Case Study Report 2 – Exercise 3

Attribution

This report is based on a quantitative time series measuring the retail turnover by cafes, restaurants and take away food services in Australia in millions of $AUD over monthly intervals from March 2011 to February 2021. The data was collected by the Australian Bureau of Statistics via monthly business surveys.

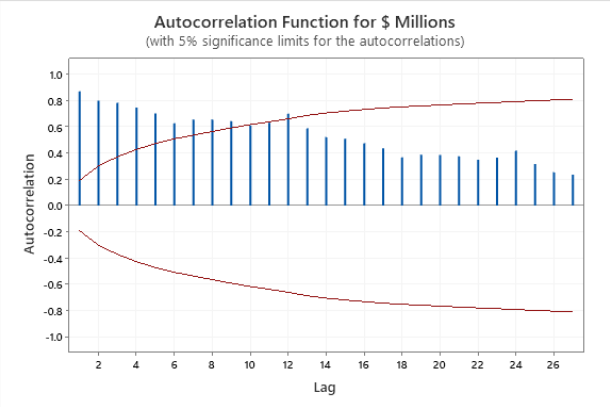
Scope

The data comprises of time series components including trend and seasonality. The trend is linear and positive. The seasonality displays larger retail turnover during December and lower turnover during February. The data shows a combination of additive and multiplicative seasonality patterns.

A multiple variable regression model was applied to the data to generate a forecast. This is a causal model which attempts to explain changes in a response variable based on the values of predictor variables. This model was used due to its ability to accurately capture and project trend and seasonal components, through dummy seasonal variables, of an additive time series.

The applied model can generate an accurate short to medium term forecast. Long term forecasts may become inaccurate due to changing relationships between the response and predictor variables or the effect of unconsidered factors.

Application



The correlogram to the left was used to identify the systematic components within the data.

A trend is evident by the successive decline of the autocorrelation coefficient. A seasonal component is evident by spikes in the chart at consistent periods (periods 1, 12 and 24).

A time variable was created from March-2011 being 1 to February-2020 being 108. 11 binomial dummy seasonal variables were created representing the months January to November. The base month was December.

The response variable, retail turnover in $AUD, was regressed over the predictor variables, time and the 11 seasonal dummy variables.

The model as applied comprised of three main components, the intercept term (blue), the predictor variable term responsible for forecasting trend (green) and the 11 dummy variables responsible for adjusting for seasonal fluctuations (red).

$$\hat{Y} = b_0 + b_1 \times time + b_{2-12} \times dummy\ variable_{2-12}$$

b_k represents the estimates of the regression coefficients calculated via the ordinary least squares method in excel.

The model was used to generate a forecast by extending the time variable past 108 and matching the relevant seasonal dummy variables to the projected time period. The validity of the model was then assessed.

