Project: Forecasting Sales

Step 1: Plan Your Analysis

A video game company wishes to forecast the next 4 months of sales data in order to synchronize supply with demand. In order to complete this task, historic time series data is required. As the supplied historic data contains monthly sales measured over a continuous time period with sequential and equal time intervals and only one data point per time unit, the historic data is sufficient to carry out a time series analysis. To assess the performance of our time series analysis, a comparison of models will be conducted on a holdout sample. This holdout sample will contain the most recent 4 months of the historical sales data. All analysis will be conducted in Alteryx 2018.3 (Colorado, USA).

Step 2: Determine Trend, Seasonal, and Error components

A decomposition plot of the time series demonstrates a seasonal component of the data that slightly increases over time. The positive trend line indicates that the time series is non stationary with a consistent increase in sales over time. Finally, the remainder component in the time series decomposition plot demonstrates an error that varies in magnitude (fig 1).

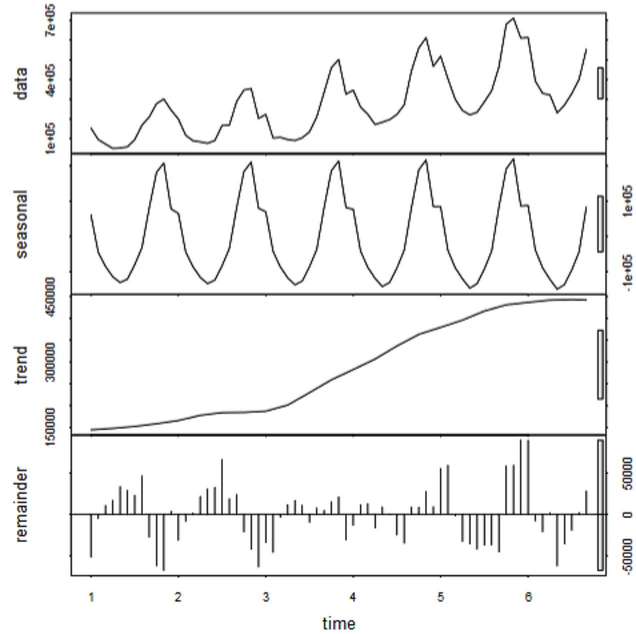

Figure (1) Time Series Decomposition plot.

Step 3: Build your Models

An ETS$_{(M,A,M)}$ model was utilized both with and without a damped trend. The multiplicative error term was chosen due to the non-constant variance over time. The additive trend term was chosen given the linear increase in trend over time. Finally, multiplicative seasonality term was utilized due to the slight increase in seasonality over time. When exploring the in-sample errors (fig 2 & 3), it can be seen that both models perform similar. Despite the undamped ETS having a lower RMSE, the damped model has lower MASE and MAE values suggesting the damped model may have a small number of extreme errors influencing it's RMSE score. The AIC values which demonstrates the trade-off between goodness of fit and model complexity suggests that the damped model is marginally better.

ME	RMSE	MAE	MPE	MAPE	MASE	ACF1	AIC	AICc	BIC
2818.2731122	32992.7261011	25546.503798	-0.3778444	10.9094683	0.372685	0.0661496	1639.7367	1652.7579	1676.7012

Fig (2) In-sample errors and information criteria for the undamped ETS model

ME	RMSE	MAE	MPE	MAPE	MASE	ACF1	AIC	AICc	BIC
5597.130809	33153.5267713	25194.3638912	0.1087234	10.3793021	0.3675478	0.0456277	1639.465	1654.3346	1678.604

Fig (3) In-sample errors and information criteria for the damped ETS model

For the ARIMA model, the data was seasonally differenced to remove the effects of seasonality. Examination of autocorrelation function plot (ACF) depicted similar autocorrelation patterns as the original time series (fig 4).

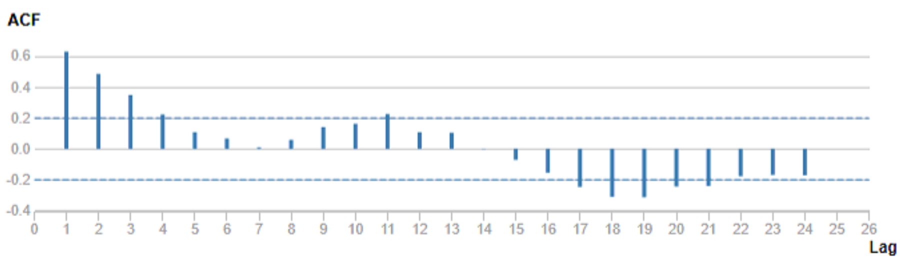

Fig (4) Autocorrelation function plot for the seasonally differenced time series

In order to make the times series stationary, the first and second seasonal difference were calculated. Examination of the first seasonal difference depicts a time series with approximately consistent mean and variance (fig 5). Furthermore, the ACF plot depicts that most of the autocorrelation has been removed from the data (fig 6). In addition the autocorrelations have no pattern and the lag-1 autocorrelation is negative suggesting that the series does not need a higher order of differencing.(1) To confirm, the second seasonal difference was examined which demonstrated an ACF value at lag-1 that was more than -0.5 negative, suggesting that the data is over differenced at the second seasonal difference.

