UNEMPLOYMENT CLAIMS FORECASTING USING TIME SERIES ANALYSIS IN R: A STATE LABOR DEPARTMENT CASE STUDY

1. Background

In the wake of post-pandemic labor market volatility, a U.S. state labor department sought to forecast unemployment insurance (UI) claims with higher accuracy. Their existing spreadsheet-based method lacked robustness during seasonal fluctuations and economic shocks. We were brought in to develop a scalable, accurate time series forecasting model using R, allowing the state to better manage fund allocation and staff capacity.

2. Objective

- To forecast monthly UI claims for the next 12 months across key counties
- To identify trend shifts, cyclical spikes, and anomalies in historical claims data
- To generate clear, validated reports for budget planning and federal reporting compliance

3. Data Used

Sources:

- State Labor Department's administrative UI claims records
- U.S. Bureau of Labor Statistics for benchmark comparison

Structure:

- Monthly time series data from Jan 2015 to Dec 2023
- 5 regions × 108 months = 540 series records

Key Variables:

- Month, Region, Unemployment_Claims, Labor_Force_Size, Industry_Employment, Federal_Stimulus_Indicator

4. Methodology

4.1 Data Preparation

- Checked time series stationarity using adf.test() from the tseries package

- Filled missing months using linear interpolation and confirmed consistency with tsclean()
- Created time series objects per region using ts() and tsibble

4.2 Exploratory Analysis

- Used ggplot2 and feasts::STL() for decomposition
- Identified seasonal peaks in January (post-holiday layoffs) and June (construction sector dips)
- Cross-referenced economic policy dates with spikes in claims

4.3 Model Selection and Forecasting

- Compared **ETS**, **ARIMA**, and **TBATS** models
- Applied auto.arima() for stable regions, ets() for volatile regions
- Used fabletools and forecast packages for cross-validation and simulation

```
fit_ets <- ets(region_ts)

forecast_ets <- forecast(fit_ets, h = 12)
```

4.4 Validation

- Used 2023 as test window for forecasting validation
- Evaluation metrics: RMSE, MAPE, MAE
- Selected ETS(A,A,N) and ARIMA(0,1,1) for most regions based on residual performance

5. Results

Region	**Best Model**	**MAPE (%)**	**RMSE**
Central	ETS(A,A,N)	4.3%	1,240
North	ARIMA(0,1,1)	5.1%	1,390
South-East	TBATS	6.8%	1,710
Mountain	ARIMA(1,1,0)	5.4%	980
Coastal	ETS(A,A,A)	3.7%	860

statssy

6. Interpretation and Recommendations

- Seasonal spikes corresponded to **industry-specific layoff cycles**; added industry dummy variables improved model precision
- Coastal region saw **sharp but predictable tourism-linked fluctuations**, ideal for ETS with seasonality
- Recommended monthly recalibration of models with updated labor force and policy indicators
- Suggested integrating forecast outputs into **internal fund planning templates** and **federal quarterly reports**

7. Reporting Output

- **R Markdown Report (PDF, 34 pages)**:
  - Region-wise claim forecast charts with confidence intervals
  - Overlay of economic stimulus markers
  - Model residual diagnostics and performance comparison
- **Excel Summary Workbook**:
  - Forecast tables for 12 months with upper/lower bounds
  - Monthly change and year-over-year comparisons
  - Regional claim ratio per 1,000 labor force for decision making
- **R Script Repository**:
  - Regional forecast generator with fable and forecast integration
  - Custom function: generate_ui_claim_forecast(region_code, months=12)
  - Batch reporting export function via rmarkdown::render()

8. Government Impact

- Used in **Q1 budget submission** to federal UI authorities for matching grants
- Helped the agency anticipate **June construction layoffs** with +3-week lead
- Enabled leadership to justify **staff reallocation during seasonal surges**
- Framework was extended to **Medicaid enrollment prediction** in 2024