Question Zero (1 page)

Referring to your last case study report: Refer to your data that you used for Question #5 in Case Study Report #1. Provide a line chart of the relevant time series for the last 4 years. Include a relevant heading for your line chart – clearly identifying your designated Australian state.

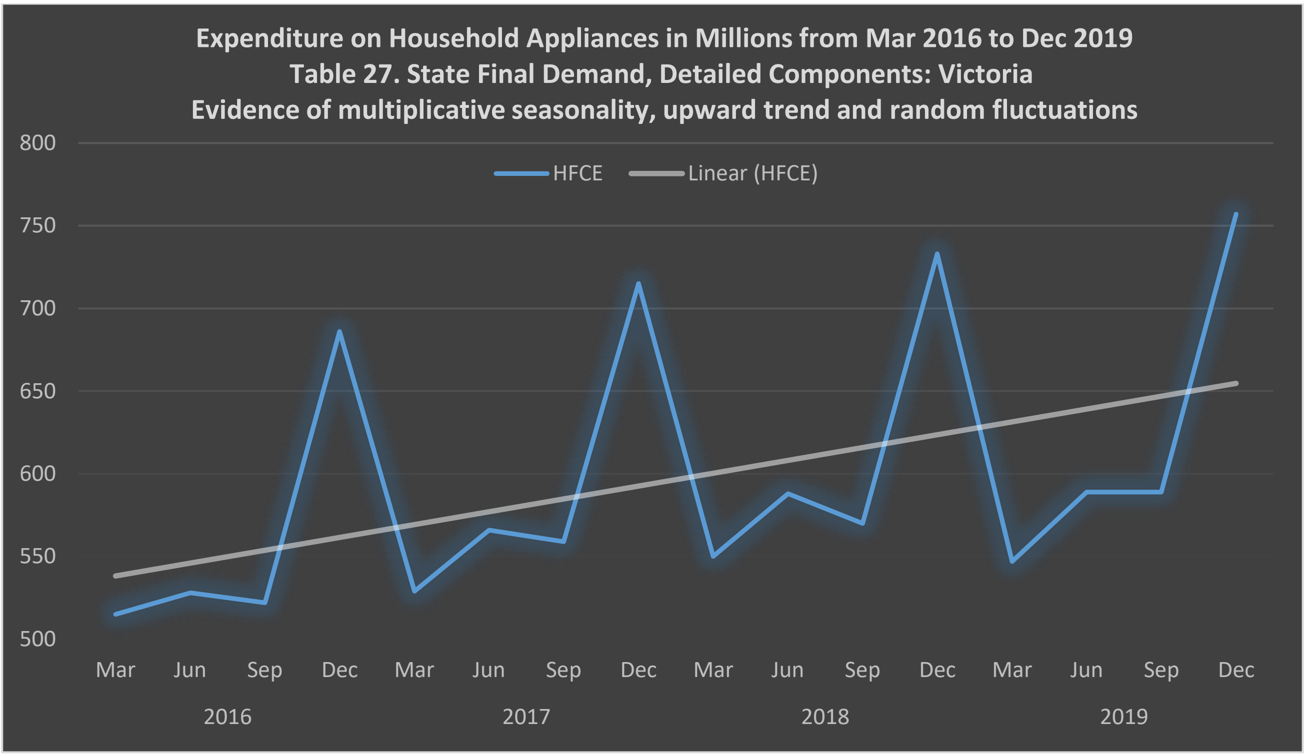


Figure 1. HFCE ($m) on Household Appliances – Current Prices in Victoria, Australia over the past 4 years.

Questions 1 & 2 (1 page) – The Regression

1. Multiple regression model to estimate appliance sales for Victoria.

In a multiple regression model, 4 quasi explanatory variables are first created. They are Time variable to explain the trend component with an index of 1, 2, 3, 4 up to 16 (representing the 16 observations excluding the forecast period) that is generated after every quarter. To capture the seasonality of the time series, 3 dummy binary variables (0,1) for each quarter are created – Mar, Jun, Sep. The dummy variables become 1 when it is their corresponding quarter, and otherwise 0. Being the final month of the quarterly cycle, December is the base, therefore having 0 in its entire row.

