Bayesian Time Series Analysis of US Unemployment Rates

STAT 447C Final Project - Project Report

Rohan Joseph (67089839)

Due Date: April 19th 2024

Key Details

- **Student(s):** Rohan Joseph (Student No.: 67089839)
- **Professor:** Alexandre Bouchard-Côté
- **Course:** STAT 447C - Bayesian Statistics
- **GitHub Repository Link:** https://github.com/RohanUBC/STAT-447C-Final-Project

Abstract

- Brief summary of objectives, methods, key results, focus on uncertainty quantification and calibration.

[To be added]

Introduction

- *Introduce the problem:* Forecasting US unemployment rates using Bayesian time series analysis.
- Provide context and the importance of understanding and predicting unemployment rates.
- The importance of forecasting and uncertainty quantification.
- *Context:* Briefly explain the significance of unemployment rates in economic planning and policy-making.

[To be added]

Problem Formulation

The central focus of this project is to model and forecast unemployment rates in the United States (US), utilizing a real-world dataset that has documented the monthly unemployment rate (from 01/1948 to 02/2024) in the US. A key element of the project is to accurately forecast these rates into the future while also quantifying the inherent uncertainty of such forecasts. The dataset that is being used is from the US Federal Reserve on monthly US unemployment rates.

Federal Reserve Economic Data (FRED): This dataset offer historical data on monthly US unemployment rates (from 1948 to 2024) and their associated economic indicators.

URL: https://fred.stlouisfed.org/series/UNRATE

We will adopt a Bayesian time series analysis approach, utilizing a Bayesian Structural Time Series (BSTS) model to capture the dynamics of the US labor market based on the provided data. This involves uncovering and modeling historical trends and seasonal patterns within the data, providing probabilistic forecasts of future unemployment rates. We opt for BSTS due to its flexibility in modeling complex time series data with trend and seasonality components. BSTS allows for the estimation of latent states, such as trend and seasonal effects, and provides a framework for forecasting future states while quantifying uncertainty. Please see Appendix C on a general overview on the BSTS model.

Key Modelling/Methodological Challenge

This project faces challenges in applying Bayesian time series analysis with BSTS to economic data, particularly in modeling the seasonal patterns and long-term trends inherent in unemployment rates. The primary hurdle lies in selecting and fine-tuning parameters and models that can accurately reflect the dynamics of the US labor market. This includes determining appropriate prior distributions, specifying the structural components of the model, and validating the model's performance through diagnostics and posterior predictive checks. Additionally, handling the irregularity in seasonal variation presents a challenge, requiring careful consideration in the modeling approach to ensure accurate representation of the data.

In addition, we aim to also include a calibration of the uncertainty of the forecast, demanding a sophisticated approach in how the model is constructed and validated. This is done to ensure that the forecasts are not only precise, but also accompanied by reliable uncertainty measures. Such measures are pivotal for economic planning and policy-making, providing a foundation for risk-aware decision-making. By accurately quantifying forecast uncertainty, this approach aims to assess potential risks effectively, thereby facilitating more informed and prudent economic decision making.

Literature Review

- Summarize key findings from existing literature on Bayesian forecasting and unemployment rate analysis.
- Highlight gaps your project aims to fill or how it builds upon previous work.
- Discussion of uncertainty quantification in economic forecasting.
- Calibration of probabilistic forecasts in existing literature.

[To be added]

Data Analysis

- *Model:* A Bayesian model is precisely described (e.g., using the ... ~ ... notation).

  – Description of BSTS for modeling unemployment rates.

- Implementation code in the appendix (e.g., using Stan)

  – R code for implementing the BSTS using Bayesian methods.

- *Motivation of prior choice:* If appropriate, several choices are compared or sensitivity analysis is performed.
- *Critical evaluation of the posterior approximation:* An appropriate combination of diagnostics, synthetic datasets and other validation strategies.

