Report

Introduction

This study centers around analyzing historical data on employee counts in New Zealand's finance and insurance industry. The dataset employed.csv, contains comprehensive records of past employment figures within this sector.

The main goal of the study is to develop a reliable forecasting model capable of predicting future employee numbers in the finance and insurance industry of New Zealand. This predictive model holds great importance for policymakers, economists, and investors within the country. By utilizing this model, they can gain valuable insights into future economic trends, enabling them to make informed decisions.

A tsibble: 73 x 2 [1Q]

Quarter <S3: yearquarter>	Financial and Insurance Services <dbl>
2003 Q1	52.1
2003 Q2	51.1
2003 Q3	56.0
2003 Q4	60.6
2004 Q1	60.2
2004 Q2	59.2
2004 Q3	60.8
2004 Q4	59.9
2005 Q1	58.3
2005 Q2	68.1

1-10 of 73 rows

Table 1 : Number of Employees in Financial and Insurance Services Over Time

Data Analysis

By visualizing the data, we can observe regarding the overall trends present in the dataset. Since 2003, the number of employees in the financial and insurance industry in New Zealand has consistently shown an upward trend. There was rapid growth from the beginning until 2005, after which the growth rate gradually stabilized. In 2010, a minor downturn occurred. However, from 2018, the industry experienced a resumption of high-speed growth. Until 2022, a downward trend emerged, but further analysis is needed to determine the specific trajectory.

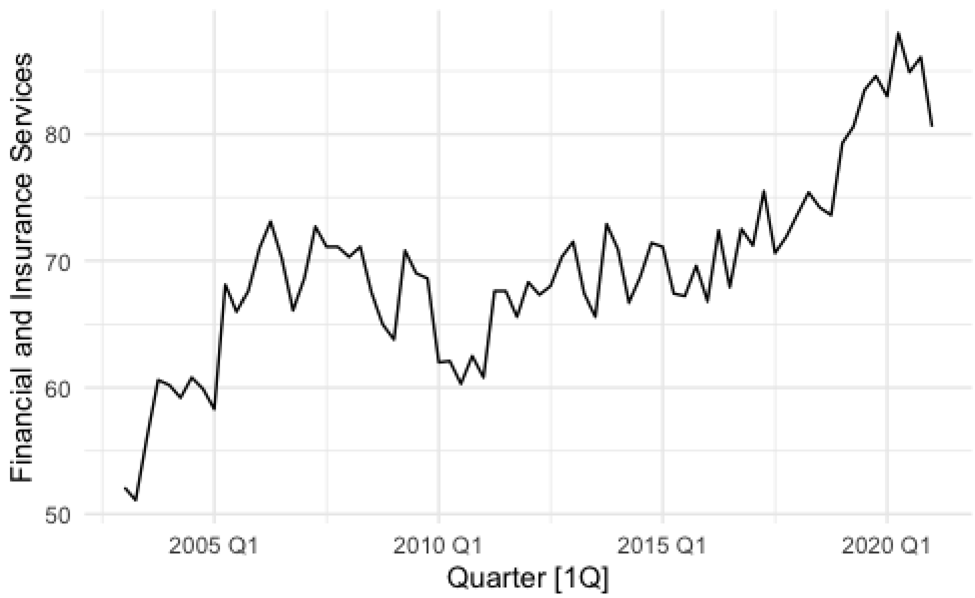

Figure 2 : Trend of Financial and Insurance Services Employees Over Time

The result of the CUSUM test also proves those breakpoints mentioned before.