Year	Quarter	HFCE	Time	Mar	Jun	Sep
2016	March	515	1	1	0	0
	June	528	2	0	1	0
	September	522	3	0	0	1
	December	686	4	0	0	0
2017	March	529	5	1	0	0
	June	566	6	0	1	0
	September	559	7	0	0	1
	December	715	8	0	0	0
2018	March	550	9	1	0	0
	June	588	10	0	1	0
	September	570	11	0	0	1
	December	733	12	0	0	0
2019	March	547	13	1	0	0
	June	589	14	0	1	0
	September	589	15	0	0	1
	December	757	16	0	0	0
2020	March		17	1	0	0
	June		18	0	1	0
	September		19	0	0	1
	December		20	0	0	0

Figure 2. Snapshot of excel working for X and Y variables.

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.994715891
R Square	0.989459705
Adjusted R Square	0.98562687
Standard Error	9.577138973
Observations	16

ANOVA

	df	SS	MS	F
Regression	4	94713	23678.25	258.1535031
Residual	11	1008.9375	91.72159091	Significance F
Total	15	95721.9375		8.60425E-11

	Coefficients	Standard Error	t Stat	P-value
Intercept	674.9375	7.18285423	93.96508385	2.475539E-17
Time	4.78125	0.535378344	8.930600295	2.260943E-06
Mar	-173.15625	6.959918476	-24.87906296	5.081933E-11
Jun	-145.4375	6.856188106	-21.21258894	2.839603E-10
Sep	-157.96875	6.793189636	-23.25398796	1.054795E-10

Figure 3. Snapshot of Summary Output.

The next step is to perform the Regression Analysis with HFCE as the Y variable and Time and dummy variables (Mar, Jun, Sep) as the X variables. The estimated regression output appears in a new worksheet. The coefficients are then used for the regression equation to provide forecasts.

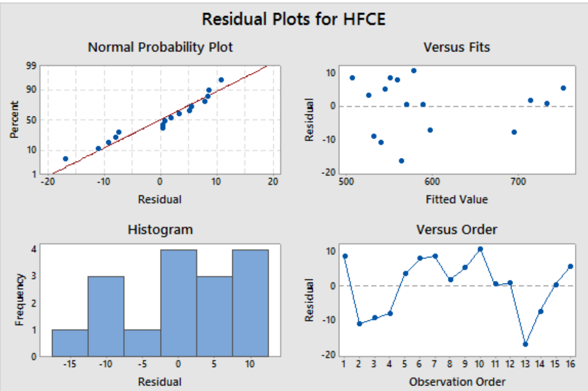


Figure 4. Residual Four in One Plots for HFCE.

The residuals appear random and relatively normally distributed as shown from the Normal Probability Plot (an approximately straight line) and Histogram (somewhat symmetric and bell-shaped). The residuals are scattered randomly with a mean close to zero in the Residuals versus Fitted (predicted HFCE) and the Observation Order plot also looks random with no discernible patterns (equal number of positive and negatives errors), deeming the linear regression model appropriate and assumption about the errors in the model valid. In addition, forecasts generated from the model should produce valid forecasts for Y (HFCE).

Regression Equation: **$Y_t = \beta_0 + \beta_1 * t + \alpha_1 * Mar + \alpha_2 * Jun + \alpha_3 * Sep + \varepsilon_t$**

- Where t is time, Y_t is the forecast at period t, β_0 is the intercept, β_1 is the time coefficient, α_k is the coefficient for the quarter k and ε_t is the error term

The overall estimated regression equation with coefficients:
HFCE = 674.9375 + 4.7813Time - 173.1563Mar - 145.4375Jun - 157.9688Sep

2. Meaning of the intercept, the coefficient of the time variable, as well as the coefficient for the September dummy variable.

