Risto B. Rushford
GSCM 451-Spr 2019
HW07

Data Analysis Concepts

Briefly comment on each of the following concepts, usually just several sentences.

1. How is an exponential smoothing forecast a type of weighted average?

Exponential smoothing is a moving weighted average, meaning that as part of the means of determining a future forecast value, the smoothing coefficient α is used to set up a weighting system. The system usually uses stronger weights for the most recent values, with the weights exponentially decreasing in value until they reach zero.

2. What does it mean to say that an exponential smoothing forecast is self-adjusting?

The moving weighted averages have an effect on an exponential smoothing forecast such that the series quickly adjusts to changes in pattern behavior. The larger the smoothing coefficient, the more reactive the series is to the changes, which can be problematic if the model still has a lot of random error.

3. What two values are needed in addition to the smoothing coefficient α to generate an exponential smoothing forecast?

Y_t which is the real current value of Y, and $\hat{Y}_t$ which is the current forecasted value of Y.

4. How do the coefficients across the time values differ in an exponential smoothing forecast for values of α near 1 versus values near 0?

For α values near one, there are fewer time value coefficients (as the associated weight of the coefficients must sum to ≤ 1. The higher α values allow for quick self-adjustment of the forecast to real data as it is entered into the model, however the trade-off for the quick adjustment is high reactivity to modeling errors.

5. What is simple exponential smoothing? What is its major limitation?

Simple exponential smoothing (SES) is the most basic form of exponential smoothing fitting the descriptions above. The limitations of SES are that it cannot account for seasonality and trend.

6. What is an important advantage of Holt-Winters exponential smoothing compared to linear regression with trend and seasonal adjustments?

Linear regression with trend and seasonal adjustments has the disadvantage of not self-adjusting automatically from newly collected data. However, the Holt-Winters' seasonal method three smoothing coefficients (ℓ_t for level, b_t for trend, and s_t for seasonality) to account for trend and seasonality, which allows for continued self-adjusting beyond the SES model.

Worked Problems

1. Australian tourist data

The following data set consists of quarterly visitor nights (in millions) spent by international tourists to Australia from the first quarter of 1999 through 2015.

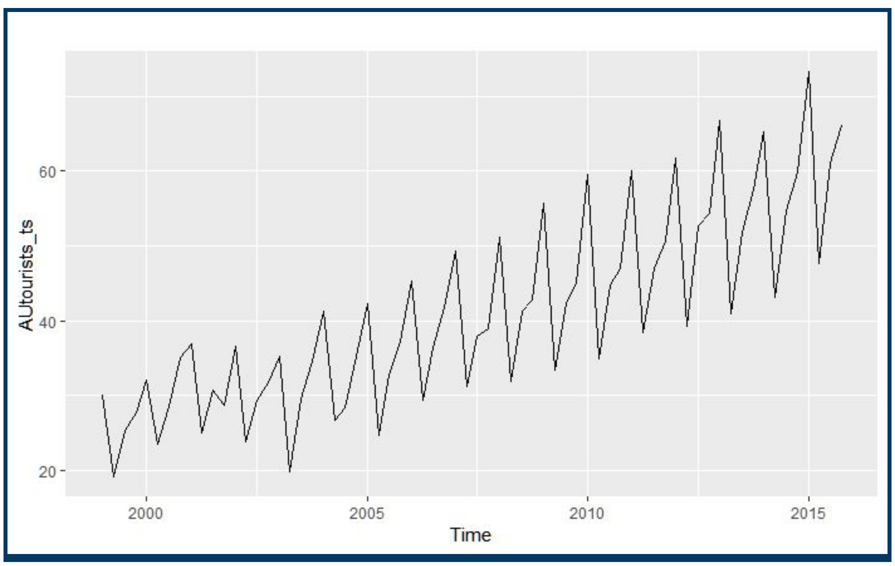
Data: http://web.pdx.edu/~gerbing/451/Data/aus_tourists.xlsx

Forecast the visitor nights through 2019.

a. Create the time series.

```
AUtourists_ts <- ts(d$tourists, frequency=4, start=c(1999,1))
```

b. Plot the time series.



c. Verbally describe the major characteristics of the time series.

