Statistics 435/535/711 16 February 2022
P. Shaman

Assignment 2—due 2 March at 11:59 pm

You may work together for this assignment. However, perform calculations and the writeup by yourself. File sharing is not permitted. Submit via Canvas. Your submission should be either a Word or a pdf file. R Markdown documents will not be graded.

Your submission should include clear focus on discussion and interpretation of the statistical results in your presentation.

The data set for this assignment gives used car sales in the U.S. monthly, in millions of dollars, for the period 1992 through 2021(11). The data are in the file UsedCarSales.txt.

To do this assignment, you need to construct the *Time* variable and convert the class for *Month* to factor class. Then add these new variables to the data frame [sales<-data.frame(sales, Time, fMonth, if the name of the data frame you read in is sales]. If you create any other new variables, for example, outlier dummies, they also need to be added to the data frame.

1. During the years 1992 to 2021 the Business Cycle Dating Committee of the National Bureau of Economic Research defined three periods of economic contraction, 2001(4) to 2001(11), 2008(1) to 2009(6), and 2020(3) to 2020(4). Make separate time series plots for (i) Sales and (ii) log Sales, and mark the contraction periods on the plots. Discuss and compare the two plots. Comment on trend structure, seasonality, and volatility. Do the plots reveal any unusual features? If yes, describe what is notable and discuss the underlying causes. Do the two plots indicate whether an additive decomposition model or a multiplicative decomposition model should be fit to model Sales? Explain your answer.

2. Form a spectral plot of the log Sales series. The purpose is to gain information about the structure of the time series and give guidance in the formation of a multiplicative decomposition model in the next part. Mark important frequencies on the plot, and explain in detail what it reveals.

3. Using the data for the years 1992 to 2019, fit a multiplicative decomposition model to the Sales data. Include a polynomial trend and a static seasonal component using month dummies. If the spectral plot suggests calendar structure is present, include relevant trigonometric variable pairs. If you find an outlier value which warrants use of a dummy for adjustment, form the dummy and include it in your model (remember to add outlier dummies you create to the data frame). If a calendar pair or an outlier dummy you have included is not significant, remove and refit.

(a) Construct estimates of the static seasonal indices, and display them in a table and a plot. Describe the indices. Explain why the seasonal peaks and valleys occur in the months indicated by the estimates.

(b) Perform a residual analysis for the fitted model. Include a normal quantile plot of the residuals and the Shapiro–Wilk test, a plot of the residuals vs. time, a residual acf plot, and a residual spectral plot. Interpret the results of each. Is there reduction to white noise by the model?

4. Visual inspection of the plots in part 1 suggests the seasonal structure is dynamic. To explore this, add the following variables to your model in part 3: *Dynamic* and *Dynamic*fMonth*. This addition will produce separate static seasonal estimates for the time spans 1992 to 2010, and 2011 to 2019. For consideration of this methodology, review the analysis of the variety stores sales data in the 7 February notes, where *wtgrant* corresponds to *Dynamic*, and *Time* corresponds to *fMonth.* There the purpose is to estimate the change in trend, and here the purpose is to estimate dynamic seasonal structure.

(a) Is the addition of the new variables a statistically significant step? Answer with an appropriate test result.

(b) Construct two sets of static seasonal index estimates, for the two time spans. Code to do this is in the following illustration.

Assume the output of the model is such that, line 1 is for the intercept, lines 8 to 18 are for *fMonth*, line 22 is for *Dynamic*, and lines 23 to 33 are for *Dynamic*fMonth.*

For the first time span:

```
b1<-coef(model)[1]
b2<-coef(model)[8:18]+b1
b3<-c(b1,b2)
seas1<-exp(b3-mean(b3))
```

For the second time span:

```
b1<-coef(model)[1]+coef(model2)[22]
b2<-coef(model)[8:18]+coef(model2)[23:33]+b1
b3<-c(b1,b2)
seas2<-exp(b3-mean(b3))
```

Tabulate and plot these two sets of estimates, and also the estimates in part 3. Describe them and discuss the differences between the two time spans. Why do you think the indicated changes have occurred?

(c) Perform a residual analysis of the model as in part 3(b).

5. Construct the lag 1 and lag 2 residuals for the model in part 4, and add them to the model. Now perform a residual analysis of this new model and discuss the results. [Be sure to add the two lagged residuals variables to the data frame.]

6. Write a summary of the results obtained from parts 3 through 5. In your discussion discuss what the analysis has revealed about U.S. used car sales.