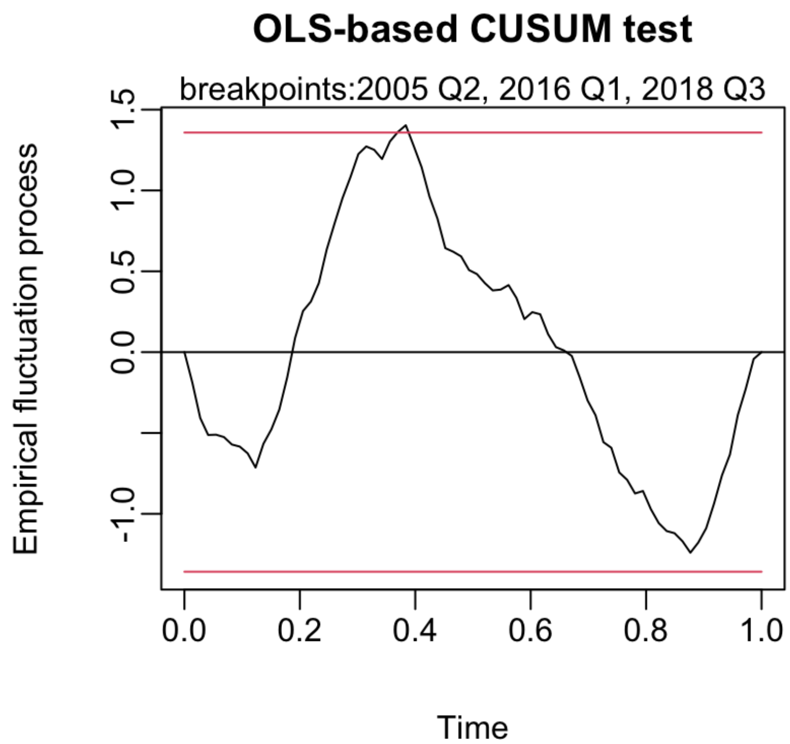

Figure 3 : Finding the breakpoints of structrual breaks

After a observing the raw data, we analyzed it using graphs and the STL method. From the decomposition results, it is clear that this time-series data shows seasonal patterns and mostly demonstrates an upward trend. However, there are downward trends around 2008 and after 2020.

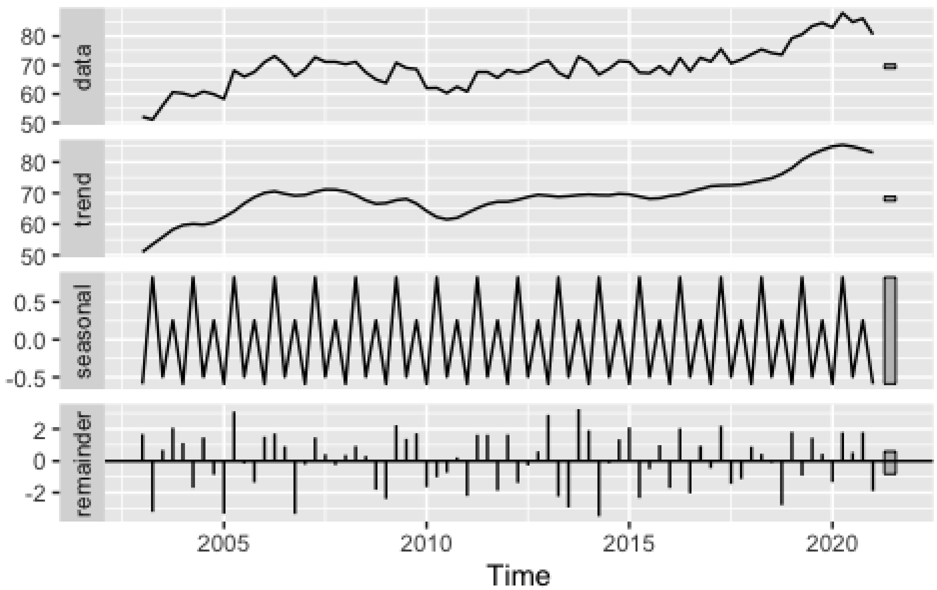

Figure 4 : STL Decomposition of Financial and Insurance Services Employees

Following that, we applied differencing to the employee count data and visualized the results. We began with seasonal differencing. After taking the first order difference, we observed whether the autocorrelation in the lagged periods weakened. From the graph, it can be seen that the mean and variance of the time series became relatively stable, no longer exhibiting significant changes over time.