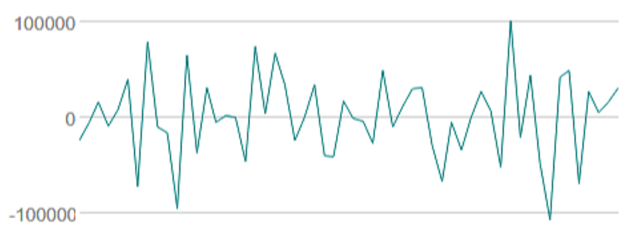

Fig (5) First seasonal difference time series

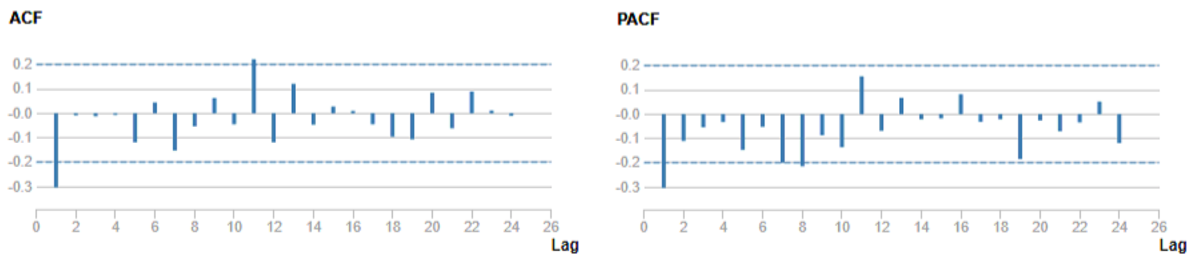

Fig (6) Autocorrelation function and Partial auto correlation function plots for the first seasonally differenced time series.

Taking the first seasonal difference data, the ACF and "Partial auto correlation function" (PACF) plots illustrate no significant correlation at the seasonal lags (12, 24 months). Therefore no seasonal autoregressive or moving average term is required (1). The negative lag-1 autocorrelation in both the ACF and PACF plots indicate a MA$_{(1)}$ model term. The resultant ARIMA model can be described as ARIMA$_{(0,1,1)(0,1,0)[12]}$.

When examining the in sample errors of the ARIMA model, the RMSE was higher in comparison to the ETS models. Despite this, the MAE, MASE and importantly the AIC were all lower.

ME	RMSE	MAE	MPE	MAPE	MASE	ACF1	AIC	AICc	BIC
-356.2665104	36761.5281724	24993.041976	-1.8021372	9.824411	0.3646109	0.0164145	1256.5967	1256.8416	1260.4992

Fig (7) In-sample errors and information criteria for the ARIMA model

Step 4: Forecast

When comparing the three models on the hold out sample, it can be observed that the ARIMA model outperformed the ETS models in every performance metric (fig 8).

Model	ME	RMSE	MAE	MPE	MAPE	MASE
ETS	-49103.33	74101.16	60571.82	-9.7018	13.9337	1.0066
ARIMA	27271.52	33999.79	27271.52	6.1833	6.1833	0.4532
ETS_Damped	-41317.07	60176.47	48833.98	-8.3683	11.1421	0.8116

Fig (8) Out of sample errors for the three models

This is again reflected in the more accurate forecast values for the ARIMA model in comparison to the ETS model (fig 9).

Actual	ETS	ARIMA	ETS_Damped
271000	248063.01908	263228.48013	255966.17855
329000	351306.93837	316228.48013	350001.90227
401000	471888.58168	372228.48013	456886.11249
553000	679154.7895	493228.48013	656414.09775

Fig (9) Actual and Forecast values for the three models on the hold out sample.

As the ARIMA model's in-sample error and forecast errors are generally smaller in comparison to the ETS models, the ARIMA model was chosen to forecast the next four periods (fig 10). The forecasted values for the next four periods are 754854.46, 785854.46, 684854.46, 687854.46.

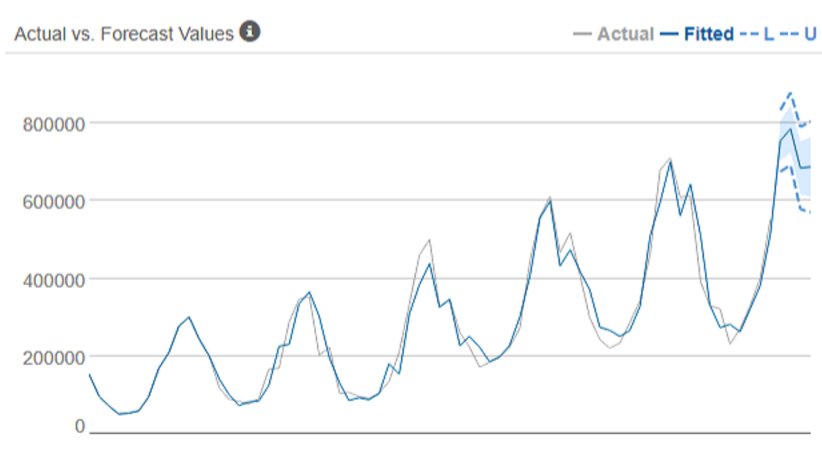

Fig (10) Actual and Future Forecasted values with 95% and 80% confidence intervals.

References

1. Hyndman R, Athanasopoulos G. *Forecasting: principles and practice*. 2018. [cited 2019 Jan 12] Available from: https://books.google.com/books?hl=en&lr=&id=_bBhDwAAQBAJ&oi=fnd&pg=PA7&dq=Forecasting:+Principles+and+Practice&ots=Thi0weZPNJ&sig=wWssrVg0DZWASqZwIwtWsl5Slnc