Abstract

This study examines Canadian wage growth from January 2001 to December 2019 using ARIMA models. The exploratory data analysis revealed an upward trend with minimal seasonal fluctuations, motivating comparison of ARIMA (1,1,2) and ARIMA (3,1,0) specifications. The one step ahead rolling forecasts and loss metrics showed no significant accuracy difference (Diebold–Mariano $p = 0.55$) and the residual diagnostics confirmed the absence of autocorrelation and conditional heteroskedasticity, hence the more parsimonious ARIMA (3,1,0) was selected.

Forecasts projected a continuation of the wage growth trend, indicating stable labor market conditions in the near term. However, residual diagnostics revealed small deviations from normality, suggesting unmodeled complexity or external shocks that were not covered by the linear assumptions. Restrictions are mentioned in the study, including potential structural breaks from economic disruptions and how future research should incorporate external covariates to consider additional robustness.

This research highlights the effectiveness of ARIMA in trend-based wage forecasting in Canada as a credible tool for policymakers and business to plan the labor market. The results highlight the value of model simplicity where there is no seasonality and contribute to evidence-based knowledge of wage dynamics in informing well-informed decision-making in economic policy and workforce management.

Contents

1 Introduction

The ability to predict future situations enables organizations to optimize their operations while making better informed decisions. Organizations that employ effective forecasting methods are able to optimize their resources while minimizing costs and achieving optimum performance in forecasting Canada's time series wages data. The ability to detect trends in the labor market as well as economic shifts and earnings trends through forecasting models makes them extremely valuable to both business strategy development as well as policy development.

Risk management is one of the principal advantages of forecasting because it recognizes economic risks which allow organizations to reduce their financial losses (Rajopadhye et al., 2001). Room demand forecasting is advantageous to the hotel industry because it allows pricing optimization which translates to revenue maximization (Rajopadhye et al., 2001).

The forecasting process comes with certain restrictions. Economic metrics face external impacts, while forecasting inaccuracies result in both operational inefficiencies and poor business decisions. Model reliability improves only through ongoing model refinement efforts.

Canada's wage time-series data analysis benefits from forecasting because it enables policymakers to understand labor market trends for economic policy development. Organizations that implement advanced predictive methods achieve better decision-making capabilities, better financial planning, and better ability to handle wage adjustment uncertainties. The application of forecasting models by policymakers and businesses enables economic planning that remains responsive while being evidence-based and labor force needs-driven.

The analysis uses Canadian wage data to conduct time series analysis through seven distinct sections. Section 2 explains the dataset while Section 3 explains model selection and Section 4 explains estimation and diagnostics and Section 5 presents the forecast and Section 6 draws conclusions with important findings.

2 Data Description

The data analyzed in this study come from Statistics Canada's Real-Time Data Tables, which contain 28 economic and social time series data with historical revisions. This study uses Table 14-10-0331-01, which displays monthly average weekly earnings with overtime pay for all employees in all industries in Canada. The data span from January 2001 to December 2019, a total of 228 months without any gaps in the data. The series have been seasonally adjusted to filter out the short-term variations so that the underlying economic patterns can be observed. When no seasonal pattern is present, unadjusted data is used to provide better representation of true trends in specific industries or time periods.

To authenticate data, quality measures are embedded into the dataset. They assess reliability based on the coefficient of variation (CV), non-response error and imputation error. Each piece of data gets a quality rating: A (Excellent), B (Very Good), C (Good), D (Acceptable), E (Use with Caution), and F (Too Unreliable to Publish or Sample Size Too Small). These ratings have significant analytical implications. For instance, high-rated A- or B-rated data are best suited for the robust trend analysis and economic forecasting, whereas E- or F-rated data should be interpreted with caution, especially when used to inform policies or labor market assessments.

The statistics conform to the 2022 North American Industry Classification System (NAICS) Version 1.0 to be in line with the current industrial classification. However, because industry classifications and definitions have changed over time, the analysis of trends by sector might be problematic, and users should be cautious in this respect.