- Intercept (β_0 = 674.9): The predicted average household final consumption expenditure for the year is 675 units at period 0.
- Coefficient of the Time Variable (β_1 = 4.7): For every 1 unit increase in Time (every quarter), there will be an underlying increase of 5 units on average to the base expenditure. This is the slope or gradient of the time series as well as all the quarter's regression lines.
- Coefficient of the September Dummy Variable (β_4 = -157.9): Relative to the base quarter December, expenditure in the September quarter on household appliances is 158 units lower on average in a given year. It is only used when the regression calculates its prediction for September that it is equal to 1 and the rest of the variables are 0.

Questions 3 & 4 (1 page) – R^2 and the Dummy Variables

3. Coefficient of determination for the regression.

Regression Statistics	
R Square	**0.989459705**
Adjusted R Square	**0.98562687**

Figure 5. Regression Output.

The $R^2 = 0.9894$ and the adjusted R^2 is 0.9856.

- The coefficient of determination measures the explanatory power of the X variables for the sample HFCE data. The closer the R^2 is to 1, the better the estimated model fits the sample data. Both statistics are high, indicating that the model explains 99% of the sample variation in HFCE. The explanatory power strongly suggests the seasonal dummy variables and time are required as a group to adequately model and forecast HFCE.
- The adjusted R^2 is marginally lower than the R^2. This is because the explanatory power of the variables increases as new variables are added, thus the R^2 adjusted accounts for this by lowering itself accordingly (Minitab Blog Editor, 2013).
- Although the R^2 is a tool in deciding the suitability of the model, it is not the most important statistic nor should it be the only indicator of how good or bad a model is. A formal hypothesis test is required to determine whether the relationship is statistically significant as the R^2 tends to be a biased estimate.

