Jingcheng Zhou

Forecasting New Zealand Members of Parliament’s Salaries: An Analysis of Predictive Models and Economic Forecasts Reliability

High-level Summary

Given the requirement for a high level of confidence in predicting MP salaries, I aimed to provide precise forecasts. However, based on my model results, no matter the linear regression model or the time series model that I used, I couldn't specify a particular confidence level. From this perspective, I would not consider my analysis to be successful.

But we have to know that for every type of forecasting methodology, we're using what we know and what we have right now to predict the future, but we've got to keep an eye out for uncertainties ahead. Things like new government rules, how the country's economy is doing, and what's happening globally should all be paid attention to. So, in my opinion, it's crucial to take a cautious approach when deciding future MP salaries. For instance, when making determinations, we may consider and make a list of certain assumptions. If there were any inaccuracies in previous predictions or if there were any violations of the assumptions, we can adjust them based on the actual situation.

Summary

In this project focused on predicting MP salaries for the next three years, it's essential to acknowledge that no model can perfectly capture the complexities of the real world. To address this, I adopted a multi-model approach, utilising various models to tackle the problem from different perspectives.

Firstly, the linear regression model allowed us to explore factors influencing MP salaries, considering a range of variables:

- Economic factors such as CPI and GDP.
- Labour market data including LCI, QES, HLFS.
- Alignment of MPs' salaries with trends in public service wages.
- Relationship between MPs' salaries and overall wage changes in the country.
- Incorporation of specific events like the COVID period and pay freeze periods influencing salary structures.

Subsequently, four distinct models were created to analyse these factors in different ways:

- Comprehensive Model: Examined all variables to provide a holistic view.
- Public Sector Focused Model: Zoomed in on the public sector with specific data considerations.
- All Sector Model: Considered data for all sectors, not just the public sector.
- Salary-Based Model: Focused on salary data and explored correlations with averages across different sectors.

Key Discoveries:

- Concern arises due to the limited dataset, potentially capturing more noise than actual patterns.
- Strong correlations among certain variables hinted at potential issues, including overfitting.

Despite these challenges, two variables—GDP and QES—consistently exhibited a strong connection with MPs' salaries. However, it's crucial to note that while GDP and QES play a significant role in influencing salaries, asserting that changes in these factors directly cause changes in MPs' salaries is uncertain. The relationship observed is one of correlation, not causation. Regarding the mention of QES, the strong significance observed might be attributed to its inclusion as a component in a formula during the years 2015-2018. This could contribute to its statistical significance in the models. The significance of GDP in predicting MP salaries likely arises from its role as a key indicator of economic activity and health. GDP growth is often perceived as a measure of overall economic well-being, influencing public sentiment and political considerations that may impact MPs' salaries.

As previously mentioned, strong correlations among certain variables raised concerns. In response, I implemented another model known as LASSO to mitigate the issue to some extent. Unfortunately, despite our efforts, the problem persisted.

Time series models like the ARIMA model were used to forecast MPs' future salaries. ARIMA model relied on past MP salary data to forecast the next three years salary. This model suggested that average salaries were expected to remain relatively consistent over the next three years, showing no significant changes compared to the most recent data.

However, it is crucial to delve into a critical aspect of our forecasts—the confidence intervals. These intervals provide a range of possible salaries, offering insights into where actual future salaries are likely to fall. Typically, we would associate a certain percentage chance, such as 80% or 95%, with these ranges under normal circumstances.

Now, here's the challenge: our model's predictions rely on past salary data, which, due to the pay freeze, exhibit no substantial increases or decreases, particularly in the most recent years. Moreover, working with a relatively small set of data points limits the model's ability to capture complex patterns. Consequently, the usual confidence intervals may not be as reliable as we would prefer due to underlying statistical assumptions in terms of the residuals not being fully met. Residuals are the difference between actual value and predicted value, which is deemed to be extremely important with the goal of prediction.

Another time series model — the ARIMAX model, which is a tool that incorporates external factors into MPs' salary predictions. In this model, I utilised forecasts from BERL to obtain the predictions.

The predictions heavily hinge on BERL's estimates for external variables. If these estimates are inaccurate, our salary forecasts may also be affected.

In simpler terms, while we can provide a likely range for future salaries, the exact level of confidence associated with these ranges is less certain. The forecast accuracy of BERL plays a pivotal role in predicting MP salaries. However, this doesn't diminish the value of our forecasts; it simply underscores the importance of exercising caution when considering the range of salaries we anticipate in the future.

With regard to the prediction assessment part, there are two main findings. First, the forecast accuracy decreases over time, with 1-year forecasting being the most accurate. Second, in terms of the labour indicator prediction, it indicated to some extent bias, which means there may be some overestimation or underestimation.

Abstract

The report provided a thorough examination of the potential approaches for informing the forecastings of Member of Parliament (MPs) salaries in New Zealand for the following three years. It made use of a dataset that included historical salary information, economic indicators, and labour market statistics. This information was evaluated using statistical models such as Linear Regression, LASSO Regression, ARIMA, and ARIMAX. The analysis highlighted both the benefits and drawbacks of each model.

Linear regression, a frequently used method, has encountered issues such as overfitting[1] and multicollinearity[2], which are especially noticeable in small sample sizes. Despite all that, the analysis found that GDP and QES data have similar statistical significance in relation to MPs salaries across all models. LASSO Regression, which was supposed to reduce multicollinearity, still did not manage to address this issue in this study.

ARIMA model, which is well-known for capturing temporal dynamics and to predict the future trends, encountered issues with overfitting and residual non-normality. The ARIMAX model, which included exogenous variables for a more complete analysis, depended significantly on the accuracy of the external forecasts. ARIMA and ARIMAX models were used to provide forecast values for the following three years with 80% and 95% confidence, but the non-normal distribution of residuals caused uncertainty regarding the predicted salary range's accuracy.

The report additionally investigated the accuracy of economic estimates issued by key external institutions such as the Reserve Bank, NZIER, and Treasury. To assess the precision and reliability of the forecasts, this study used methodologies such as direct and cross-sectional comparisons utilising RMSE and MAE measurements, as well as the Diebold-Mariano test. We discovered that short-term(1-year) projections were more precise than long-term forecasts. This disparity was attributed to the difficult nature of long-term projections, which was exacerbated by the inclusion of the unexpected COVID-19 era in the dataset. Despite reported changes in forecasting accuracy across time horizons, the DM test found no statistically significant differences.

[1]Overfitting: Overfitting refers to a modelling condition where a machine learning algorithm captures noise or random fluctuations in the training data, leading to poor generalisation performance on new, unseen data.

[2]Multicollinearity: Multicollinearity occurs when two or more independent variables in a regression model are highly correlated, potentially causing instability in coefficient estimates and reduced interpretability.

1. Introduction

1.1. Organisation

The Remuneration Authority is a critical body in New Zealand, entrusted with assessing key public officials' salaries. This includes determining salaries, allowances, and superannuation contributions for a wide range of positions, including the Governor-General, members of Parliament (MPs), judicial officers, and elected members of local governments. Their responsibilities also include determining annuities for former high-ranking officials and handling travel allowances for MPs families. The Authority operates under the context of the Remuneration Authority Act of 1977 and related legislations, promoting fairness and transparency in public office holder compensation (Remuneration Authority, 2021).

1.2. Goals

The project's goal was to help the Authority use data and analytical tools to make informed choices about the salary and allowances of MPs. The following were the key research questions that guided this study:

1. What information is available to assist the Authority in establishing the salaries and allowances payable in each of the three future years (2024, 2025, and 2026) at the beginning of July?

2. Where can the necessary information be obtained, and how will the Authority have access to it in three years and beyond?

3. What models/tools can the Authority use to estimate wage movement/growth over the next three years and make projections beyond that time frame?

4. What additional calculations and analyses are necessary to make the identified information useful in meeting the Authority's statutory obligations regarding MPs' salaries and allowances?

5. How to assess the prediction performance of economic indicators from external agencies?

2. Data Source

2.1. Salary Prediction

The data was obtained from various different sources and was mainly public available data.

- Historical MPs Salary Data

The dataset included historical information on the salaries of Members of Parliament (MPs). This dataset ranged from 1999 to 2023(before the 2023 General Election). This data originated directly from the Remuneration Authority.

- Average Salary in the Public Service

This dataset provided information on the base salaries of staff in the public service, which includes both the average and median salary for the public service. This dataset ranged from 2000 to 2023. The source of this data was derived from the Public Service Commission.

- Economic Data

Consumer Price Index (CPI) and Gross Domestic Product (GDP) were used to account for inflation and changes in the prices of goods and services over time, as well as the overall economic performance. The CPI and GDP data was sourced from Stats NZ. The CPI data ranged from 1914 to 2023 while GDP data ranged from 1987 to 2023.

- Labour Market Statistics

The labour market statistics information was released by Stats NZ, and provided a whole picture of the New Zealand labour market.

The Household Labour Force Survey (HLFS) Income is released annually (as at June), which produces a comprehensive range of income statistics, including the average/median weekly and hourly earnings, ranging from 1998 to 2023.

The Quarterly Employment Survey (QES) estimates the demand for labour by New Zealand businesses – the levels and changes in jobs average hourly and average weekly earnings by sector. The data is released quarterly,ranging from 1989 to 2023(third quarter).

The Labour Cost Index (LCI) measures changes in salary and wage rates for a fixed quantity and quality of labour input. It is a measure of wage inflation, reflecting changes in the rates that employers pay to have the same job done to the same standard. The data is released quarterly, ranging from 1989 to 2023(third quarter).

- Forecasts from external agency(BERL)

This dataset includes the three-year forecasts of CPI, LCI(all sector and public sector), average ordinary time hourly earnings from QES(all sector and public sector) and average hourly earnings from HLFS from BERL.

Below is a table summarising the descriptive statistics for each data source:

Table 1: *Descriptive Statistics for Salary Prediction Data Source*

Data	Time Range	Quantity	Nature	Source
Historical MPs Salary Data	1999-2023	24 years of data	Quantitative salary figures	Remuneration Authority
Average Salary in Public Service	2000-2023	23 years of data	Quantitative, average and median salaries	Public Service Commission
CPI Data	1914-2023	109 years, quarterly updates	Quantitative, economic indicators	Stats NZ
GDP Data (production measure)	1987-2023	36 years, quarterly updates	Quantitative, economic indicators	Stats NZ
Household Labour Force Survey (HLFS) Income	1998-2023	25 years of data	Quantitative, income statistics	Stats NZ
Quarterly Employment Survey (QES) (public and all sector average ordinary hourly earnings)	1989-Q3 2023	34 years, quarterly updates	Quantitative, employment and earnings by sector	Stats NZ

Labour Cost Index (LCI) (public and all sector ordinary salary rates)	1989-Q3 2023	34 years, quarterly updates	Quantitative, measures wage rates	Stats NZ
Forecasts of CPI,LCI,QES,HLFS	2024-2026	3 years, quarterly updates	Quantitative	BERL

For the "Historical MPs Salary Data" dataset, which spans from 1999 to 2023, it is important to note a couple of unique aspects. Firstly, in the year 2000, there was an amendment made to MPs' salaries. As a result, the dataset for this year includes two distinct entries, reflecting the changes before and after the amendment. Secondly, from 2017 through to 2023, the dataset indicated that there have been no changes in the MPs' salaries due to the pay freeze and COVID-19 pandemic.

The pay freeze mentioned refers to the period between July 1, 2018, and June 30, 2019, during which MPs' salaries and allowances were kept at the 2017 levels. This decision was a result of a provision to the Remuneration Authority Act 1977, freezing any salary adjustments during that time frame. The initial reason behind this freeze was the perception that MPs' salary was considered too high, prompting legislative action in 2018 to halt any increases. The freeze was implemented to address concerns about the rate of salary growth for MPs and to reevaluate the method used for calculating their remuneration. In 2015, a law change had linked MPs' pay adjustments to movements in average public sector salaries, intending to moderate increases. However, this formula resulted in a higher-than-expected pay raise. The pay freeze in 2018-2019 was a deliberate measure to reassess and potentially adjust the system for determining MPs' remuneration.

Then at the start of the COVID-19 pandemic, the authority decided to maintain MPs' salaries and allowances at the levels of the 2017 Determination.

2.2. External Agencies Prediction Performance Assessment

The data was obtained from various different sources and was all publicly available data.

- Treasury

  This data was directly obtained from the official website of the New Zealand Treasury and covered three specific updates:

  1. Budget Economic and Fiscal Update (BEFU) for the years 2016 to 2023: released in the second quarter of each year.

  2. Half Year Economic and Fiscal Update (HYEFU) for the years 2016 to 2023: released in the fourth quarter of each year.

  3. Pre-election Economic and Fiscal Update (PREFU) for 2017 and 2023: issued 4 to 6 weeks before the start of an election.

  The forecasted and actual value for the Consumer Price Index (CPI), Gross Domestic Product (GDP), and average ordinary time hourly wages in the private sector were extracted from the above three updates.

- Reserve Bank

Monetary Policy Statement(MPS) for the years 2014 to 2023 was directly obtained from the official website of the Reserve Bank. The forecasted and actual value for the Consumer Price Index (CPI), Gross Domestic Product (GDP), and Labour Cost Index(LCI) in the private sector were extracted from the MPS. To maintain consistency in our analysis, we selectively extracted the forecast data that was reported in the first quarter of each year.

- New Zealand Institute of Economic Research (NZIER)

  Quarterly Predictions and CPI Forecasts for the years 2018 to 2023 were directly obtained from the membership publications of the official website of the NZIER. The forecasted and actual value for Labour Cost Index(LCI) in all sectors and average ordinary time hourly earnings in the private sector were extracted from the Quarterly Predictions while the forecasted and actual index value for the Consumer Price Index (CPI) was extracted from CPI Forecasts. To maintain consistency in our analysis, we selectively extracted the forecast data that was reported in the first quarter of each year.