2.1 Exploratory Data Analysis

The analysis of Canada's monthly average wages from January 2001 to December 2024 shows a stable wage structure that grows steadily. The wages spanned between 9.6 million and 12.7 million units throughout this period. The mean wage stood at 11.07 million which was slightly above the median of 11.05 million indicating a mild right skew in the distribution possibly due to wage revisions or high-earning months. The inter-quartile range (Q1 = 10.39 million; Q3 = 11.70 million) indicates that most wage values are moderately dispersed, indicating that the bulk of monthly earnings clustered around the central trend with relatively few extreme values.

Statistic	Value
Minimum	9,649,220
1st Quartile (Q1)	10,389,409
Median	11,050,133
Mean	11,071,012
3rd Quartile (Q3)	11,695,698
Maximum	12,661,663

Table 1: Exploratory Data Analysis (2001-2019).

The time series plot (Figure 1) shows a consistent upward trajectory from 2001 to 2019, reflecting Canada's economic growth and gradual improvement in labor market conditions over the years. This upward momentum, however, may have experienced fluctuations beyond 2019 due to external shocks such as the COVID-19 pandemic, which would be observable in the later segments of the dataset.

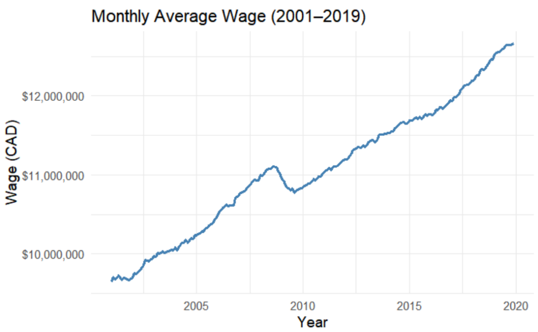

Figure 1: Time Series Plot of Average Wage (CAD)

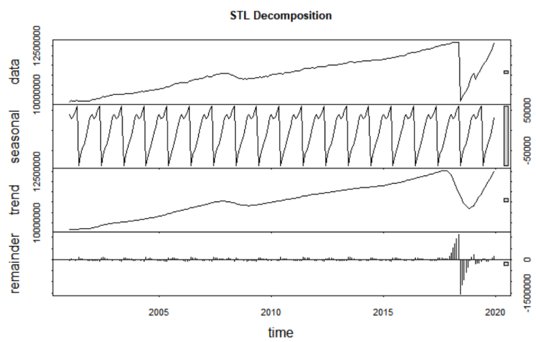

Figure 2: STL decomposition chart

Figure 2 presents a decomposition of the time series into trend, seasonal, and residual components. The trend component continues to affirm a clear upward growth pattern. Interestingly, the seasonal component exhibits a repetitive zigzag structure that remains stable over time. However, the amplitude of this seasonal fluctuation is minimal, suggesting that any month-to-month variation is not statistically significant. This implies that the data lacks strong or consistent seasonal behavior, which is essential when selecting appropriate forecasting models.

The Seasonal Subseries Plot shown in Figure 1 in Appendix A further supports this observation. Wage levels across months show little deviation from one year to the next, reinforcing the notion that Canadian average wages are not subject to pronounced seasonal shifts. In contrast to industries where wages are sensitive to seasonal employment cycles (such as agriculture or retail), the aggregated data across all industries smooths out such effects, leaving a predominantly trend-driven structure.

Given the limited presence of seasonality, seasonal models like SARIMA may not provide a significant improvement in forecasting accuracy. Non-seasonal models, particularly ARIMA, which specialize in modeling trend and auto regressive components, are more appropriate for this dataset. This also increases the reliability of long-term projections, policy assessments, and economic modeling derived from this wage series.

3 Model Specification

3.1 Initial Time Series Assessment

To commence the analysis, the data must first be tested for seasonality. This step is essential to determine whether a transformation is required to enhance model accuracy and ensure the selection of the most appropriate forecasting approach.