4. Write out a separate regression equation for each of the quarters: Mar, Jun, Sept, & Dec

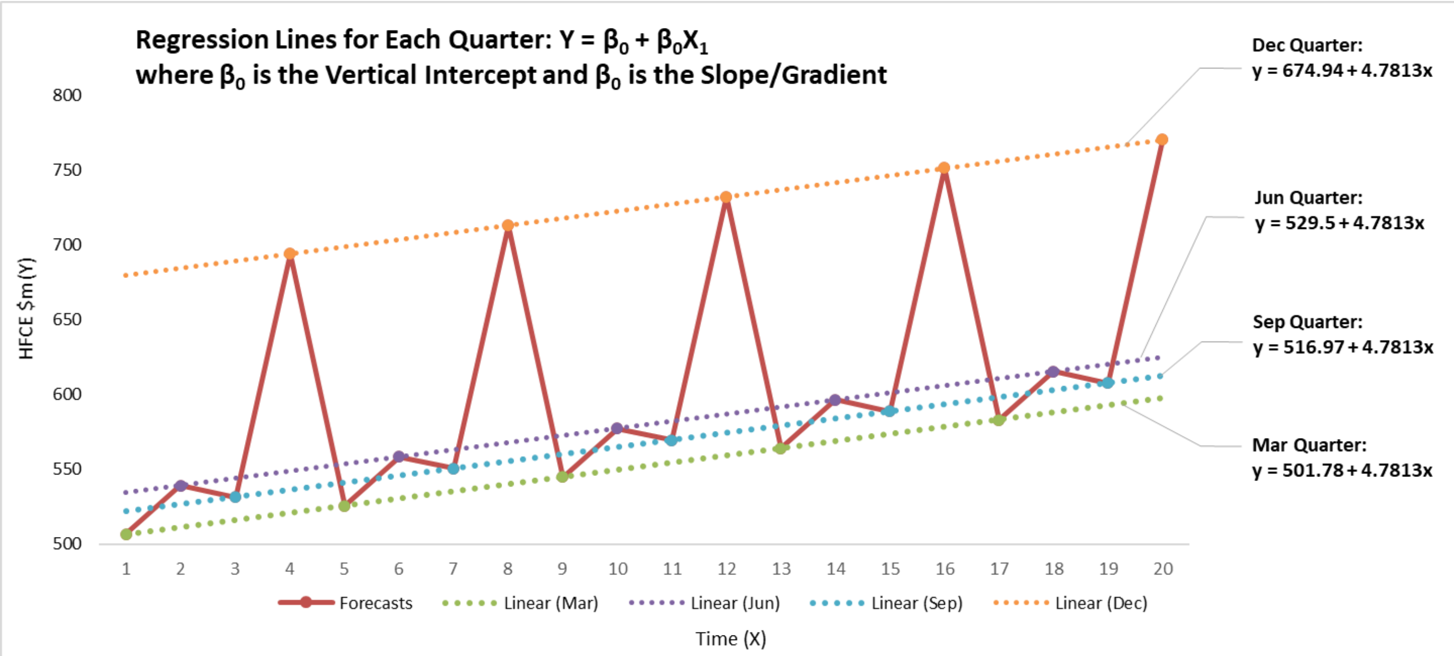


Figure 6. The time series with forecasts for each quarter's regression lines. The Dec line appears to show the peak while the Mar line shows the trough of the seasonality, with the other 2 lines representing the middle of the seasons.

For each quarter, the equation is $Y_t = (\alpha_k + B_0) + B_1{*}t + \varepsilon_t$. The lines have an identical gradient, only the y-intercept changes. Regression equations where α_k is the respective quarter's coefficient:

- March Quarter: HFCE = (-173.1563 + 674.9375) + 4.7813Time + ε_t
  Expenditure in the Mar Quarter is 173 units lower than the Dec quarter.
- June Quarter: HFCE = (-145.4375 + 674.9375) + 4.7813Time + ε_t
  Expenditure in the Jun Quarter is 145 units lower than the Dec quarter.
- September Quarter: HFCE = (-157.9688 + 674.9375) + 4.7813Time + ε_t
  Expenditure in the Sep Quarter is 158 units lower than the Dec quarter.
- December Quarter: HFCE = 674.9375 + 4.7813Time + ε_t

	Coefficients
Intercept	674.9375
Time	4.78125
Mar	-173.15625
Jun	-145.4375
Sep	-157.96875

Figure 7. Coefficients for the intercept, Time, Mar, Jun, Sep

Questions 5 & 6 (one page) – Tests

5. What is the appropriate test to test the overall significance of a regression model?

ANOVA				
	df	*SS*	*MS*	*F*
Regression	4	94713	23678.25	**258.1535031**
Residual	11	1008.9375	91.72159091	**Significance F**
Total	15	95721.9375		**8.60425E-11**

Figure 8. ANOVA table output from the full model

Figure 9. Residuals Plot for the Model.

The F test is most appropriate for examining the overall significance of a regression model.

- H_0: $\beta_1 = ... = \beta_4 = 0$;
- H_1 : at least one β_i is **not** = 0;
- F statistic: 258.154
- P-value: 8.60425E-11 < 0.05 (reject H_0)

The F statistic is large (258.154), and its associated p-value is smaller than the significant level, giving evidence that at least one X variable affects Y. Rejection of the null hypothesis at the 5% level indicates the full model is significant in explaining the variations in HFCE.

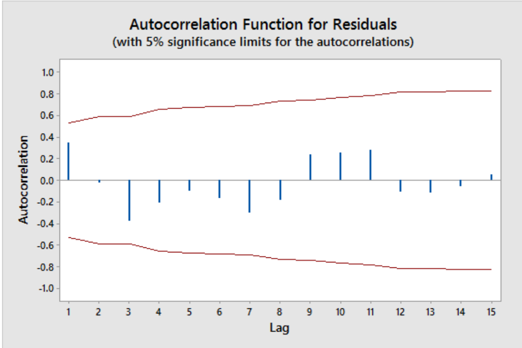


Figure 10. ACF for Full Model shows a possible pattern as it falls in the middle of the time series before increasing again.