  – Techniques used for uncertainty quantification and calibration of forecasts.
  – Model diagnostics and validation strategies.

- *Methodological/Theoretical aspect:*

  – Assessment of the approach's robustness and creativity.

– Discussion on the choice of BSTS and Bayesian calibration methods.
- *Calibration of uncertainty:*
  – Explanation of calibration tests for predictive intervals.
  – Presentation of calibration test results.
  – Analysis of how well the model's uncertainty measures align with observed data.

[To be added]

BSTS Model for US Unemployment Rates

BSTS Model Specification

Observation Model: The observed unemployment rate at time t, y_t, is modeled as:

$$y_t = \mu_t + \tau_t + \epsilon_t$$

where:

- μ_t represents the trend component.
- τ_t denotes the seasonal component.
- ϵ_t is the observation error.

The observation error ϵ_t follows a Normal distribution with mean 0 and variance σ_y^2:

$$\epsilon_t \sim N(0, \sigma_y^2)$$

Local Level (Trend) Component: The trend component μ_t evolves over time as:

$$\mu_{t+1} = \mu_t + \eta_{\mu,t}$$

where $\eta_{\mu,t}$ represents the trend shock at time t.

$\eta_{\mu,t}$ follows a Normal distribution:

$$\eta_{\mu,t} \sim N(0, \sigma_\mu^2)$$

Seasonal Component: The seasonal component τ_t adjusts for seasonality, following a similar evolution:

$$\tau_{t+1} = -\sum_{s=1}^{S-1} \tau_{t-s} + \eta_{\tau,t}$$

where:

- S is the seasonal period (in this case $S = 12$ for monthly.
- $\eta_{\tau,t}$ represents the seasonal shock.

$\eta_{\tau,t}$ follows a Normal distribution:

$$\eta_{\tau,t} \sim N(0, \sigma_\tau^2)$$

Hyperparameters and Hyperpriors:

Observation Noise Variance σ_y^2:

- Hyperprior: $\sigma_y^2 \sim \text{Inverse-Gamma}(\alpha_y, \beta_y)$
- Where, $\alpha_y = 2$ and $\beta_y = 0.5228194$, based on the ARIMA residuals.

Trend Variance σ_μ^2:

- Hyperprior: $\sigma_\mu^2 \sim \text{Inverse-Gamma}(\alpha_\mu, \beta_\mu)$
- Where, $\alpha_\mu = 2$ and $\beta_\mu = 7.724585$, derived from the decomposition of the trend component.

Seasonal Variance σ_τ^2:

- Hyperprior: $\sigma_\tau^2 \sim \text{Inverse-Gamma}(\alpha_\tau, \beta_\tau)$
- Where, $\alpha_\tau = 2$ and $\beta_\tau = 0.007432969$, derived from the decomposition of the seasonal component.

Discussion

- *Project Theme:* Assess the soundness and creativity of the approach within the context of Bayesian forecasting and time series analysis. Discuss the model's ability to handle the intricacies of unemployment data and its forecasting accuracy.
- *Summarizing results:* Summarize key findings and their implications for understanding and forecasting US unemployment rates.

    - Interpretation of the forecasting results.

- Implications of calibrated uncertainty for economic decision-making.
- Discuss the limitations of your study, such as data constraints, model assumptions, or potential biases.
- Suggest areas for future research or methodological improvements.

[To be added]

Conclusion

- Recap of the project's contributions to Bayesian forecasting and calibration.
- Reflect on the value of Bayesian methods in economic time series forecasting.
- Final thoughts on policy implications and the value of calibrated forecasting.

