File Edit View Run Kernel Tabs Settings Help

Demand Forecasting for a Fa

Markdown

Python 3 (ipykernel)

Supply Chain and Logistics, Imperial College Business School

Demand Forecasting for a Fast Food Restaurant

CID: 02005361

Introduction

This assignment is tasked with forecasting how much lettuce should be procured for a fast food restaurant, Subway, in Berkeley California, USA. The data available contains records of over 7,000 orders placed between 15th March 2015 and 15th June 2015, at store 46673. The data tables of interest to us are tables 1 and 2 for order information, tables 4, 5, 6, and 7 for recipe information, tables 8 and 9 for subrecipe information, and tables 10 and 11 for ingredients. Using this dataset our objective is to calculate a forecast of the demand for lettuce over the following two weeks, June 16-29th 2015, using the ARIMA model and the Holt-Winters model.

Data Import & Prep

•••

•••

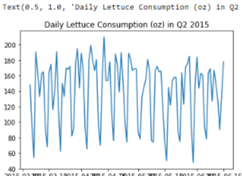
```
No duplicate records.
```

•••

Timeseries of Lettuce Consumed from March-June 2015 (oz)

•••

[4]:
```
Text(0.5, 1.0, 'Daily Lettuce Consumption (oz) in Q2 2015')
```



ARIMA Analysis

The Autoregressive Integrated Moving Average (ARIMA) forecasting model utilizies autoregressive and moving average components and is a flexible kind of model that can be applied to a broad set of case problems and industries. There are three parameters that we need to pay attention to for this model: p, the count of lagged observations, q, the moving average, and d, the differentiation. We can identify these parameters through the three methods below.

Stationarity

•••

•••

```
ADF Statistic: -2.242017815518828
p-value: 0.19135022390069556
Interpretation: The series is Non-Stationary.
```

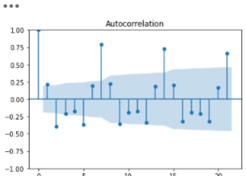
Observations:

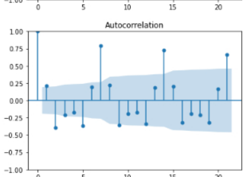
We do not have stationarity, which we also could have inferred from the timeseries graph showing a vast discrepancy on a given day. But this check gives us the evidence to estimate our model where d=1. If we did find there to be stationarity, we would have set d=0.

Autocorrelation Function (ACF)

•••

[7]:



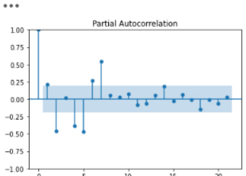


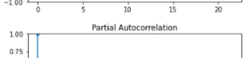
The ACF calculates the correlation between a time series and its lagged values, capturing the degree of similarity between a given time and the time minus lag. We run ACF to get the p parameter for our ARIMA model estimation. In practice, we must count the number of lags before a massive spike cutoff in the chart, and in this case we stop at the first lag, so p=1.

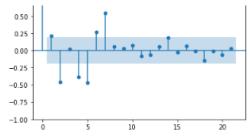
Partial Autocorrelation Function (PACF)

•••

[8]:







The PACF accounts for the effects of all intermediate lags before calculating the correlation between a time series and its lagged values. We run PACF to get the q parameter for our ARIMA model estimation. Here we also stop at the first lag, so q=1.

ARIMA Model

•••

•••

```
BIC Model Fit: 1062.7167106720797
ARIMA avg forecasted lettuce per day (oz): 140.59346942206795
```

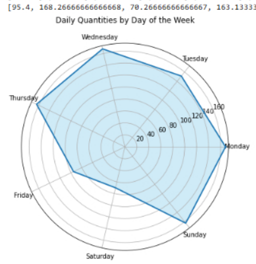
Holt-Winters Analysis

Also known as triple exponential smoothing, which refers to the value level, trend, and seasonality behaviors of the time series. The major advantage of the Holts-Winter model over ARIMA is that it can account for seasonality, and does not need too much historical data to be accurate. For our weekly seasonality, we will need to estimate our model with the cadence of seasonality set at 7.

Seasonality - lettuce consumption by day of the week (oz)

•••

```
['Friday', 'Monday', 'Saturday', 'Sunday', 'Thursday', 'Tuesday', 'Wednesday']
[95.4, 168.26666666666668, 70.26666666666667, 163.13333333333333, 164.0, 151.28571428571428, 167.92857142857142]
```



Observations:

As suggested in our timeseries graph, we anticipated that there would be a difference in average lettuce consumption depending on the day. We can see in the polar chart above that there is a distinct decrease in average daily consumption on Fridays and Saturdays, whereas the demand on the other days or consistent. From this information we can gather that there is weekly seasonality to the demand of lettuce, and this will need to be taken into account in the Holt-Winters model.

