Question Zero (1 page)

<u>Time Series</u>

This report bases itself on the data of 'Households; final consumption expenditure - household appliances: current prices' in Tasmania, Australia from the period of March 2016 to December 2019, as reported quarterly. According to the Australian Bureau of Statistics, final consumption expenditure at current prices refers to the current price estimates, which are valued at the prices of the period from which the observation is made.

The data was obtained from the Australian Bureau of Statistics, 2019, *Australian National Accounts: National Income, Expenditure and Product Dec 2019*, 'Table 31. spreadsheet, series ID A3605894L, viewed 26 April 2020, time series https://www.abs.gov.au/AUSSTATS/abs@.nsf/DetailsPage/5206.0Dec%202019?OpenDocument

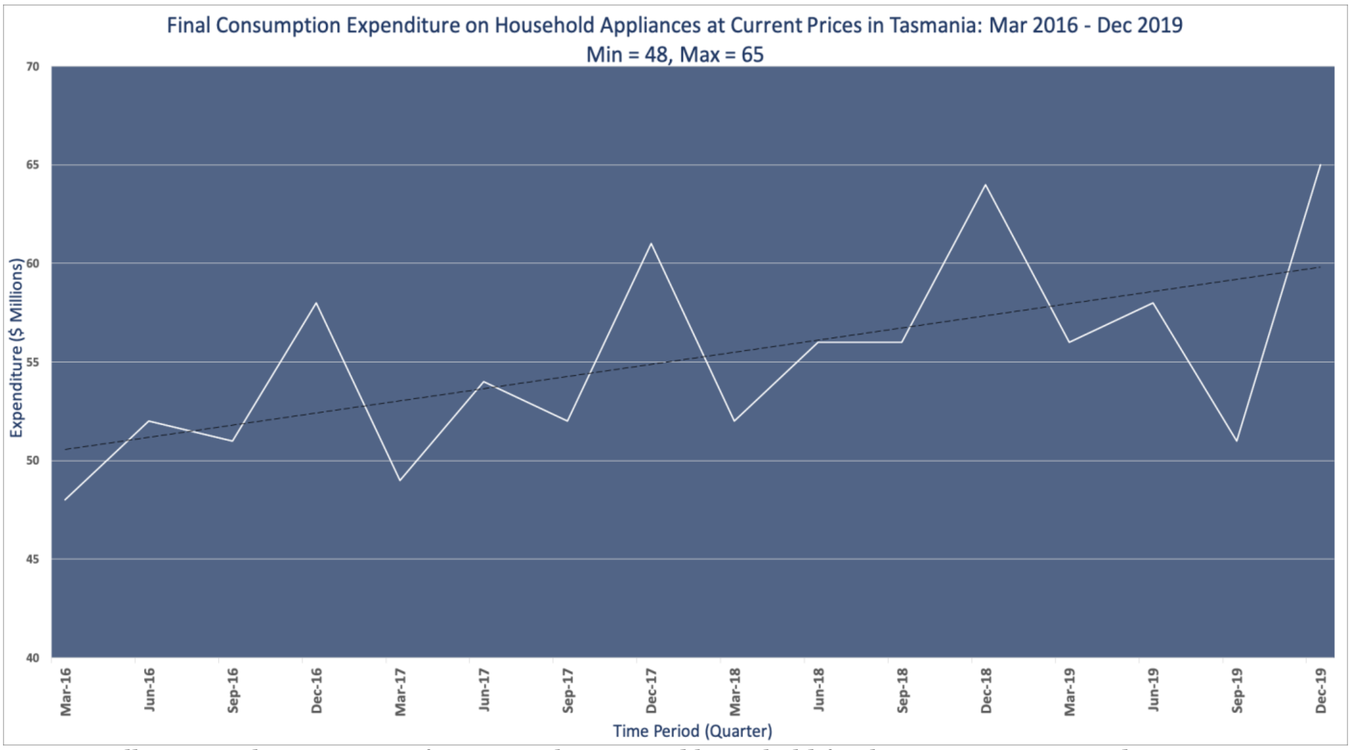


Figure 1. illustrates the time series for quarterly reported household final consumption expenditure on household appliances at current prices in Tasmania from March 2016 to December 2019. The time series contains a positively increasing linear trend, additive seasonality and random components.

Questions 1 & 2 (1 page) – The Regression

Multiple Regression Model

A multiple regression model was applied to estimate sales on household appliances for Tasmania. The quarterly data reported for the last four years i.e. from March 2016 to December 2019 was considered in the model. The model contains the following components:

- Time variable: time indexes were appropriately assigned to each quarterly period using time *t* for the observed period and time *t+1* for the following period. That is, $time = 1$ for Mar-16, $time = 2$ for Jun-16, $time = 3$ for Sep-16, … , and $time = 16$ for December 2019, as shown in figure 1.1
- Dummy variables: as only one quarter is to be omitted from the number of seasons, three dummy variables, *DumMar, DumJun and DumSep* were identified for the quarters of March, June and September, with December being the baseline. Each of the dummy variables were calculated for each observed period by employing $D_i = 1$ if the observation is season i or $D_i = 0$ otherwise i.e. if observation is not season i (see figure 1.2)
- A regression analysis was then performed through the following: *Data* > *Data Analysis* > *Regression*. Expenditure (*CELLS $B$1:$B$17*) was inputted into the *Input Y Range* and the time variable as well as the three dummy variables for March, June and September (*CELLS $C$1:$F$17*) were entered into the *Input X Range* > *OK*, as shown in figure 1.3. A summary output was then generated by Excel as shown in figure 1.4. From the summary output the intercept, the coefficient for the time variable and the coefficients for each of the three dummy variables were identified, see figure 1.5.