[To be added]

References

[To be added]

Appendix

Appendix A - Preliminary Data Analysis

**Reading in the Dataset**

```
dat <- read_csv("UNRATE.csv")
dat$DATE <- as.Date(dat$DATE)
head(dat)
```

```
## # A tibble: 6 x 2
##   DATE       UNRATE
##   <date>      <dbl>
## 1 1948-01-01    3.4
## 2 1948-02-01    3.8
## 3 1948-03-01    4
## 4 1948-04-01    3.9
## 5 1948-05-01    3.5
## 6 1948-06-01    3.6
```

Historical US Unemployment Rates from 01/1948 to 02/2024

```
plot(dat$DATE, dat$UNRATE,
     type = "o",
     pch = 20,
     col = "red",
     xlab = "Date",
     ylab = "Unemployment Rate (in %)",
     main = "Figure 1: Historical US Unemployment Rate Over Time")
lines(dat$DATE, dat$UNRATE, col = "black")
```


**Figure 1: Historical US Unemployment Rate Over Time**
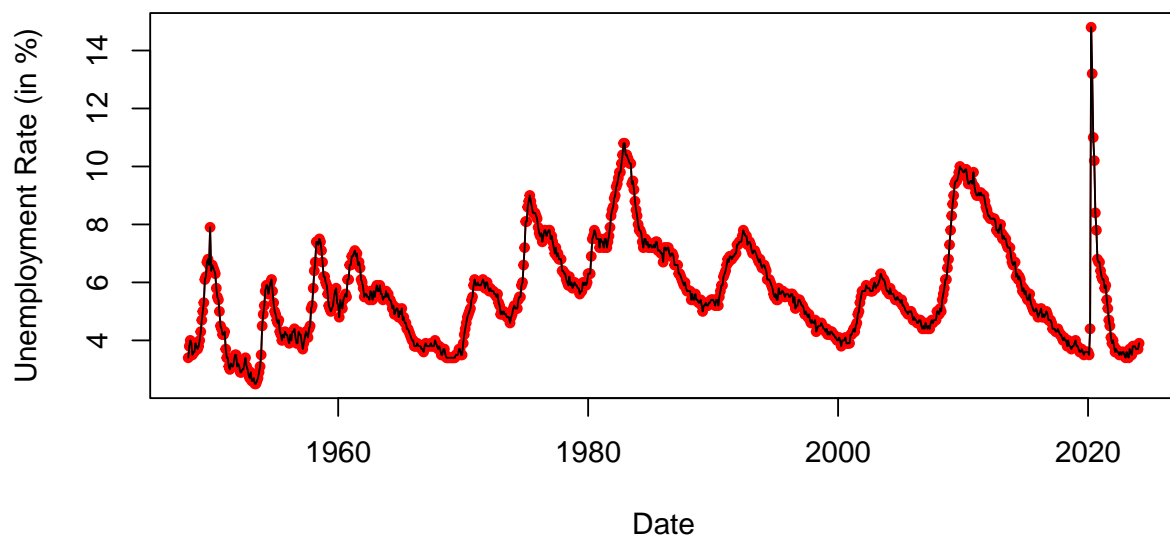
Average Yearly US Unemployment Rates from 01/1948 to 02/2024

```r
avg_yearly_unrate <- dat |>
  mutate(YEAR = format(DATE, "%Y")) |>
  group_by(YEAR) |>
  summarise(Average_Unemployment_Rate = mean(UNRATE, na.rm = TRUE))

avg_yearly_unrate$YEAR <- as.numeric(as.character(avg_yearly_unrate$YEAR))

plot(avg_yearly_unrate$YEAR, avg_yearly_unrate$Average_Unemployment_Rate,
     type = "o",
     pch = 20,
     col = "red",
     xlab = "Year",
     ylab = "Average Unemployment Rate (in %)",
     main = "Figure 2: Average Yearly Unemployment Rate in the US")
lines(avg_yearly_unrate$YEAR, avg_yearly_unrate$Average_Unemployment_Rate, col = "black")
```


