**Stat 510 Week 8 Homework**

**Submit answers in a Word document or a pdf to the designated ANGEL Drop Box in the Drop Boxes folder.**

1. Use the xhw8prob1.dat and yhw8prob1.dat datasets from the Week 8 folder. The tasks will be the same that we carried out in Lesson 8.1, so use the code provided there for guidance.

1. Plot the data. Do an ordinary simple linear regression relating  to. Give the estimated intercept and slope along with their standard errors as the answer to this part.

**Coefficients:**

**Estimate Std. Error t value Pr(>|t|)**

**(Intercept) 463.492 36.289 12.77 <2e-16 \*\*\***

**x 5.575 0.117 47.64 <2e-16 \*\*\***

**---**

**Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1**

**Residual standard error: 66.92 on 98 degrees of freedom**

**Multiple R-squared: 0.9586, Adjusted R-squared: 0.9582**

**F-statistic: 2270 on 1 and 98 DF, p-value: < 2.2e-16**

1. Examine the ACF and PACF of the residuals from the regression in part A. For this part, show the graphs and write a brief explanation of what ARIMA model may be appropriate for the residuals.



**Tapering pattern in the ACF along with single spike at lag 1 in PACF indicates AR(1) model for residuals.**

1. Let’s go through the Cochrane-Orcutt procedure. Estimate the coefficient(s) of the ARIMA model identified in part B. Remember – use the residuals from part A as the variable. Look at Lesson 8.1 for guidance. Give the estimated coefficient(s) as the answer to this part AND discuss whether the model is a suitable fit to the residuals.

**=0.6382 (from R sarima with no intercept).**

**Value might differ ever so slightly from this for students who used R with an intercept.**

**AR(1) model for residuals is suitable. Diagnostics such as Ljung-Box-Pierce tests and ACF of residuals look good.**

1. Carry out the necessary “adjustment regression.” Give the adjusted intercept and slope of the regression relationship between and . Also give the standard errors of these estimates. See Step 5 of the first example in Lesson 8.1 for guidance.

**Coefficients:**

**Estimate Std. Error t value Pr(>|t|)**

**(Intercept) 174.801 21.445 8.151 1.28e-12 \*\*\***

**xnew 5.502 0.186 29.580 < 2e-16 \*\*\***

**---**

**Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1**

**Residual standard error: 51.55 on 97 degrees of freedom**

**(2 observations deleted due to missingness)**

**Multiple R-squared: 0.9002, Adjusted R-squared: 0.8992**

**F-statistic: 875 on 1 and 97 DF, p-value: < 2.2e-16**

**Adjusted intercept = 174.801 / (1 - .6382) = 483.18**

**SE for adjusted intercept = 21.445/(1-.6382) = 59.28**

**Adjusted slope = 5.502**

**SE for adjusted slope = 0.186**

1. Use the Cochrane.orcutt command to iterate this procedure and obtain final estimates.

**Adjusted intercept = 483.62**

**SE for adjusted intercept = 59.68**

**Adjusted slope = 5.5**

**SE for adjusted slope = 0.187**

1. Using the arima command with the xreg option, estimate the coefficient(s) of the regression model and the ARIMA model for the residuals identified in part (B) with maximum likelihood. Give the estimated coefficient(s) and their test statistics as the answer to this part.

**Coefficients:**

**ar1 intercept xreg**

**0.6420 495.9328 5.4674**

**s.e. 0.0768 56.7209 0.1802**

**sigma^2 estimated as 2592: log likelihood = -535.16, aic = 1078.32**

**Test Statistics:**

**ar1 intercept xreg**

**8.362832 8.743385 30.334427**

1. Discuss whether the model in part (F) is a suitable fit. If not, repeat step (F) with an amended ARIMA model and refit using the arima command with the xreg option until the fit is suitable. You need only provide output for your final model.

**Diagnostics for the model in F look fine: insignificant Ljung-Box-Pierce tests and residual autocorrelations.**

****

2. Use the sales8.dat and lead8.dat datasets from the Week 8 folder. Each file contains 78 consecutive monthly values. The variable in the sales8.dat file is monthly sales of a product for 78 months. The variable in the lead8.dat file is a company indicator (predictor) of future sales measured in the same 78 months. For this problem, = sales and  = leading indicator. Use Lesson 8.2 for guidance on parts C to G.

1. Create a time series plot of the sales variable. Describe any noteworthy features.

**There appears to be an overall upward trend in sale. This trend is particularly steep at times. There is no evidence of seasonality, non-constant variance, or outliers.**



1. Create a time series plot of the leading indicator variable. Describe any noteworthy features.

**There appears to be a downward trend in the lead indicator. There is no evidence of seasonality, non-constant variance, or outliers.**



1. Determine the CCF for the relationship between  = leading indicator and = sales indicator. Give the plot as the answer for this part. Notice that it’s a little difficult to identify the main peak(s) because quite a few correlations are relatively large.



1. The plots in A and B suggest trends, so create variables that are the first differences for the sales variable and the leading indicator variable. Commands will be something like:

diffsales = diff(sales,1)

difflead = diff(lead,1)

Exact commands will depend upon what you called the variables when you read them in.

Next, determine the CCF for the relationship between the two sets of differences. The differenced leading indicator variable is the x-variable. Give the plot AND write a brief interpretation.

