Homework 1

Christian Steinmetz
MATH 8090-Spring 2018

February 6, 2018

Problem 7. Consider fitting a model to the log of the Australian wine sales. Fit a model allowing for a term for each month. Fit the model with linear trend and monthly variables, plot the data and overlay the estimated values from the fitted model. Does it appear to provide a better fit than the sin/cos model from ppt? Use the Box test to test if there is significant autocorrelation in the residuals.

Solution The new model is shown in Figure 1 and appears to fit the data better than the sine and cosine model from class. The results from the box text are shown below.

```
Box-Ljung test
data:  lsfit$res
X-squared = 71.29, df = 15, p-value = 2.629e-09
```

A $p - value << 0.05$ indicates that there is still autocorrelation present in the fitted model.

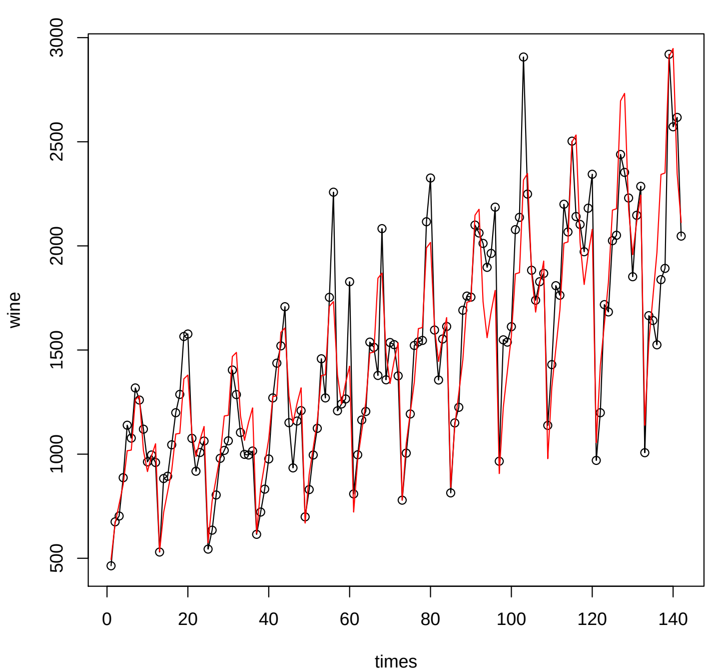

Figure 1: Model with linear trend and monthly variables.

Problem 8. Consider the global temperature deviations. Allow for a quadratic trend with AR(1) errors. You can mimic the R code for the lake Huron data.

(a) Are the coefficients for both linear and quadratic terms statistically significant? If not, drop that term and refit with just a linear trend.

(b) Make a 95% confidence interval for the rate of change and interpret the limits of this interval in terms of temperature increase per century.

Solution

(a) The quadratic model was fitted first and is shown in Figure 2. The results from the ARIMA model are shown below.

```
Coefficients:
          ar1   intercept    times     tsqr
       0.4180     -0.5108   0.0110   -1e-04
s.e.   0.1752      0.0615   0.0028    2e-04

sigma^2 estimated as 0.01477:
log likelihood = 74.28,  aic = -138.56
```

Since $|\frac{-1e-04}{2e-04}| = \frac{1}{2} < 1.96$, the quadratic term is not significant. Therefore we refit the with a linear model, dropping the quadratic term. The output of this model is shown in Figure 3, and we observe that the fit is similar. The output of the new ARIMA model is shown below. Notice that the times values are clearly significant.

```
Coefficients:
          ar1   intercept    times
       0.4877     -0.4126   0.0056
s.e.   0.0836      0.0459   0.0007

sigma^2 estimated as 0.01534:
log likelihood = 72.18,  aic = -136.35
```

(b) Based on our fitted model, we estimate with 95% confidence that the average rate of change of the global temperature is increasing by at least $(0.0056 - (1.645 * 0.0007)) = 0.0044485$ degrees per year or 0.44485 degrees per century.