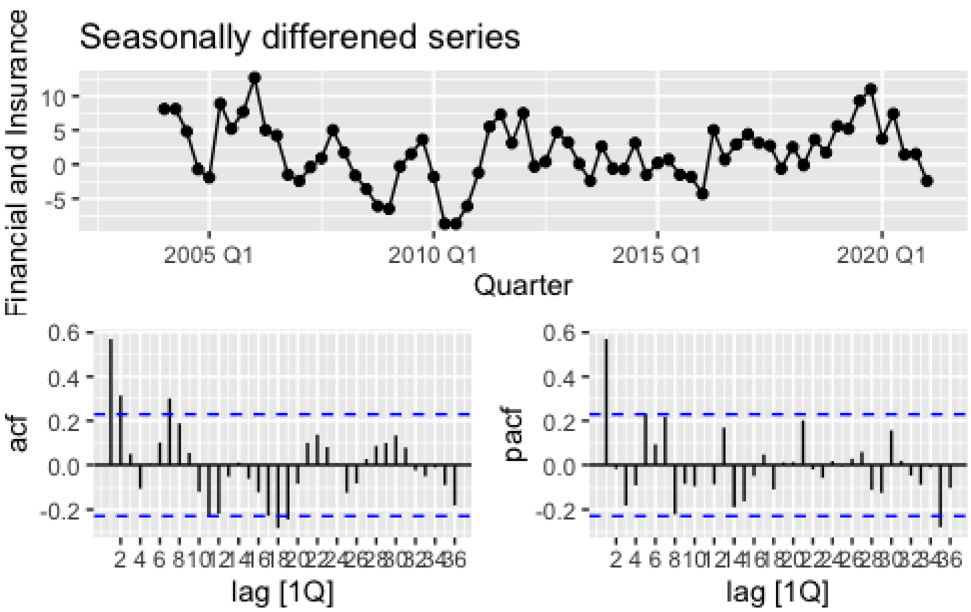

Figure 5: Partial Autocorrelation of First Order Difference of Financial and Insurance Services

Furthermore, we performed second-order differencing and visualized the results. From the graph, we can observe that the time series plot exhibits a more stable range of fluctuations. In the ACF plot, as the lag increases, the correlation between data points decreases rapidly until it approaches zero. This indicates that the time series after double differencing has become stationary. With this stationary data, we can proceed with further analysis and modeling tasks.

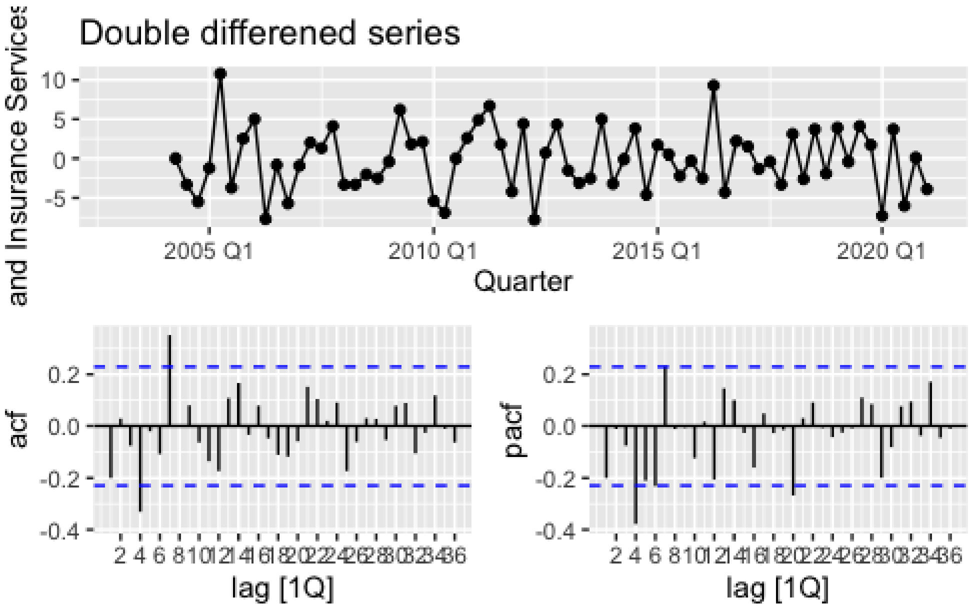

Figure 6: Partial Autocorrelation of Double Difference of Financial and Insurance Services

Time Series Model Selection

ARIMA

Next, we proceeded with automatic and manual selection of ARIMA and ETS models, followed by evaluating and comparing the models.

Firstly, we performed automatic selection of the optimal ARIMA model. This was done using the `ARIMA` function with `stepwise = FALSE` to conduct a full parameter space search, and then using `stepwise = TRUE` for a stepwise search. Finally, we utilized the `glance()` function to obtain performance metrics of the models and ranked them based on the AICc. Here are the results.

```
Series: Financial and Insurance Services
Model: ARIMA(0,1,1)(1,0,0)[4]

Coefficients:
          ma1    sar1
      -0.2730  0.2355
s.e.   0.1298  0.1212

sigma^2 estimated as 9.905:  log likelihood=-183.85
AIC=373.7   AICc=374.05   BIC=380.53
```

Figure 7: Model Report of ARIMA for Financial and Insurance Services

Then, we manually selected two ARIMA models: m1 and m2. The m1 model is an ARIMA(1,1,0)(1,0,0) model, while the m2 model is an ARIMA(0,1,1)(1,0,0) model. Both models take into account the seasonal influence, but they differ in their AR, I, and MA parameters. From the provided results, we can see that the AICc value of model m2 (374.0548) is smaller than the AICc value of model m1 (374.5775). A lower AICc value indicates better predictive performance for the model. Therefore, we choose model m2 as the optimal model. The advantage of model m2 is that it can fit the data well while maintaining relatively lower model complexity.