This time series shows a semi-biennial seasonality with an upward trend.

d. Given the data would you forecast with simple exponential smoothing, Holt's method, or the Holt-Winters method? Why?

Given this information, I would forecast with Holt-Winters' method. Seeing that the data exhibits a clear seasonality (with one max peak and min trough in each yearly period) and an upward overall trend, the correct method of forecasting must take each of these factors into account.

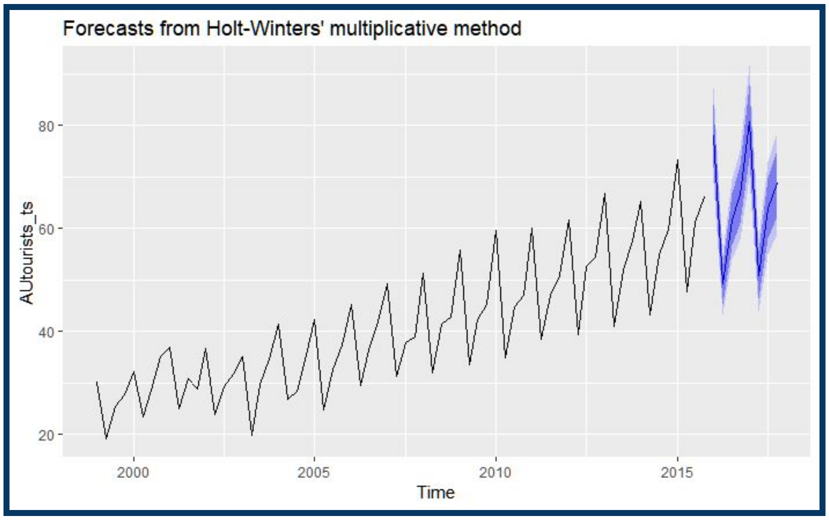
e. Given the data, would you forecast with an additive or multiplicative model?

Another factor to consider in the data is the growing gap between the local maximums and minimums, suggesting that a multiplicative model would be best to use.

f. Generate the numerical values of the forecast and the corresponding prediction intervals.

	Point Forecast	Lo 80	Hi 80	Lo 95	Hi 95
2016 Q1	77.97306	71.94180	84.00433	68.74903	87.19709
2016 Q2	49.13475	45.20183	53.06768	43.11986	55.14964
2016 Q3	61.64632	56.55395	66.73868	53.85822	69.43442
2016 Q4	66.49844	60.84265	72.15422	57.84866	75.14822
2017 Q1	80.82401	73.72354	87.92448	69.96478	91.68324
2017 Q2	50.91502	46.32846	55.50158	43.90049	57.92956
2017 Q3	63.85989	57.97028	69.74950	54.85250	72.86728
2017 Q4	68.86502	62.37164	75.35841	58.93425	78.79580

g. Plot the data, the forecasted values, and the prediction intervals.



h. What is the standard deviation of the residuals for the fitted data?

```
sigma:  0.0604
```

i. Comment on the size of the prediction intervals. Why are they this size?

For a two-year period, the prediction intervals begin at Q1 2016 with a point forecast of 77.97±6.03 (80% interval) or 77.97±9.23 (95% interval). At Q1 2017, the forecast is 80.82±7.19 (80% interval) or 80.82±10.86 (95% interval). The reason for the widening intervals is my choice of using the multiplicative model to forecast. The model uses a multiplier to steadily increase the range of the seasonal effects in the forecast.

2. Sales data

Consider the following quarterly sales data beginning in the first quarter of 2001.

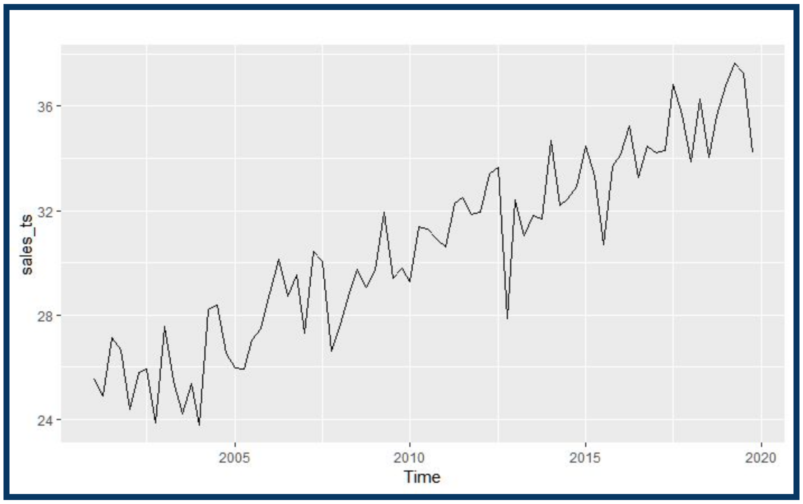
Data: http://web.pdx.edu/~gerbing/451/Data/HW9_2.xlsx

