

FORECASTING VIDEO GAMES SALES

The monthly sales data meets the criteria of time series dataset. The dataset is continuous over a time range, starting from 2008 to 2013, where the data is given only has two variables: Month and Monthly Sales (in USD). The data is categorized by months and it contains sales value for each month that is one month apart. The most recent 4 months of data was used as a holdout sample to forecast the next 4 months.

Determining Trend, Seasonal, and Error components

The characteristics of Error, Trend and Seasonality were determined using decomposition plot.

Image of decomposition plot



The seasonal graph looks constant at first but upon a closer inspection, there is a small increase each year. This suggests that any ARIMA models used for analysis will need seasonal differencing and any ETS models will use a multiplicative method.

The trend graph shows an upward linear trend. This shows that any ETS model will need to use an additive method.

The remainder of the sales data or the error component shows a general trend of increasing in error over time. This shows that a multiplicative method needs to be applied in an ETS model.

Building Models

ETS Model

The ETS model chosen for this analysis is ETS(M,A,M). From the graph above, the seasonality and error are increasing over time and the trend is increasing in a linear fashion.

With ETS(M,A,M), the in-sample error results were obtained:

In-sample ETS error

Summary of Time Series Exponential Smoothing Model ETS						
Method: ETS(M,A,M)						
In-sample error measures:						
ME	RMSE	MAE	MPE	MAPE	MASE	ACF1
3729.2947922	32883.8331471	24917.2814212	-0.9481496	10.2264109	0.3635056	0.1436491
Information criteria:						
AIC	AICc	BIC				
1634.6435	1645.9768	1669.4337				
Smoothing parameters:						
Parameter	Value					
alpha	0.765251					
beta	0.000103					
gamma	0.001046					

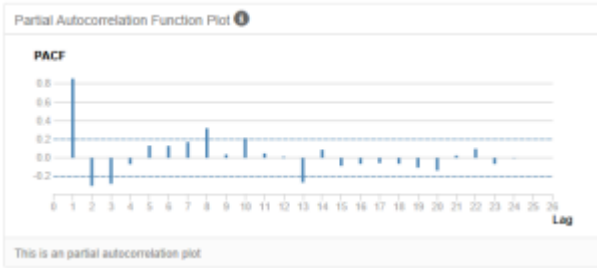
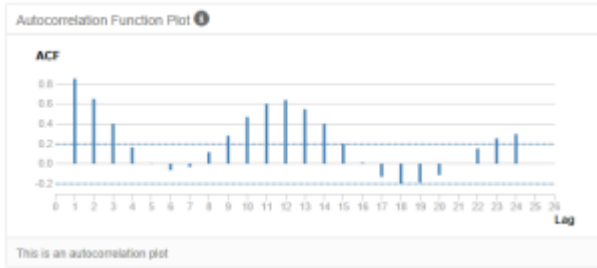
MASE measures is one of the best metric of error measurement of the model because it does not have any percentage errors and offers a scale-independent measurement. It measures the relative reduction in error compared to a naïve model. This ETS(M,A,M) results in a 0.3635 of MASE, which is the ideal MASE value (less than 1). This means that ETS(M,A,M) forecasting method yields an in-sample absolute error that is 0.3635 times as large as those of a naïve random walk model, which is a relatively small error.

RMSE is a great measurement to show how many deviations from the mean the forecasted values fall, which means that it measures the residual error. However, it is hard to decide whether a RMSE of 32883.83 is high or low because the model needs to be validated against the holdout sample first. However, looking at MPE, which measures the average percentage difference between the actual and forecasted value at -0.948, the ETS(M,A,M) might have a tendency to under-predict the sales value with a forecast that is off by 10.23% (MAPE).