6. What is the appropriate test to test the individual significance of a coefficient in your regression?

	Coefficients	*Standard Error*	*t Stat*	*P-value*
Intercept	674.9375	7.18285423	93.96508385	2.475539E-17
Time	4.78125	0.535378344	8.930600295	2.260943E-06
Mar	-173.15625	6.959918476	-24.87906296	5.081933E-11
Jun	-145.4375	6.856188106	-21.21258894	2.839603E-10
Sep	-157.96875	6.793189636	-23.25398796	1.054795E-10

Figure 11. Summary of the t-test and associated p-values used to examine the individual variable significance of a coefficient in a regression, indicating support for H_0 or H_1

Figure 12.

- H_0 : $\beta_j = 0$; H_1: β_j is **not** = 0;
- <u>Time</u>
  T statistic: 8.93
  P-value: 2.260943E-06 < 0.05 (reject H_0)
- <u>September</u>
  T statistic: -23.25
  P-value: 1.054795E-10 < 0.05 (reject H_0)

Time has a positive T statistic while the September dummy variable has a negative T statistic. For both X variables, the p-values are less than 0.05. Rejection of the null hypothesis at the 5% level indicates that Time and September variables are significant and relevant in explaining HFCE.

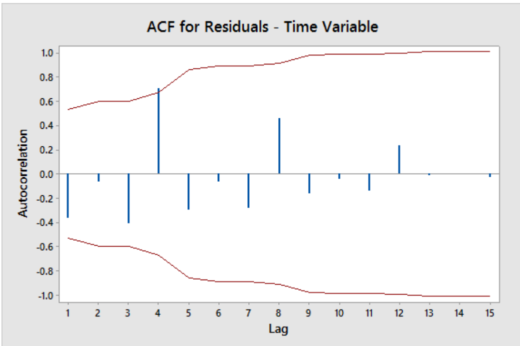


Figure 13.

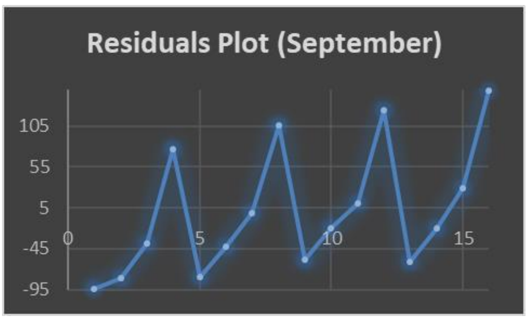


Figure 14.

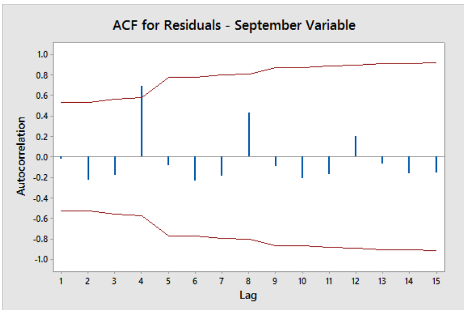


Figure 15.

Figures 12 & 14. The Time and Sep residuals plot show a seasonal quarterly pattern; hence they are not random.

Figures 13 & 15. Declining values show that the trend is captured. Spikes every 4 lags captures seasonality, substantially larger in the Sep ACF as it is more closely related to the quasi variable as compared to time.