The overall regression equation is: $Y_t = B_0 + B_1X_{1t} + B_2X_{2t} + \cdots + B_kX_{kt} + \varepsilon_t$. In the context of the time and dummy variables as well as assuming additive seasonality and a linear trend, the model is postulated:

$$Y_t = \beta_0 + \beta_1 \times t + \alpha_1 \times D_1 + \alpha_2 \times D_2 + \alpha_3 \times D_3 + \varepsilon_t$$
$$Expenditure_t = \beta_0 + \beta_1 time + \alpha_1 Mar + \alpha_2 Jun + \alpha_3 Sep + \varepsilon_t$$
$$Expenditure_t = 57.3125 + 0.46875t + (-9.24275)Mar + (-6.0625)Jun + (-9.03125)Sep + \varepsilon_t$$

- Where:
  - Y_t = expenditure on household appliances in Tasmania for time t
  - β_0 = the intercept
  - β_1 = the coefficient for the time variable *t*
  - α_i = the coefficient for the dummy variable D_i
  - ε_t = the random component i.e. the margin of error within the regression model

Interpreting the components

- The intercept: is the expected mean of the dependent Y variable i.e. *expenditure*, when all independent X variables, i.e. the *coefficient for the time variable* and the coefficients for each of the dummy variables, *DumMar, DumJun and DumSep*, are all equal to zero. When graphed against the X variables, it is the given Y value at the point in which the graphed relation meets the Y-axis. According to the regression summary output, when all independent variables as mentioned above are equal to zero, the expected mean value of expenditure on household appliances in Tasmania is approximately $57.3125 million.
- The coefficient for the time variable: is the number of units on average that the dependent Y variable increases or decreases for each unit of time increase. According to the output, for each increasing unit of time i.e. per quarter, expenditure of household appliances in Tasmania increases by 0.46875 units on average, holding all other variables constant.
- The coefficient for the September dummy variable, *DumSep*: is the number of units higher or lower relative to the baseline i.e. December quarter – the group with all $D_i = 0$ – for the September quarter. Corresponding to the summary output, expenditure for the September quarter is approximately on average 9.03125 units lower than the average expenditure for the December quarter, holding all other variables constant.

Questions 3 & 4 (1 page) – R² and the Dummy Variables

The coefficient of determination

The coefficient of determination, also known as R^2 is a statistical measurement that explains how the percentage variation in one variable i.e. the dependent Y variable is explained by the variation in another variable that is, the independent X variable, in predicting the outcome of an event. R^2 lies between zero and one, and is calculated using the following equation:

$$R^2 = \frac{SSR}{SST} = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y})^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2}$$

The closer R^2 is to one, the better the model fits the sample data. Given the regression summary output (figure 1.4), 90.73% of the variation in the expenditure of household appliances in Tasmania is explained by the variation in the time variable and the dummy variables for the March, June and September quarters, which is quite high and affirms the accuracy of the model employed. R^2 also assesses the strength of the linear relationship between the expenditure variable and the time and dummy variables. However, it should not be the benchmark indicator of strength as it is flawed and can easily be influenced. As it assumes that the variation of every single independent variable explains the variation of the dependent variables, it does not take into consideration the number and relevancy of the independent variables. Hence, it does not accurately indicate the strength of the model. *Adjusted* R^2 provides a solution to this problem, considering the number and relevancy of the independent variables and giving a penalty for adding independent variables that are not appropriate for the model. Thus, the adjusted R^2 refers to the variation in the dependent variable that is explained by only the independent variables truly influencing the dependent. The regression summary output conveys that 87.36% of variation in the expenditure of household appliances in Tasmania is explained by the variation in the independent variables actually affecting expenditure.