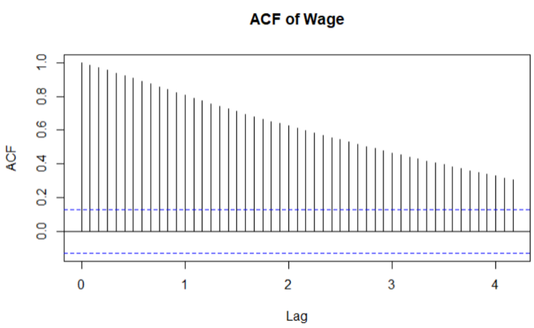

Figure 3: ACF plot of Wage

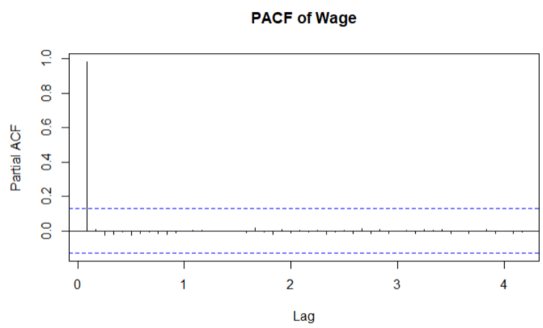

Figure 4: PACF plot of Wage

The downward trend observed in the wages time series indicates that wages are not stationary over time, a characteristic that exerts a significant influence on model choice and estimation. The smooth and linear decline in the ACF plot in Figure 3 presents evidence of a trend or drift in the time series. This drift indicates that the series lacks a constant mean or variance violating an extremely crucial requirement to be satisfied prior to time series model usage.

Partial Autocorrelation Function (PACF) plot in Figure 4 indicates an exceptionally high peak at lag 1, of approximately 0.2, well beyond the 95% boundary. All from there then reduces to close to zero and remains within constraints until lag 50. This is what a stationary series, however, shows as typical with this pattern when the AR (1) model, where the current value heavily depends on the previous one right before it.

But the gradual decline of the ACF implies non-stationarity, a long-run trend that affects the entire series. The peak lag 1 spike in the PACF is thus suspect, perhaps an indication of the underlying trend. The lack of strong autocorrelation past lag 1 lends itself to this explanation.

3.2 Stationarity Testing: Augmented Dickey-Fuller (ADF) Test

The ADF test served to statistically evaluate stationarity. The ADF test checks whether a unit root exists in time series data which indicates non-stationarity. The null hypothesis gets rejected when the p-value remains below 0.05 which shows that the series demonstrates stationarity.

The ADF test produced a p-value of 0.3177 when applied to the original wage series which exceeds the required significance threshold. The results show that the series was non-stationarity i.e., its mean, variance and covariances change over time. Non-stationary data are unpredictable and cannot be modeled or forecasted since it changes over time, so to proceed, the original wage series needs to be transformed before modeling.

3.3 Variance Stabilization: Box-Cox Transformation

Box-Cox transformation was applied to address potential heteroskedasticity (non-constant variance). The Box-Cox method identifies an optimal power parameter λ that stabilizes variance and makes the series more amenable to modeling. In our analysis, the optimal lambda was estimated at 0.5050505.

However, since log-transformation corresponds to a lambda of 0, and this value was within the 95% confidence interval of the estimated lambda, the log transformation was chosen. This choice ensures interpretability and is commonly used in economic time series. A Box-Cox diagnostic plot supporting this decision is included in Figure 2 of Appendix A.

3.4 Differencing for Stationarity

The stationarity of the series was re-evaluated after applying the log transformation, and the resulting p-value was 0.1996, indicating the series remained non-stationary even after the transformation. To eliminate the trend and achieve stationarity, first-order differencing was applied.