**Figure 2: Average Yearly Unemployment Rate in the US**
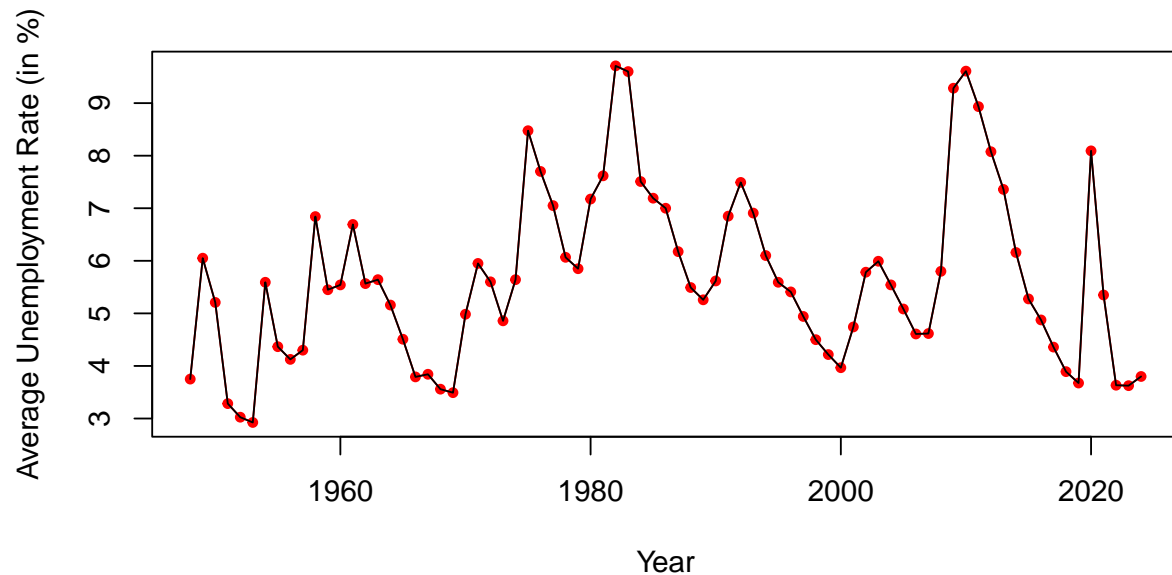
Appendix B - Time Series Analysis

Decomposition of the Time Series

```
ts_unrate <- ts(dat$UNRATE, start = c(1948, 1), frequency = 12)
decompose_of_ts = stl(ts_unrate, s.window = "periodic")
plot(decompose_of_ts,
     main = "Figure 3: Decomposition of the Time Series of US Unemployment Rates")
```