Regression equation

March:

$$Expenditure_{Mar} = 57.3125 + (0.46875 \times t) + (-9.34375 \times 1) + (-6.0625 \times 0) + (-9.03125 \times 0) + \varepsilon_t$$

June:

$$Expenditure_{Jun} = 57.3125 + (0.46875 \times t) + (-9.34375 \times 0) + (-6.0625 \times 1) + (-9.03125 \times 0) + \varepsilon_t$$

September:

$$Expenditure_{Sep} = 57.3125 + (0.46875 \times t) + (-9.34375 \times 0) + (-6.0625 \times 0) + (-9.03125 \times 1) + \varepsilon_t$$

December:

$$Expenditure_{Sep} = 57.3125 + (0.46875 \times t) + (-9.34375 \times 0) + (-6.0625 \times 0) + (-9.03125 \times 0) + \varepsilon_t$$

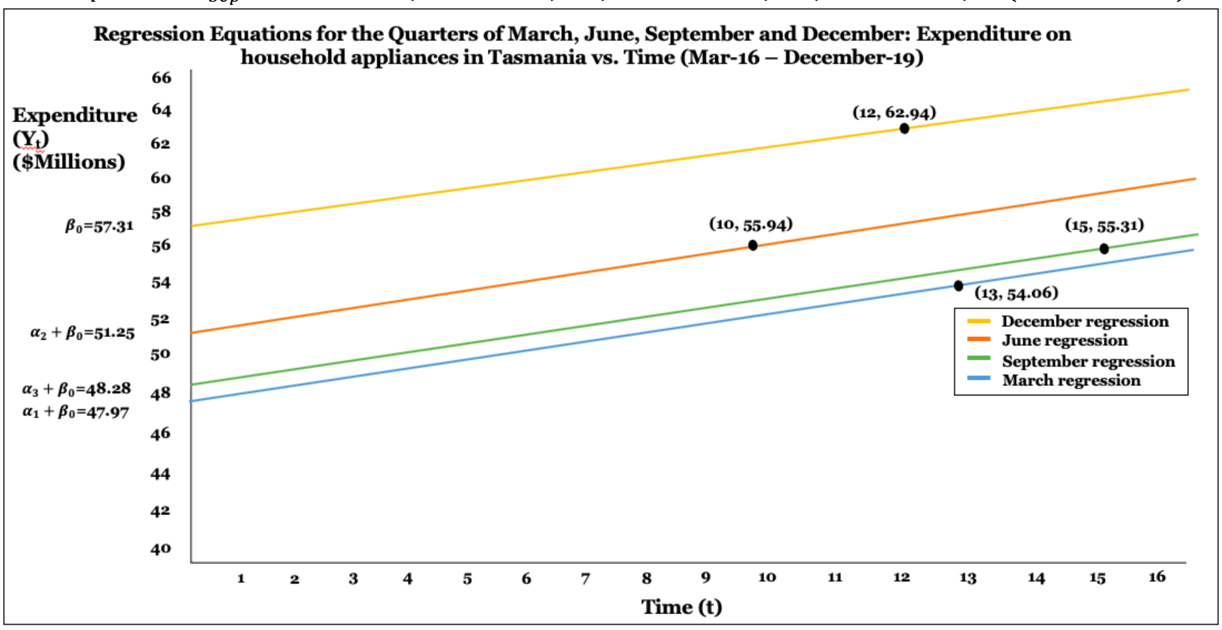


Figure 2. Expenditure vs time; multiple regression equations graphed for the quarters of March, June, September and December from Mar-16 (time 1) to Dec-19 (time 16).

Questions 5 & 6 (one page) – Tests

Testing the overall significance of the model

The appropriate test to test for the overall significance of the model is the *F test*, which tests the explanatory power of the model. It follows the following hypothesis: $H_0: B_1 = B_2 = \cdots = B_k = 0$ $(no\ relationship)$; $H_1: at\ least\ one\ B_i \neq 0$ $(at\ least\ one\ independent\ variable\ affects\ Y)$

The F-test suggests that if the p-value of the test statistic falls below the significance level of $\alpha = 0.05$, then the null hypothesis H_0 should be rejected. Rejection of the null hypothesis suggests some explanatory power of the model selected. In particular, it implies that the model has some predictive ability. The non-rejection of the null hypothesis i.e. all of the population slope coefficients are conjointly zero is indicative that collectively, the independent variables employed, and the model selected have no significant explanatory power. The p-value for the F-test, which is obtained from the ANOVA table in the regression summary output is approximately 124571E-05 (0.00001246). As this is less than 0.05 and is insignificant, the null hypothesis is rejected at the 5% significance level. The p-value 124571E-05, is not the probability that H_0 is true but indicates there is not enough evidence to state that there is no relationship between the independent variables and the dependent variable. As the p-value is very low, there is more evidence against H_0, hence, exhibiting that there is a relationship between the independent and dependent variables – at least one independent variable may affect expenditure – and that the regression model is statistically significant and may have some predictive ability in forecasting expenditure on household appliances in Tasmania.

Testing the individual significance of the model:

The appropriate test to assess the individual significance of the model is the *t test*, which separately tests the explanatory power of each independent variable. It follows the following hypothesis: $H_0: B_j = 0$ $\left(no\ linear\ relationship\ exists\ between\ X_j\ and\ Y\right)$; $H_1: B_j \neq 0$ $(linear\ relationship\ exists\ between\ X_j\ and\ Y)$

The t-test proposes that if the individual p-value of the test statistic falls below the significance level of $\alpha = 0.05$, then the null hypothesis H_0 should be rejected. Rejection of the null hypothesis suggests some explanatory power that the variable selected has some explanatory power and may influence the dependent variable. The non-rejection of the null hypothesis i.e. the population slope coefficients (B_j) being zero, is indicative that the independent variable being assessed does not influence the dependent variable and highlights the possibility of the variable being removed from the model due to its irrelevancy.

