Project: Forecasting Sales

Complete each section. When you are ready, save your file as a PDF document and submit it here: https://classroom.udacity.com/nanodegrees/nd008/parts/edd0e8e8-158f-4044-9468-3e08fd08cbf8/project

Step 1: Plan Your Analysis

Look at your data set and determine whether the data is appropriate to use time series models. Determine which records should be held for validation later on (250 word limit).

Answer the following questions to help you plan out your analysis:

1. Does the dataset meet the criteria of a time series dataset? Make sure to explore all four key characteristics of a time series data.

   Yes the dataset meets the time series dataset criteria. They are continuous over a time interval, one data point at each time unit, sequential across intervals and equal spacing between each interval

2. Which records should be used as the holdout sample?

   As per the business problem, we need to forecast for 4 months period, so the hold out sample would be from June 2013 to September 2013.

Step 2: Determine Trend, Seasonal, and Error components

Graph the data set and decompose the time series into its three main components: trend, seasonality, and error. *(250 word limit)*

Answer this question:

1. What are the trend, seasonality, and error of the time series? Show how you were able to determine the components using time series plots. Include the graphs.

Fig. 1

From the above graph, Fig. 1 we can see that there is an increasing trend in seasonality and a linear trend is observed in trend so we can apply multiplication for seasonality and addition for trend whereas for error plot, there is no trend hence multiplication can be applied respectively.

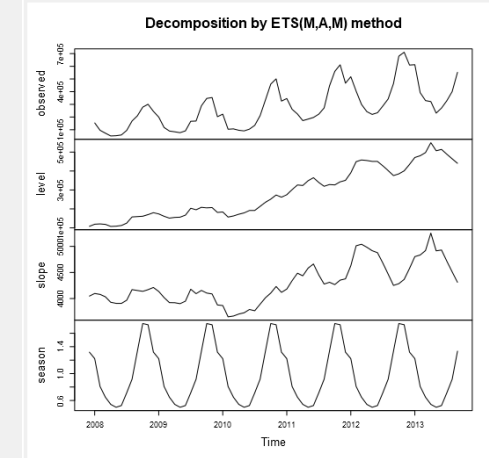
Step 3: Build your Models

Analyze your graphs and determine the appropriate measurements to apply to your ARIMA and ETS models and describe the errors for both models. (500 word limit)

Answer these questions:

1. What are the model terms for ETS? Explain why you chose those terms.
    a. Describe the in-sample errors. Use at least RMSE and MASE when examining results

    For ETS model MAM (Multiplication, Addition and Multiplication) term is taken into consideration we can see in the Fig. 1.



Method:
ETS(M,A,M)

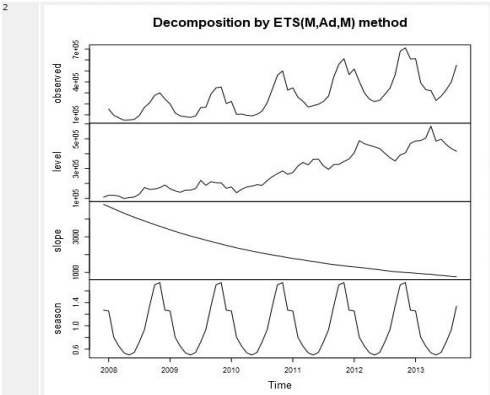
In-sample error measures:

ME	RMSE	MAE	MPE	MAPE	MASE	ACF1
2172.6846897	35853.2829726	27104.508479	-0.9328833	10.4804416	0.4003516	0.3589597

Information criteria:

AIC	AICc	BIC
1750.87	1762.87	1788.8498

Fig. 2



Method:
ETS(M,Ad,M)

In-sample error measures:

ME	RMSE	MAE	MPE	MAPE	MASE	ACF1
4554.4079516	32286.2730969	24978.2677011	-0.017139	9.9907989	0.3689456	0.1066282

Information criteria:

AIC	AICc	BIC
1744.6597	1758.3397	1784.8736

Fig. 3

By comparing the RMSE and AIC value of Dampened model (32286.27, 1744.66) vs Non-Dampened model (35853.28, 1750.89) it is evident that the Dampened value performs better since the RMSE and AIC value is lower than the other

2. What are the model terms for ARIMA? Explain why you chose those terms. Graph the Auto-Correlation Function (ACF) and Partial Autocorrelation Function Plots (PACF) for the time series and seasonal component and use these graphs to justify choosing your model terms.

b. Describe the in-sample errors. Use at least RMSE and MASE when examining results
c. Regraph ACF and PACF for both the Time Series and Seasonal Difference and include these graphs in your answer.

For building ARIMA model we need to find out the model terms. Fig. 4 gives the ACF and PACF graph without any difference.