Differencing involves subtracting the previous observation from the current observation (i.e., $Y_t = Y_t - Y_{t-1}$). This operation removes linear trends and stabilizes the mean over time. After differencing the log-transformed series, the ADF test yielded a p-value of 0.0219, which is below the 0.05 threshold – confirming that the differenced series is now stationary. This transformed and differenced series serves as the basis for our subsequent model identification and estimation steps.

Visual inspection of the differenced log-transformed wage data, along with its ACF and PACF plots supports this conclusion. The differenced series appears to fluctuate around a constant mean with stabilized variance. The ACF shows a sharp spike at lag 1 followed by values within the 95% confidence interval, indicating reduced autocorrelation and confirming the effectiveness of the transformation. The PACF also shows a significant spike up to lag 3, and minor fluctuations thereafter, providing initial guidance for ARMA model identification.

The ACF plot shown in Figure 3 of Appendix A shows that the autocorrelations taper off slowly rather than cutting off sharply after lag 1, while the PACF in Figure 4 of Appendix A shows a significant spike at lag 1, then essentially zero afterwards. That pattern slowly decaying ACF + PACF that cuts off after lag 3 is the hallmark of an AR (3) on the differenced series.

Following log transformation and first-order differencing, the data was reassessed for any remaining seasonality. Seasonal plots, subseries plots, and the ACF showed no clear monthly patterns or significant autocorrelation at seasonal lags, indicating that seasonality had been effectively removed. Although STL decomposition suggested a minor seasonal component, its magnitude was negligible. Overall, the transformed series appeared stationary with no strong seasonal structure, supporting the use of a non-seasonal ARIMA model.

3.5 Model Identification Using EACF

Extended Autocorrelation Function (EACF) table to indicate potential ARMA (p, q) model structures. EACF is a matrix method which is intermediate in characteristics to both ACF and PACF and assists in identifying the correct p and q. An 'o' in the table indicates a potential model, and 'x' eliminates it. The subsequent models were indicated to be potential by the outcomes of the EACF as ARMA (1,1,1), MA (0,1,3), ARMA (1,1,3), AR (3,1,0), ARMA(3,1,1) and ARMA (3,1,3). These models were shortlisted for further evaluation in the parameter estimation and diagnostic stage.

4 Parameter Estimation and Diagnostic Check

This section discusses the estimation of the parameters of the models via Maximum Likelihood Estimation (MLE) followed by residual diagnostic tests to check if each fitted model is satisfactory. The intention is to have each of the candidate models fit statistically and satisfy key assumptions before model selection and forecasting.

4.1 Model Estimation (MLE)

Maximum Likelihood Estimation (MLE) was used to estimate the parameters of the candidate models. MLE is widely preferred in time series analysis due to its statistical efficiency and asymptotic properties. By maximizing the likelihood function, MLE ensures that the parameter values chosen are those most likely to have generated the observed data. Using this approach, the following ARIMA-type models were estimated - ARIMA (1,1,1), ARIMA (0,1,3), ARIMA (1,1,2), ARIMA (3,1,0), ARIMA (1,1,3), ARIMA (3,1,1), ARIMA (3,1,3). Estimation was performed using the `arima()` function with `method = "ML"`.

4.2 Statistical Significance of Coefficients

To ensure the robustness and interpretability of our models, all estimated parameters were evaluated using the z-test (based on standard errors from Maximum Likelihood Estimation). Models with at least one statistically insignificant coefficient (p-value > 0.05) were flagged for rejection or further scrutiny. Table 2 shows the result with the actual result available in Table 1 of Appendix A.