<u>Coefficient for the time variable</u>: Conducting a t-test for the individual significance of the coefficient for the time variable, the p-value as per the ANOVA table is 0.0006806, which is less than 0.05 and is insignificant. On this basis, the null hypothesis is rejected at the 5% significance level. The p-value exhibits evidence against H_0, denoting a possible linear relationship between the time variable and expenditure i.e. the time variable may in a linear manner influence the expenditure on household appliances in Tasmania. Thus, this independent variable should remain in the model as it is statistically significant may have some explanatory power and relevance

<u>Coefficient for the September dummy variable</u>: Conducting a t-test for the individual significance of the coefficient for the September dummy variable, the p-value according to the ANOVA table is 0.00002009, which is also insignificant and is less than the significance level of 0.05. As a result, the null hypothesis is rejected at the 5% significance level, suggesting the probable linear relationship between the September dummy and the expenditure on household appliances in Tasmania. The p-value is much lower than that of the coefficient for the time variable, hence this individual t-test provides more evidence against H_0, as compared to the t-test of the coefficient for the time variable. Similarly, this independent variable should remain in the model as it is statistically significant and may have some explanatory power and weighting.

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

Employing the multiple regression model, forecasts were generated for March, June, September and December for 2020, as shown in figure 1.6. Overall, the model has generated an effective forecast, still representing a positive linear trend and a seemingly additive seasonality component (see figure 3). As shown, the December quarter exhibits a peak in expenditure with March continuing to show a trough. Collectively, this reasonably supports the accuracy and validity of the forecasts produced, assisting HomeAppCo to deliberate on a possible expansion to Australia. An example forecasting calculation is shown for expenditure on household appliances in Tasmania for the quarter of September 2020:

$$Expenditure_{Sep-20} = 57.3125 + (0.46875 \times 19) + (-9.34375 \times 0) + (-6.0625 \times 0) + (-9.03125 \times 1)$$
$$Expenditure_{Sep-20} = 57.3125 + 8.90625 + 0 + 0 - 9.03125$$
$$Expenditure_{Sep-20} = 57.1875$$

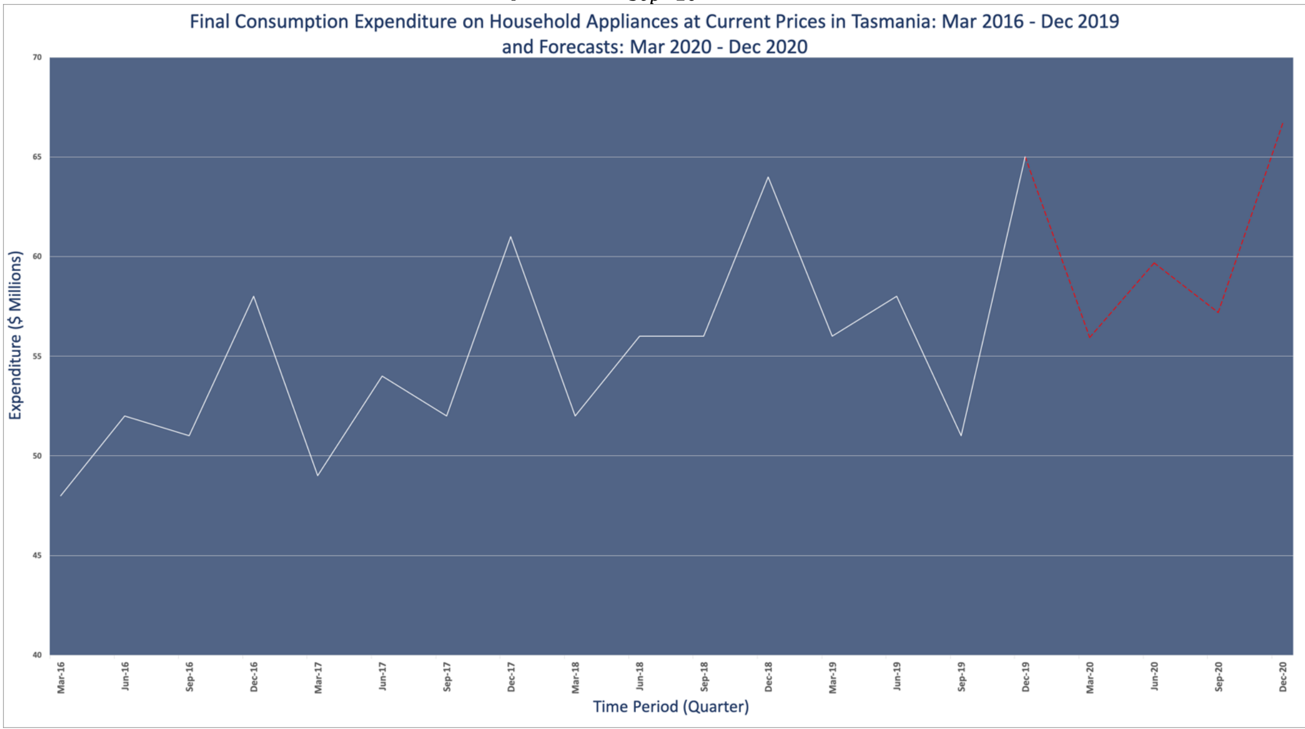


Figure 3. Time series for quarterly reported household final consumption expenditure on household appliances at current prices in Tasmania (Mar-16 - Dec-19) as well as the forecasts (Mar20 - Dec-20)