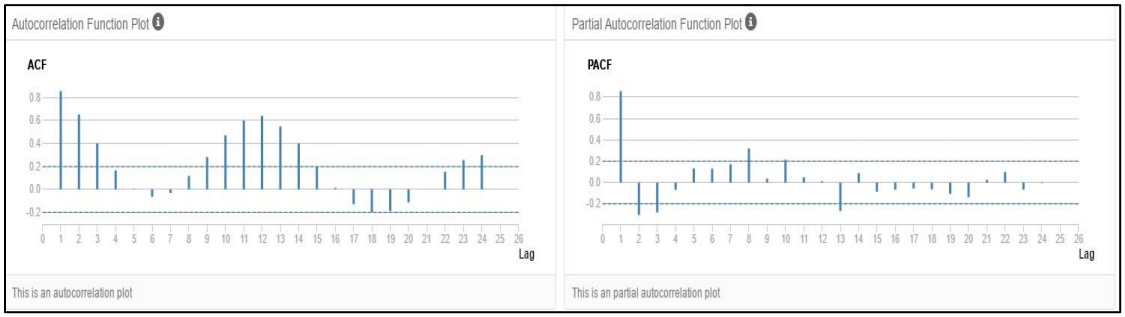


Fig. 4 ACF and PACF (Intial)

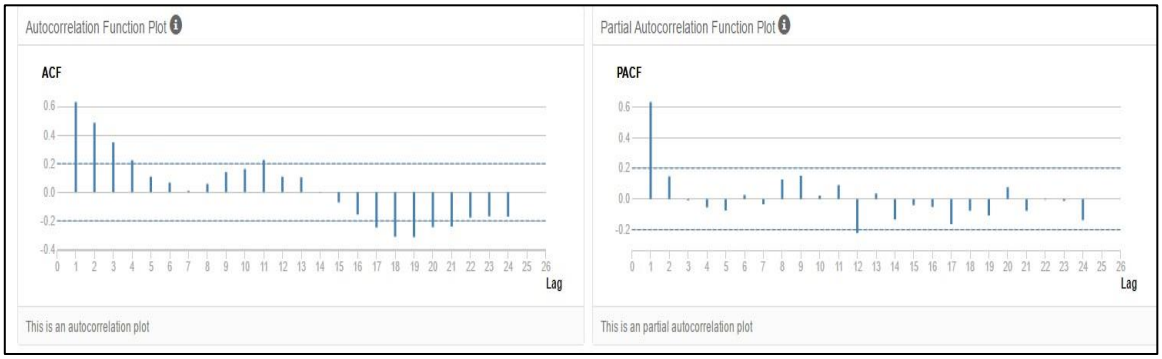


Fig. 5 ACF and PACF with seasonal difference

From the Fig. 5 we can see that after seasonal difference there is still strong correlation in ACF plot.

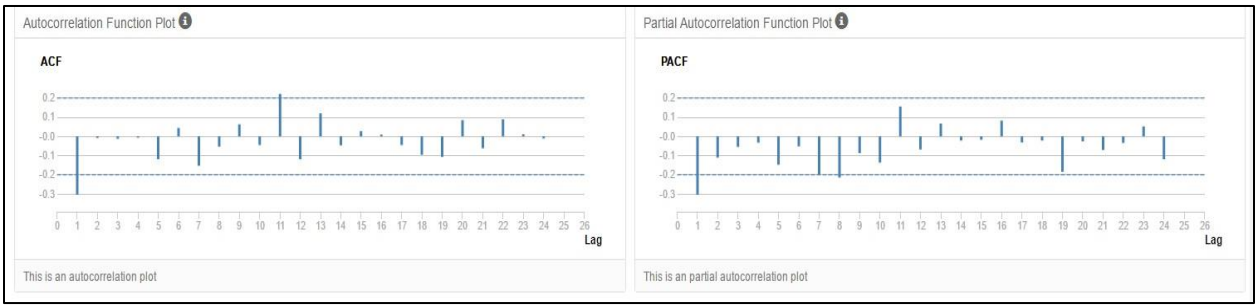
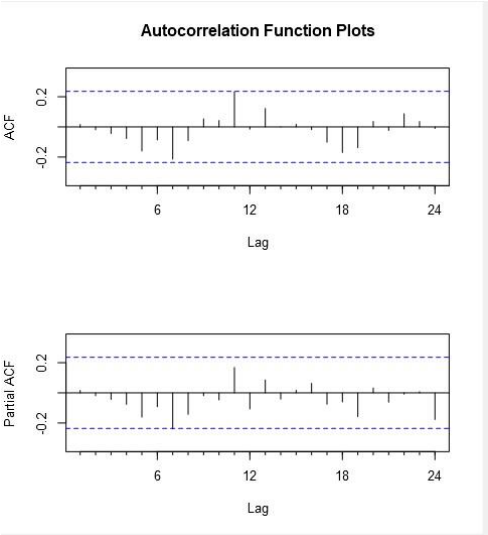


Fig. 6 ACF and PACF after normal difference

From the above graph Fig. 6 we can see that now there is no strong correlation in ACF as well as PACF graph.