Holt-Winters Model

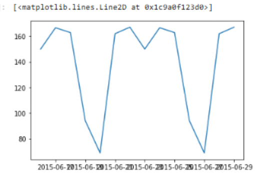
•••

•••

```
Holt-Winters avg forecasted lettuce per day (oz): 138.6948327146055
BIC Model Fit: 631.233538669927
```

•••

[14]: `[<matplotlib.lines.Line2D at 0x1c9a0f123d0>]`



Observations:

In forecating the following two weeks of lettuce demand, the Holt-Winters model is able to project for the Fridays and Saturdays that are going to have much lower consumption than the rest of the weekdays. Based on the clear observed seasonality it makes the Holt-Winters model a good candidate for our final forecast.

Forecast for June 16th-30th 2015

•••

```
            ARIMA Forecast (oz)  Holt-Winters Forecast (oz)
Date
2015-06-16           148.482754                  149.840835
2015-06-17           141.767396                  166.483012
2015-06-18           140.239610                  162.727629
2015-06-19           139.892029                   94.103952
2015-06-20           139.812952                   68.958094
2015-06-21           139.794962                  161.824664
2015-06-22           139.790869                  166.925642
2015-06-23           139.789938                  149.840835
2015-06-24           139.789726                  166.483012
2015-06-25           139.789678                  162.727629
2015-06-26           139.789667                   94.103952
2015-06-27           139.789664                   68.958094
2015-06-28           139.789664                  161.824664
2015-06-29           139.789663                  166.925642
```

Observations:

Side-by-side we can observe that the ARIMA model provides a constant baseline whereas the Holt-Winters forecast provides a much better prediction with the lower or higher demand figures on a given weekday. The implication of choosing a better model for the forecast lies in saving costs by not buying too much lettuce, in guaranteeing customer satisfaction by not running out of lettuce, and in keeping the ingredient as fresh as possible. Based on these considerations, it is clear that the Holt-Winters model is the optimal forecast to accomodate for the seasonality of the weekdays.

Business Recommendation:

Knowing that Fridays and Saturdays has much less demand for lettuce than any other days of the week, weekly shipments could be delivered on Satuday evenings or Sunday mornings. This ensures the freshest stock for the most consumption during the week, and also leaves the least fresh lettuce for the days where there is much less consumption.

Residual Analysis

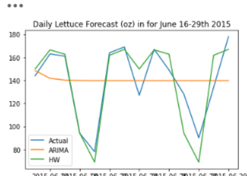
Mean Squared Error (MSE) is a measure of the average squared difference of the actual versus predicted values. In other words, this helps us get a better idea of how well our model fits the data. Models with better fits will have lower MSEs.

•••

```
Mean Squared Error for ARIMA: 954.875532440156
Mean Squared Error for Holt-Winters: 238.50066902974095
```

Model Fit

•••



To finalize our forecast, we choose the Holt-Winters model for two reasons. First, the residual analysis suggests it is a far better fitting model than ARIMA, and second, triple exponential smoothing is the optimal method to deal with the heavy weekday seasonality of lettuce consumption. It also gives confidence in our decision after having vizualized the demand forecast on a timeseries chart and observe how the Holt-Winters predictions closely match past consumption.

Final Forecast

•••

[18]:

	Store	California 1 (ID:46673)
0	2015-06-16	149.840835
1	2015-06-17	166.483012
2	2015-06-18	162.727629
3	2015-06-19	94.103952
4	2015-06-20	68.958094
5	2015-06-21	161.824664
6	2015-06-22	166.925642
7	2015-06-23	149.840835

8	2015-06-24	166.483012
9	2015-06-25	162.727629
10	2015-06-26	94.103952
11	2015-06-27	68.958094
12	2015-06-28	161.824664
13	2015-06-29	166.925642

•••

Improving the ARIMA model with revised ACF/PACF lags

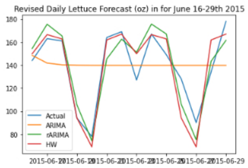
In the ACF and PACF charts we noticed that the first lag cutoff is at p=1 and q=1, however, if we look past this we can see than the next noticeable lag is at 7, which matches with the days in the week. When we update our model with p=7 and q=7, we see that the revised ARIMA (rARIMA) forecast is much more accurate, where the bic model fit is much lower and the mean squared error is very close to the Holt-Winters residuals.

•••

•••

```
BIC Model Fit: 975.3148162899488
rARIMA avg forecasted lettuce per day (oz): 140.18749536544556
Mean Squared Error for Revised ARIMA: 204.59568929447957
```

•••



•••

[23]:

Date	ARIMA Forecast (oz)	rARIMA Forecast (oz)	Holt-Winters Forecast (oz)
2015-06-16	148.482754	154.378726	149.840835
2015-06-17	141.767396	175.516418	166.483012
2015-06-18	140.239610	165.351343	162.727629
2015-06-19	139.892029	105.513538	94.103952
2015-06-20	139.812952	74.287818	68.958094
2015-06-21	139.794962	145.336834	161.824664
2015-06-22	139.790869	162.517929	166.925642
2015-06-23	139.789938	151.116668	149.840835
2015-06-24	139.789726	175.667900	166.483012
2015-06-25	139.789678	166.998263	162.727629
2015-06-26	139.789667	105.988431	94.103952
2015-06-27	139.789664	75.159173	68.958094
2015-06-28	139.789664	143.298128	161.824664
2015-06-29	139.789663	161.493766	166.925642

•••