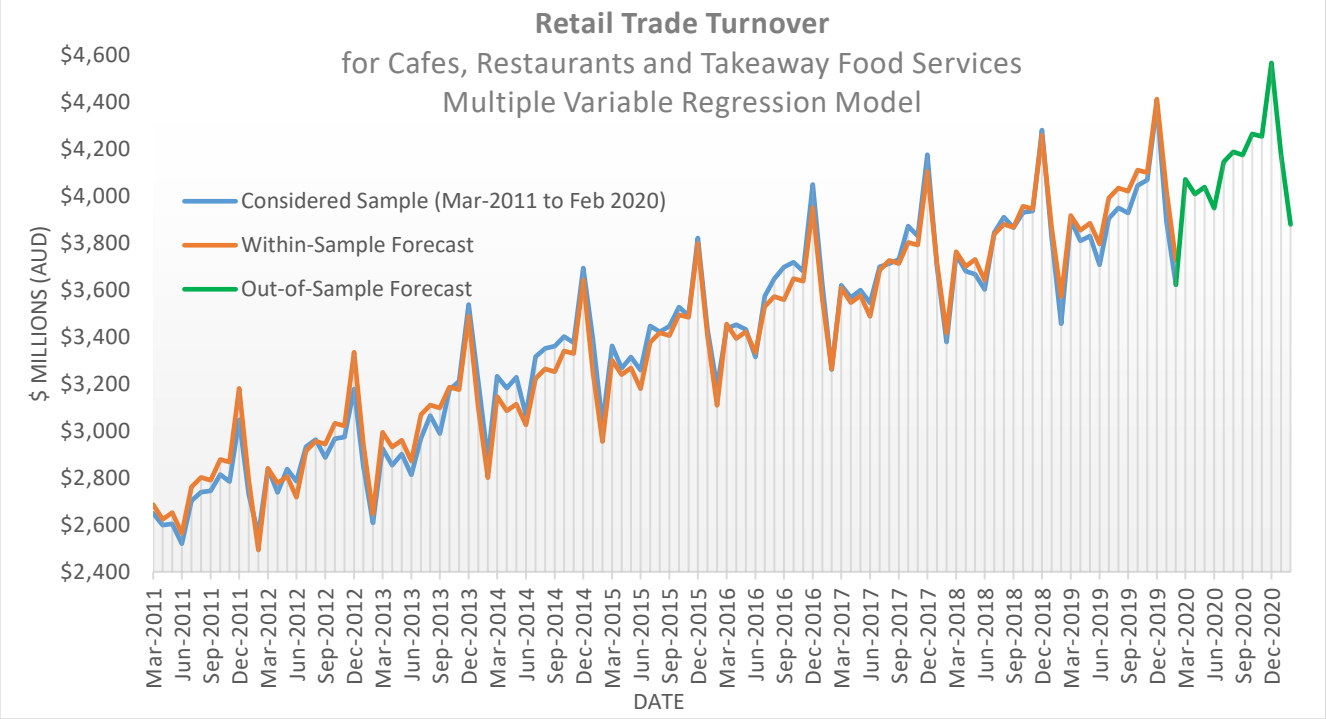
Analysis

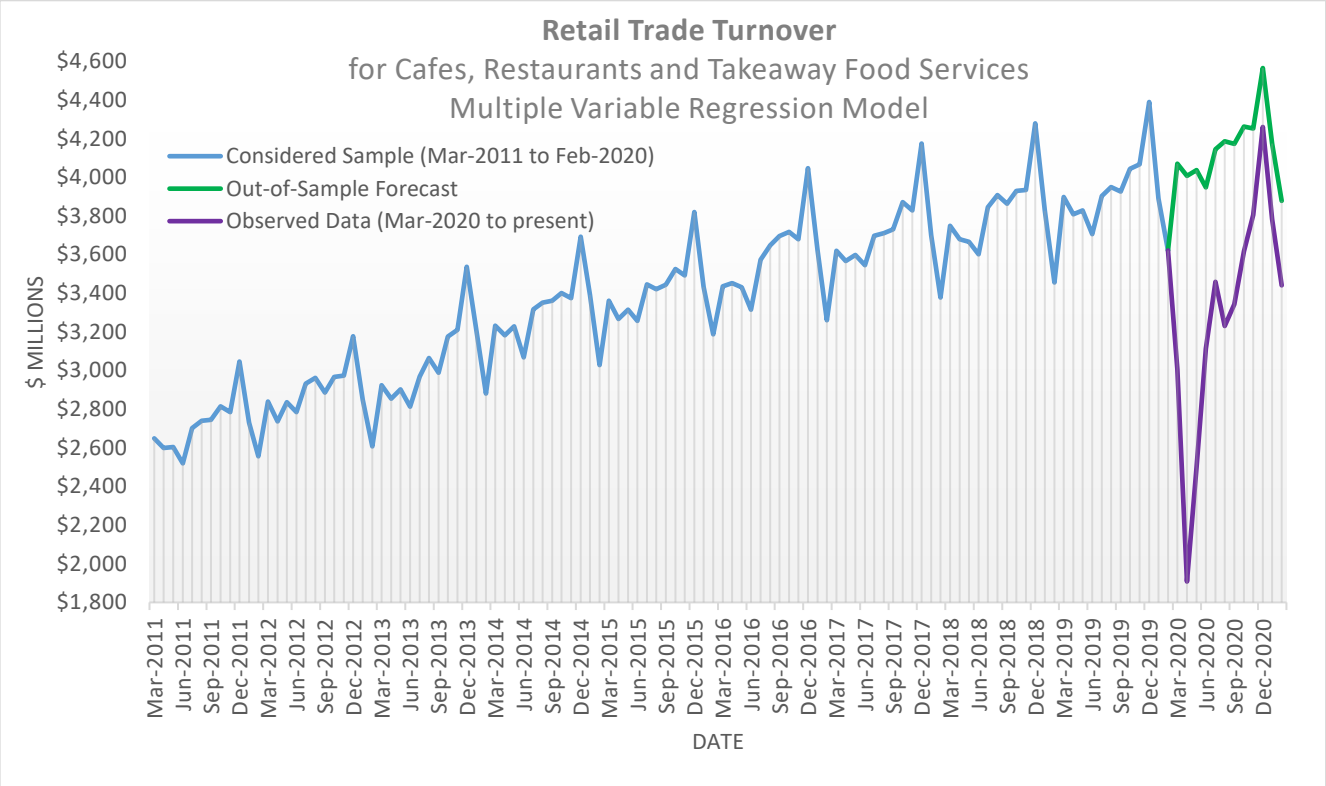
The graph below displays the considered data, within sample and out of sample forecasts using the following regression equation:

$$\begin{aligned}\hat{Y} = {} & 3052.52 + (12.82 \times time) + (-403.44 \times Jan) + (-713.44 \times Feb) + (-379.90 \times Mar) + (-455.15 \times Apr) \\ & + (-438.50 \times May) + (-540.45 \times Jun) + (-356.10 \times Jul) + (-327.54 \times Aug) \\ & + (-353.12 \times Sep) + (-276.59 \times Oct) + (-300.68 \times Nov)\end{aligned}$$

The model seems to capture the positive linear trend of the data. For every unit increase in time there is a $12.82 million increase in turnover. The model does not capture seasonal fluctuations accurately. As the considered data oscillates from additive to multiplicative, the model can be seen to overestimate during early periods, under-estimate during the middle periods and again overestimate in the later periods.

The forecast post February-2020 will provide some indicative value however will most likely overestimate the actual data should influencing variables remain constant.





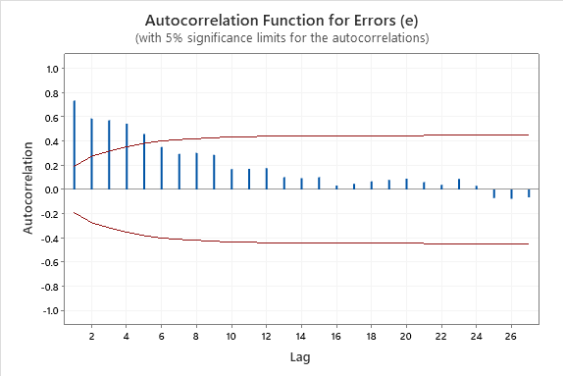
The graph above displays the considered sample, out-of-sample forecast and the observed data for the period of March-2020 to present. As evident by the large difference in the forecast and actual

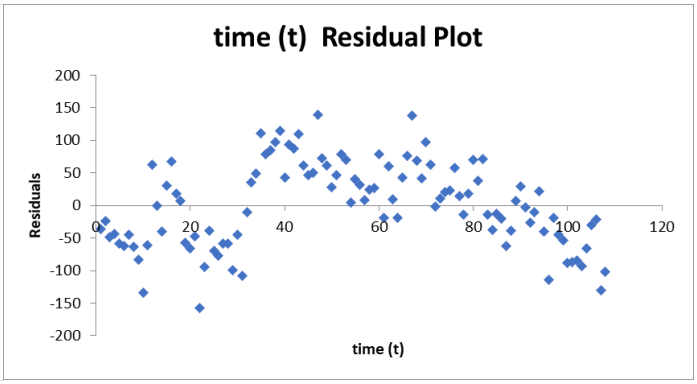
data, an event (Covid 19 restrictions) has occurred to influence consumer behavior causing a large overestimation in turnover.

Articulation of Issues

The validity of the model was assessed, and the linearity assumption, limited multicollinearity and normal distributions of residuals were found to hold true. The variance of the residual plots for the dummy seasonal variables could be considered equal.

The residual plot for the time variable showed a concave down pattern and suggests that the model does not accurately capture the trend component of the data. This is supported by the ACF test of the errors which also indicates a non-captured trend component.





The P-Values of the coefficients were all found to be significant.

The R^2 suggests 97.82% of the variation in retail turnover is explained by the variation in time and seasonal dummy variables. The average residual of the regression was 68.92.

Critique

Both the WES and multi-variable regression model failed to accurately forecast the impact of Covid-19 on consumer behavior and therefore overestimated retail turnover from early 2020 onwards.

The regression model does not capture the trend of the data well, possibly due to the trend component not being constant. The WES model more accurately captures the seasonal component however would consistently overestimate low periods.

The MSE for the regression model was 4177.71 while the MSE for the WES model was 5656.11. This suggests the regression model is more accurate.

Analysis of the predictor variable coefficient values, within the regression model, provide greater insight into the impact the variable has on the forecasted value.

Position

The multi-variable regression model is the favoured model to forecast retail turnover. The model has a lower MSE error term when compared to the WES model. The benefits of the regression model over the WES model include the ability to gain insight of the effect a certain variable has on the response variable and an ability to generate longer term forecasts.

To improve the accuracy of the model, investigation of additional relevant predictor variables should be undertaken. Also, leading indicators of large downturn in retail turnover should be explored to better forecast situations such as the Covid-19 lockdowns.