Table 2: *Summary for Agency Prediction Performance Assessment Data Source*

Data Source	Reports	Year Base	Data Extracted
Treasury	• BEFU (2016-2023, Q2), • HYEFU (2016-2023, Q4), • PREFU (2017, 2023, pre-election)	June(Since 2016)	• CPI • GDP(production measure)(Annual average % change) • Average ordinary-time hourly wages in private sector
Reserve Bank	Monetary Policy Statement (MPS) (2014-2023,Q1)	March	• CPI • GDP(production measure)(Annual % change) • LCI in private sector
New Zealand Institute of Economic Research (NZIER)	• Quarterly Predictions (2018-2023,Q1), • CPI Forecasts • (2018-2023,Q1)	March	• CPI • LCI in all sectors • Average ordinary-time hourly earnings in private sector

3. Methodology & Results

3.1. Salary Prediction

3.1.1. Data retrieval and cleaning

These data were directly downloaded from the official websites corresponding to the different datasets, and the file formats are either CSV or XLS.

The following preparation steps were undertaken to enable a robust and coherent data analysis:

- Time Span Alignment: All variables were adjusted to cover the period from 2000 to 2023.

- Quarterly to Annual Conversion: Converted quarterly data to annual format by averaging the four quarterly values for each year.

- Sector Distinction: Separated all-sector and public-sector data in the Labour Cost Index (LCI) and Quarterly Employment Survey (QES), isolating data relevant to MPs within the public sector.

- Focused Data Extraction: Concentrated on extracting average ordinary hourly earnings from QES and salary and ordinary time wage rates from LCI.
- Annual Earnings Calculation: Utilised 'Average Hourly Earnings * 40 * 52' to compute average annual earnings in the Household Labour Force Survey (HLFS).
- Primary Key Extraction: Extracted 'year' information from each dataset, serving as the primary key for dataset merging and analysis.
- Dummy variable: Used dummy variable to account for the COVID years and pay freeze years.

3.1.2. Methods

- Salary and Indicator Trend and Correlation Analysis

Before settling on the specific models to be utilised in our analysis, we did a detailed preliminary data analysis to investigate salary and indicator trends, as well as their correlations with various economic and labour market variables. It entailed graphing historical salary information across various parliamentary roles and indicator trends over multiple years to detect overall trends and patterns. A thorough correlation study was also performed to investigate the correlations between salaries and a variety of potential predictors, such as economic indices like CPI and GDP, as well as labour market data.

- Linear Regression Model

Linear regression is a basic statistical approach for forecasting a dependent variable based on one or more independent variables. It is highly regarded for its ease of use, interpretability, and effectiveness in revealing correlations between variables. In this study, linear regression was used as a key method to investigate how numerous economic and policy-related factors influence MPs' salaries on average.

The dependent variable in our analysis was the average salary of MPs. The 'average salary of MPs' refers to the mean salary calculated across various positions held by MPs. In this way it incorporated the diversity of roles and responsibilities, resulting in a more inclusive and reflective measure of the compensation structure within the parliamentary setting. To predict this, we selected a range of independent variables based on their relevance and potential impact. These included:

- Economic Indicators: CPI and GDP, as they reflected the general economic condition and inflation, which were likely to influence government salary scales.
- Labour Market Data: Data from the QES and LCI provided insights into employment earnings, hours worked, and changes in wage rates within the economy.
- Public Service Wages: The average and median wages in public service were included to assess the alignment of parliamentary salaries with broader public service wage trends.
- General Wage Trends: The average annual wage calculated from the HLFS and QES helped in understanding how overall wage changes in the country might correlate with parliamentary salaries.
- Policy Impact Variables: Two dummy variables representing the COVID period and pay freeze periods were included to account for specific policy impacts on salary structures.

Four distinct linear regression models were formulated to analyse the relationship between the chosen independent variables and the average salary of MPs:

- Comprehensive Model: This model included all selected variables (CPI, GDP, QES, LCI, public service wages, HLFS data, and policy impact dummy variables) to provide an all-encompassing analysis.

- Public Sector Focused Model: Focusing on the public sector, this model included public sector-specific LCI and QES data, along with CPI, GDP, and the policy impact variables.

- All Sector Model: This model took a broader approach by including LCI and QES data for all sectors, as well as CPI, GDP, and the policy impact variables.

- Salary-Based Model: Concentrating specifically on salary data, this model included average public service wages, HLFS average annual wage data, QES average annual wage data and the policy impact variables.

We began by using the four distinct models to identify the variables that influence MPs' salaries based on historical data. Subsequently, we constructed the model using these selected variables, extracting their coefficients. With these coefficients in hand, we then utilized the predicted values of these variables for the next three years to estimate the future salaries of MPs.

We employed several performance metrics such as R-squared, Adjusted R-squared and p-values for each predictor. These metrics helped assess each model's explanatory power and the significance of individual predictors.

Specifically, R-squared is a measure of the proportion of the variance in the dependent variable that is explained by the independent variables in a regression model. It ranges from 0 to 1, where 0 indicates that the model does not explain any variance, and 1 indicates that the model explains all the variance. A higher R-squared value suggests that a larger proportion of the variability in the dependent variable is accounted for by the model. However, it does not indicate the goodness of fit on its own and can be misleading when adding more predictors.

Adjusted R-squared adjusts the R-squared value for the number of predictors in the model. It penalises the addition of irrelevant predictors that do not improve the model significantly. Unlike R-squared, it can decrease if adding a new predictor does not improve the model sufficiently. A higher adjusted R-squared is generally preferred as it indicates that the added predictors contribute meaningfully to explaining the variability in the dependent variable.

In the context of a regression model, p-values assess the statistical significance of each predictor's coefficient. A small p-value suggests that the predictor is likely to be a meaningful addition to the model because changes in the predictor's value are related to changes in the response variable. If the p-value for a predictor is less than the chosen significance level (e.g., 0.05), it is often considered statistically significant. However, it's important to interpret p-values alongside other metrics, as statistical significance does not necessarily imply practical significance.

- LASSO Regression Model

LASSO is a regression analysis method that improves the predictability and interpretability of the prior statistical model through variable selection and regularisation (James et al., 2021). Regularisation, in particular, can aid in the prevention of overfitting and multicollinearity, particularly in cases with small sample sizes or when the number of predictors is large in comparison to the number of observations. Furthermore, feature selection can discover the most important variables from a vast number of predictors.

- ARIMA Model

The Autoregressive Integrated Moving Average (ARIMA) model is a statistical method for forecasting time data. Its major parameters are p, d, and q. To be more specific, p represents the number of lag observations included in the model, d represents the number of times the raw observations are differenced, where the current observation is subtracted from the previous one, resulting in a stationary time series, and q represents the size of the moving average window, indicating the number of lagged forecast errors that should be included in the model. The values of p,d, and q can be determined using autocorrelation and partial autocorrelation plots (Hayes, 2023).

The forecast accuracy of the model was measured using statistical methods. The Mean Error (ME), which indicates the least amount of bias, is calculated as the sum of forecast errors divided by the number of observations:

$$\text{ME} = \frac{1}{n}\sum_{i=1}^{n}(Y_i - \widehat{Y}_i)$$

Where:

- n is the number of observations
- Y_i is the actual observed value for each observation i
- $\widehat{Y}_i$ is the predicted value for each observation i

A ME close to zero indicates minimal bias, where positive values suggest overestimation and negative values indicate underestimation.

The Root Mean Squared Error (RMSE), revealing the magnitude of forecast errors, is calculated as the square root of the mean of squared differences between the observed and predicted values:

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(Y_i - \widehat{Y}_i)^2}$$

Where:

- n is the number of observations
- Y_i is the actual observed value for each observation i
- $\widehat{Y}_i$ is the predicted value for each observation i

A lower RMSE signifies better accuracy, with values close to zero indicating minimal prediction errors.

The Mean Absolute Error (MAE), representing a straightforward average of absolute prediction errors, is calculated as:

$$\text{MAE} = \frac{1}{n}\sum_{i=1}^{n}\left|Y_i - \widehat{Y}_i\right|$$

Where:

- n is the number of observations
- Y_i is the actual observed value for each observation i
- $\widehat{Y}_i$ is the predicted value for each observation i

Like RMSE, lower MAE values indicate more accurate predictions.

The Mean Percentage Error (MPE), providing relative accuracy, is computed as the mean of percentage errors:

$$\text{MPE} = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{Y_i - \widehat{Y}_i}{Y_i}\right) \times 100$$

Where:

- n is the number of observations
- Y_i is the actual observed value for each observation i
- $\widehat{Y}_i$ is the predicted value for each observation i

A MPE close to zero signifies accurate predictions, with positive values indicating overestimation and negative values indicating underestimation.

The Mean Absolute Percentage Error (MAPE), another measure of relative accuracy, is calculated as the mean of absolute percentage errors:

$$\text{MAPE} = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{Y_i - \widehat{Y}_i}{Y_i}\right| \times 100$$

Where:

- n is the number of observations
- Y_i is the actual observed value for each observation i
- $\widehat{Y}_i$ is the predicted value for each observation i

Lower MAPE values indicate better accuracy, with zero representing a perfect prediction.

The Mean Absolute Scaled Error (MASE), evaluating the model against a simple baseline, is computed as the MAE of the forecast divided by the MAE of a naive baseline:

$$\text{MASE} = \frac{MAE}{\frac{1}{n-1}\sum_{i=2}^{n}\left|Y_i - Y_{i-1}\right|}$$

Where:

- n is the number of observations
- Y_i is the actual observed value for each observation i

- $\widehat{Y}_i$ is the predicted value for each observation i

A MASE value less than 1 suggests the model outperforms the baseline, while values greater than 1 indicate the baseline is a more suitable predictor.

The Autocorrelation of Residuals at Lag 1 (ACF1) assesses residual correlations. If ACF1 is close to zero, it indicates that the model's residuals do not exhibit significant correlation.

These metrics collectively provided a comprehensive evaluation of the model's forecasting performance across various dimensions.

- ARIMAX Model

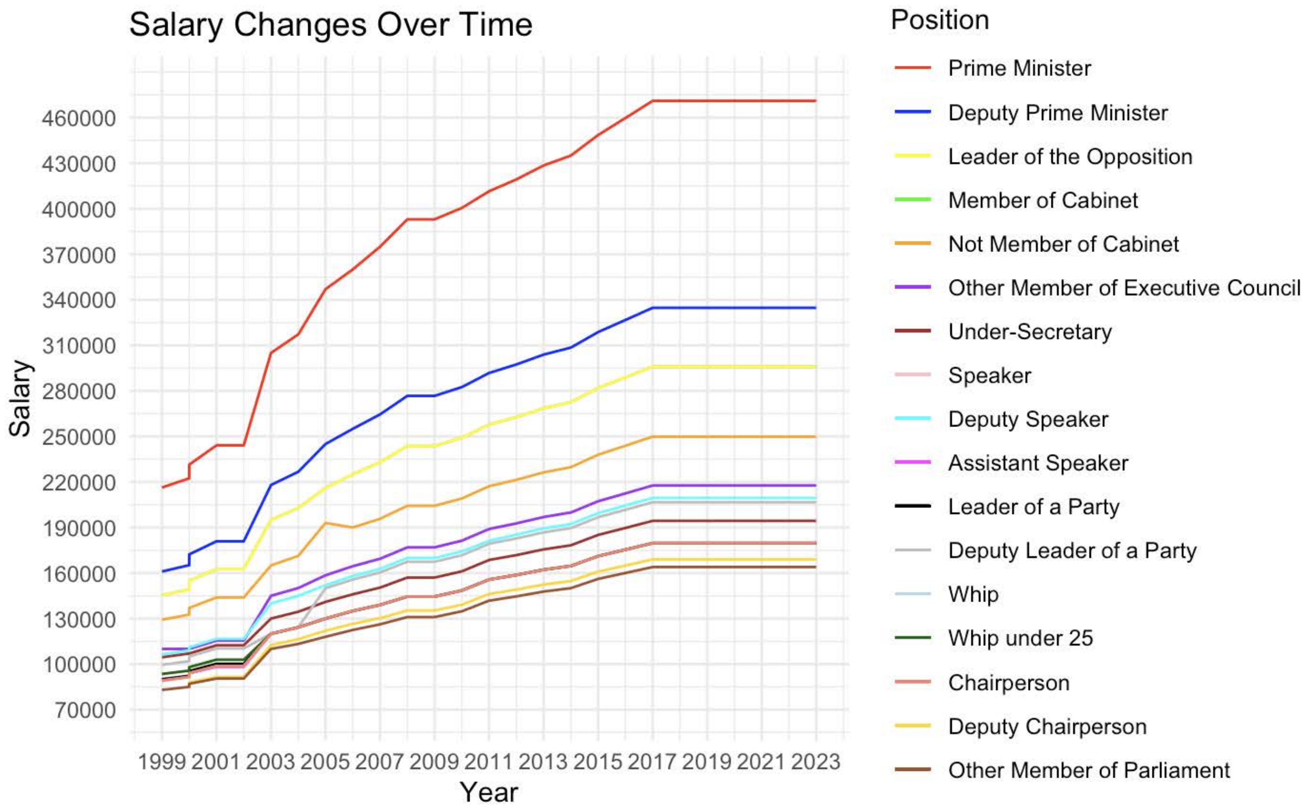
The AutoRegressive Integrated Moving Average with exogenous variables (ARIMAX) model extends the capabilities of the classic ARIMA model by integrating extra variables that may affect the projected variable (Smarten)—in our case, MP salary. Exogenous variables could include indicators such as the CPI, LCI, and QES. The forecast accuracy measurements are identical to those of the ARIMA model.