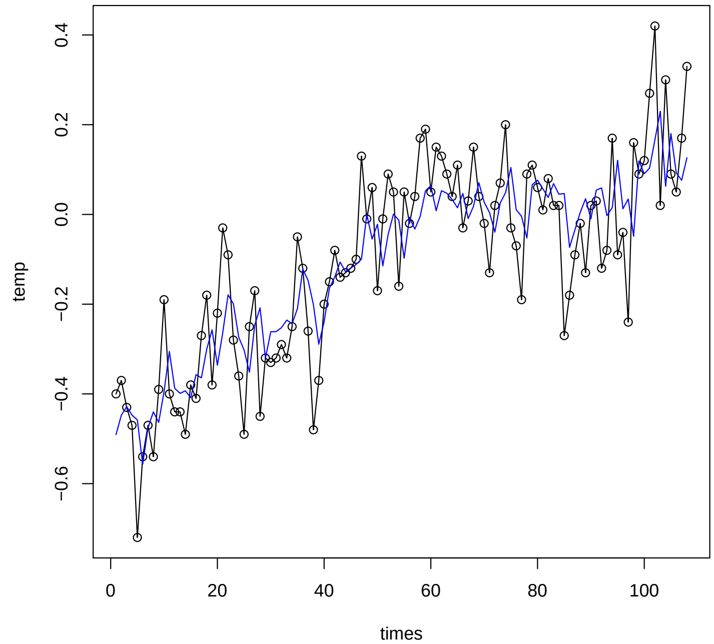

Figure 2: Quadratic model with AR(1) errors.

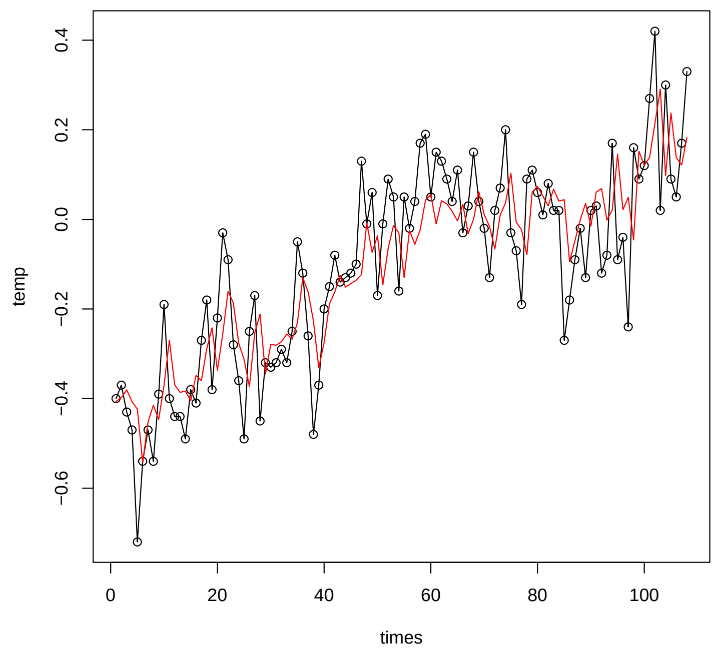

Figure 3: Linear model with AR(1) errors.

Problem 9. Use R to

(a) fit a cubic trend to the lake data: $X_t = \beta_0 + \beta_1 t + \beta_2 t^2 + \beta_3 t^3 + Y_t$.

(b) plot the lake data with estimated cubic trend overlay.

Find residuals from the estimated trend and plot the sample ACF. Use the Ljung- Box test to check for statistically significant autocorrelation.

Solution

(a) The fitted trend is shown from the output below.

```
              Estimate    Std. Error  t value  Pr(>|t|)
(Intercept)   1.131e+01   4.327e-01   26.129    <2e-16 ***
Xtimes       -8.972e-02   3.766e-02   -2.383    0.0192 *
X             6.415e-04   8.814e-04    0.728    0.4686
X             2.288e-07   5.854e-06    0.039    0.9689
```

(b) The cubic trend is shown in Figure 4.