Questions 7 & 8 (one page) – Forecasts and Performance

7. Forecast for 2020 September Quarter: (-157.9688 + 674.9375) + 4.7813Time + ε_t
   When Time = 19 and all other quarters are 0: (-157.9688 + 674.9375) + 4.7813*19 = **607.813**

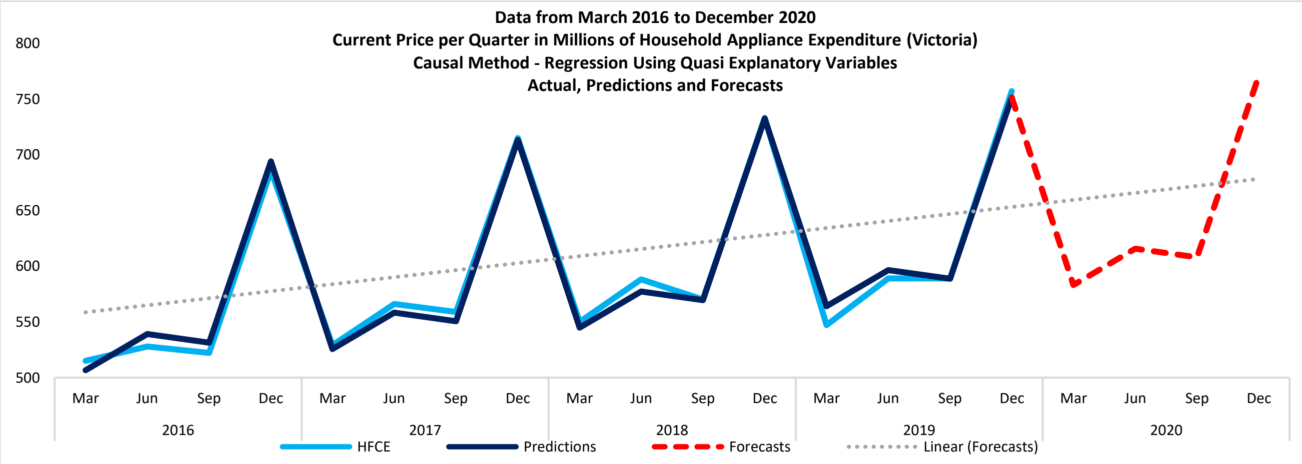


Figure 16. Causal Method – Regression Using Quasi Explanatory Variables. The forecasts fit the data quite well and captures the relevant systematic components of the time series.

8. The table below is a comparison of the accuracy and adequacy criteria for all appropriate models. According to *The Analysis Factor*, a lower RMSE is the best measure of model fit, indicating a better forecast model (Martin, 2020). Based on the error analysis and using the lower RMSE as the main criteria here, the best model to use is the Decomposition Multiplicative model (RMSE: 7.582). The Decomposition model outperformed the regression model in all criteria as well as being faster and simpler to execute. It also has the least overestimation or underestimation.

	Model	MAE	MSE	RMSE	MAPE
Time Series Method					
1	WES Multiplicative	13.106	232.613	15.252	2.249
2	WES Multiplicative - SOLVER	10.547	165.763	12.875	1.863
3	WES Additive	10.833	188.249	13.72	1.934
4	WES Additive - SOLVER	10.622	168.823	12.993	1.86
5	**Decomposition Multiplicative**	**6.413**	**57.481**	**7.582**	**1.147**
6	Decomposition Additive	7.006	67.144	8.194	1.233
Causal Method					
7	Regression	6.602	63.059	7.941	1.160

One modification I would recommend is to collect data on regular explanatory variables such as Weather, Population Size, National Income, and Price of Home Appliances instead of using quasi variables to forecast. This improved regression model will allow *HomeAppCo* to forecast a range of values based on various possibilities for the regular variables, enabling them to plan for various contingencies and strategies.

A timeframe of only 4 years of data is also unsuitable as it disregards the cyclic behavior of the time series (the duration of these fluctuations usually lasts several years). Data over the past few decades is thus required to understand the effect and cyclic pattern of appliance expenditure. In *HomeAppCo's* case, key time series such as sales of household appliances are impacted by macroeconomic factors (e.g. income, consumer expectations, and interest rates). Economic or business cycles can have an impact on consumer spending and subsequently spending on *HomeAppCo's* sales. Hence, leading indicators can be used to judge cyclical components for broad economic activity and specifically any turning points of these cycles, helping the business to adjust and improve the forecasts effectively. For instance, Housing Market, Daily Valuations, New Residential Listings, and Stock Market are reasonably composite leading indicators of appliance production and sales because new homes require new appliances. A decline in housing demand decreases a homeowner's wealth, thereby reducing household expenses. At a more micro level, the company can also use the Consumer Survey Data as an indicator of consumer willingness to spend, increasing the precision of estimated impacts.