3.1.3 Results

- Salary and Indicator Trend and Correlation Analysis

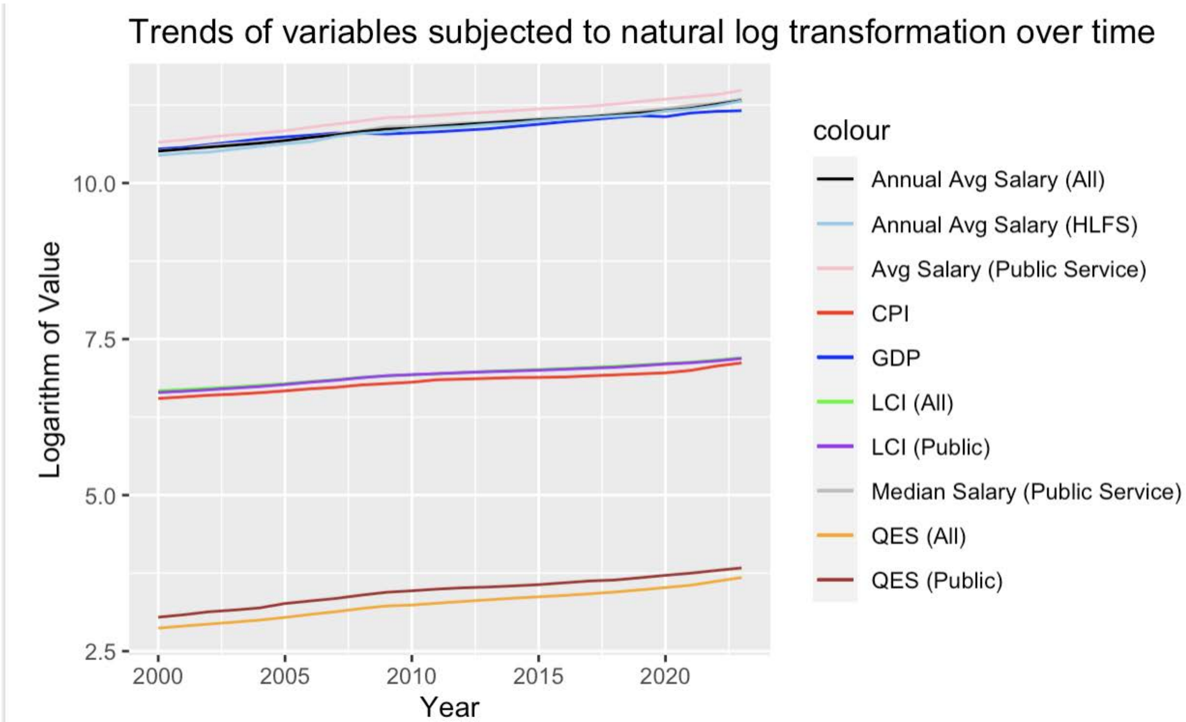
As shown in Figure 1, the salary data for various parliamentary positions demonstrated a steady upward trend over time, indicating that salary calculations for these roles were most likely based on a consistent set of standards or criteria. Notably, the statistics indicated a pay freeze beginning in 2017, after which there were no major adjustments in salary levels.

Figure 1 *: Salary Trends Over Time*



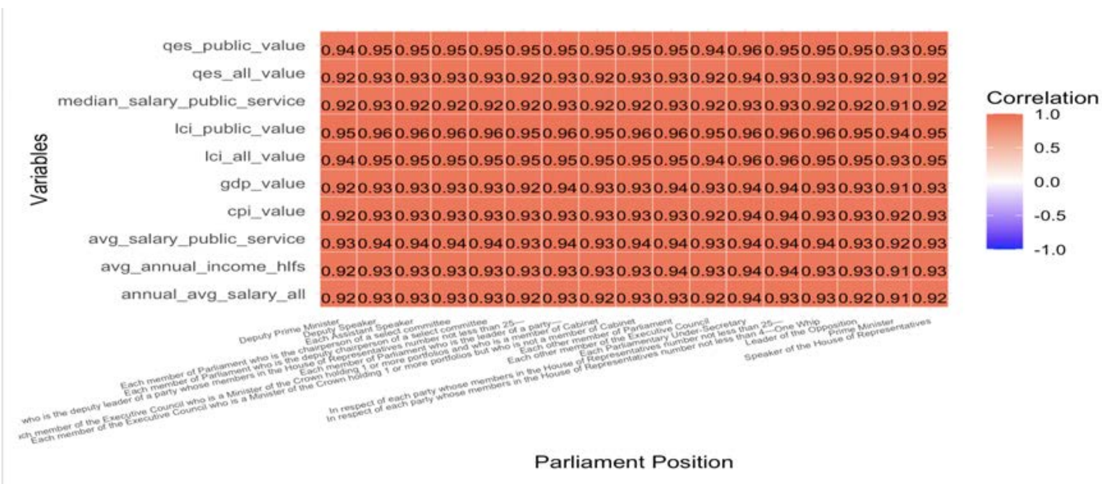
We discovered uniformity in the long-term trends of many variables when we analysed their extended trends (see Figure 2). This consistency revealed that the economic factors determining parliamentarian pay have followed similar growth patterns, indicating that they reflected overarching economic trends rather than isolated incidents. When plotting the trends of multiple variables, a logarithmic transformation was used to account for the discrepancies in their scales.

Figure 2 *: Trends of Variables Subjected to Natural Log Transformation Over Time*



Furthermore, the correlation coefficients, which were predominantly greater than 0.9, indicated a strong positive correlation between the salaries of various parliamentary positions and the indicators(see Figure 3). In simpler terms, when one of these indicators increases, the salaries across various parliamentary roles tend to increase together.

Figure 3 : *Correlation Coefficients Between Different Parliamentary Positions and Various Indicators*



- Linear Regression Model

When the results of the four linear regression models were examined, evidence suggested that overfitting might have occurred. This conclusion was reached after making the following observations:

- R-squared Values: The R-squared values, which were mostly above 0.9 and in some cases close to 1, indicate that the model is doing a good job explaining a significant portion of the variability in the data. In simpler terms, the model is capturing a large part of the patterns and trends present in the data. However, it's important to note that, especially when dealing with a small sample size, extremely high R-squared values can be misleading. In this context, the concern is that the model might not only be capturing the meaningful patterns in the data but also picking up on random fluctuations or noise. So, while the R-squared suggests a strong explanatory power, there is a need for caution due to the limited amount of data, as the model might be overly sensitive to individual data points.

- Assumptions of Linear Regression: The residuals, which represent the differences between the actual and predicted values, were not normally distributed and showed heteroscedasticity, according to diagnostic testing such as the histogram analysis of residuals, residuals vs. fitted values plot, and scale-location plot. Heteroscedasticity suggests that the variability in the residuals changes across different levels of the predicted values. In simpler terms, the spread of prediction errors is not consistent, and it may increase or decrease as the predicted values change.These situations violated linear regression assumptions, which could lead to erroneous coefficient estimates and wider confidence ranges.

- Multicollinearity: An initial examination of the coefficients of correlation between variables and salaries revealed strong positive correlations. Subsequent Variance Inflation Factor (VIF) calculations verified the models' considerable multicollinearity. Except for the two dummy variables, all variables had VIF values more than the generally accepted threshold of 10, suggesting a significant degree of multicollinearity. With only 25 observations, such extreme multicollinearity, together with inflated R-squared values, clearly suggested overfitting.

However, GDP and QES data revealed constant statistical significance(see Table 3), indicating a strong correlation with MP wages across all models examined. The strength and consistency of the correlation suggested that these variables might be used to forecast the variation in MP salaries. Regarding the mention of QES, the strong significance observed might be attributed to its inclusion as a component in a formula during the years 2015-2018. This could contribute to its statistical significance in the models.

It was crucial to highlight that the appearance of a substantial correlation does not imply the presence of a cause-and-effect relationship. Despite the recurring importance of GDP and QES in our models, we were unable to conclude that changes in these indicators are causal factors for changes in MP salaries.

Table 3*: Summary of Linear Regression Model Results*

Model	R-squared	Adjusted R-squared	Significant Predictors and Their P-values
Comprehensive Model	0.9942	0.9894	• GDP: 0.00137 (**) • QES_all_sector: 0.05517 (.)
Public Sector Focused Model	0.9726	0.9634	• CPI: 0.00187 (**) • GDP: 0.03774 (*) • Pay_freeze: 0.01475 (*)
All Sector Model	0.9909	0.9879	• CPI: 0.075637 (.)

			• GDP: 0.000192(***) • LCI_all_sector:0.00000000158(***) • QES_all_sector:0.00000029353(***)
Salary-Based Model	0.9337	0.9163	• Avg_salary_public_service:0.00328(**) • Avg_salary_all:0.01267(*)

Note: Significance Level: '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1

The significance level indicates the confidence we have in the statistical significance of a predictor in a regression model. For example, in the Comprehensive Model, the p-value for GDP is 0.00137 (**) which is less than the significance level of 0.01 represented by '**'. This implies that we can confidently conclude that GDP significantly influences the changes in MPs salary.

- LASSO Regression Model

The LASSO regression model indicated that the problem of multicollinearity persisted. The VIF values for the non-zero coefficients, namely GDP, LCI, and Median_salary_public_service, were significantly higher than the generally used threshold of 10, indicating a high degree of multicollinearity. As a result, the factors were not only connected to the MP salary but also to each other, potentially inflating the regression coefficients.

- ARIMA Model

One of the ARIMA model's fundamental assumptions is that the data be stationary, meaning that its statistical properties such as mean and variance remain constant over time with no existence of trends. The original data had non-stationary features, according to preliminary examination utilising the Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF) plots. In order to address this, the data was first differenced by subtracting each data point from the one before it, which resulted in stationarity.

The ACF plot showed a substantial spike at the initial lag and no more significant spikes after differencing, leading to the choice of q=1 for the moving average component of the model. As a result, two ARIMA models were created: ARIMA(0,1,1) and ARIMA(0,1,0).

A comparison of the two models found significant changes in performance measures across the training and test sets (see Table 4). Such metric fluctuations indicate potential overfitting difficulties, in which the model may not generalise well to new data.

The residuals were validated further to confirm the model's assumptions. The ACF plot confirmed that autocorrelation in the residuals was not an issue, indicating that the temporal structure had been sufficiently represented. The residuals, however, were discovered to be non-normally distributed in the histogram(see Appendices Figure 13 & 14). The residuals' non-normality might have an impact on the confidence intervals and hypothesis tests related to the model's predictions.

A log transformation was performed to the data to minimise the difficulties of overfitting and non-normally distributed residuals in our ARIMA models. Log transformation is a frequent approach used to stabilise the variance of time series data, which can assist make the data more compatible with stationarity and homoscedasticity assumptions. This transformation is very useful when dealing with skewed distributions since it can compress large values while expanding tiny ones, normalising the residuals (Feng et al., 2014).

Two new models were fitted after the log transformation: Log-Transformed ARIMA(0,1,1) and Log-Transformed ARIMA(0,1,0). The performance parameters of these influenced models improved noticeably, with differences between the training and test sets significantly reduced, implying better generalisation to unknown data (see

Table 4). However, despite the log transformation, the issue of residual non-normality persisted, as indicated by Figure 16 and 17 in the Appendices.

Table 4: *Summary of ARIMA Model Results*

Model	Dataset	ME	RMSE	MAE	MPE	MAPE	MASE	ACF1	AIC
ARIMA(0,1,1)	Train	4832.413	7578.599	5152.578	2.9593	3.1733	0.8324	-0.4551	336.39
	Test	8396.74	8580.834	8396.74	3.77953	3.7795	1.3565	NA	
ARIMA(0,1,0)	Train	5832.601	7967.425	5832.601	3.5692	3.5692	0.9422	-0.2153	335.84
	Test	9951.419	10107.234	9951.419	4.4805	4.4805	1.6076	NA	
Log-Transformed ARIMA(0,1,1)	Train	0.0322	0.053	0.0339	0.2688	0.2837	0.8625	-0.3999	-43.63
	Test	0.0396	0.0404	0.0396	0.3219	0.3219	1.0061	NA	
Log-Transformed ARIMA(0,1,0)	Train	0.0378	0.0549	0.0378	0.3151	0.3151	0.9586	-0.1875	-44.55
	Test	0.0459	0.0466	0.0459	0.3726	0.3726	1.1646	NA	

The forecasting model was chosen to be the Log-Transformed ARIMA(0,1,1). Normally, a lower value of the Akaike Information Criterion (AIC) is preferred when selecting a forecasting model. The AIC is a measure that considers both the goodness of fit and the complexity of the model. In essence, it penalises models for being too complex and rewards them for explaining the data well.

Despite the fact that the AIC for this model was -43.63, which was somewhat higher than the AIC for the Log-Transformed ARIMA(0,1,0), which was -44.55, other factors were regarded as more important in our selection. Primarily, the Log-Transformed ARIMA(0,1,1) model had a MASE of 1.006, which was slightly lower than the Log-Transformed ARIMA(0,1,0) model's MASE of 1.1646. In terms of estimating average MP wages, the MASE is a critical measure. Although a MASE score of one or below indicates that a model outperforms a naive baseline, a MASE slightly above one can still be appropriate. As a result, the Log-Transformed ARIMA(0,1,1) model was chosen as the preferable option because of its balance of performance indicators.

The forecasting results from the Log-Transformed ARIMA(0,1,1) model for the years 2025 through 2027 show no change in the predicted average MP pay, remaining steady at the level reported in our dataset (see Table 5). This result was mostly caused by of two factors:

- Lack of Trend or Seasonality: Since the pay freeze was implemented in 2017, the dataset has revealed no identifiable trends in MP salary. As a result, the ARIMA model, which forecasts future values using historical patterns, projected this lack of movement forward, yielding a static projection.