Compared to the winters' exponential smoothing (WES) method, the multiple regression model proves to be more appropriate, as it provides a causal explanation of why the forecast is the way it is. As overall and individual tests for significance saw the rejection of the null hypothesis, the model does not need changing nor do specific independent variables need removing i.e. no stepwise regression is needed. The standard error of the regression indicates the average distance the observed values are from the regression line (Frost, J, n.d.). In figure 1.4, the standard error of the model is 1.7949 while the standard errors of the time coefficient and the coefficients for March, June and September were 0.1003, 1.3044, 1.2849 and 1.2731 respectively, which aren't highly significant. As such, the coefficient for time variable proves to have the smallest error. Figure 1.7 and 1.8 also show overall residuals and residuals of each independent variable appearing to be random and exhibiting a mean that is about zero, a better result than the WES method. Hence, any errors here are likely owing to randomness. Although an additive WES model would be suitable in smoothing randomness and minimising the error in its forecasts, the regression model allows HomeAppCo to identify which factors actually influence the forecasts of sales on household appliances in Tasmania and which ones don't, a quality, not evident in the WES model. This would allow HomeAppCo to maximise their sales and thus consider its expansion to Australia. Hence the multiple regression model could provide scenario-based (best and worst case) analyses and prediction, which would allow HomeAppCo's management to identify sources of risk in their forecasts when considering a possible expansion.

Questions 9 & 10 (one page) – Judgmental Forecasting

Subjective Assessment Methods

Subjective assessment methods rely on the weight and assessment of relevant or subjective information to yield forecasts. This short to medium term group of methods is a form of judgemental forecasting, which is based on appropriate information performed in a quasi-systematic manner. There are three components of subjective assessment methods: *sales force composite forecasting*, an aggregation of the assessments done by individual salespeople or sales management about prospective sales; *jury of executive opinion* is a representative forecast collectively made on the basis of forecasts or judgements made by individuals; and, *subjective probability assessments*, integrating individual judgement by assigning probabilities to different outcomes of an unknown event. Subjective assessment methods along with quantitative methods, may aid HomeAppCo in providing qualitative value in forecasting the sales of household appliances in Tasmania, as quantitative models can neglect certain changes in the data. Subjective assessment methods may support the forecasting of expenditure on household appliances and the decision of HomeAppCo to expand to Australia, as they allow for rapid changes and uncertain futures. For instance, employing the sales force composite forecasting method particularly the Grass Roots Approach would provide salespeople and sales management with the opportunity to express their views about the future sales outlook, in their field of expertise. Hence, during a period as uncertain as the COVID-19 pandemic, sales staff and distributors are able to better predict future outcomes based on their knowledge in the field. However, the COVID-19 pandemic proved to be an unknown experience to every professional, including forecasters. Due to the unavailability of accurate and reliable data about the effects of the virus, it is unlikely even the most experienced of forecasters and modellers, even the government, could predict the depth, duration and the impacts of COVID-19 on the expenditure on household appliances. Although subjective assessment methods allow individuals to forecast, whether or not this alone, is a method that can truly address the business problem of HomeAppCo is uncertain and thus must be used alongside quantitative methods.

Scenario Analysis

Contrary to subjective assessment methods, *explanatory methods* explore multiple potential forecast outcomes and are used when there is an absence of other (quantitative) information for medium to long-term forecasting. *Scenario analysis* is an explanatory method, which identifies plausible ranges of forecast outcomes and the consequences of those outcomes. Scenarios are developed by analysing the costs and magnitude of each individual scenario on different outcomes and the probability of each event occurring. The seven stages of identifying scenarios include: *identifying the focal decision*, *determining the circumstances influencing this particular decision; determining the driving forces influencing these circumstances; ranking these circumstances and its driving forces by uncertainty and significance*; *establishing alternative scenarios affecting the decision*; *evaluating the decision in the posited future scenarios*; *and determining the indications and signposts for each scenario*. Scenarios allows an organisation to take into consideration and plan for the possible future outcomes they may encounter, whether positive or negative. As scenarios allow the combination of all social, economical, technological and political data, scenario analysis is undoubtedly useful for aiding HomeAppCo in being more prepared for periods of uncertainly, particularly the adverse effects of COVID-19. For instance, by identifying the decision to expand HomeAppCo to Australia, scenario analysis would enable management to possibly consider a pandemic as a future alternative scenario similar to that of COVID-19. Identifying the social, economical, technological and political effects caused by the pandemic that may influence the sales would allow management to prioritise this scenario as one that is highly uncertain and unknown. As a result, in a time of uncertainty and high risk, indicators and signposts as determined in step seven would be identified much earlier allowing for a more accurate and timelier forecast. Hence, scenario analysis highlights to organisations the importance of not relying on forecasts alone, but also being malleable in considering adaptable strategies that could work for several different future outcomes, possibly even a pandemic. In the context of judgemental forecasting, scenario analysis would greatly aid forecasters in generating accurate, valid and timely forecasts for long-term periods of uncertainty.