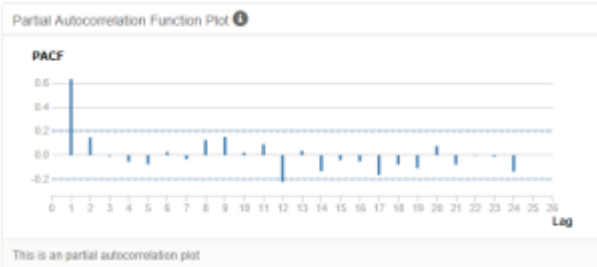
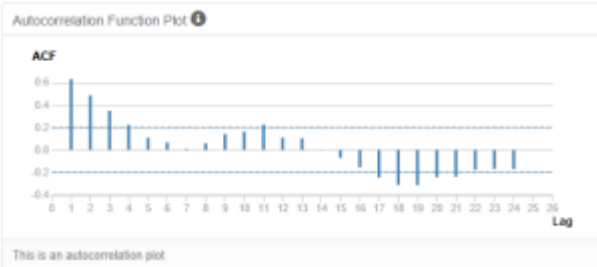
ARIMA Model

The ARIMA model chosen was ARIMA(0,1,1)(0,1,0)[12]. Such conclusion is arrived after inspecting the ACF and PACF plots after first seasonal differencing. Without seasonal difference, ACF and PACF do not exhibit stationary behaviour. ACF slowly decay towards 0 with increasing seasonal lags. After seasonal differencing, ACF and PACF did not present much changes. After the first seasonal difference, ACF and PACF show a strong negative correlation at lag 1 (and decaying towards 0 at PACF). This suggested an MA(1) term. After adding MA(1), there is no significant correlation lags in the ACF and PACF plots. Thus, resulting in ARIMA(0,1,1)(0,1,0)[12].

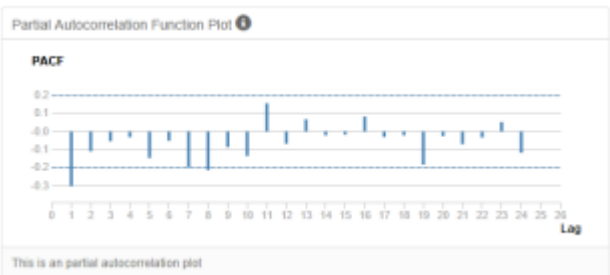
ACF/PACF before seasonal



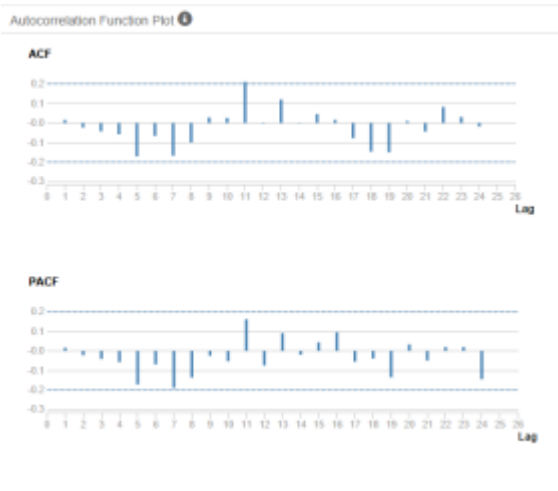
After seasonal



After first seasonal difference



After MA term



This ARIMA model results in a 0.3646 of MASE, which is the ideal MASE value (less than 1). This means that this forecasting method yields an in-sample absolute error that is 0.3646 times as large as those of a naïve random walk model, which is a relatively small error.

Again, it is hard to decide whether a RMSE of 36761.53 is high or low because the model needs to be validated against the holdout sample first. However, looking at MPE, which measures the average percentage difference between the actual and forecasted value at -1.802, the ARIMA might have a tendency to under-predict the sales value with a forecast that is off by 9.82% (MAPE).

In-sample ARIMA model

Report

Summary of ARIMA Model ARIMA

Method: ARIMA(0,1,1)(0,1,0)[12]

Call:
auto.arima(Monthly.Sales)

Coefficients:

ma1
Value -0.378032
Std Err 0.146228

sigma² estimated as 1689257799.31927: log likelihood = -626.29834

Information Criteria:

AIC	AICc	BIC
1256.5967	1256.8416	1260.4992

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

Forecast

Looking at the in-sample error measures alone isn't enough to determine which model is more appropriate. However, the AIC has shown that ARIMA might be a better choice

In-sample error measures:

ETS

In-sample error measures:

ME	RMSE	MAE	MPE	MAPE	MASE	ACF1
3729.2947922	32883.8331471	24917.2814212	-0.9481496	10.2264109	0.3635056	0.1436491

ARIMA

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

And after looking at the Accuracy measures of both models against the hold-out samples, ARIMA was chosen to predict the next 4 months of sales because it has lower MAPE, MASE, MAE and RMSE. If Moreover, its forecast values are a lot closer to the actual values than ETS's.