- Sample Size Limitations: The dataset only had 25 observations, which may not be enough for the model to recognise and learn from more complicated patterns. This limitation may prohibit the model from producing dynamic projections, causing it to project the most recent known salary value into the future.

The accompanying forecast table, described in Table 5, included point projections for the expected average MP salary, as well as 80% and 95% confidence ranges. While these intervals were designed to quantify forecast uncertainty, their dependability was undermined by the model's non-normal residual distribution (Cohen et al., 2013). Non-normality could indicate that the actual forecast error distribution was more spread out, with larger tails or skewness, than a normal distribution would imply. As a result, the true confidence intervals may be wider, signifying a greater likelihood that future values would fall outside of the predicted ranges.

Table 5 *: Forecast Average MPSalary for the Next Three-year and the Confidence Interval Using ARIMA*

Year	Forecast_Average_MPSalary	Lower_80_Confidence_Interval	Upper_80_Confidence_Interval	Lower_95_Confidence_Interval	Upper_95_Confidence_Interval
2024	222,475.2	210,034.2	235,653.0	203,732.6	242,942.1
2025	222,475.2	203,377.1	243,366.7	193,940.1	255,208.8
2026	222,475.2	198,676.6	249,124.5	187,126.9	264,500.7

- ARIMAX Model

The system chose the ARIMAX model coefficients p, d, and q automatically after identifying an ARIMA(0,0,0) model enhanced with exogenous variables. This specification suggested that there were no autoregressive or moving average components, implying that the model relied solely on exogenous variables to explain the variation in MPs salaries.

While this method allowed for the incorporation of crucial external variables, our analysis revealed that it did not account for the temporal dynamics that may be present in salary data. Because the model relied solely on exogenous factors, its ability to perform was dependent on the accuracy of the forecasts for these external predictors provided by our economic forecasting partner, BERL.

Furthermore, the disparities in training and test set metrics were notable (see Table 6), raising questions about the model's prediction accuracy. Although the log transformation addresses the issue of error disparity to some extent, the residuals' persistent non-normality calls the accuracy of the confidence intervals into question (see Appendices Figure 18). As a result, the anticipated confidence intervals for average MP salaries from 2025 to 2027 (see Table 7) should be considered with care.

The prediction intervals, particularly the broader 95% intervals, indicated a significant range for potential salary values, but the non-normality of residuals suggested that actual salaries are more likely to fall outside these intervals than would be expected under a normal distribution. The reliance on the predicted outcomes of the external variables added to the uncertainty. If BERL's estimates were incorrect, it would have a direct impact on the accuracy of our salary forecasts.

Table 6*: Summary of ARIMAX Model Results*

Model	Dataset	ME	RMSE	MAE	MPE	MAPE	MASE	ACF1
ARIMAX(0,0,0)	Train	-10.7674	5635.815	4467.493	-0.2135	2.9699	0.7217	0.1397
	Test	-25615.697	27324.816	25615.698	-11.5309	11.5309	4.1382	NA
Log-Transformed ARIMAX(0,0,0)	Train	0.0000017	0.0389	0.0289	-0.001	0.2425	0.7346	0.0506
	Test	-0.1031	0.111	0.1030	-0.8371	0.8371	2.6164	NA

Table 7 *: Forecast average MPSalary for the Next Three-year and the Confidence Interval Using ARIMAX*

Year	Forecast_Average_MPSalary	Lower_80_Confidence_Interval	Upper_80_Confidence_Interval	Lower_95_Confidence_Interval	Upper_95_Confidence_Interval

2024	231,293.6	219,714.1	243,483.4	213,820.8	250,194.2
2025	242,510.4	227,668.8	258,319.6	220,183.4	267,101.4
2026	255,575.0	238,676.5	273,669.9	230,188.0	283,761.7

Our evaluation of BERL's historical predicting accuracy faced numerous limits that prevented it from being applied to this report:

- Archival Availability: BERL's online report collection runs from Autumn 2019 to Winter 2023. A few years ago, a system change resulted in the non-retention of past editions of the Berl Economic View (BEV) prior to Autumn 2019. As a result, the available sample size for testing prediction performance was insufficient to allow for a meaningful study.

- Forecast Table Simplification: BERL simplified the contents of their web forecast tables in 2021, standardising all forecasts to match the June year, rather than the prior March year. Concurrently, the forecast tables stopped including forecasts for Wages (Average hourly wages), which is a focus of our forecast interest. The recent format's lack of these precise salary projections hindered our ability to evaluate previous forecasting effectiveness for this economic indicator.

- Metric Adjustment for GDP: The way by which GDP (production measure) is anticipated has changed. Prior to 2021, the forecast tables presented an annual percent change in GDP as an average. This statistic was changed to an annual percent change after 2021. The metric's unreliability hindered direct comparison of GDP projections across revised time frames.

In summary, because the predictive data employed in our model was obtained from BERL, our inability to assess their historical forecast accuracy placed doubt on the credibility of our model's projections.

3.2. External Agencies Prediction Performance Assessment

3.2.1. Data retrieval and cleaning

These reports were directly downloaded from the respective official websites, and the required data were manually extracted from the reports due to the nature of the data presentation and saved in the XLSX file format for subsequent analysis.

The Remuneration Authority, which is in charge of setting salaries for Members of Parliament, established salary ranges for a future three-year term. In accordance with this, a three-year review term was employed when examining the accuracy of economic indicator estimates made by external agencies such as the Treasury, Reserve Bank, or NZIER. For example, forecasts for the Consumer Price Index (CPI) for 2015, 2016, and 2017 were pulled from the 2014 report to compare the forecast to actual economic outcomes. This methodology ensured that the evaluation of predictive accuracy was directly comparable to the time frame for which the salaries were set.

Several considered steps were taken to ensure consistency and accuracy in the forecast evaluation. Firstly, since 2016, the Treasury adjusted the year base from March to June. Consequently, data extraction has been limited to the period from 2016 onwards. This was done to maintain consistency in the comparison of subsequent years' data.

Second, type of 'actual outturn' data, referring to the actual values of the variable being forecasted, should be decided to be used for accuracy assessment. In contrast to the CPI, which is rarely revised, original GDP estimates may be subject to several revisions due to new information, methodological modifications, new weights, or rebasing. Because forecasts at any given time were based on the information available at the time (including the methodology) and were frequently evaluated against the first available data outturn, the initial outturn was used as the criterion for assessing forecast accuracy in this study.

Third, because the evaluation of different institutions' forecasting accuracy involved comparing the predicted values for the next three years to their actual figures, the data frames used for analysis from different organisations were specific to the periods they cover: Reserve Bank (2014-2019), Treasury (PREFU 2017, BEFU, HYEFU 2016-2019), and NZIER (2018-2019).

3.2.2. Methods

- Direct comparison

The first method used was a direct comparison of predicted and actual economic outcomes. This method involved a simple yet effective assessment of the disparities between predicted and actual values for key economic indicators.

The Root Mean Squared Error (RMSE) and Mean Absolute Error (MAE) were used to quantify forecast accuracy (The Data Scientist,2022). The square root of the average of the squared differences between predicted and actual values is measured by RMSE. The root mean square error (RMSE) is a standard statistic used to quantify the difference between predicted and observed values, demonstrating how near the forecasted data points are to the actual data points. A smaller RMSE number indicates that the discrepancy between the anticipated and actual values is minimal, meaning that the model is more accurate. However, because it squares the errors before averaging them, RMSE is particularly sensitive to outliers, resulting in a greater penalty for big mistakes. As a result, an increased RMSE may indicate the presence of significant inconsistencies in some predictions. The formula for RMSE is:

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(Y_i - \widehat{Y}_i)^2}$$

Where:

- n is the number of observations
- Y_i is the actual observed value for each observation i
- $\widehat{Y}_i$ is the predicted value for each observation i

The Mean Absolute Error (MAE) was used in conjunction with RMSE to calculate the average magnitude of errors in a series of predictions without taking into account their direction. MAE is the average of the absolute discrepancies between predicted and actual values. In contrast to RMSE, MAE treats all mistakes linearly, which means that all deviations, big or small, are weighted equally in the calculation. A lower MAE score indicates greater forecast accuracy. MAE gives a more balanced depiction of error magnitudes over the whole dataset since it is less sensitive to outliers than RMSE. The formula for MAE is:

$$\text{MAE} = \frac{1}{n}\sum_{i=1}^{n}\left|Y_i - \widehat{Y}_i\right|$$

Where:

- n is the number of observations

- Y_i is the actual observed value for each observation i
- $\widehat{Y}_i$ is the predicted value for each observation i

The Mean Error was generated in addition to the RMSE and MAE to provide a simple average of the predicting mistakes. This statistic assists in determining the presence of systematic bias in forecasts, revealing whether the model tends to overestimate or underestimate actual values. A positive mean error implies that the model tends to overpredict, whereas a negative mean error shows that the model tends to underpredict. Mean Error can be used to discover bias in a forecasting model. Mean Error is calculated as:

Mean Error = $\frac{1}{n}\sum_{i=1}^{n}(Y_i - \widehat{Y}_i)$

Where:

- n is the number of observations
- Y_i is the actual observed value for each observation i
- $\widehat{Y}_i$ is the predicted value for each observation i

Confidence Intervals were built around the Mean Error to test its accuracy. The 95% CIs are a statistical range within which the true mean error is predicted to lie with 95% certainty, assuming the error distribution is normal. These intervals are very important for illustrating forecast precision and showing potential error fluctuation. The Confidence Intervals are calculated as:

Upper CI = Mean Error + (1.96× $\frac{Standard\ Deviation}{\sqrt{n}}$)

Lower CI = Mean Error - (1.96× $\frac{Standard\ Deviation}{\sqrt{n}}$)

Where:

- n is the sample size
- Standard deviation is the square root of the average of the squared differences between the forecasted values and the actual values

If the confidence interval range is narrow, the forecast is more precise, with a more consistent error pattern. A larger interval, on the other hand, indicates greater fluctuation in forecasting accuracy. The occurrence of zero inside this interval may imply that there is no major bias in the forecasts, whereas its absence may indicate that there is a systematic bias that needs to be addressed.

- Cross-sectional comparison

This method allowed for a comparison of an agency's forecasting performance to that of other agencies forecasting the same economic indicators. One of the most important aspects of this comparative analysis was confirming that each agency uses consistent measuring methods for the same economic indicators and bases its forecasts on data from the same time periods.

To keep a consistent methodology, this comparison just examined the Reserve Bank and NZIER's CPI forecasts for the next three years, as published in their 2018-2019 reports. Furthermore, because consensus CPI estimates were given in an index format, the present study included a comparison of NZIER's forecasts with consensus forecasts. The Root Mean Square Error (RMSE) and Mean Absolute Error (MAE) were the major performance metrics.

- Diebold-Mariano (DM) test

The Diebold-Mariano (DM) test was a statistical method used to determine whether observed differences in forecasting accuracy between two models were statistically significant or simply by chance. This test was notable for not depending on assumptions about forecast error distribution and for accounting for temporal autocorrelations and probable interactions between the two forecast series being compared (Gilleland and Roux, 2015).

We used the DM test in our analysis to examine the accuracy of several forecasting models implemented by different organisations. The DM test statistic was critical in determining the statistical significance of the differences in forecast performance observed. The DM test's p-value was critical in understanding the results. A p-value expresses the likelihood that the observed difference in forecasting accuracy occurred by chance. A lower p-value suggests that the performance difference between the two models is statistically significant. Many analyses use a p-value threshold (such as 0.05) to determine if a difference is statistically significant. If the p-value was less than the selected threshold, it indicated that one forecasting approach was statistically superior to the other. A high p-value, on the other hand, indicated that there was no significant difference in the forecasting accuracy of the two models.

3.2.3. Results

Direct comparison

- *Reserve Bank*

For analytical purposes, the Reserve Bank's monetary policy statement projections have been divided into three distinct periods: the overall period from 2014 to 2019, the pre-COVID period from 2014 to 2016, and the COVID-impacted period from 2017 to 2019.

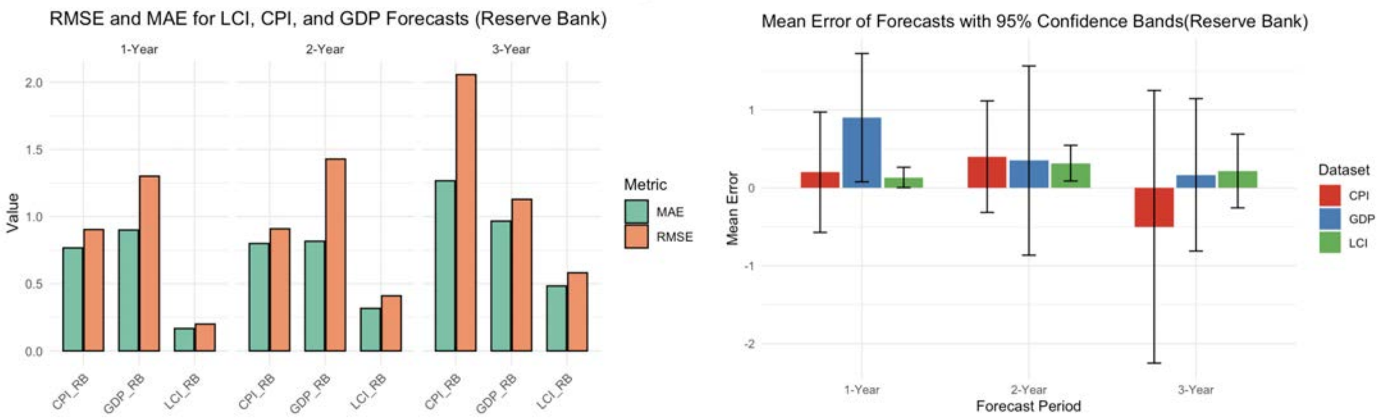
1) Overall Forecast Accuracy (2014-2019)

Figure 4 gave various insights into the total forecast accuracy between 2014 and 2019. For starters, there was a noticeable tendency in which forecast accuracy decreases as the forecast horizon extends. This pattern corresponded to conventional forecasting issues, as evidenced by the lower precision of the 3-year forecasts compared to the 1-2 year projections. Conventional forecasting issues typically refer to challenges associated with predicting outcomes over longer time horizons, implying that the forecasting model may struggle to capture and accurately project trends, uncertainties, or external factors that become more prevalent over longer time periods. (FasterCapital, 2023)

The lower RMSE and MAE values indicate that LCI projections were more accurate than CPI and GDP estimates. Furthermore, the close relationship between RMSE and MAE values for LCI demonstrated a lower incidence of major forecasting mistakes, reflecting consistency in prediction quality without notable outliers.