A tibble: 2 x 6

.model <chr>	sigma2 <dbl>	log_lik <dbl>	AIC <dbl>	AICc <dbl>	BIC <dbl>
m2	9.905198	-183.8509	373.7019	374.0548	380.5319
m1	9.977118	-184.1123	374.2246	374.5775	381.0546

2 rows

Table 2: Comparison of ARIMA Models for Financial and Insurance Services

ETS

Next, we used the ETS model to select the appropriate model for the data. We employed both automatic selection and manual selection methods, resulting in respective models and forecasted results.

Firstly, we employed the ETS model for the automatic selection of the optimal model. Using the `model()` function, we created a model by applying the ETS function to the time series data. Then, we used the `report()` function to generate the model's output and employed the `forecast()` function to generate forecasts for the next 8 quarters.

```
Series: Financial and Insurance Services
Model: ETS(A,N,N)
  Smoothing parameters:
    alpha = 0.6938722

  Initial states:
   l[0]
 52.378

  sigma^2:  10.3025

     AIC     AICc      BIC
487.4396 487.7874 494.3110
```

Figure 8: Model Report of ETS for Financial and Insurance Services

During the manual selection process, we compared the forecasting performance of three ETS models: ANN, AAN, and AAdN. We found that the ANN model had the smallest AICc value (487.79), indicating its superior performance. Consequently, we selected the ANN model as the best fit. Subsequently, we further optimized the parameters of this model. By comparing the models with different alpha values, we determined that the model with an alpha value of 0.7 yielded the smallest AICc value, indicating the best fit. This outcome provides crucial insights for our future decision-making process.

A tibble: 4 × 9

.model <chr>	sigma2 <dbl>	log_lik <dbl>	AIC <dbl>	AICc <dbl>	BIC <dbl>
0.1	53.79935	-301.0499	606.0997	606.2711	610.6806
0.4	26.55380	-275.2776	554.5553	554.7267	559.1362
0.7	10.68423	-242.0479	488.0958	488.2672	492.6767
0.9	11.08198	-243.3820	490.7640	490.9354	495.3449

4 rows | 1-6 of 9 columns

Table 3: Comparison of ETS Models for Financial and Insurance Services

Next, we conducted residual analysis and the Ljung-Box test for both the ETS and ARIMA models. Based on the obtained results, the innovation residuals appear to have constant variance with no outstanding outlier for EST and ARIMA model. The innovation residuals appear to be consistent with white noise as there are only 1 slightly significant autocorrelations at lag 11. This is further backed up by the Ljung-Box test that does not reject the null hypothesis of independence, because the p values for both models are larger than 0.05, this indicates that the autocorrelation between the observed values is not significant under the given lag order and degrees of freedom.

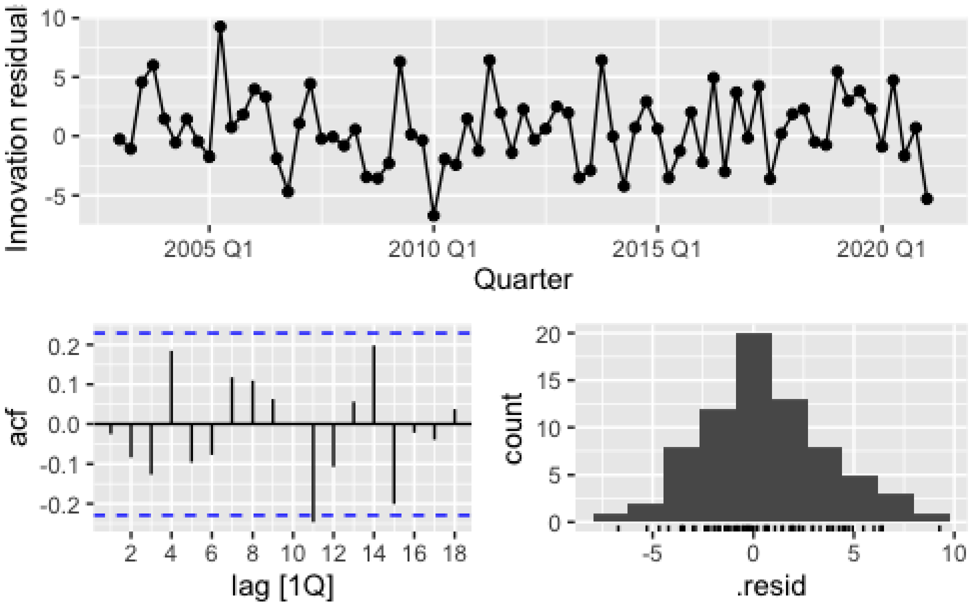

Figure 9: Residual Plot of ETS Model for Financial and Insurance Services

A tibble: 1 × 3

.model <chr>	lb_stat <dbl>	lb_pvalue <dbl>
ETS(`Financial and Insurance Services`)	8.298978	0.3069701