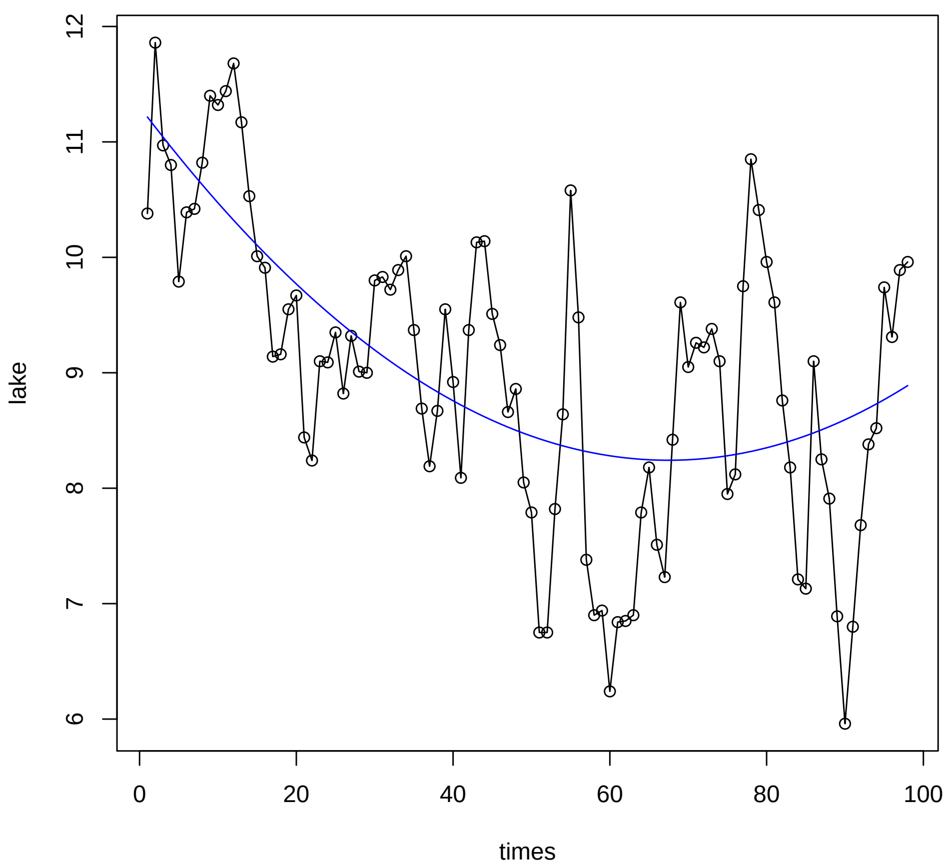

Figure 4: Cubic model for the Lake Huron data.