Questions 9 & 10 (one page) – Judgmental Forecasting

9. Subjective Assessment Methods

Under Judgmental Forecasting, Subjective Assessment Methods use relevant and non-quantitative information such as surveys, then assessing and weighing the information to generate short to medium-term forecasts. Throughout the models explored, a small area of improvement could be made such as periods of consistent error. They can be rectified to a certain extent using subjective methods by finding the unknown factors and adjusting the forecasts accordingly. Although we have many years of quantitative data and an effective decomposition model (for short to medium-term forecasts), qualitative data is required due to the impact of COVID-19 which may potentially cause structural breaks and disruption of the regular trend for the appliance sales forecast. The three main types of Subjective Assessment Methods include Sales Force Composite Forecasting, Jury of Executive Opinion, and Subjective Probability Assessments.

Sales Force Composite Forecasting could assist in addressing COVID-19 as relevant salespeople/managers can make individual assessments and estimate the demand for appliances. It can lead to political and workplace biases, however, affecting the salespeople's feedback. It would also require more time for sales reps to familiarize themselves with a new territory – Victoria, making it harder to gauge sales. The Jury of Executive Opinion is vulnerable to workplace pressures and groupthink, which could skew the data, leading to inaccurate forecasts for appliance sales. The most effective method is Subjective Probability Assessment as the forecaster is in the best independent position to judge and incorporate the factors affecting appliance sales during the pandemic, and their weights including the economic and population trends. It is also based on professional judgment rather than experimental observation. Experts would need to look back in histories such as the Global Financial Crisis and 9/11 terrorist attack to better understand and model the relationship of uncertainty in the external environment to making business decisions and performance (e.g. HomeAppCo's future appliance sales) during the most recent COVID-19. Hence, assisting in creating more accurate and reliable forecasts. However, the high cost of collecting qualitative data, certain biases, and individual psychological influences could negate the likely gain in explanatory power.

10. Scenario Analysis

Scenario Analysis is one of the exploratory methods that is useful for medium to long term forecasting for this business problem as it can assist in planning for competitive, economic, and other changes in the environment. It is essentially a combination of intuition and rationality where several situations are considered during the COVID-19 pandemic and weighed in a matrix in terms of likelihood and effect. The first step is to identify plausible scenarios and assess the relative likelihood of occurrence for: 1) Quick recovery – best case, 2) Global slowdown and strict lockdown – base case, 3) Global recession/depression – worst-case. For each scenario, environmental factors that may impact key organizational factors such as sales revenue, profit and loss, pricing, and analysis of percentage variances on individual product lines should be determined. From social distancing and banning of public events to shutdowns of many businesses, there has been a rapid shift from in-person purchases to e-commerce. According to IBISWorld's analysts, COVID-19 is likely to impact the Online Large Appliance Sales in the Australia industry (IBISWorld, 2020). Business sales are expected to increase because of higher demand from consumers who are panic buying industry products such as fridges and freezers to store the food. The technological shift is thus one of the driving forces that will affect HomeAppCo's future appliance sales.

Scenario Analysis considers these outcomes and variables which will assist in future branching and longer-term forecasts. As the environment changes, quantitative forecasting will become less reliable. It is also the most useful despite forecasts being produced solely based on estimates (a downside of this method), as various scenarios for HomeAppCo will be carefully considered. Therefore, it allows the forecaster to better account for uncertainty, judge factors effectively in the macro level, and give appropriate weights according to opportunities and threats. The process of identifying and ranking driving forces, as well as creating a matrix for COVID-19 will prove useful in the early years of HomeAppCo's business expansion to Australia during this pandemic.