Another significant finding was the Reserve Bank's systematic overestimation of labour costs, as evidenced by the positive confidence interval bars for the LCI in the 1-2 year projections. This frequent over-prediction indicated that the Reserve Bank's LCI forecasting process had an optimistic bias.

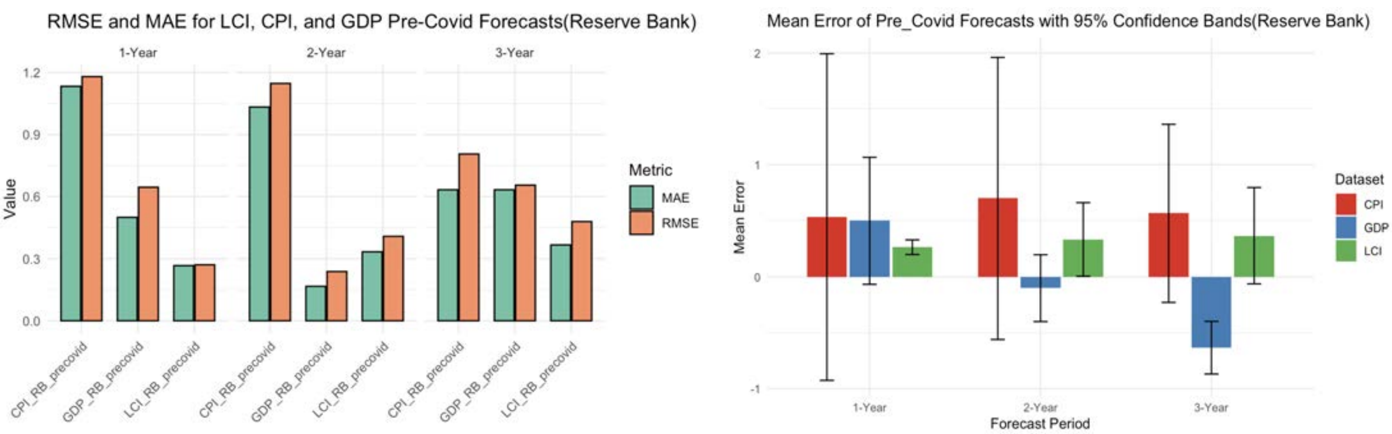
Figure 4*: RMSE,MAE, Mean Error for LCI, CPI and GDP Forecasts From 2014 to 2019 by Reserve Bank*



2) Pre-COVID Forecast Accuracy (2014-2016)

Higher forecast accuracy was observed in the months preceding the COVID-19 pandemic. This was obvious from the decreased error values over these years (see Figure 5) compared to the following pandemic-affected period (see Figure 6). The little discrepancy between MAE and RMSE revealed that there were no greater errors in the forecasts, lending credence to the idea of a more stable and predictable economic climate prior to COVID-19. According to the data, the Reserve Bank's forecasting models were more aligned with actual economic outcomes over this time period, demonstrating the relative economic stability.

Figure 5*: RMSE,MAE, Mean Error for LCI, CPI and GDP Forecasts From 2014 to 2016 by Reserve Bank*

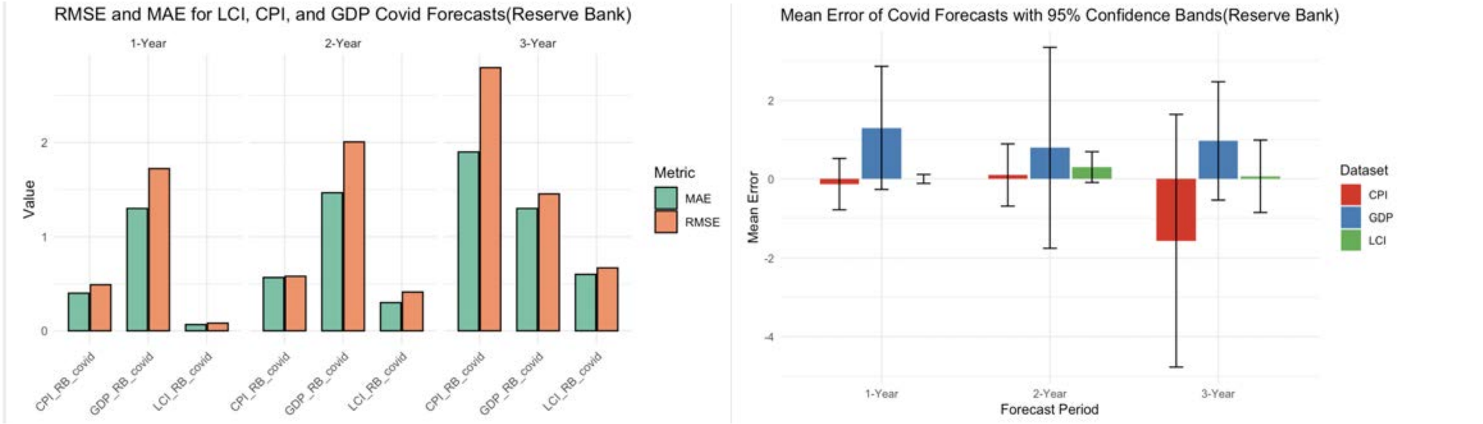


3) COVID-impacted Forecast Accuracy (2017-2019)

When compared to other variables, the LCI projections were indeed less affected by the COVID-19 epidemic. This was obvious from the decreased RMSE and MAE values for LCI projections throughout the COVID period, especially when compared to GDP and CPI forecasts (see Figure 3).

The very low rise in forecast errors for LCI during the pandemic showed that labour cost trends were more predictable and less susceptible to COVID-19-induced economic disruptions. This resiliency in LCI forecasting during a period of substantial economic uncertainty could be attributed to the potential stabilisation measures such as COVID-19 Leave Support Scheme implemented by the government that maintained a certain level of wages during the epidemic. (Work and Income, 2020)

Figure 6*: RMSE,MAE, Mean Error for LCI, CPI and GDP Forecasts From 2017 to 2019 by Reserve Bank*



- *Treasury*

From 2016 to 2019, the Treasury's BEFU and HYEFU reports provided useful information for analysing forecast accuracy. Throughout this time span, forecast precision remained steady, with no apparent shifts between the two reports (see Figures 7 and 8).

Short-term (1-Year) predictions showed to be more accurate than their 2-Year and 3-Year counterparts, with lower RMSE and MAE values, in line with frequent forecasting issues. This pattern represents the inherent rise in uncertainty and the possibility of errors with longer forecast periods. (FasterCapital, 2023)

When compared to CPI and QES projections, GDP forecasts had higher RMSE and MAE values across all periods, showing a greater difficulty in reliably estimating economic production. Furthermore, the QES mean error confidence intervals were continuously less than zero, indicating a systematic underestimating of employment estimates. This development could be attributed to unforeseen labour market dynamics or exogenous shocks that were not accounted for in forecasting models.

Figure 7 *: RMSE,MAE, Mean Error for QES, CPI and GDP Forecasts From 2016 to 2019 by Treasury(BEFU)*

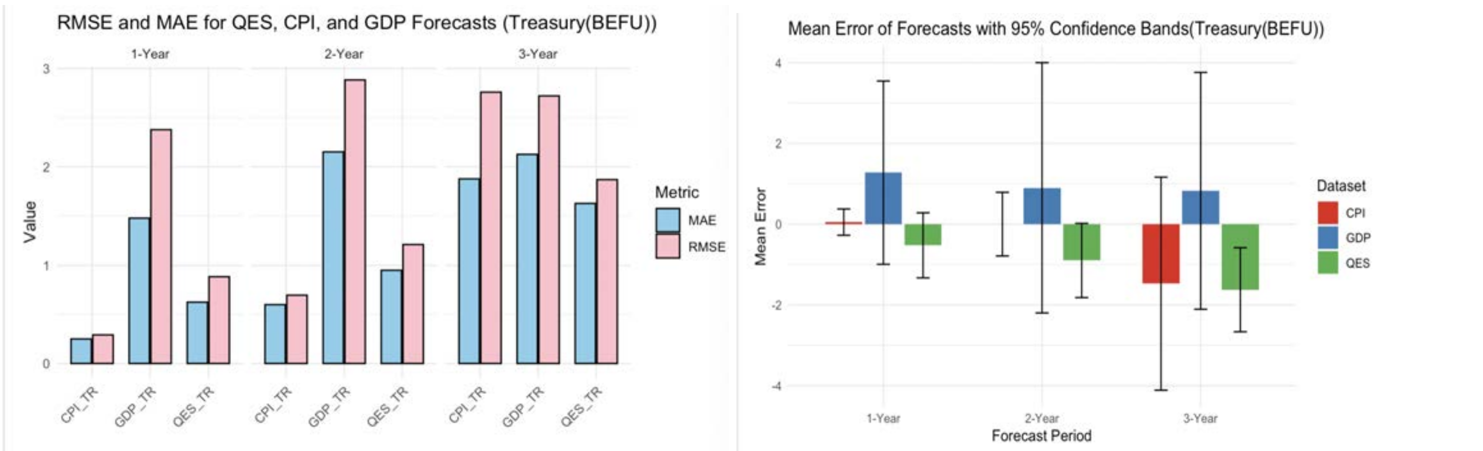
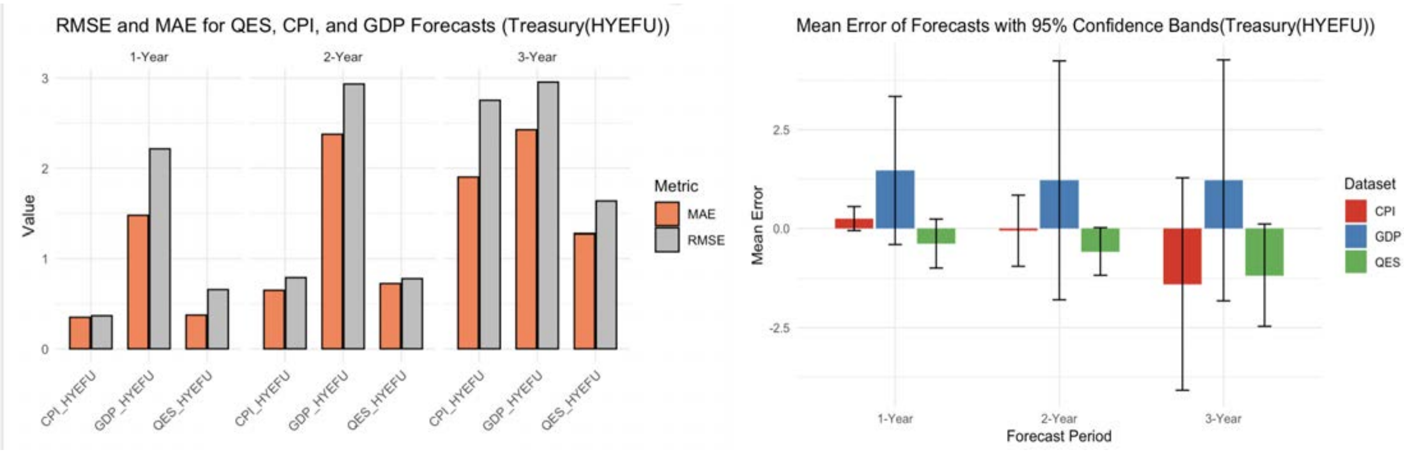


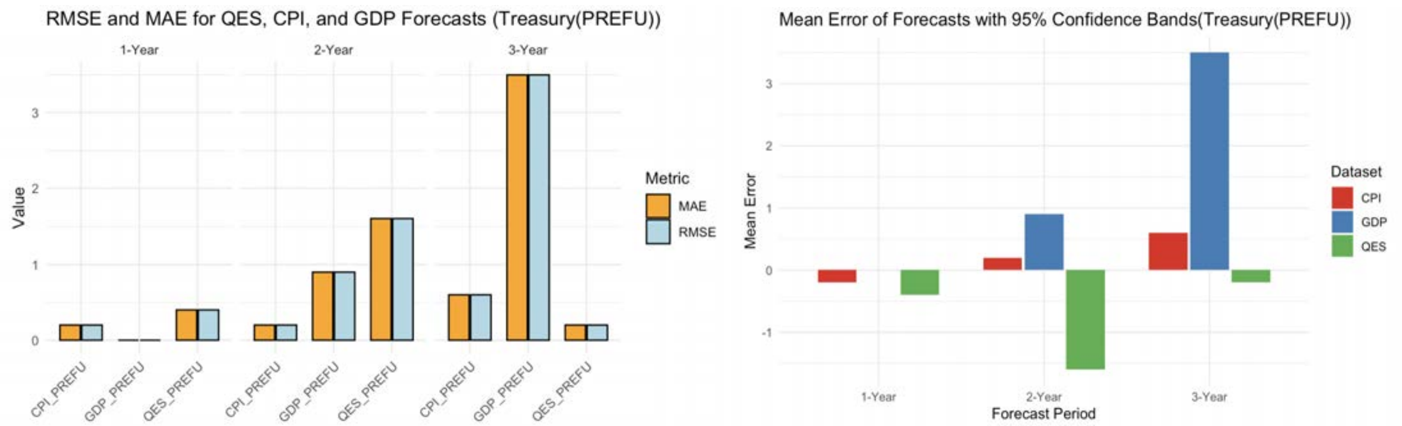
Figure 8 *: RMSE,MAE, Mean Error for QES, CPI and GDP Forecasts From 2016 to 2019 by Treasury(HYEFU)*



The 2017 PREFU results showed a considerable rise in RMSE and MAE values for the 3-Year forecast horizon, which far outperforms the 1-Year and 2-Year predictions (see Figure 9). This disparity highlighted COVID-19's enormous impact on economic predictability, particularly for long-term forecasts.

