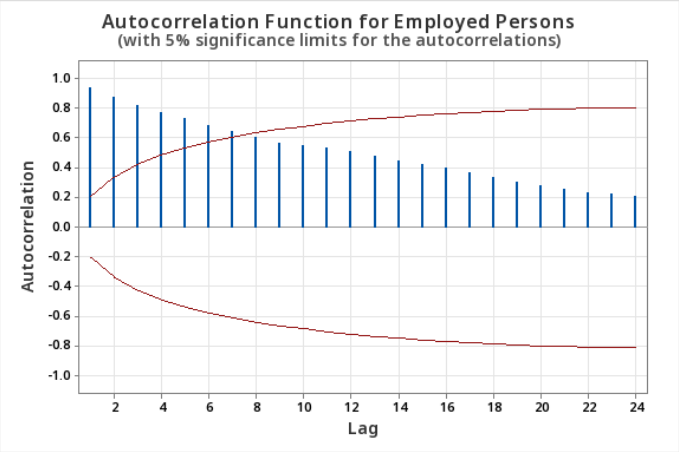
# Case Study Report 2 – Exercise 3

## Attribution

This report will use both the seasonally adjusted (A84423043C) and original (A84423043C) datasets for total employment in Australia. This data has been collected from the ABS website, more specifically the dataset in Table 1 with catalogue number 6202001. The two datasets show the monthly data for the total number of employed Australians. For this report, a sample period from August 2014 to July 2022 will be used to generate forecasts regarding employment within Australia.

## Scope

Multiple regression analysis is a causal forecasting model used to observe the relationship between a single dependent variable (Y) and various independent variables (X).


The data on total employment in Australia reveals seasonal and trend components, with higher values in certain periods. A simple regression model should not be used for forecasting future trends, as it doesn't account for seasonal aspects. A multiple regression analysis, specifically a quasi-explanatory model, is more effective as it can help accurately predict datasets that have seasonal components.

## Application

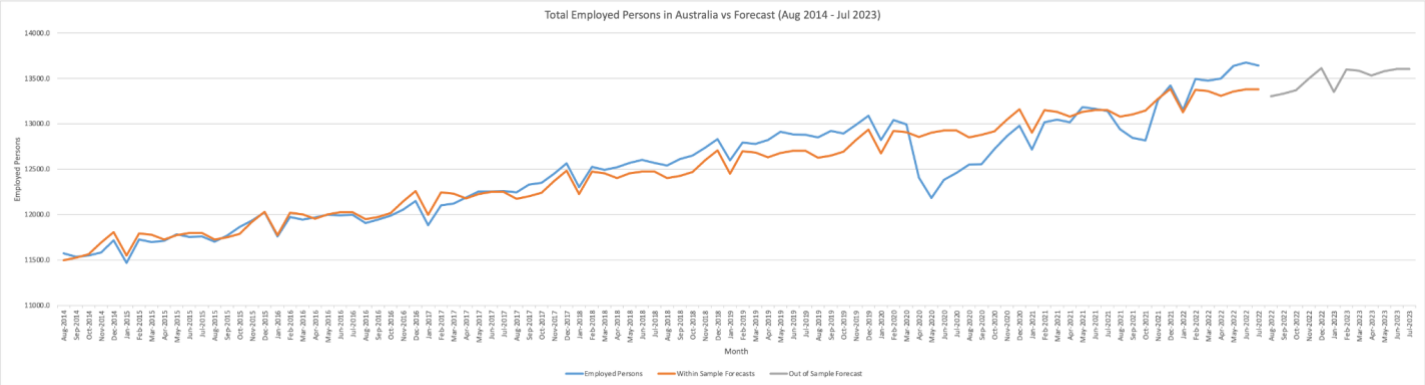
Utilising a quasi-explanatory model as part of the multiple regression analysis can help provide a more accurate forecast in predicting future levels of total employment in Australia. Through the use of this model, explanatory variables (X) such as the time index "t" and seasonal dummies have been constructed to model the seasonal effects that could influence the time-series dataset. Since the dataset is monthly, 11 dummy variables from August to July have been used to model the seasonality effects in the data, as the original dataset has not been seasonally adjusted.

To generate forecasts of future time periods (Yt) beyond the sample period, the multiple linear regression equation below was utilised.