1 row

Table 4: Ljung-Box Test Results for ETS Model Residuals

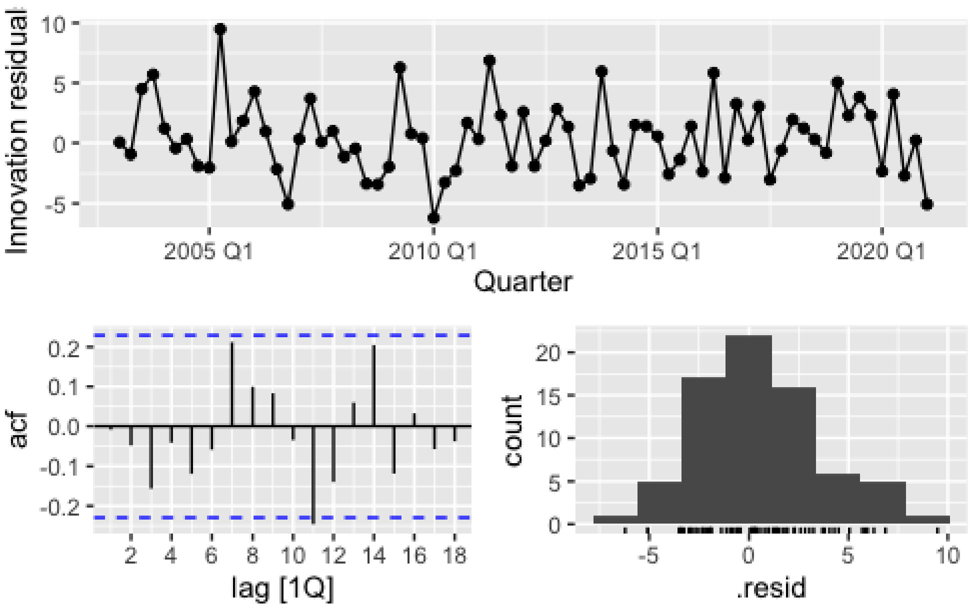

Figure 10: Residual Plot of ARIMA Model for Financial and Insurance Services

A tibble: 1 × 3

.model <chr>	lb_stat <dbl>	lb_pvalue <dbl>
stepwise	8.921881	0.2583134

1 row

Table 5: Ljung-Box Test Results for ARIMA Model Residuals

Model Comparison

This is a comparison of the forecast accuracy of the fitted ETS model and the fitted ARIMA model. The accuracy of the forecasts is evaluated using a range of different measures:

1. Root Mean Squared Error (RMSE): This measures the average magnitude of the forecast error, giving higher weight to larger errors.
2. Mean Absolute Error (MAE): This measures the average magnitude of the forecast error, regardless of direction.
3. Mean Percentage Error (MPE): This measures the average percentage error in the forecasts.
4. Mean Absolute Percentage Error (MAPE): This measures the average percentage error in the forecasts, regardless of direction.

The forecast accuracy results are as follows:

Row “arima” presents the result for best ARIMA model and row “ets” presents the result for best ETS model:

A tibble: 2 × 5

.model <chr>	RMSE <dbl>	MAE <dbl>	MPE <dbl>	MAPE <dbl>
arima	4.996727	3.976938	1.821260	5.464120
ets	5.362064	4.240695	1.017845	5.816873

2 rows

Table 6: Comparison of Performance Metrics for ARIMA and ETS Models

In general, lower values for these metrics indicate more accurate forecasts. Based on these results, the ARIMA model is performing slightly better than the ETS model across all four measures of forecast accuracy.