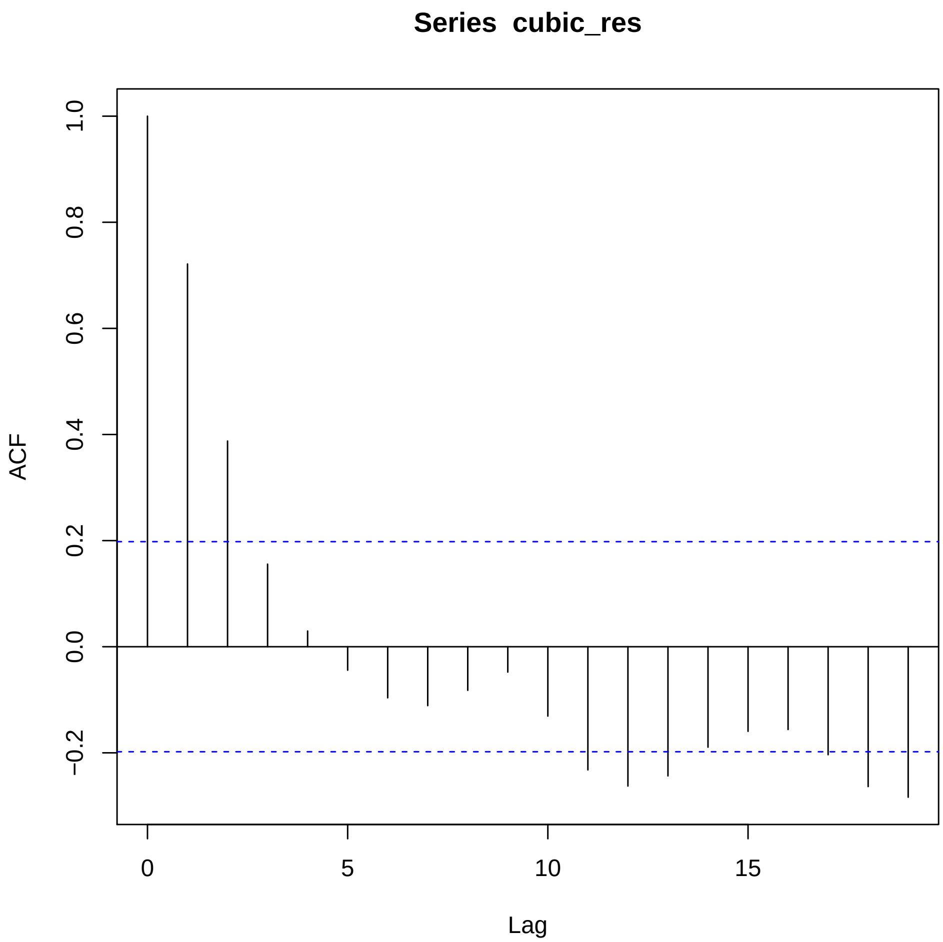

Figure 5: Cubic model for the Lake Huron data.

A box plot of the residuals from the model is shown in Figure 5. Significant autocorrelation is observed at multiple lags after lag 0. The Ljung Box Test confirms this. Its output is shown below.

```
Box-Ljung test
data:  cubic_res
X-squared = 103.9, df = 15, p-value = 2.331e-15
```

A $p - value << 0.05$ indicates that there is still autocorrelation present in the fitted model.

Problem 10. Rather than estimate first order trend and seasonal components, we can use differencing. In R we can difference using diff(data,lag,times) where lag determines the lag at which we difference (at the period for seasonal data) and times determines how many times (k to remove a kth order polynomial).

(a) Difference the log of wines sales at lag=12. Plot the differenced data. Does there appear to be any remaining trend? Test for autocorrelation in differences using Box-Ljung. Notice the differencing removes first order, but not second order effects!

(b) Difference the baseball data at lag=1. Plot the differenced data and test for autocorrelation.

Solution

(a) The differenced wine data is shown in Figure 6 and the result of the Ljung Box Test is shown below.

```
Box-Ljung test
data:  y_diff
X-squared = 63.636, df = 15, p-value = 5.917e-08
```

A $p - value << 0.05$ indicates that there is still autocorrelation present in the fitted model, even once we performed first order differencing on the data.

(b) The differenced wine data is shown in Figure 7 and the result of the Ljung Box Test is shown below.

```
Box-Ljung test
data:  baseball_diff
X-squared = 13.502, df = 15, p-value = 0.5636
```

A $p - value > 0.05$ indicates there is likely no autocorrelation present in the fitted model once we performed first order differencing on the data.