From the above graphs after seasonal difference and normal difference we can build an ARIMA model based on these terms ARIMA (0,1,1)(0,1,2)12 since lag 1 is negative and period is of12 months.



ACF AND PACF after ARIMA (0,1,1(0,1,2)12 term has been added

Information Criteria:

AIC	AICc	BIC
1350.0112	1350.2376	1354.0619

In-sample error measures:

ME	RMSE	MAE	MPE	MAPE	MASE	ACF1
718.4459867	36073.5331869	24594.4458891	-1.4412026	9.5014904	0.3632763	0.016125

Ljung-Box test of the model residuals:
Chi-squared = 20.5601, df = 23, p-value = 0.607935

Table 2

From the above Table 2 we can see that AIC value of ARIMA model is 1350.01 and RMSE value is 36073.53 and MASE value of 0.36 respectively which is falling below the generic 1.0, this is the commonly accepted value for model accuracy.

Step 4: Forecast

Compare the in-sample error measurements to both models and compare error measurements for the holdout sample in your forecast. Choose the best fitting model and forecast the next four periods. (250 words limit)

Answer these questions.

1. Which model did you choose? Justify your answer by showing: in-sample error measurements and forecast error measurements against the holdout sample.

Actual and Forecast Values:

Actual	MAM_dampened
271000	255966.17855
329000	350001.90227
401000	456886.11249
553000	656414.09775

Accuracy Measures:

	Model	ME	RMSE	MAE	MPE	MAPE	MASE	NA
MAM_dampened		-41317.07	60176.47	48833.98	-8.3683	11.1421	0.8116	NA

Table 3 Actual vs ETS

Actual and Forecast Values:

Actual	ARIMA_HOLDOUT
271000	263228.48013
329000	316228.48013
401000	372228.48013
553000	493228.48013

Accuracy Measures:

	Model	ME	RMSE	MAE	MPE	MAPE	MASE	NA
ARIMA_HOLDOUT		27271.52	33999.79	27271.52	6.1833	6.1833	0.4532	NA

Table 4 Actual vs ARIMA

In-sample error measures:

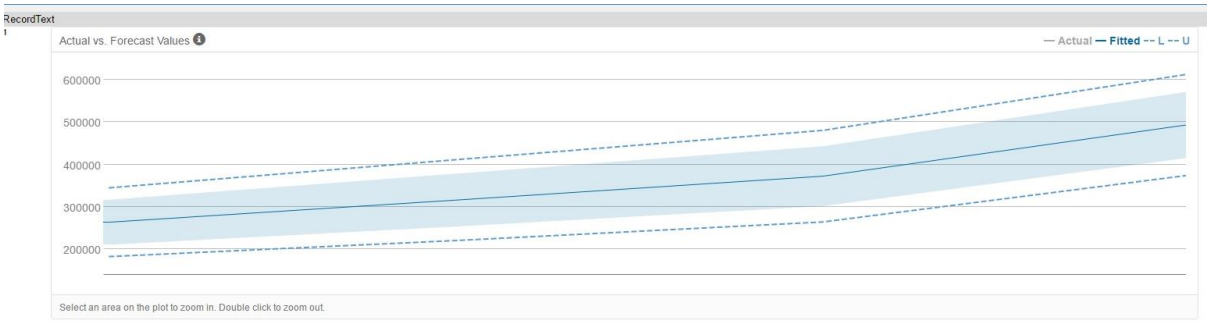
ME	RMSE	MAE	MPE	MAPE	MASE	ACF1
5597.130809	33153.5267713	25194.3638912	0.1087234	10.3793021	0.3675478	0.0456277

In-sample error measures:

ME	RMSE	MAE	MPE	MAPE	MASE	ACF1
-356.2665104	36761.5281724	24993.041976	-1.8021372	9.824411	0.3646109	0.0164145

From the above Table 3 & 4 we can conclude that ARIMA model is better model since the RMSE value (33999.79) is less than ETS model (60176.47) and MASE value of ARIMA (0.45) is lesser than MASE value of ETS model (0.81) and looking at the MPE (-1.80) and ME (-356.27)value in the in-sample error measures of ARIMA model, it is lower than the ETS model, which suggests that ARIMA model misses its forecast by lesser amount.

2. What is the forecast for the next four periods? Graph the results using 95% and 80% confidence intervals.



Period	Sub_Period	forecast	forecast_high_95	forecast_high_80	forecast_low_80	forecast_low_95
2013	10	754854.460048	834046.21595	806635.165997	703073.754099	675662.704146
2013	11	785854.460048	879377.753117	847006.054462	724702.865635	692331.166979
2013	12	684854.460048	790787.828211	754120.566407	615588.35369	578921.091886
2014	1	687854.460048	804889.286634	764379.419903	611329.500193	570819.633462

The forecasted values are 75854.46, 785854.46, 684854.46 and 687854.46

Before you Submit

Please check your answers against the requirements of the project dictated by the rubric here. Reviewers will use this rubric to grade your project.

Reference

Udacity forum