**There are significant correlations at lags 3 and 4 of the lead variables indicating that** *xt-3* **and** *xt-4* **most likely are predictors of. The CCF follows:**



1. Use the lag2.plot command to create scatter plots of the differenced sales data (as y) and the differenced leading indicator data (as x) for 6 lags. Describe noteworthy features. See Lesson 8.2 for guidance.



1. Carry out a regression with the original undifferenced sales variable as the y-variable and lags 3 and 4 of the original undifferenced leading indicator variable as the predictor variables. Give the regression results and for each of the predictor variables, indicate whether the variable has a statistically significant relationship with the y-variable.

*Teaching Note*: In parts D and E we used differenced data to identify a possible relationship. This is a variation of pre-whitening, a technique that we’ll discuss further next week. Pre-whitening removes common elements in the two series and that allows us to see the relationship more clearly. Usually, the relationship seen in the pre-whitened data holds for the original data. Thus we’re using the undifferenced data in part F.

**All terms are statistically significant.**

**Coefficients:**

**Estimate Std. Error t value Pr(>|t|)**

**(Intercept) 372.3698 2.1676 171.791 <2e-16 \*\*\***

**leadlag3 -0.3733 0.1553 -2.404 0.0189 \***

**leadlag4 -1.9107 0.1582 -12.078 <2e-16 \*\*\***

**---**

**Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1**

**Residual standard error: 1.692 on 70 degrees of freedom**

**Multiple R-squared: 0.9815, Adjusted R-squared: 0.981**

**F-statistic: 1855 on 2 and 70 DF, p-value: < 2.2e-16**

1. For the residuals from the model in part F, determine the ACF and PACF (of the residuals). Discuss a possible model for the residuals.

**Residuals appear to have an AR(1) or MA(1) pattern. Here are the ACF and PACF graphs:**



1. Using the arima command with the xreg option, estimate the coefficient(s) of the regression model and the ARIMA model for the residuals identified in part G with maximum likelihood. Give the estimated coefficient(s) and test statistics as the answer to this part. Because you cannot supply a data set to the arima command, you must access the columns of alldata as follows:

trend=time(alldata[,1])

adjreg = arima (alldata[,1], order = c(p,d,q), xreg = cbind(trend, alldata[,2], alldata[,3]))

*The brackets retrieve values from a matrix. The first entry corresponds to rows and the second to columns. Leaving the first entry blank tells R to retrieve all rows in the corresponding column. “Sales” is the first column in my alldata matrix. You may include more or fewer external regressors in the xreg option as you see necessary.*

**Coefficients:**

**ar1 intercept alldata[, 3] alldata[, 4]**

**0.3845 372.3176 -0.5384 -1.7454**

**s.e. 0.1132 3.0644 0.1767 0.1787**

**sigma^2 estimated as 2.367: log likelihood = -135.11, aic = 280.23**

**Test statistics:**

**ar1 intercept alldata[, 3] alldata[, 4]**

**3.395682 121.498230 -3.046761 -9.769658**

1. Discuss whether the model in part H is a suitable fit (e.g. acf/pacf, Ljung-Box, QQ-plot). If not, repeat step H and refit using the arima command with the xreg option until the fit is suitable.

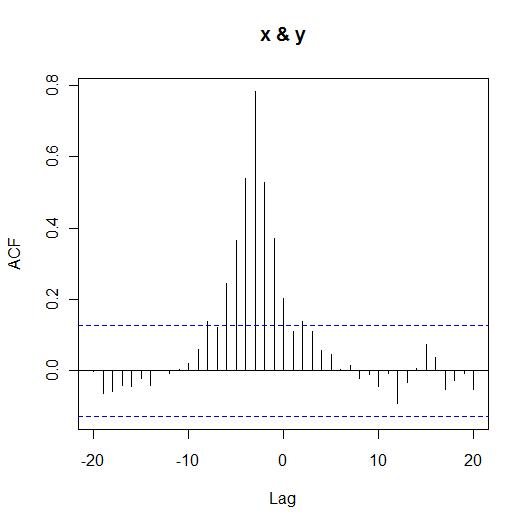
**An AR(1) is suitable as there are no remaining autocorrelations in the residuals and all Ljung-Box p-values are insignificant.**

**

3. For this problem, use the files xvar-week8prob3.dat and yvar-week8prob3.dat from the Week 8 folder. The files contain n = 240 simulated observations for a relationship between two time series. The *x*-variable is in xvar-week8prob3.dat. The *y*-variable is in yvar-week8prob3. Use Lesson 8.2 of this week for guidance on this problem.

A. Determine the CCF for the relationship between x and y. Write a brief interpretation – for instance, where is the peak value? What’s happening in the neighborhood of the peak value?

**Peak in CCF is at lag 3 (or -3, in R). The values taper in both directions from there.**



B. Carry out the regression. That is use lag 1 of y and lag 3 of x to predict y. For each of the two predictors, indicate whether the variable has a statistically significant relationship with the y-variable.

**Lag 1 of y is not statistically significant. Lag 3 of x is.**

**Coefficients:**

**Estimate Std. Error t value Pr(>|t|)**

**(Intercept) 0.28178 0.88636 0.318 0.751**

**ylag1 0.05484 0.04738 1.157 0.248**

**xlag3 1.56627 0.09788 16.001 <2e-16 \*\*\***

C. Now leave the lag 1 of y variable out of the model and carry out a regression with only lag 3 of x as a predictor. For the residuals from this model, determine the ACF and PACF (of the residuals), and interpret them.

**ACF and PACF of the residual show no significant correlations. Residuals look to be white noise.**