GDP and CPI predictions were significantly impacted, with their RMSE and MAE values for the third year reflecting the pandemic's increased uncertainty and economic volatility. The significant departure in these longer-term estimates could be linked to the COVID-19 crisis' unusual nature, which altered typical economic patterns and added significant new variables into the forecasting process.

Figure 9 *: RMSE,MAE, Mean Error for QES, CPI and GDP Forecasts in 2017 by Treasury(PREFU)*



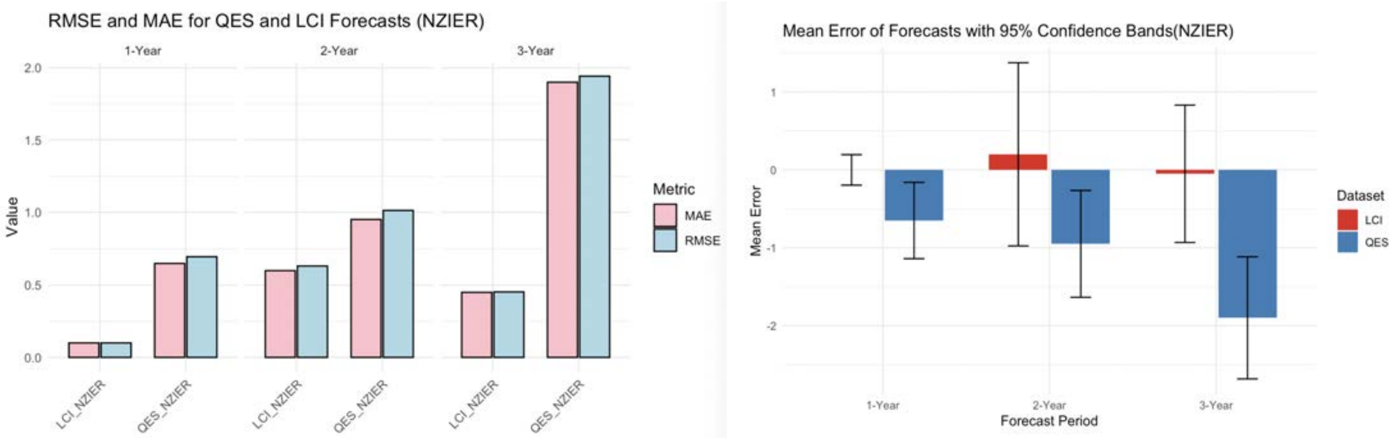
- *NZIER*

According to the NZIER's Quarterly Predictions reports from 2018 to 2019, forecasts for 1-year and 2-year horizons have lower RMSE and MAE than 3-year forecasts (see Figure 10). This pattern indicated that NZIER's short-term forecasting models were more accurate, which corresponded to the common hypothesis that predictability reduces as the prediction horizon lengthened (FasterCapital, 2023).

There was only a minor discrepancy between the RMSE and MAE figures during the 2018-2019 reporting period, which covered the beginning phase of the COVID-19 pandemic. This finding indicated that, despite the economic disruptions caused by the pandemic, the forecast accuracy for QES and LCI remained generally consistent, with no substantial mistakes.

Further examination of the QES mean error confidence intervals revealed that they were constantly less than zero, showing that NZIER's forecasts routinely underestimate actual employment levels.

Figure 10 *: RMSE,MAE, Mean Error for QES and LCI Forecasts from 2018 to 2019 by NZIER*



Cross-sectional comparison

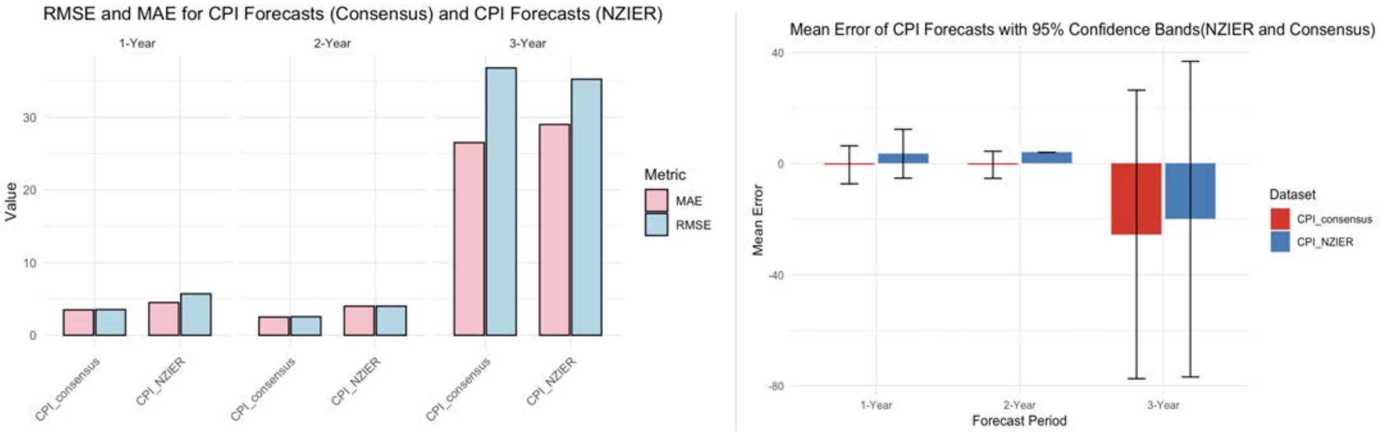
- *NZIER & Consensus CPI forecasts*

From 2018 to 2019, the CPI projections reports showed the consensus CPI index forecasts from several financial and economic agencies. When compared to NZIER and Consensus projections, the 1-year and 2-year RMSE and MAE values were much lower than those for the 3-year forecasts (see Figure 11). The COVID-19 pandemic, which put a high degree of unpredictability and volatility into economic conditions, was blamed for the pronounced disparity in accuracy for the longer horizon (FasterCapital, 2023).

When the two sets of estimates were compared, it was discovered that the consensus forecasts had lower RMSE and MAE values than the NZIER forecasts. This disparity in accuracy could be explained by the nature of consensus forecasting. Consensus forecasts, which are averages of estimates from several financial and economic agencies, inevitably encompass a wide range of perspectives and approaches. Individual forecast extremes and variations were effectively smoothed down by this aggregation, potentially resulting in a more balanced and less volatile projection.

In contrast to projections from a single body, such as NZIER, which may be based on a certain set of models and assumptions, consensus forecasts benefited from a variety of input sources. Each agency involved in the agreement may use different models, respond to various indicators, or prioritise economic concerns differently. This collaborative method might result in a more robust projection, particularly during times of economic uncertainty, such as the period hit by the COVID-19 pandemic. Thus, the average of consensus projections might have aided in countering unique biases or inaccuracies that individual models from different agencies, including NZIER, may have had.

Figure 11 *: RMSE,MAE, Mean Error for CPI Index Forecasts From 2018 to 2019 by NZIER and Other Financial and Economic Agencies*

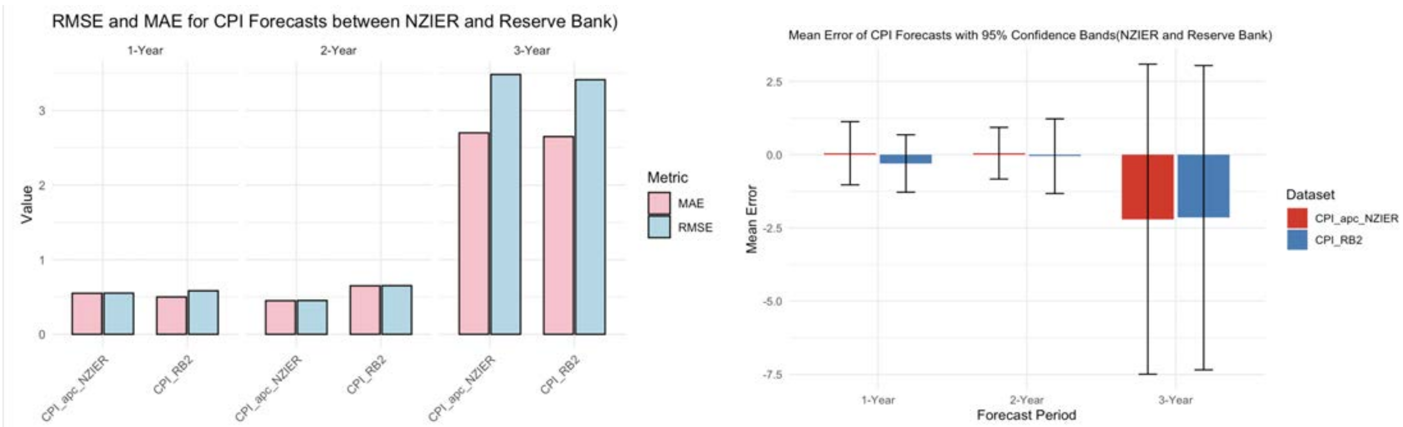


- *NZIER & Reserve Bank CPI forecasts*

NZIER and the Reserve Bank's CPI predictions from 2018 to 2019 revealed that the RMSE and MAE values for the 1-year and 2-year forecast horizons were much lower than those for the 3-year horizon (see Figure 12). This significant discrepancy in accuracy over longer projection periods could be related to the COVID-19 pandemic's unanticipated impact. The pandemic created significant uncertainty and volatility in economic conditions, which altered projections for the years 2021 and 2022 in particular.

The comparison analysis revealed a small variation between the RMSE and MAE values of the NZIER and Reserve Bank CPI projections. This resemblance indicates that both institutions had almost identical CPI predicting accuracy over this time period.

Figure 12 *: RMSE,MAE, Mean Error for CPI Forecasts From 2018 to 2019 by NZIER and Reserve Bank*



Diebold-Mariano (DM) test

The Diebold-Mariano (DM) Test results for comparing the forecast accuracy of NZIER and the Reserve Bank about the Consumer Price Index (CPI) revealed no statistically significant differences across the datasets analysed. In particular, a marginal difference in accuracy was noticed in the first test, which had a DM statistic of 1.027 and a p-value of 0.38, although it was not statistically significant. The second test, which produced a DM statistic of 0.6946 and a p-value of 0.5591, confirmed this conclusion, suggesting that there was no significant difference in forecast accuracy. Similarly, the third test revealed the same trend (see Table 8), with a DM statistic of 1 and a p-value of 0.5. The p-values in all of these examples were substantially over the conventionally recognised standards for statistical significance, which are typically set at 0.05. As a result, no evidence was found to imply that one forecasting model was superior to the other for CPI over the periods studied. These findings showed that the forecasting performances of NZIER and the Reserve Bank were statistically indistinguishable for the datasets evaluated.

Table 8 *: Diebold-Mariano Test of CPI Forecast Accuracy Between the NZIER and Reserve Bank*

CPI (NZIER & Reserve Bank)		
Forecast horizons	DM statistic	P-value
1 year	1.027	0.38
2 year	0.6946	0.5591
3 year	1	0.5

Similarly, none of the p-values for the 1-year, 2-year, and 3-year forecast horizons were less than the traditional limits for statistical significance (0.05) based on the Diebold-Mariano test results shown in Table 9. We determined that there were no statistically significant variations in CPI forecast accuracy between NZIER and Consensus estimates for any of the forecast horizons studied, with p-values of 0.3368, 0.4407, and 0.6643, respectively.

Table 9 *: Diebold-Mariano Test of CPI Forecast Accuracy Between the NZIER and Consensus Forecasts*

CPI (NZIER & Consensus)		
Forecast horizons	DM statistic	P-value
1 year	1.1408	0.3368
2 year	-0.95412	0.4407
3 year	-0.58225	0.6643

The small sample size could be a common factor for the lack of statistical significance, as revealed by the Diebold-Mariano (DM) test results. The power of a test in statistical analysis, which is the chance of successfully rejecting a false null hypothesis, is heavily reliant on sample size. With fewer observations, the test may lack the capacity to identify a true difference in forecast accuracy, even if one exists.

4. Conclusions

To summarise, it is critical to recognise that no model can precisely reflect the complexities of the actual world, and estimating the salaries of Members of Parliament is no exception. Each prediction model used in this study, including Linear Regression, LASSO Regression, ARIMA, and ARIMAX, has benefits and drawbacks.

Linear Regression, despite being widely used, suffered from overfitting and multicollinearity, especially with small sample sizes. Because of the interrelated nature of components and the possibility of overfitting, LASSO Regression, which was developed to solve multicollinearity, created interpretation issues. ARIMA models, which are good at capturing temporal dynamics through differencing and log transformation, faced difficulties such as overfitting and were concerned about residual non-normality. The ARIMAX model included exogenous variables for a more complete analysis, but it excluded autoregressive and moving average components, relying too heavily on external forecasts.