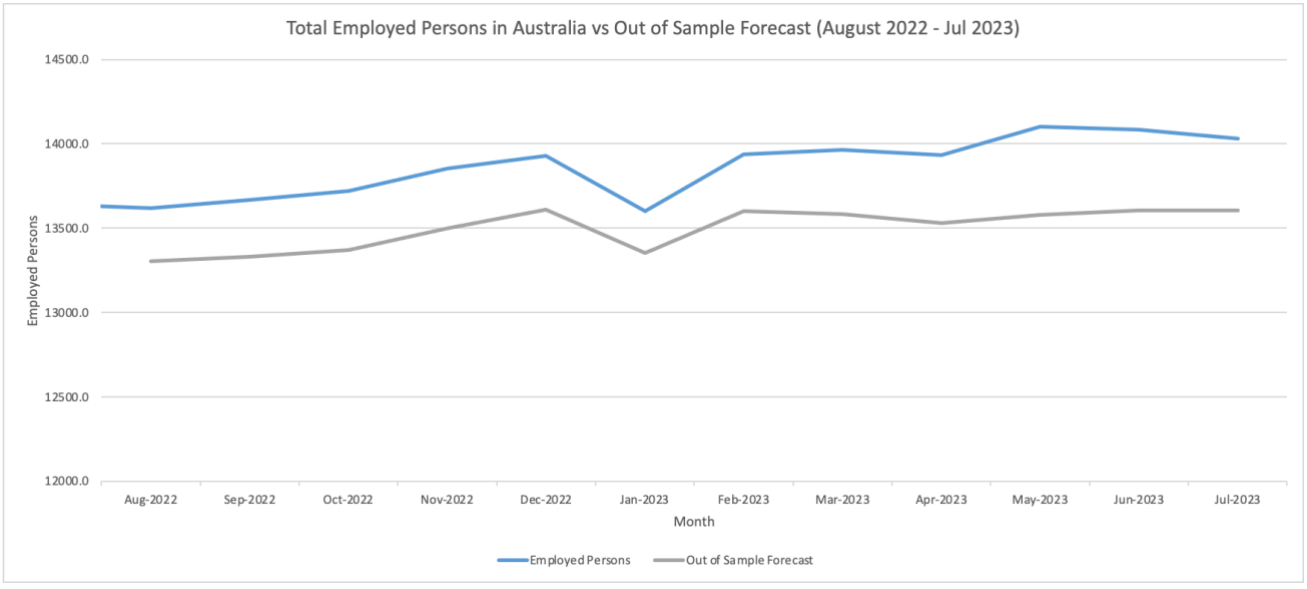
$$Y_t = \beta_0 + \beta_1 * t + \alpha_1 * D_1 + \alpha_2 * D_2 + \alpha_3 * D_3 + \varepsilon_t$$

B0 and B1 represent the intercept coefficient and the time coeffcient (X1) respectively, these coefficients were generated from the regression analysis conducted in Excel. The time index "t" acts as a quasi explanatory variable to explain and forecast Yt, which is used to estimate B0 and B1. The coefficients for each of months used in the analysis are multiplied by the corresponding monthly dummy variable. This equation helps determine the slope of the trend line, which helps forecast future time series values (Yt)

**Analysis**




The above figure shows the recorded data for employment in Australia in the specified sample period, as well as the within sample forecasts for employment. Additionally, the figure also shows the predicted future values for total employment in Australia, or the out-of-sample forecast. The forecasts are consistent from August 2014 to February 2020. The forecasts from this graph are significantly different from the actual recorded values from March to October 2020 due to the effects of the unprecedented global COVID-19 pandemic, where there was a decrease in employment of 6.1% from Australia's first recorded COVID-19 case to the trough in the labour market in May 2020.




The figure above shows the comparison between the actual recorded total employment in Australia and the out-of-sample forecast for August 2022 - July 2023. Whilst there is a slight difference in terms of the values

recorded in the actual data and the forecast, the model is fairly accurate in predicting the future values of employment in Australia. The typical seasonality effects are also present in this figure, which indicate that employment in Australia continues to have similar patterns throughout each seasonal period.