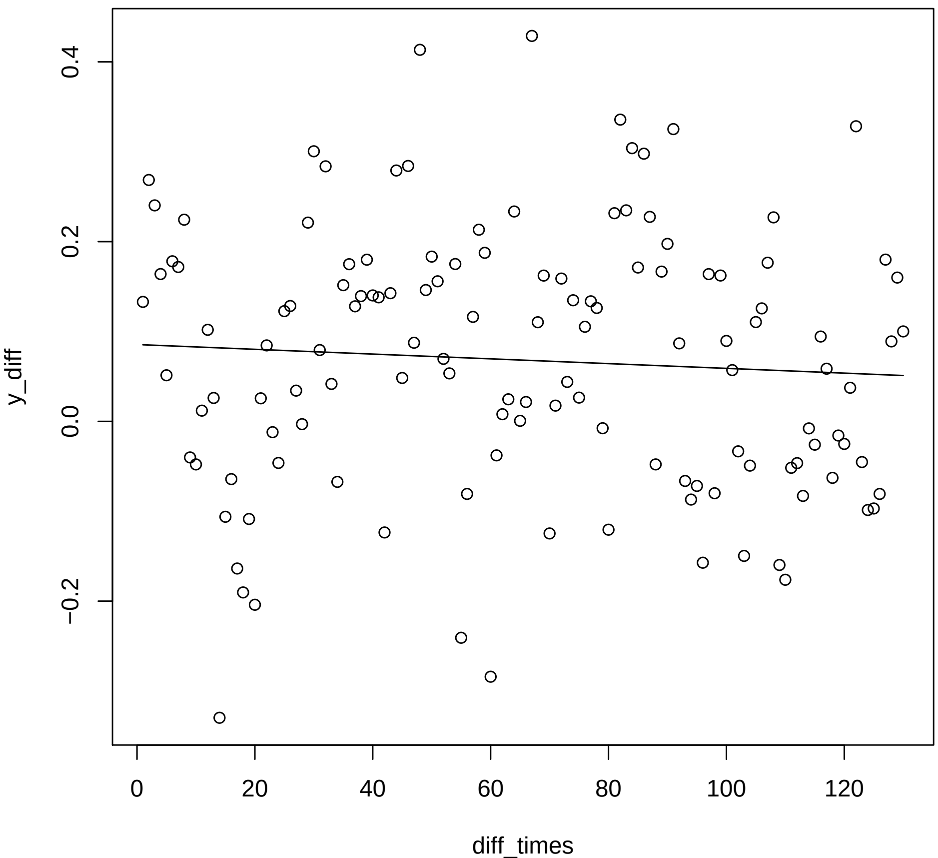

Figure 6: First order differncing on wine data with lag=12.

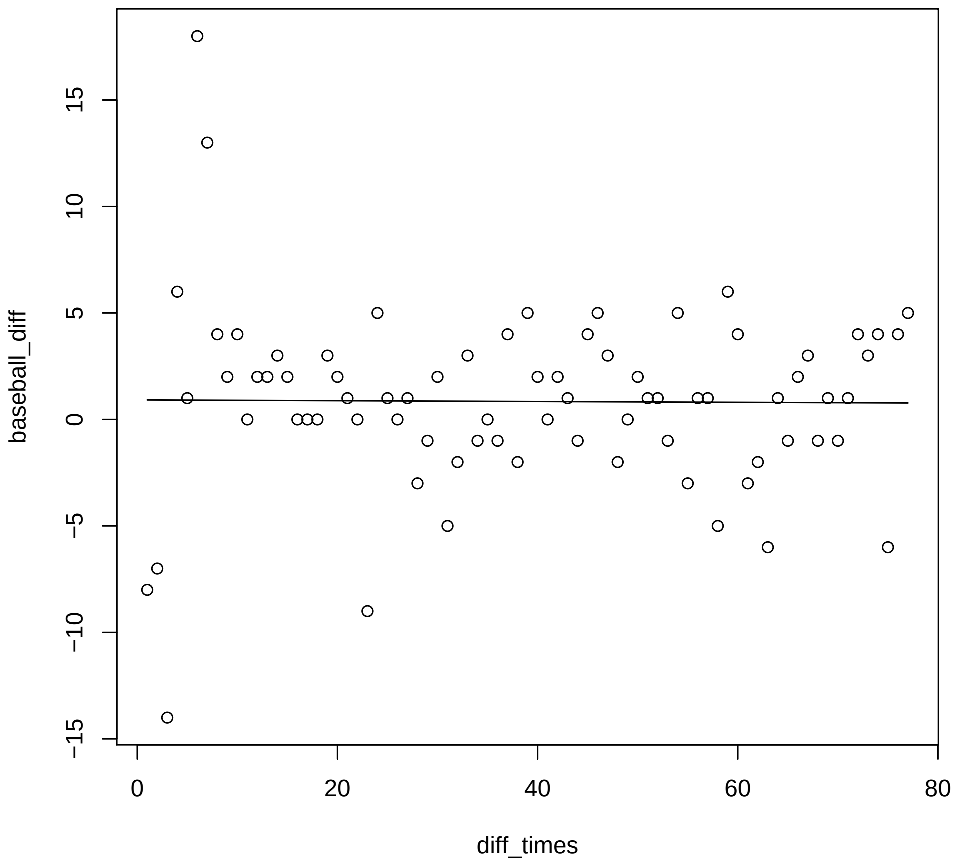

Figure 7: First order differncing on baseball data with lag=1.