the data and let the estimated β_1 be the initial level estimate l_0.
True or False? (Circle the correct answer.)

d. The U.S. beverage manufacturer monthly product shipments were recorded for 15 years, and the resulting time-series plot of monthly beverage shipments (in millions of dollars) is given below. Let y_t : Millions of dollars of beverage shipments in month t. Propose a time series regression model that will take into account a linear trend and seasonal (monthly) components by defining the appropriate variables.



e. A 95% prediction interval for y_t was calculated as [−7.78, 13.78] as a result of a Box-Jenkins model. Calculate a 99% prediction interval for y_t.