Generate the forecast for four years after the last data point.88

a. Create the time series.

```
sales_ts <- ts(d$Sales, frequency=4, start=c(2001,1))
```

b. Plot the time series.



c. Verbally describe the major characteristics of the time series. Is the data seasonal? Justify your answer.

The data does seem to have some seasonality, but there is a lot of variability in the model which seasonality can't fully account for. I do see a consistent number of local max and min values in each five year period.

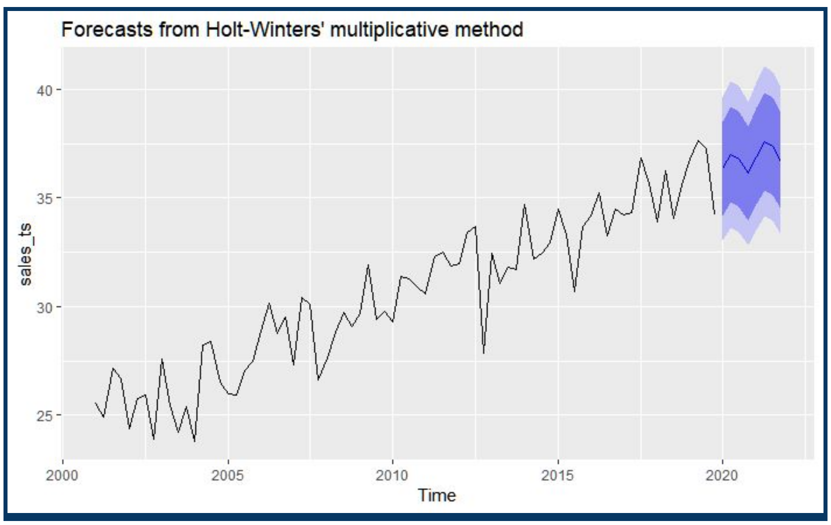
d. Given the data would you forecast with simple exponential smoothing, Holt's method, or the Holt-Winters method? Why?

Because there is a clear upward trend and there is some discernible seasonality, I would predict again using the Holt-Winters method.

e. Generate the numerical values of the forecast and the corresponding prediction intervals.

	Point Forecast	Lo 80	Hi 80	Lo 95	Hi 95
2020 Q1	36.33990	34.18280	38.49701	33.04090	39.63891
2020 Q2	37.00833	34.80557	39.21109	33.63950	40.37716
2020 Q3	36.79664	34.60058	38.99270	33.43806	40.15522
2020 Q4	36.12697	33.96513	38.28880	32.82072	39.43321
2021 Q1	36.91946	34.70436	39.13456	33.53175	40.30717
2021 Q2	37.59620	35.33461	39.85779	34.13739	41.05501
2021 Q3	37.37884	35.12451	39.63317	33.93114	40.82653
2021 Q4	36.69632	34.47749	38.91514	33.30292	40.08972

f. Plot the data, the forecasted values, and the prediction intervals.



g. Comment on the size of the prediction intervals. Why are they this size?

The prediction intervals range as they do to account for the random error in the model, thus giving a greater range that the future sales values will fall into with an 80 or 90% probability.

h. What is the standard deviation of the residuals?

```
sigma:  0.0463
```

i. What is another forecasting technique we have previously studied that could also be applied to forecasting from this time series?

Another forecasting technique that we've studied and could use for this analysis is the STL (Seasonal and Trend decomposition using Loess) method. For a data set like this, STL would work well because it can compensate for significant outliers, such as the sharp drop midway between 2010-2015.