The forecasting process was inherently uncertain, influenced by factors such as overfitting, multicollinearity, and the accuracy of external predictor forecasts. Consequently, caution was advised in interpreting and relying solely on the presented forecasts.

The report also examined the reliability of key external economic forecasts, providing insights into their importance in improving salary prediction models. According to the data, short-term economic forecasts were often more accurate than long-term ones. This trend might be linked to the relative stability of forces influencing the economy in the short run, which allows models to capture and anticipate short-term changes more successfully.

As the Remuneration Authority requires a high level of confidence in predicting MPs' salaries for the next three years, our analysis faced challenges in determining the accuracy of the forecasted values due to issues such as overfitting, multicollinearity, and non-normality of residuals in the models mentioned above. Therefore, to some extent, I consider our analysis not entirely successful.

Despite the models not providing highly confident predictions, they still offer valuable insights for the Remuneration Authority. They can utilise this information in decision-making, adopting a conservative approach when determining future salaries, given the uncertainties surrounding the accuracy of the predictions. They may adopt a reset button, especially after the next election when they must make another determination. If there were any inaccuracies in their previous predictions, they can adjust them based on the actual circumstances of the future role holders.

In the end, we delve into addressing the key research questions(refer to 1.2 goals).

1. What information is available to assist the Authority in establishing the salaries and allowances payable in each of the three future years (2024, 2025, and 2026) at the beginning of July?

The available information includes historical MPs salary data, economic indicators(CPI, GDP), labour market statistics(LCI,QES,HLFS) and forecasts from the external agency (BERL), as outlined in the report section 2.1.

2. Where can the necessary information be obtained, and how will the Authority have access to it in three years and beyond?

The necessary information can be obtained from Stats NZ and BERL. The information is usually published quarterly or annually, which can be downloaded directly from their official website.

3. What models/tools can the Authority use to estimate wage movement/growth over the next three years and make projections beyond that time frame?

Regression models such as the linear regression model and time series models such as ARIMA and ARIMAX can be used to estimate the wage movement over the next three years. Each model has its strengths and limitations, as outlined in the report section 3.1.2 and 4.1.

4. What additional calculations and analyses are necessary to make the identified information useful in meeting the Authority's statutory obligations regarding MPs' salaries and allowances?

We're using what we know and what we have right now to predict the future, but we've got to keep an eye out for uncertainties ahead. Things like new government rules, how the country's economy is doing, and what's happening globally should all be paid attention to.

5. How to assess the prediction performance of economic indicators from external agencies?

The prediction performance of economic indicators from external agencies can be assessed through methodologies such as direct and cross-sectional comparisons using RMSE and MAE measurements.

The Diebold-Mariano test can also be employed to determine the statistical significance of differences in forecasting accuracy, as outlined in the report section 3.2.2 and 4.2.

In the future, RMSE and MAE can also be calculated by using Excel instead of using the R programming language. The formula for calculating RMSE and MAE can be referred to the report section 3.2.2. Let's take RMSE as an example.

Step1:Prepare Data

Ensure you have two sets of data - the actual values(Column A) and the predicted values(Column B).

Step2:Calculate the Squared Differences

In an empty column (let's say column C), subtract each observed value from the corresponding predicted value, then square the result. You can use the following formula in cell C2 and drag it down: =POWER(B2 - A2, 2)

Step3: Calculate the Square Root of Mean of Squared Differences

You can use the following formula =SQRT(AVERAGE(C2:Cn))

5. Future work

In advancing our forecasting capabilities, two strategic initiatives stand out:

Firstly, with the passage of time, there is an opportunity to update and incorporate fresh data into our analysis. This involves ensuring the availability of a more extensive and diverse dataset, which can significantly contribute to refining and retraining our prediction models.

Secondly, collaboration with experts in economic forecasting is pivotal. Engaging with professionals in this field can offer valuable insights and expertise, particularly in improving the precision of external predictor forecasts. By working closely with economic forecasting specialists, we can enhance the reliability of economic indicators, addressing potential limitations and refining our models to better capture the intricacies of the economic landscape. This collaborative effort ensures that our predictions are grounded in robust economic insights, contributing to a more accurate and informed forecasting process.

6. Acknowledgements

This report was made possible by the combined efforts of individuals and organisations. We would like to convey our heartfelt gratitude to the Remuneration Authority for their collaboration and support throughout this study project. Also, a heartfelt thank you to our academic supervisors from the University of Canterbury, James Williams and Hazel Halton. Their experience, encouragement, and consistent assistance have been invaluable throughout the study process.

7. References

Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2013). *Applied Multiple Regression/Correlation Analysis for the Behavioral Sciences* (J. Cohen, P. Cohen, S. G. West, & L. S. Aiken, Eds.). Taylor & Francis Group. https://doi.org/10.4324/9780203774441

The Data Scientist. (2022). *Performance measures: RMSE and MAE*. The Data Scientist. Retrieved January 3, 2024, from https://thedatascientist.com/performance-measures-rmse-mae/

FasterCapital. (2023, December 7). *Forecast horizon: Evaluating the Influence of Forecast Horizon on Accuracy Measurements*. FasterCapital. Retrieved January 3, 2024, from https://fastercapital.com/content/Forecast-horizon--Evaluating-the-Influence-of-Forecast-Horizon-on-Accuracy-Measurements.html

Feng, C., Wang, H., Lu, N., Chen, T., He, H., Lu, Y., & Tu, X. M. (2014). Log-transformation and its implications for data analysis. *Shanghai Arch Psychiatry*, *26*(2), 105–109. 10.3969/j.issn.1002-0829.2014.02.009

Gilleland, E., & Roux, G. (2015). A new approach to testing forecast predictive accuracy. *22*(3), 534-543. Retrieved 1 3, 2024, from https://doi.org/10.1002/met.1485

Hayes, A. (2023, September 29). *Autoregressive Integrated Moving Average (ARIMA) Prediction Model*. Investopedia. Retrieved January 3, 2024, from https://www.investopedia.com/terms/a/autoregressive-integrated-moving-average-arima.asp

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2021). *An Introduction to Statistical Learning with Applications in R* (Second ed.). Springer. https://www.statlearning.com/

NZIER. (2023). NZIER. Retrieved December 20, 2023, from https://www.nzier.org.nz/member-publications

Public Service Commission. (2023, November 2). *Raraunga Ohumahi - Ia utu Workforce Data - Wage trends*. Te Kawa Mataaho Public Service Commission. Retrieved November 29, 2023, from https://www.publicservice.govt.nz/research-and-data/workforce-data-remunerationpay/wage-trends/

Remuneration Authority. (2021, June 4). *Our responsibilities*. Remuneration Authority. Retrieved December 21, 2023, from https://www.remauthority.govt.nz/footer/about-us/our-responsibilities/

Reserve Bank of New Zealand. (2023). *All Monetary Policy Statements*. Reserve Bank of New Zealand. Retrieved December 18, 2023, from https://www.rbnz.govt.nz/monetary-policy/monetary-policy-statement/monetary-policy-statement-filtered-listing-page#sort=%40computedsortdate%20descending

Smarten. (NA, NA NA). *What is ARIMAX Forecasting and How is it Used for Enterprise Analysis?* Smarten Augmented Analytics. Retrieved January 3, 2024, from https://www.smarten.com/blog/arimax-forecasting-enterprise-analysis/

Stats NZ. (2023). *Home Large datasets CSV files for download*. Stats NZ. Retrieved November 29, 2023, from https://www.stats.govt.nz/large-datasets/csv-files-for-download/

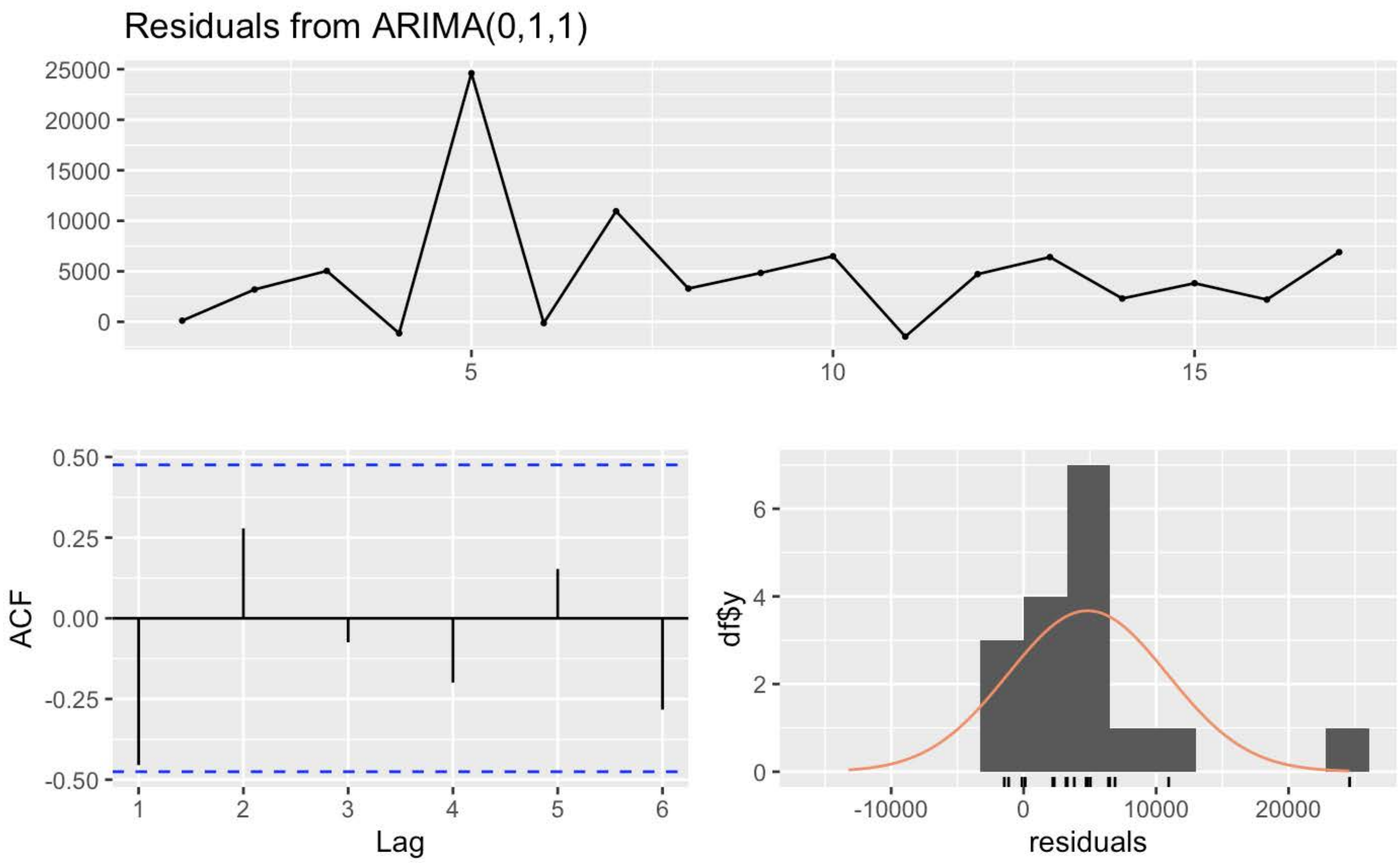
The Treasury New Zealand. (2023). The Treasury New Zealand. Retrieved December 18, 2023, from https://www.treasury.govt.nz/

Work and Income. (2020). *COVID-19 Leave Support Scheme*. Work and Income. Retrieved January 5, 2024, from https://www.workandincome.govt.nz/covid-19/previous-payments/leave-support-scheme.html

8. Appendices

ARIMA Model Residual Check Result

Figure 13 *: Residuals From ARIMA(0,1,1)*



Residuals from ARIMA(0,1,1):

This graph displays the residuals, which are the differences between the predicted values and the actual values from the ARIMA model. Ideally, the residuals should appear as random scattered points around zero. If there's a noticeable pattern or trend, it suggests that the model might not be capturing certain aspects of the data.