Model	All Coefficients Significant?	Reason for Acceptance/Rejection
ARIMA (1,1,1)	Yes	Passed all significance tests
ARIMA (0,1,3)	Yes	Outperformed MA(2) in AIC and residual tests
ARIMA (1,1,3)	No	AR (1) coefficient not significant (p = 0.0583) and ARMA (1,2) was better in residuals test and AIC
ARIMA (1,1,2)	Yes	All coefficients highly significant
ARIMA (3,1,0)	Yes	
ARIMA (3,1,1)	No	3 coefficients not significant
ARIMA (3,1,3)	No	4 coefficients not significant

Table 2: ARIMA Component Model

The shortlist from this stage is ARIMA (1,1,1), ARIMA (0,1,3), ARIMA (1,1,2) and ARIMA (3,1,0). These models exhibited complete parameter significance and proceeded to the residual diagnostic phase.

4.3 Residual Diagnostics

To validate model adequacy, residual diagnostic tests was conducted focused on autocorrelation and normality.

The Ljung-Box Test was used to detect autocorrelation in the standardized residuals. This test evaluates whether the autocorrelations of residuals at multiple lags are jointly zero. The null hypothesis is that residuals are independently distributed (i.e., no autocorrelation). A p-value greater than 0.05 indicates that the standardized residuals behave like white noise. In our analysis, the test was applied to standardized residuals to ensure comparability and account for scale.

The results were as shown in Table 4 shows that ARMA (1,1,2) and AR (3,1,0) behave like white noise.

Model	Ljung-Box Value	Pass/Fail	Notes
ARMA (1,1,1)	0.0000495	Fail	Strong residual autocorrelation
MA (0,1,3)	0.05322	Borderline	
ARMA (1,1,2)	0.0765	Pass	No significant autocorrelation
AR (3,1,0)	0.1394	Pass	Residuals behave like white noise

Table 3: Ljung-Box test

The Shapiro-Wilk test is a widely used test for normality. However, it was not employed in this analysis due to its sensitivity to outliers. Specifically, outliers in January 2001 and November 2006 caused the test to falsely indicate non-normality. Instead, visual inspections using Q-Q plots and histograms were performed.

A visual inspection of the Q-Q plot for ARIMA (1,1,2) and ARIMA (3,1,0) showed that most points lie almost exactly on the 45° reference line between roughly the 10th and 90th percentiles, indicating the bulk of your residuals follow a normal distribution closely. All models visually passed the normality checks. These diagnostic results were pivotal in informing the model

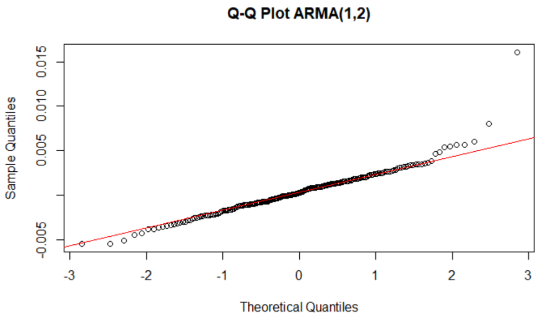

Figure 5: Q-Q plot ARIMA (1,1,2)

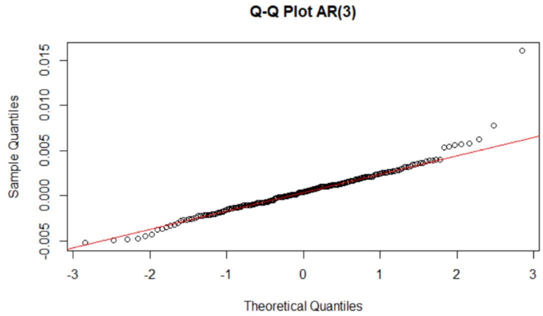

Figure 6: Q-Q plot ARIMA (3,1,0)

comparison and final model selection steps discussed in the subsequent section.

4.4 Outlier Detection and Influence Assessment

Outlier detection was performed to assess whether specific data points exert disproportionate influence on the model fit or residual diagnostics. Standardized residuals were computed for all fitted models, and any residual exceeding ± 3 standard deviations was flagged as an outlier.

Across the two retained models (ARIMA (1,1,2), and AR (3,1,0)), two consistent outliers were detected:

- January 2001: This likely corresponds to post-Y2K labor market adjustments or early 2000s economic shifts.
- November 2006: A potential economic anomaly warranting further investigation.