References

Frost, J, n.d., *Standard Error of the Regression*, viewed 19 May 2020, https://statisticsbyjim.com/glossary/standard-error-regression/

Macquarie University, May 2020, Part 1 - Judgmental Forecasting I - (Report 2 - Question 9 Research), viewed 16 May 2020, https://ilearn.mq.edu.au/pluginfile.php/6230553/mod_resource/content/1/Assignment%202%20Question%209%20Research.pdf

Macquarie University, May 2020, Part 2 - Judgmental Forecasting II - (Report 2 - Question 10 Research), viewed 16 May 2020, https://ilearn.mq.edu.au/pluginfile.php/6230554/mod_resource/content/1/Assignment%202%20Question%2010%20Research.pdf

StatisticsHowTo, n.d., *Adjusted R-squared: What is it used for?*, viewed 19 May 2020, https://www.statisticshowto.com/adjusted-r2/

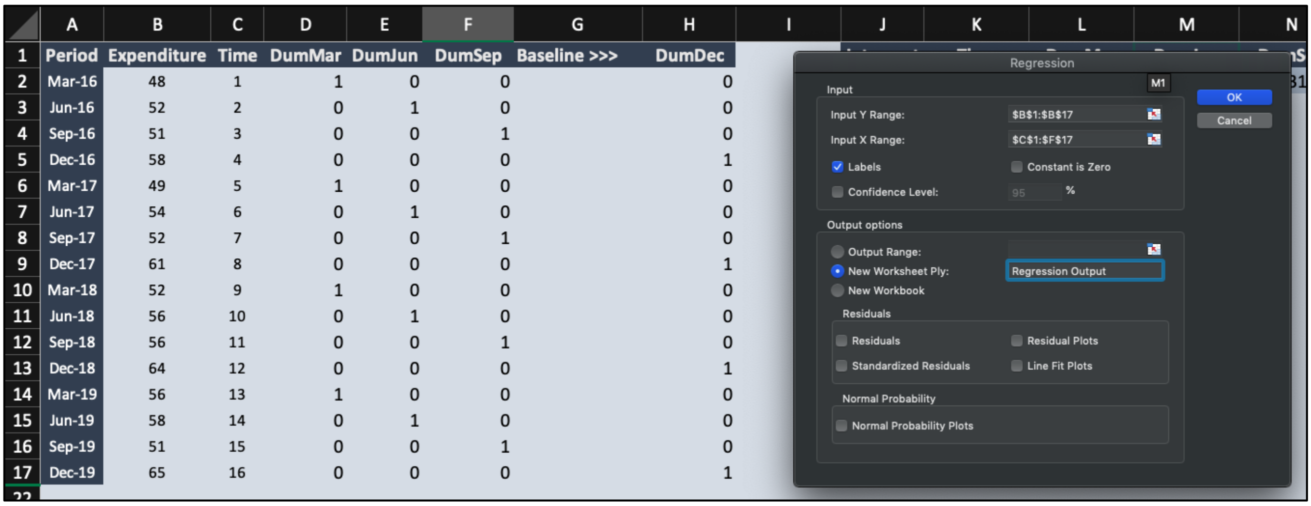
Appendices

Period	Expenditure	Time
Mar-16	48	1
Jun-16	52	2
Sep-16	51	3
Dec-16	58	4
Mar-17	49	5
Jun-17	54	6
Sep-17	52	7
Dec-17	61	8
Mar-18	52	9
Jun-18	56	10
Sep-18	56	11
Dec-18	64	12
Mar-19	56	13
Jun-19	58	14
Sep-19	51	15
Dec-19	65	16

Figure 1.1 The time variable, as allocated to each quarterly period.

Period	Expenditure	Time	DumMar	DumJun	DumSep	Baseline >>>	DumDec
Mar-16	48	1	1	0	0		0
Jun-16	52	2	0	1	0		0
Sep-16	51	3	0	0	1		0
Dec-16	58	4	0	0	0		1
Mar-17	49	5	1	0	0		0
Jun-17	54	6	0	1	0		0
Sep-17	52	7	0	0	1		0
Dec-17	61	8	0	0	0		1
Mar-18	52	9	1	0	0		0
Jun-18	56	10	0	1	0		0
Sep-18	56	11	0	0	1		0
Dec-18	64	12	0	0	0		1
Mar-19	56	13	1	0	0		0
Jun-19	58	14	0	1	0		0
Sep-19	51	15	0	0	1		0
Dec-19	65	16	0	0	0		1

Figure 1.2 The dummy variables for each of the three quarters, DumMar, DumJun and DumSep as calculated for each observed period with the baseline being DumDec.