**Articulation of Issues**

From the regression analysis, it can be determined that the model is a good fit for the sample data, as the R2 is very close to 1. The overall significance of the model can also be assessed using the multiple regression model in the form of an f-test. The F Significance (or p-value) is lower than 0.05, indicating that the model has some explanatory power and potential predictive ability.

| | A | B | C | D | E | F | G | H | I |
|---|---|---|---|---|---|---|---|---|---|
| 1 | SUMMARY OUTPUT | | | | | | | | |
| 2 | | | | | | | | | |
| 3 | | *Regression Statistics* | | | | | | | |
| 4 | Multiple R | 0.949909224 | | | | | | | |
| 5 | R Square | 0.902327533 | | | | | | | |
| 6 | Adjusted R Sq | 0.888206212 | | | | | | | |
| 7 | Standard Erro | 187.543257 | | | | | | | |
| 8 | Observations | 96 | | | | | | | |
| 9 | | | | | | | | | |
| 10 | ANOVA | | | | | | | | |
| 11 | | *df* | *SS* | *MS* | *F* | *Significance F* | | | |
| 12 | Regression | 12 | 26969509.76 | 2247459.147 | 63.8982403 | 1.0458E-36 | | | |
| 13 | Residual | 83 | 2919315.281 | 35172.47326 | | | | | |
| 14 | Total | 95 | 29888825.04 | | | | | | |

A t-test can be conducted to examine the effect of individual variables (X) on the dependent variable. From analysing the individual variables, it can be determined that time is significant as the p-value < 0.05, however the monthly seasonal dummies don't seem to have any significance or predictive ability on total employment.

| 16 | | *Coefficients* | *Standard Error* | *t Stat* | *P-value* | *Lower 95%* | *Upper 95%* | *Lower 95.0%* | *Upper 95.0%* |
|---|---|---|---|---|---|---|---|---|---|
| 17 | Intercept | 11573.00306 | 76.22164317 | 151.8335551 | 2.919E-103 | 11421.4013 | 11724.6048 | 11421.4013 | 11724.6048 |
| 18 | Time (x1) | 18.80053873 | 0.696153386 | 27.00631658 | 7.1042E-43 | 17.4159176 | 20.1851598 | 17.4159176 | 20.1851598 |
| 19 | August | -92.79673594 | 94.08378442 | -0.986320188 | 0.32684258 | -279.9256 | 94.3321272 | -279.9256 | 94.3321272 |
| 20 | September | -86.08902028 | 94.02968292 | -0.915551532 | 0.36255453 | -273.11028 | 100.932237 | -273.11028 | 100.932237 |
| 21 | October | -64.60866128 | 93.98070711 | -0.687467282 | 0.4937049 | -251.53251 | 122.315185 | -251.53251 | 122.315185 |
| 22 | November | 46.02439923 | 93.936865 | 0.489950343 | 0.62546161 | -140.81225 | 232.861046 | -140.81225 | 232.861046 |
| 23 | December | 140.0548282 | 93.89816379 | 1.491560884 | 0.13960539 | -46.704843 | 326.814499 | -46.704843 | 326.814499 |
| 24 | January | -137.7232963 | 93.86460984 | -1.467254768 | 0.14608588 | -324.41623 | 48.9696375 | -324.41623 | 48.9696375 |
| 25 | February | 90.25431949 | 93.83620865 | 0.961828283 | 0.33892965 | -96.382125 | 276.890764 | -96.382125 | 276.890764 |
| 26 | March | 55.60223637 | 93.81296493 | 0.592692454 | 0.55499839 | -130.98798 | 242.19245 | -130.98798 | 242.19245 |
| 27 | April | -14.9301012 | 93.79488249 | -0.159178207 | 0.87391517 | -201.48435 | 171.624148 | -201.48435 | 171.624148 |
| 28 | May | 14.82760285 | 93.78196433 | 0.15810719 | 0.87475637 | -171.70095 | 201.356158 | -171.70095 | 201.356158 |
| 29 | June | 18.86086371 | 93.77421258 | 0.201130601 | 0.84108853 | -167.65227 | 205.374001 | -167.65227 | 205.374001 |

**Critique**

The multiple regression analysis model enables forecasters to analyse best and worse case scenarios, which can help provide greater strategy and scenario-based prediction. This allows the model to determine sources of risk to total employment in Australia, especially if there are any external factors that could influence the dataset. The WES (Multiplicative) model on the other hand does not account for any external factors that could influence the dataset such as the unprecedented COVID-19 pandemic. It can be said however that the regression method is expensive and time-consuming to conduct, which may not ultimately result in greater forecasting accuracy.

**Position**

When forecasting and analysing the dataset for total employment in Australia, it can be determined that multiple regression models provide the most effective method for forecasting. The model is an accurate predictor for this dataset due to the relatively high R2 and adjusted R2. Additionally, the p-value from the f-test is statistically significant, indicating that the model is a effective predictor. The model also provides greater strategy and scenario-based prediction compared to other forecasting models, which can be highly beneficial for forecasters and businesses.