These outliers were not removed but were acknowledged in the interpretation of diagnostic tests. Their presence supports the decision to avoid statistical tests like Shapiro-Wilks, which are highly sensitive to extreme values.

None of the outliers appeared to severely distort auto-correlation structure or compromise model adequacy, as the Ljung-Box tests on standardized residuals still indicated acceptable white noise behavior for most models.

4.5 ARCH Effect Diagnostics

ARCH (Autoregressive Conditional Heteroskedasticity) effects occur when the variance of a time series is not constant, but rather changes over time depending on past squared residuals. In financial and economic data, this phenomenon is often referred to as "volatility clustering" periods of high variability followed by periods of low variability.

ARCH effects were tested using both graphical and formal approaches. The ACF and PACF plots of the squared standardized residuals were examined to visually detect any significant auto-correlation. Additionally, two formal statistical tests were conducted:

1. **Ljung-Box Test on Squared Residuals**: Applied at lags 4, 8, and 24. This checks for autocorrelation in the squared residuals, which may indicate time-varying volatility.
2. **ARCH LM Test** (from Cryer & Chan, Chapter 8): This test evaluates the null hypothesis of no ARCH effects against the alternative that conditional heteroskedasticity is present. The test statistic follows a chi-squared distribution.

The results for ARIMA (1,1,2) and ARIMA (3,1,0) are summarized below:

- **ARIMA (1,1,2)**:
  - Ljung-Box (lag 4, 8, 24): p-values = 0.5941, 0.8768, 0.9966
  - ARCH LM Test (lag 12): p = 0.7724
- **ARIMA (3,1,0)**:
  - Ljung-Box (lag 4, 8, 24): p-values = 0.699, 0.8873, 0.9966
  - ARCH LM Test (lag 12): p = 0.8127

In both models, all p-values were well above 0.05, suggesting the squared residuals are uncorrelated and that the conditional variance is stable. Thus, there is no significant evidence of ARCH behavior.

These findings confirm that our residuals exhibit homoscedasticity, satisfying a key assumption for reliable inference and forecasting.

4.6 Model Comparison and Selection

4.6.1 Model Selection Criteria

Model selection was guided by the Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC), which balance model fit and complexity. Lower values of AIC and BIC indicate a better-performing model.

Model	AIC	BIC
ARIMA (1,1,2)	-2126.1	-2112.4
AR (3,1,0)	-2131.2	-2117.5

Table 4: AIC and BIC for both models

Although both models passed all diagnostic tests, ARIMA (3,1,0) was selected as the final model due to its lower AIC and BIC values. This model offered the best trade-off between parsimony and fit.

4.6.2 Final Model Deficiencies

While the ARIMA (3,1,0) model emerged as the most statistically sound candidate based on AIC/BIC and residual diagnostics, several limitations remain:

- **Non-significant AR (1) term**: One of the autoregressive coefficients (AR1) had a p-value above the 0.05 threshold ($p \approx 0.1273$). While not statistically significant, its inclusion marginally improved the model's information criteria, suggesting possible redundancy—an issue discussed in Cryer & Chan, Chapter 8, in the context of parameter overfitting and identifiability.

- **Lack of seasonal terms**: The model does not include seasonal components. Although our diagnostics did not suggest strong residual seasonality, visual decomposition and subseries plots showed subtle cyclical fluctuations. A seasonal ARIMA (SARIMA) model might have improved interpretability or captured potential latent seasonality more explicitly.

- **Outlier sensitivity**: Although residuals passed normality and whiteness tests, known outliers (e.g., January 2001 and November 2006) may still bias inference or inflate error metrics. Cryer & Chan emphasize the importance of addressing such influential observations to prevent distortion of parameter estimates and forecast intervals.

