Business Forecasting Project

Meng Meng

**Forecasting Unemployment Trends in Alaska: A Comprehensive Analysis**

1. **Introduction and Forecasting Question:**

This project focuses on forecasting unemployment trends in Alaska, a critical economic indicator that carries significant implications for policy making and resource allocation. Understanding these trends is vital for several reasons: it enables better economic planning at both state and local levels, facilitates more effective resource allocation, helps identify potential periods of economic distress, and allows for better preparation of social services. The ability to accurately forecast unemployment can provide valuable insights for businesses, government agencies, and social service organizations in their strategic planning processes.

1. **Data Description:**

The analysis utilizes a comprehensive dataset of unemployment statistics covering all U.S. states, with specific focus on Alaska. The dataset spans from 1976 to 2022, providing monthly observations of key employment metrics. These metrics include Total Civilian Non-Institutional Population, Total Civilian Labor Force, Employment and Unemployment numbers, and various percentage metrics such as labor force participation, employment rate, and unemployment rate. This longitudinal dataset offers a rich source of information for analyzing employment trends and patterns over nearly five decades, capturing various economic cycles and structural changes in Alaska's economy.

1. **Exploratory Data Analysis Insights:**

The exploratory analysis revealed several significant patterns and relationships within the data. A strong correlation (0.91) exists between Total Civilian Non-Institutional Population and Total Unemployment, suggesting a close relationship between population dynamics and unemployment levels. The data exhibits notable structural breaks, particularly during the 2021-2022 period coinciding with the COVID-19 pandemic. Seasonal patterns are evident in the unemployment data, reflecting Alaska's unique seasonal economic characteristics. The relationship between population and unemployment shows a wave-like pattern, indicating complex underlying dynamics that simple linear models might not fully capture. Additionally, the presence of outliers, especially in higher population ranges, increases the difficulty of model selection and training.

1. **Accuracy Measure Selection:**

The analysis employed multiple accuracy metrics to provide a comprehensive evaluation of forecasting performance. Mean Square Error (MSE) served as our primary metric, chosen for its sensitivity to large errors and its usefulness in comparing different models. Other key metrics included Mean Error (ME), Root Mean Square Error (RMSE), Mean Absolute Error (MAE), Mean Percentage Error (MPE), and Mean Absolute Percentage Error (MAPE). For Alaska's unemployment data, the training set showed an ME of 3.67, indicating a slight positive bias where models typically overestimate by about 4 units. The RMSE of 938.34 and MAE of 263.19 suggest that while most errors are moderate, some large errors are driving up the RMSE. The MAPE of 1.13% indicates relatively good percentage accuracy in the forecasts. The ACF1 (first-order autocorrelation of residuals) value of -0.0107 suggests minimal remaining serial correlation in the residuals, indicating good model fit.

1. **Forecasting Methods and Residual Analysis:**

The analysis implemented and evaluated multiple forecasting approaches to capture different aspects of unemployment patterns. The automatic ARIMA selection identified an ARIMA(0,1,4) model as optimal, with significant MA components ranging from 0.2203 to -0.5043. This model's residual analysis through the Ljung-Box test yielded a p-value of 0.9832, strongly indicating independently distributed residuals. Moving Average models, particularly MA(12), demonstrated effectiveness in smoothing short-term fluctuations and showed robust performance across different variables. The Holt-Winters method incorporated both trend and seasonal components, proving particularly effective with percentage-based metrics. Naive methods, including Seasonal Naive and Random Walk with Drift, served as important baseline comparisons and showed surprising effectiveness for labor force predictions. Exponential Smoothing demonstrated good adaptive capability to changing trends and balanced historical data with recent observations. The residual analysis revealed that most models handled the pre-2020 data well but struggled with the structural break during the COVID-19 period, indicating the need for separate modeling approaches for different time periods. The ACF plots and Ljung-Box tests confirmed that the selected models effectively captured most of the data's systematic patterns, with minimal remaining autocorrelation in the residuals.

1. **Prediction Results and Accuracy**

The comparative analysis of different forecasting methods revealed varying levels of success across different metrics. The Moving Average (MA5) model emerged as the strongest performer for several key metrics, achieving the lowest MSE for Total Employment (2,246,929), Employment Rate (0.246), and Total Unemployment (10,152,643). The Holt-Winters method showed particular strength in forecasting Population Percentage metrics with an MSE of 0.030, while the Naive method proved most effective for Labor Force predictions with an MSE of 491,136. These results highlight the importance of selecting appropriate models for specific aspects of unemployment forecasting.

The comparison of accuracy measures across models showed that while simpler models like MA5 often performed best for specific metrics, more complex models like ARIMA and Holt-Winters provided better overall stability and reliability in forecasts. The residual analysis revealed that most models handled the pre-2020 data well, but struggled with the structural break during the COVID-19 period, suggesting the need for separate modeling approaches for different time periods. The ACF plots and Ljung-Box tests confirmed that the selected models effectively captured most of the data's systematic patterns, with minimal remaining autocorrelation in the residuals.

1. **Analysis-Based Decision:**

Based on the comprehensive analysis of various forecasting methods and their performance metrics, we can predict that Alaska's unemployment trends will likely show moderate increases in the short term, with forecasted values indicating a 5-period ahead prediction range of approximately 20,000-22,000 unemployed individuals. The Moving Average (MA5) model emerges as the most reliable overall forecasting method for this dataset, particularly for short-term predictions, with an MSE of 10,152,643 for Total Unemployment. The analysis suggests that different variables benefit from different modeling approaches - with the Holt-Winters method excelling at percentage-based predictions and Naive methods performing well for labor force forecasts - indicating that a hybrid approach might be optimal for comprehensive unemployment forecasting. The presence of structural breaks, particularly during the COVID-19 period, necessitates careful consideration in model application and interpretation. Looking at the confidence intervals from our ARIMA(0,1,4) model, we can be 95% confident that unemployment levels will remain within ±938 individuals of our point forecasts. The predicted trend suggests a gradual stabilization of unemployment rates, though with notable seasonal fluctuations typical of Alaska's economy. However, these predictions should be interpreted with caution given the recent structural breaks in the data and the potential for external economic factors to influence unemployment patterns.

1. **Recommendations for Improvement:**

Several opportunities exist for enhancing the forecasting accuracy and reliability. First, incorporating external economic indicators such as GDP, inflation rates, and regional economic factors specific to Alaska could provide additional context and improve prediction accuracy. Developing separate models for pre- and post-COVID periods could better account for structural changes in the economy. The implementation of ensemble methods combining multiple forecasting approaches could leverage the strengths of different models. Additionally, including intervention analysis for significant events and developing seasonal adjustments specific to Alaska's unique economic patterns could further refine the forecasts. The use of cross-validation techniques would also contribute to more robust model selection and validation.