Forecast

ARIMA

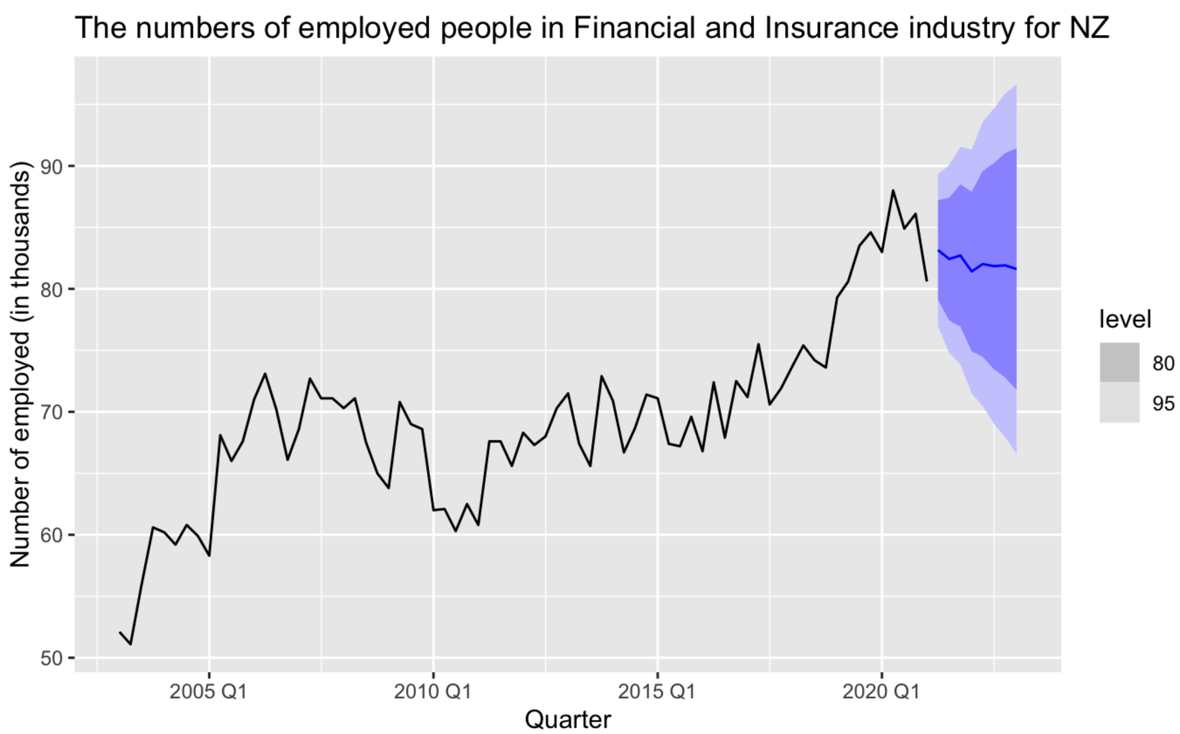

Figure 11: Forecast of Financial and Insurance Services Employees for 8 Quarters using ARIMA Model

prediction intervals:

A tsibble: 8 x 4 [1Q]

Quarter <S3: yearquarter>	.mean <dbl>	95% <S3: hilo>	80% <S3: hilo>
2021 Q2	83.16041	[76.99191, 89.32891]95	[79.12704, 87.19378]80
2021 Q3	82.43051	[74.80431, 90.05670]95	[77.44401, 87.41701]80
2021 Q4	82.71305	[73.86617, 91.55993]95	[76.92839, 88.49771]80
2022 Q1	81.41806	[71.49961, 91.33652]95	[74.93274, 87.90339]80
2022 Q2	82.02092	[70.46154, 93.58030]95	[74.46265, 89.57919]80
2022 Q3	81.84906	[69.03065, 94.66747]95	[73.46755, 90.23057]80
2022 Q4	81.91559	[67.95121, 95.87997]95	[72.78477, 91.04640]80
2023 Q1	81.61068	[66.58749, 96.63387]95	[71.78754, 91.43382]80

8 rows

Table 7: Mean and Prediction Intervals for Forecasted Financial and Insurance Services Employees using ARIMA Model

A tibble: 1 × 5

.model <chr>	RMSE <dbl>	MAE <dbl>	MPE <dbl>	MAPE <dbl>
arima	4.996727	3.976938	1.82126	5.46412

1 row

Table 8: Performance Metrics of ARIMA Model for Forecasting Financial and Insurance Services Employees

The ARIMA model's performance was evaluated using four metrics - RMSE, MAE, MPE, and MAPE. Each of these provide a different perspective on the model's forecasting ability.

1. **RMSE (Root Mean Squared Error)**: An RMSE of 4.99 in this context indicates that on average, the model's predictions deviate by approximately 5 units from the actual values.