Period	Expenditure	Time	DumMar	DumJun	DumSep	Baseline >>>	DumDec
Mar-16	48	1	1	0	0		0
Jun-16	52	2	0	1	0		0
Sep-16	51	3	0	0	1		0
Dec-16	58	4	0	0	0		1
Mar-17	49	5	1	0	0		0
Jun-17	54	6	0	1	0		0
Sep-17	52	7	0	0	1		0
Dec-17	61	8	0	0	0		1
Mar-18	52	9	1	0	0		0
Jun-18	56	10	0	1	0		0
Sep-18	56	11	0	0	1		0
Dec-18	64	12	0	0	0		1
Mar-19	56	13	1	0	0		0
Jun-19	58	14	0	1	0		0
Sep-19	51	15	0	0	1		0
Dec-19	65	16	0	0	0		1

Figure 1.3 Performing the regression analysis; cells $B$1:$B$17 was inputted into the Input Y Range and cells $C$1:$F$17 were entered into the Input X Range.

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.952542808
R Square	0.9073378
Adjusted R Square	0.873642455
Standard Error	1.794879079
Observations	16

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	4	347	86.75	26.92768959	1.24571E-05
Residual	11	35.4375	3.221590909		
Total	15	382.4375			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	57.3125	1.346159309	42.57482721	1.46226E-13	54.34962334	60.27537666	54.34962334	60.27537666
Time	0.46875	0.100336791	4.671765922	0.000680588	0.247910212	0.689589788	0.247910212	0.689589788
DumMar	-9.34375	1.304378281	-7.163374413	1.83686E-05	-12.21466724	-6.472832761	-12.21466724	-6.472832761
DumJun	-6.0625	1.284937874	-4.718126939	0.000631354	-8.890629193	-3.234370807	-8.890629193	-3.234370807
DumSep	-9.03125	1.27313115	-7.093731072	2.0094E-05	-11.83339277	-6.229107231	-11.83339277	-6.229107231

Figure 1.4 Regression Summary Output as generated by Excel.

Intercept	Time	DumMar	DumJun	DumSep
57.3125	0.46875	-9.34375	-6.0625	-9.03125

Figure 1.5. The intercept, coefficient for the time variable as well as the coefficients for each of the dummy variables, DumMar, DumJun and DumSep are identified from the coefficients column (bottom left).

B20 =$J$2+($K$2*$C$20)+($L$2*$D$20)+($M$2*$E$20)+($N$2*$F$20)

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	Period	Expenditure	Time	DumMar	DumJun	DumSep	Baseline >>>	DumDec		Intercept	Time	DumMar	DumJun	DumSep
2	Mar-16	48	1	1	0	0		0		57.3125	0.46875	-9.34375	-6.0625	-9.03125
3	Jun-16	52	2	0	1	0		0						
4	Sep-16	51	3	0	0	1		0						
5	Dec-16	58	4	0	0	0		1						
6	Mar-17	49	5	1	0	0		0						
7	Jun-17	54	6	0	1	0		0						
8	Sep-17	52	7	0	0	1		0						
9	Dec-17	61	8	0	0	0		1						
10	Mar-18	52	9	1	0	0		0						
11	Jun-18	56	10	0	1	0		0						
12	Sep-18	56	11	0	0	1		0						
13	Dec-18	64	12	0	0	0		1						
14	Mar-19	56	13	1	0	0		0						
15	Jun-19	58	14	0	1	0		0						
16	Sep-19	51	15	0	0	1		0						
17	Dec-19	65	16	0	0	0		1						
18	Mar-20	55.9375	17	1	0	0		0						
19	Jun-20	59.6875	18	0	1	0		0						
20	Sep-20	57.1875	19	0	0	1		0						
21	Dec-20	66.6875	20	0	0	0		1						
22														

Figure 1.6 Forecasts generated for March 2020, June 2020, September 2020 and December 2020. Calculation for September 2020 forecast shown as example.