When looking at both ETS and ARIMA models, the ME of ETS is a lot larger than ARIMA's. ETS's ME is at 3729.230 while ARIMA's at -356.27. ME measures a forecast's bias and can indicate a systematic bias but it's not a necessarily a good measure of how 'good' a forecast is. If the

forecasts are low, the ME can be positive and vice versa. A mix of both positive and negative errors can result in an ME of zero.

The MPE of ARIMA is almost double of the MPE of ETS. Like ME, MPE is a metric of high or low bias. It measures the average percentage difference between the actual and forecasted values. Both ME and MPE will help indicate whether the forecasts are biased to be disproportionately positive or negative.

Besides these two major differences, RMSE, MAE, MASE, MAPE and ACF1 are not very different between ETS and ARIMA models.

Comparison of Time Series Models

Actual and Forecast Values:

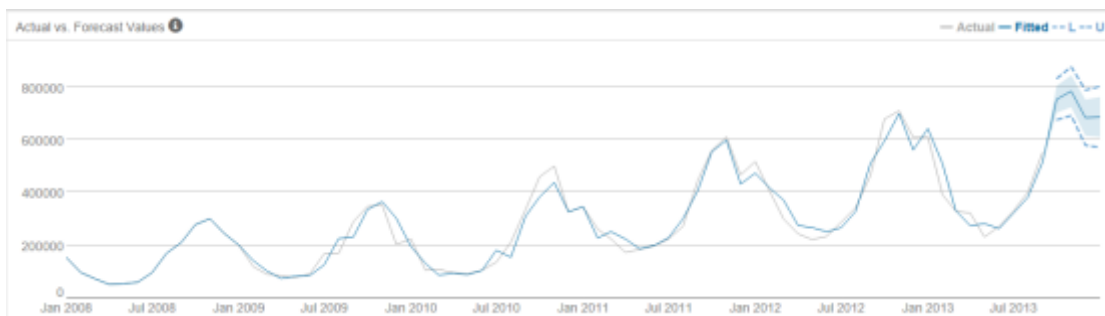
Actual	ETS	ARIMA
271000	268729.50166	263228.48013
329000	378187.04023	316228.48013
401000	488199.64792	372228.48013
553000	691913.69155	493228.48013

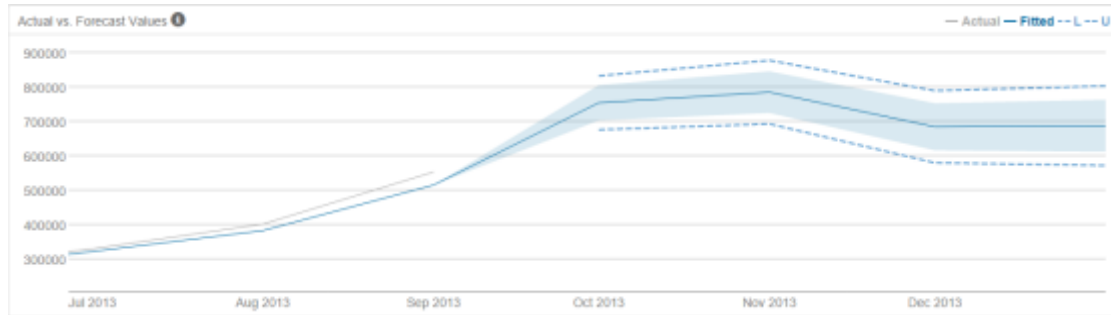
Accuracy Measures:

Model	ME	RMSE	MAE	MPE	MAPE	MASE	NA
ETS	-68257.47	85623.18	69392.72	-15.2446	15.6635	1.1532	NA
ARIMA	27271.52	33999.79	27271.52	6.1833	6.1833	0.4532	NA



ARIMA forecast Graph





ARIMA forecast

Year	Month	forecast	forecast_high_95	forecast_high_80	forecast_low_80	forecast_low_95
2013	10	754854.46	833335.8561	806170.6867	703538.2334	676373.064
2013	11	785854.46	878538.8376	846457.5171	725251.403	693170.0825
2013	12	684854.46	789837.5928	753499.2409	616209.6792	579871.3273
2014	1	687854.46	803839.4698	763692.9816	612015.9385	571869.4503