References

Australian Bureau of Statistics (December, 2019). 5206.0 - Australian National Accounts: National Income, Expenditure and Product, Table 27. State Final Demand, Detailed Components: Victoria, time series spreadsheet. Retrieved May 22, 2020, from https://www.abs.gov.au/AUSSTATS/abs@.nsf/DetailsPage/5206.0Dec%202019?OpenDocument

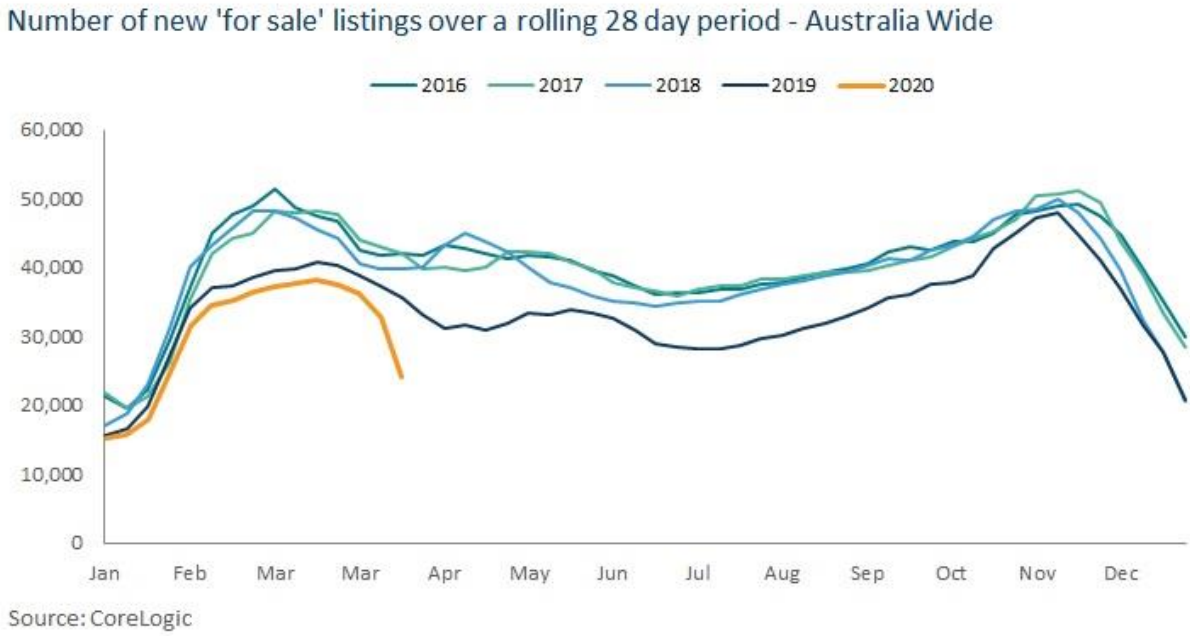
The Minitab Blog 2013, Multiple Regression Analysis: Use Adjusted R-Squared and Predicted RSquared to Include the Correct Number of Variables. Retrieved May 22, 2020, from https://blog.minitab.com/blog/adventures-in-statistics-2/multiple-regession-analysis-use-adjusted-r-squared-and-predicted-r-squared-to-include-the-correct-number-of-variables

MARTIN, K. (January 16, 2020). Assessing the Fit of Regression Models. Retrieved May 22, 2020, from https://www.theanalysisfactor.com/assessing-the-fit-of-regression-models/

Owen, E. (April 16, 2020). Early Market Indicators Show The First Signs Of A Housing Market Slowdown. Retrieved May 22, 2020, from https://www.corelogic.com.au/news/early-market-indicators-show-first-signs-housing-market-slowdown

IBISWorld. (April, 2020). Industry Market Research, Reports, and Statistics. Retrieved May 23, 2020, from https://www.ibisworld.com/au/industry/online-large-appliance-sales/5460/

Appendices



Appendix 1. Evidence to support the drop in listings Australia Wide