2. **MAE (Mean Absolute Error)**: The MAE of 3.98 suggests that the ARIMA model's predictions are typically off by around 4 units.

3. **MPE (Mean Percentage Error)**: With an MPE of approximately 1.82%, it can be inferred that the model's predictions are off by about 1.82% on average, which is quite low.

4. **MAPE (Mean Absolute Percentage Error)**: A MAPE of 5.46% indicates that the model's predictions are off by about 5.46% on average, a relatively low error rate.

ETS

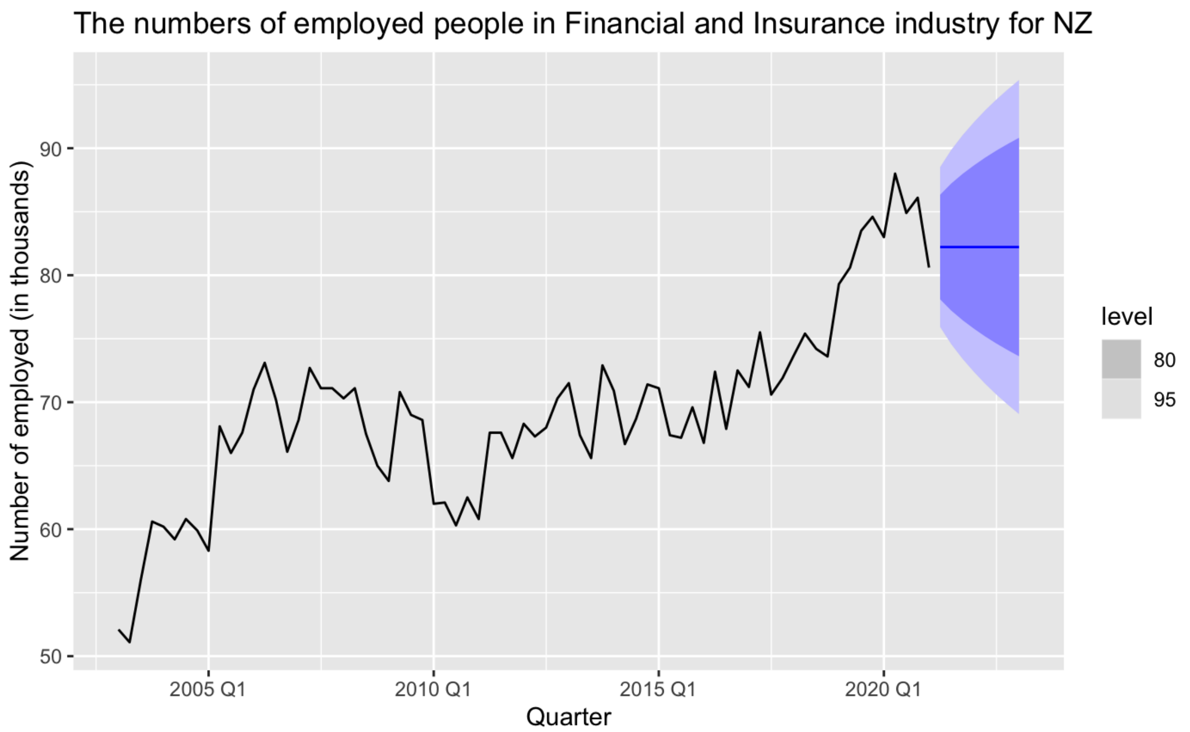

Figure 12: Forecast of Financial and Insurance Services Employees for 8 Quarters using ETS Model

Prediction intervals:

A tsibble: 8 x 4 [1Q]

Quarter <S3: yearquarter>	.mean <dbl>	95% <S3: hilo>	80% <S3: hilo>
2021 Q2	82.2187	[75.92772, 88.50969]95	[78.10525, 86.33216]80
2021 Q3	82.2187	[74.56162, 89.87579]95	[77.21201, 87.22540]80
2021 Q4	82.2187	[73.40477, 91.03264]95	[76.45559, 87.98182]80
2022 Q1	82.2187	[72.38307, 92.05434]95	[75.78753, 88.64988]80
2022 Q2	82.2187	[71.45793, 92.97948]95	[75.18261, 89.25480]80
2022 Q3	82.2187	[70.60627, 93.83114]95	[74.62574, 89.81167]80
2022 Q4	82.2187	[69.81294, 94.62447]95	[74.10701, 90.33040]80
2023 Q1	82.2187	[69.06737, 95.37004]95	[73.61951, 90.81790]80

Table 9: Mean and Prediction Intervals for Forecasted Financial and Insurance Services Employees using ETS Model

Accuracy:

A tibble: 1 × 5

.model	RMSE	MAE	MPE	MAPE
<chr>	<dbl>	<dbl>	<dbl>	<dbl>
ets	4.935332	3.898897	−0.7326943	5.736581

Table 10: Performance Metrics of ETS Model for Forecasting Financial and Insurance Services Employees

The ARIMA model's performance was evaluated using four metrics - RMSE, MAE, MPE, and MAPE. Each of these provide a different perspective on the model's forecasting ability.