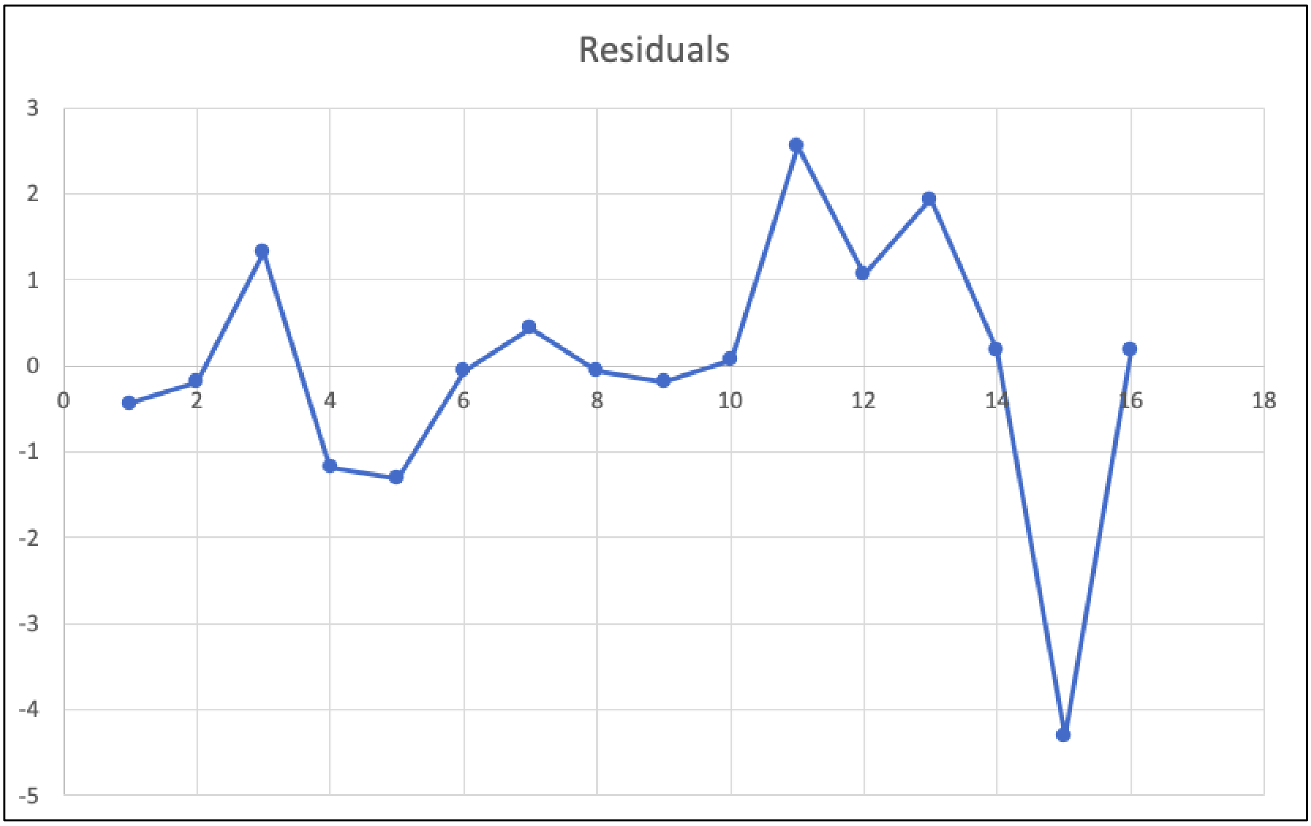


Figure 1.7 Overall residual plot for multiple regression model appears to be random and has a mean about zero.

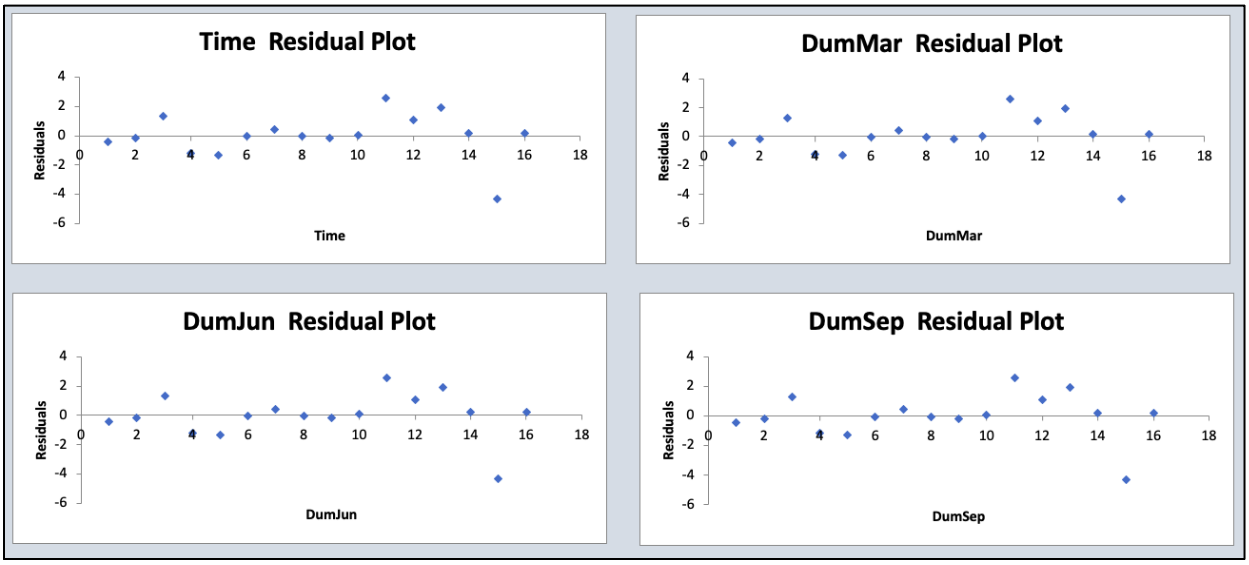


Figure 1.8 Residual plots for each of the independent variables all appear to be random and has a mean about zero.