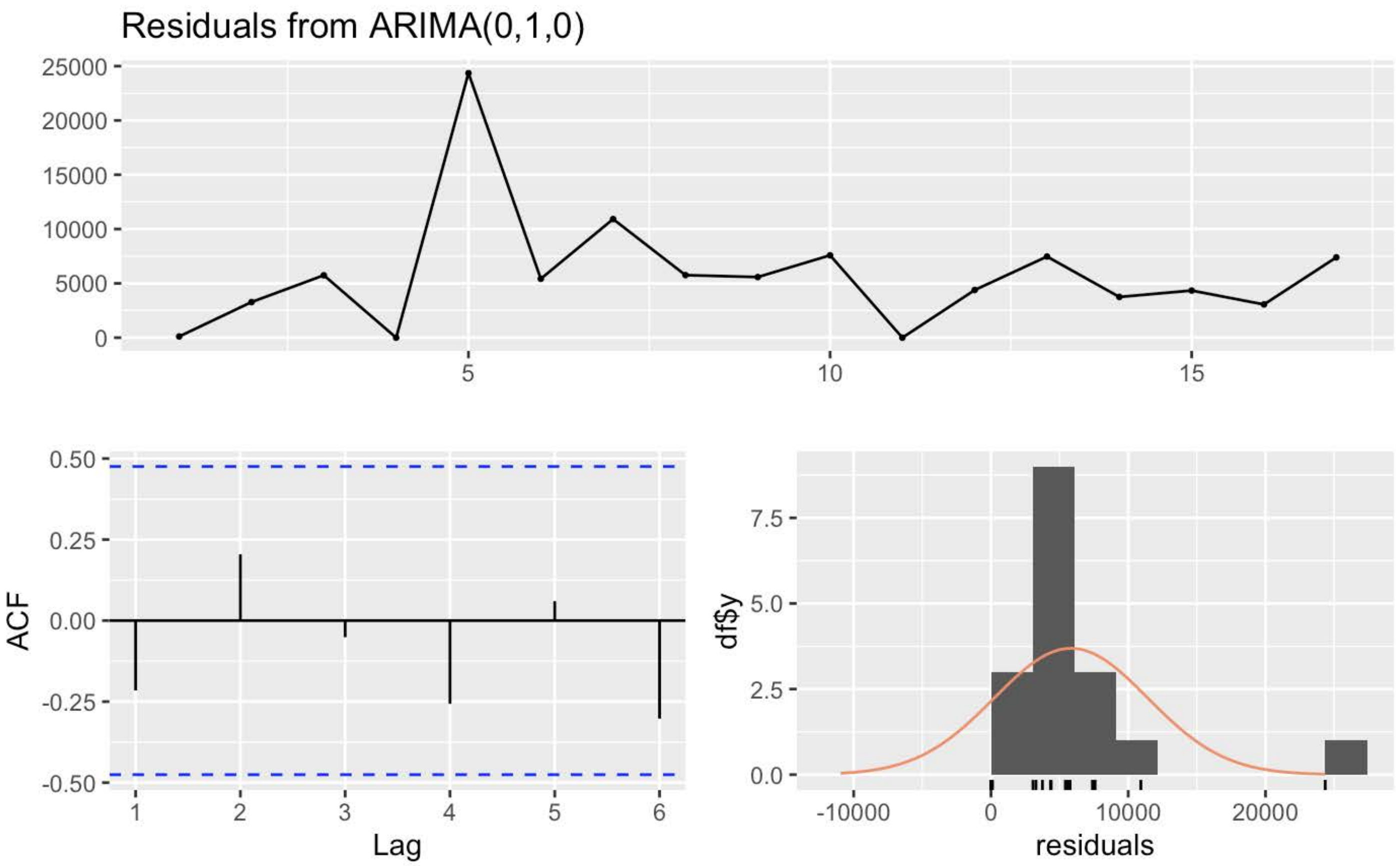
ACF (Autocorrelation Function) of Residuals:

The ACF plot shows the autocorrelation of the residuals at different lags (time intervals). In an ideal situation, the autocorrelation values at different lags should be close to zero, indicating that the residuals are not correlated across time. If there are spikes or patterns outside the confidence interval(the blue dash line), it may suggest remaining patterns in the residuals that the model has not captured.

Histogram of Residual Distribution:

This histogram visualises the distribution of the residuals. A normal distribution (bell-shaped curve) is desirable, indicating that the residuals are centred around zero. If the histogram is skewed or has unusual shapes, it suggests deviations from normality. Deviations from normality might imply that the model assumptions are not fully met.

Figure 14 *: Residuals From ARIMA(0,1,0)*



ARIMA Model Residual Check Result(After Log Transformation)

Figure 15 *: Residuals From Log-transformed ARIMA(0,1,1)*

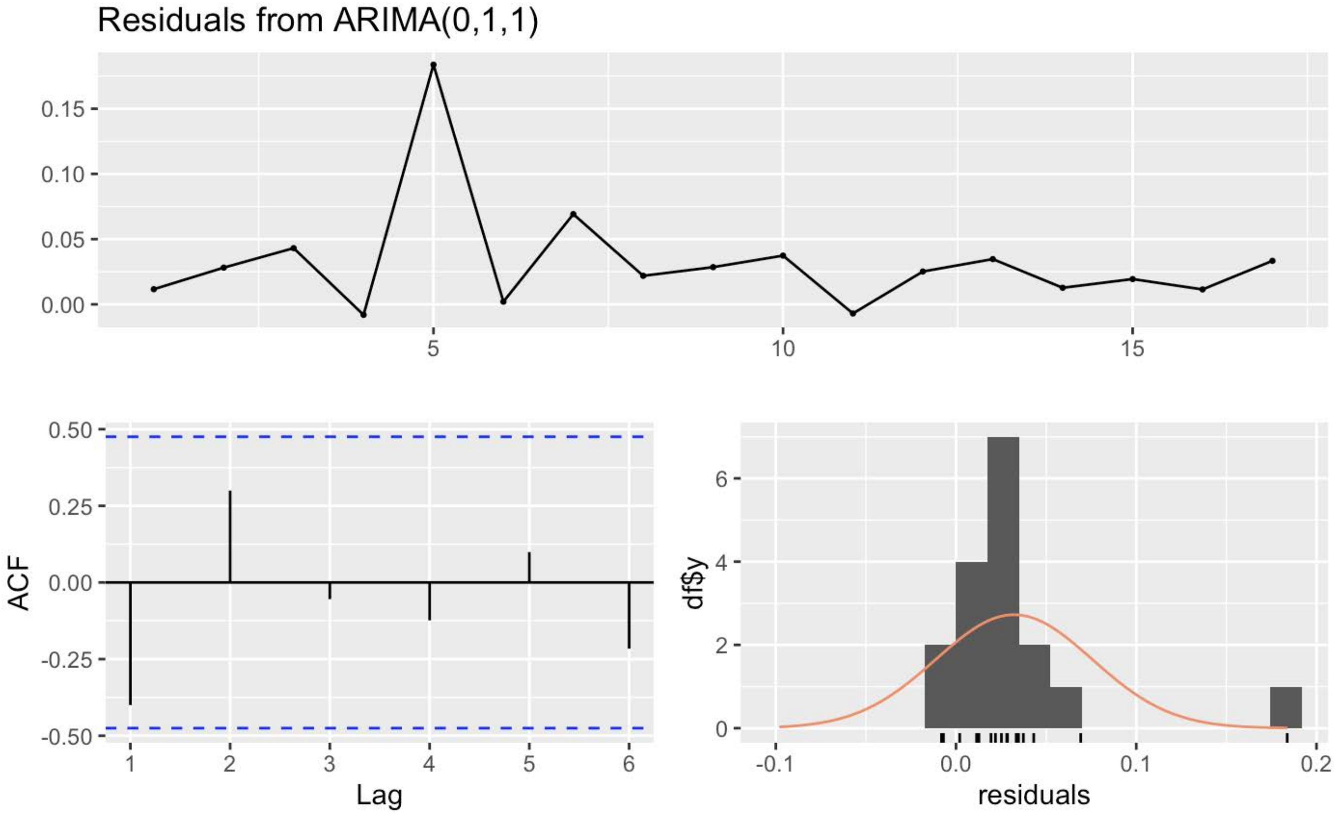
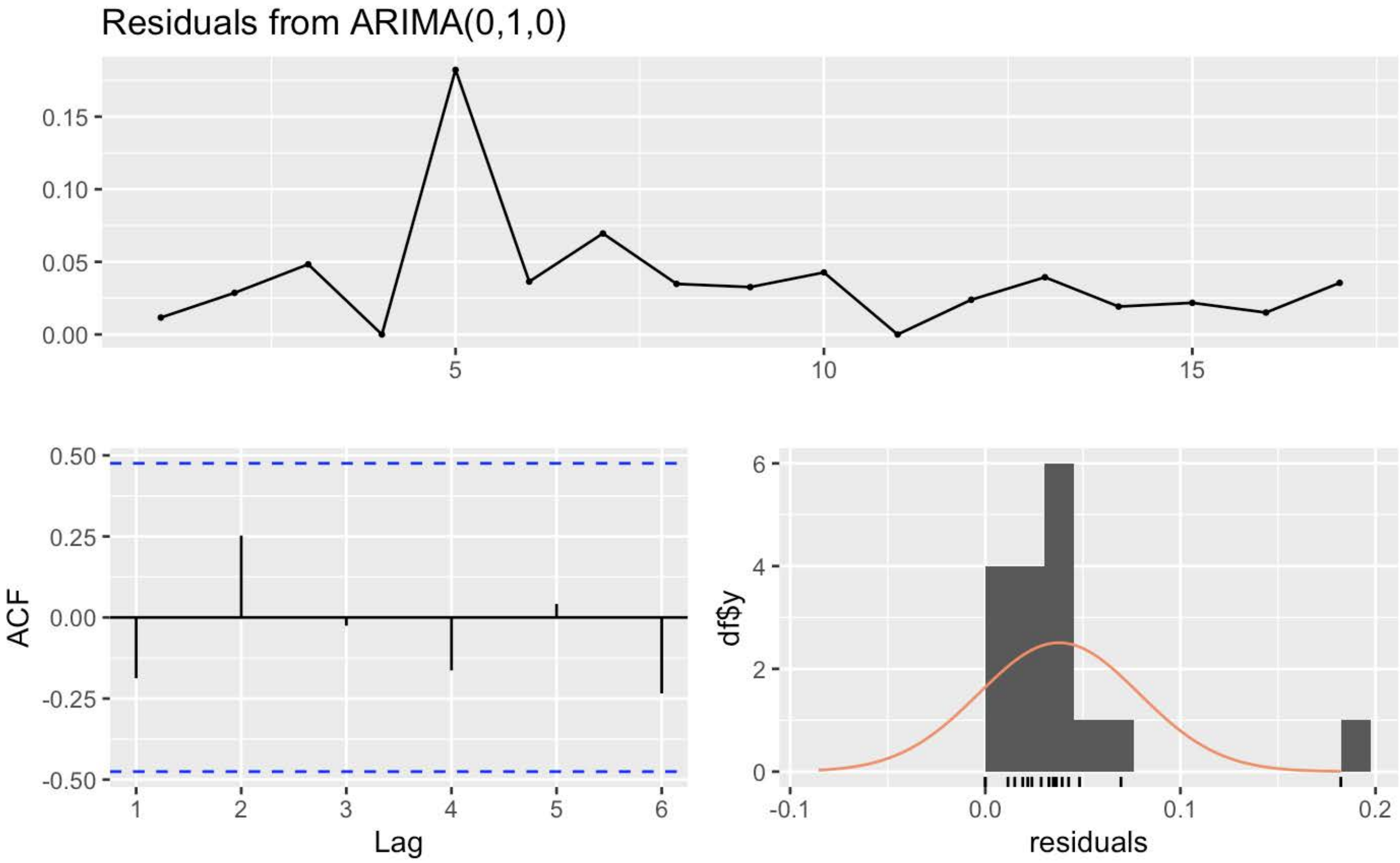
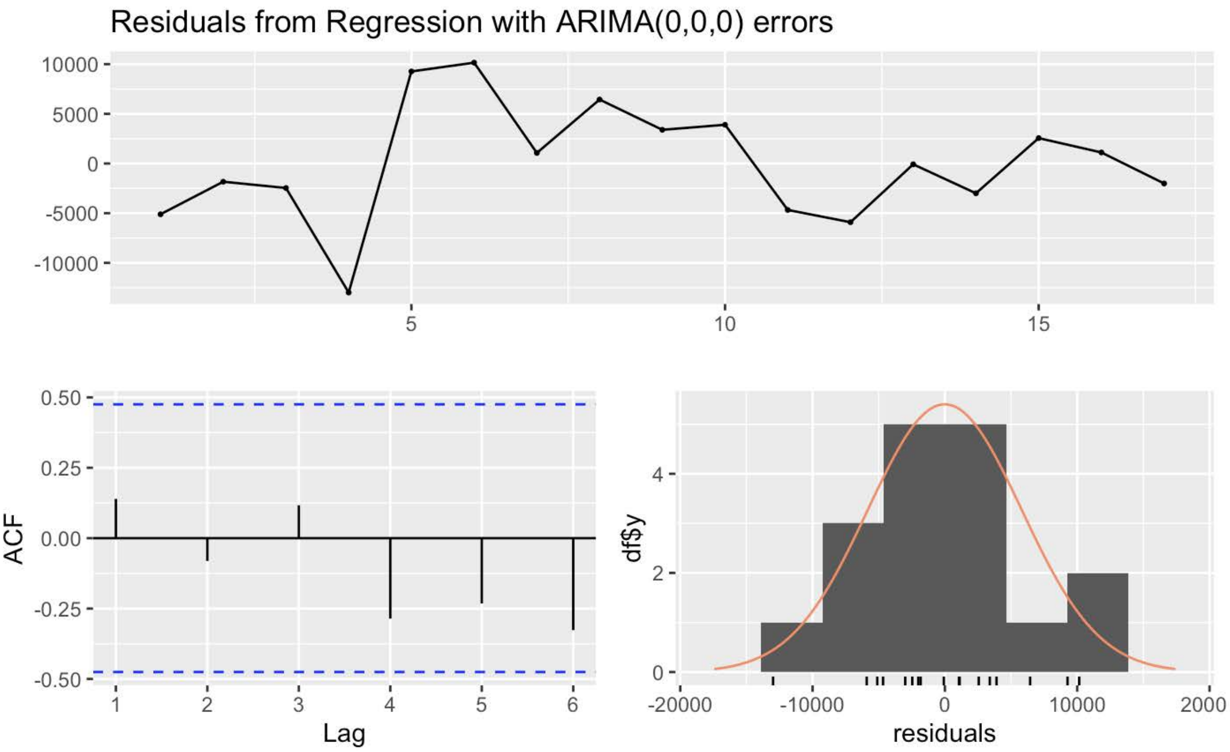


Figure 16 *: Residuals From Log-transformed ARIMA(0,1,0)*



ARIMAX Model Residual Check Result

Figure 17 *: Residuals From ARIMAX(0,0,0)*



ARIMAX Model Residual Check Result(After Log Transformation)

Figure 18 *: Residuals From Log-transformed ARIMAX(0,0,0)*