- **Temporal limitations**: The dataset only extends to 2019, meaning the model does not capture structural shifts related to the COVID-19 pandemic or post-2020 labor market disruptions. The model may need updating to maintain relevance in future forecasting scenarios.

Despite these issues, all residual diagnostics support the adequacy of the ARIMA (3,1,0) model. There was no indication of autocorrelation or ARCH effects, therefore, ARIMA (3,1,0) remains a reliable and defensible choice for forecasting wages.

5 Forecasting

Forecasting is a critical aspect of time series analysis, particularly when it is intended to support policy-making decisions, investment policy, or economic planning. Through the use of past trends in data, forecasting methods offer visibility into future actions, allowing decision-makers to make informed choices. In wage forecasting, correct projections can affect labor deals, government planning, and cost-of-living estimates.

This section explains the technique used to generate forecasts and compare model precision. Specifically, compare ARIMA (1,1,2) and ARIMA (3,1,0) model forecasting performance using rolling forecast analysis, loss functions, and the Diebold-Mariano test.

5.1 Forecasting Methodology

First, the log-transformed wage series was divided into a training set (January 2001–December 2010) and a test set (January 2011–December 2019). To assess model accuracy, rolling forecast evaluation was employed, which mimics real-time prediction by continually updating the model as new data becomes available. The steps in a 1-step ahead rolling forecast are:

1. Fit the model to the first m observations.
2. Forecast the next value.
3. Compare the forecast to the actual value.
4. Slide the training window forward by one observation, removing the first observation to maintain same width.

This iterative process helps evaluate model stability and adaptability across different time periods. It is more realistic than static training/testing splits because it reflects how models behave in real-world deployment.

5.2 Forecast Accuracy Metrics

The following error metrics for both models were computed using the residuals from the rolling forecasts:

- **MSFE (Mean Squared Forecast Error)**: This measures the average of the squared differences between the predicted values and the actual values.
- **MAFE (Mean Absolute Forecast Error)**: Measures average absolute difference between the predicted values and the actual values. It is less sensitive to outliers than MSFE.

The results from the rolling forecast error analysis are as shown in Table 6 below:

Model	MSFE	MAFE
ARIMA (1,1,2)	4.109367e-06	0.00138189
ARIMA (3,1,0)	3.651826e-06	0.001420892

Table 5: Forecast Errors

5.3 Diebold-Mariano Test

The Diebold-Mariano (DM) test was used to statistically compare forecast performance, which evaluates whether two forecasting models have significantly different predictive accuracy.

- Null Hypothesis (H_0): Both models have equal predictive performance.
- Alternative Hypothesis (H_1): One model performs significantly better.

The DM test yielded a p-value of 0.5487, hence, we fail to reject the null hypothesis and conclude that there is no statistically significant difference in forecasting performance between ARIMA (1,1,2) and ARIMA (3,1,0).

Given this result and the slightly simpler structure of ARIMA (3,1,0), it was chosen as the final model for forecasting.

5.4 Impact of Forecasting

Reliable forecasts of wage trends empower decision-makers across public and private sectors. For instance:

- Policymakers can anticipate inflation-adjusted income changes.
- Businesses may align hiring and compensation strategies.
- Analysts and economists can better assess labor market health.

Thus, time series forecasting not only provides data-driven foresight but also enhances planning accuracy and economic preparedness.

The forecasting exercise shows that ARIMA (3,1,0) and ARIMA (1,1,2) deliver virtually identical predictive performance. Over the 108-month rolling window, their MSFE and MAFE all lie within a hair's breadth of one another. The Diebold–Mariano test confirms this: with a p-value of 0.5487 ($\gg 0.05$), we cannot reject the null of equal forecast accuracy. In the rolling-forecast plots both models track the test-period log-wage path about as closely; neither enjoys a clear edge in capturing the small up-and-down deviations in 2011–2019.