1. **RMSE (Root Mean Squared Error)**: An RMSE of 4.93 in this context indicates that on average, the model's predictions deviate by approximately 5 units from the actual values.

2. **MAE (Mean Absolute Error)**: The MAE of 3.98 suggests that the ARIMA model's predictions are typically off by around 4 units.

3. **MPE (Mean Percentage Error)**: With an MPE of approximately 0.73%, it can be inferred that the model's predictions are off by about 0.73% on average, which is quite low.

4. **MAPE (Mean Absolute Percentage Error)**: A MAPE of 5.73% indicates that the model's predictions are off by about 5.73% on average, a relatively low error rate.

Overall, these metrics suggest that the ARIMA model and ETS model has performed relatively well in forecasting. The errors, both in absolute and percentage terms, are fairly low. However, the performance of the model can also depend on the specific context (like the pandamic in several quarters before forecasting) and the relative importance of accuracy in the given application. We can see from the given full data, the number of the employed people of financial and insurance industry is growing just like the before-pandamic, but the forecast producted give a slowly decreasing trend of the mean(ARIMA).

Benefits and Limitations

ARIMA

Benefits:

1. ARIMA models are highly flexible and can capture a wide range of time series patterns.
2. The model's parameters directly correspond to the components of a time series: trend, seasonality, and noise, making it interpretable.
3. ARIMA models can handle data with a trend or seasonal components through differencing.

Limitations:

1. ARIMA models assume a linear relationship, which may not always hold true in real-world data.
2. The model requires that the time series is stationary, which might need transformations and thus complicates the modelling process.
3. Choosing the appropriate AR, I, and MA terms can be quite challenging.

ETS

Benefits:

1. ETS models are intuitive, simple to understand and implement, making them a popular choice for time series forecasting.
2. These models can capture various time series patterns including trends and seasonality.
3. ETS can provide reliable forecasts for many types of business and economic data.

Limitations:

1. ETS models assume that future patterns will resemble past patterns, which may not always be the case.
2. These models may not work well when there are abrupt changes or outliers in the data.
3. ETS models assume an additive or multiplicative error structure, which may not always be appropriate.

ARIMA and ETS are great for predicting time series, but it's important to remember that there isn't a one-size-fits-all model. The optimal model often hinges on the unique traits of the time series data being analyzed. Thus, selecting a model should be underpinned by comprehensive testing and validation, employing performance metrics such as AIC and RMSE for accuracy assessment.

Conclusion

In this project, we focused on creating reliable forecasting models, specifically ARIMA and ETS, to predict future employment numbers in New Zealand's finance and insurance industry. The data showed that since 2003, there has generally been a growth in employment, apart from slight dips observed around 2008 and post-2020. When comparing the forecasting accuracy of both models, ARIMA had a slight edge over ETS. Interestingly, despite the upward trend observed in the data, the forecast hinted at a slow decrease in the future. Both models have their merits, but their effectiveness can depend on the unique attributes of the analyzed data. For future research, it might be beneficial to consider different forecasting models or delve into external factors that could influence employment trends.

Contribution

Yifan Wang

- Data transformation and processing
- Time series statistical characteristics Explaination
- Build Suitable ARIMA and ETS Models
- Conducte Ljung-Box Test on the models
- Model accuracy evaluation
- Compare and selecte the suitable model
- Formatting of the Report

Jialu Xing

- Time series statistical characteristics Explaination
- Model accuracy evaluation
- Discuss the fulfillment of model assumptions
- Compare and selecte the suitable model
- Formatting of the Report

Shuo Feng

- Group project management
- Explain the statistical characteristics of the original time series
- Performe predictions using the selected models
- Model accuracy evaluation
- Compare and selecte the suitable model
- Discussion of Model Advantages and Limitations
- Formatting of the Report

Shuyi Chen

- Data transformation and processing
- Explain the transformations and differencing used
- Compare and selecte the suitable model
- Discussion of Model Advantages and Limitations
- Formatting of the Report