Although ARIMA (1,1,2) has one extra moving-average term, that extra flexibility does not translate into materially better out-of-sample fit. By contrast, ARIMA (3,1,0) is more parsimonious (fewer parameters) and makes no unneeded assumptions about moving-average dynamics. Moreover, Canadian wages exhibit no obvious seasonality, so the simpler ARIMA (3,1,0) structure is fully adequate to model the persistent upward drift in log-wages without over-fitting.

Finally, when we project 24 months ahead with ARIMA (3,1,0), and the forecast continue the long-term trend: Canadian wages are expected to rise steadily through the next two years as shown in Figure 5 of Appendix A. This stability reflects the dominant role of macroeconomic growth, labor-market tightness and inflation in driving wage levels factors that do not fluctuate seasonally. The combination of indistinguishable forecast accuracy and greater simplicity makes ARIMA (3,1,0) the clear choice for short-term wage projection in this study.

6 Conclusion and Discussion

6.1 Study Overview and Key Finding

This study analyzed Canadian wage dynamics from 2001 through 2019, fitting and comparing two non-seasonal ARIMA models ARIMA (1,1,2) and ARIMA (3,1,0)—on the log-transformed series. Both specifications captured the clear upward trend and absence of strong seasonal patterns. Although ARIMA (1,1,2) exhibited marginally lower AIC/BIC values, its additional moving-average term offered no statistically significant gain in forecast accuracy over ARIMA (3,1,0), as evidenced by a Diebold–Mariano p-value of 0.5487. Coupled with superior parsimony and well-behaved residual diagnostics, ARIMA (3,1,0) emerged as the preferred model. Its one-step-ahead rolling forecasts and back-transformed projections onto the original wage scale underscore a continuation of steady wage growth driven by economic fundamentals.

6.2 Strengths of the ARIMA (3,1,0) Approach

Selecting ARIMA (3,1,0) balanced model simplicity with robust performance. By limiting the specification to three autoregressive lags and a single difference, avoiding overfitting while preserving the essential trend structure. Rigorous diagnostic tests confirmed that the standardized residuals were uncorrelated, homoscedastic, and approximately Gaussian. The rolling-forecast framework delivered realistic, out-of-sample validation, demonstrating that ARIMA (3,1,0) remains stable and reliable across shifting sample windows. Moreover, back-transformation with bias correction-maintained interpretability in the original wage units.

6.3 Limitations and Areas for Future Research

Despite its robustness, the ARIMA (3,1,0) model has some weaknesses. First, slight departures from normality of the residual distribution indicate possible unmodeled nonlinear influences or infrequent external shocks. Second, the linear-Gaussian setup can possibly not cover structural breaks resulting from extreme economic events like the post-2019 pandemic period because our dataset ends at December 2019. Third, the model excludes exogenous variables (e.g., inflation, unemployment) that might provide improved explanatory power and forecast accuracy. Future work could extend this analysis by exploring regime-switching or threshold autoregressive models, integrating relevant macroeconomic predictors, and applying formal structural-break tests to bolster resilience under changing economic conditions.

6.4 Practical Implications

The ARIMA (3,1,0) forecasts offer an easy tool for stakeholders. Policymakers can use projected wage trajectories to inform social-program budgets and labor-market policy initiatives. Employers and compensation committees can align salary planning with the anticipated wage trajectory, and financial analysts can incorporate the forecasts into more broad macroeconomic models. The model's parsimony ensures ease of updating as new data emerge, facilitating timely decision support without sacrificing accuracy.

6.5 Conclusion

In conclusion, the ARIMA (3,1,0) model provides a simple, reliable framework to predict Canadian wages with no apparent seasonality. Its in-sample fit, diagnostic soundness, and comparable out-of-sample performance compared to higher-order counterparts render it a suitable model to employ. Regular re-estimation particularly using data subsequent to 2019 and ongoing diagnostic tests will be required to maintain prediction accuracy and adapt to structural shifts in the future. Overall, this study brings a robust, evidence-based approach to wage forecasting, enabling informed planning for public policy as well as business decision-making.