Figure 3: Decomposition of the Time Series of US Unemployment Rates
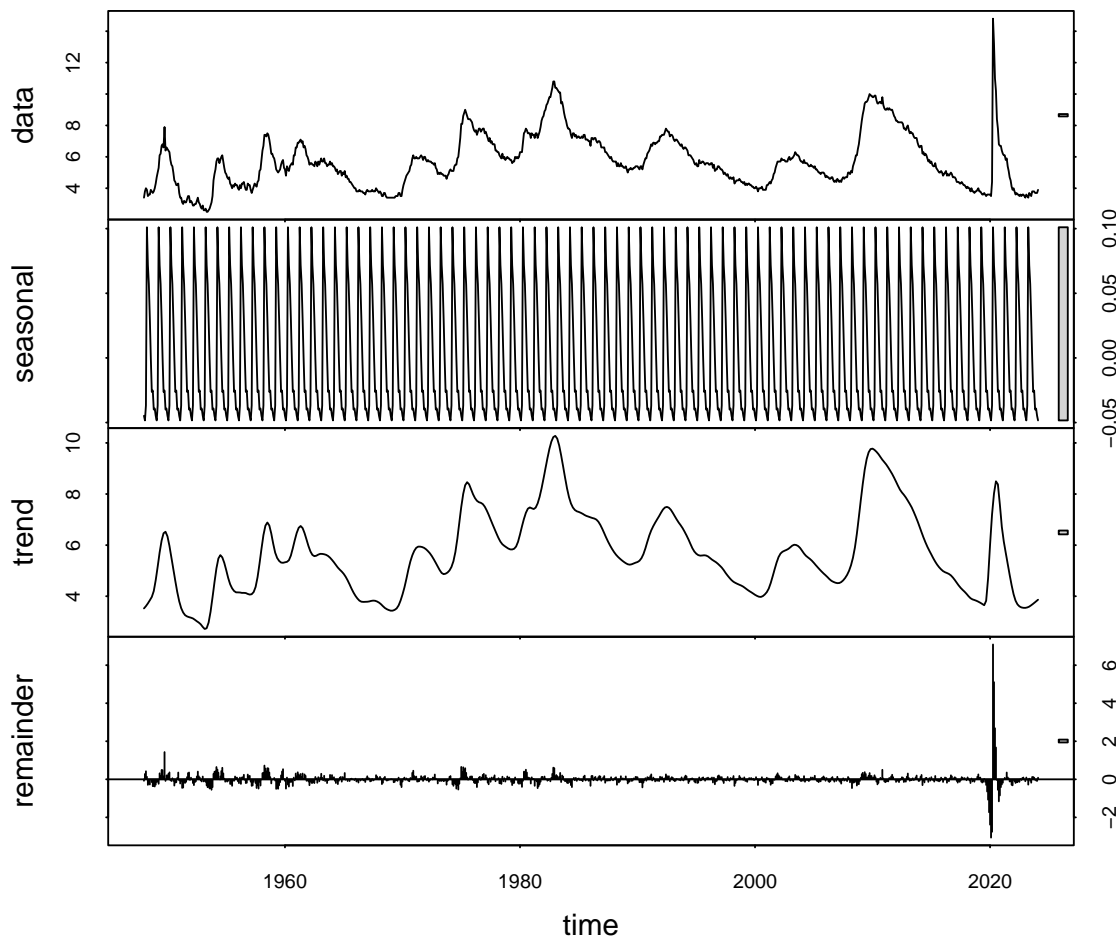
Trend: The trend plot in the decomposition suggests that there doesn't seem to be any particular long-term trend in the US unemployment rates, however, there are periods where the unemployment rate exhibits some trends, which seem to be non-linear and varying with time.

Seasonality: The seasonal plot in the decomposition suggests that there is seasonal variation in the US unemployment rates, and there is a consistent interval in the variation of the seasonality. The consistent interval in the seasonality suggests a stable seasonal effect that repeats annually.

Remainder/Residual: The remainder/residual plot of the decomposition appears to be relatively small compared to the seasonal and trend components, suggesting that the model has captured most of the systematic structure in the data.

Appendix C - General Overview on the Bayesian Structural Time Series (BSTS) Model

The Bayesian Structural Time Series (BSTS) model is a statistical method that serves several applications such as feature selection, time series forecasting, and causal inference analysis. It operates on time series data to ascertain underlying patterns and forecast future data points.

BSTS models are composed of three primary elements:

1. *Kalman Filter:* A method for decomposing time series into components like trend and seasonality, allowing state variables to be modeled dynamically over time.
2. *Spike-and-Slab Prior:* A technique for feature selection that identifies which predictors in a regression are most informative.
3. *Bayesian Model Averaging (BMA):* A process where multiple models are averaged together to produce predictions or infer parameters, accounting for model uncertainty. In the BSTS framework, BMA is utilized extensively to generate samples from repeated Markov Chain Monte Carlo (MCMC) simulations into final model outputs, providing a comprehensive prediction that encompasses model variability.

A general BSTS model consists of two set of equations,

1. *Observational equation:* Response variable as a function of predictors and/or latent variables.
2. *State equations:* How the parameters evolve over the time.

BSTS models are usually implemented using the `bsts` package in R, but can also be implemented using `rstan`. Stan's advanced MCMC algorithms enhance the flexibility and scalability of BSTS models, making it a powerful tool for time series analysis.

References

"Bayesian Structural Time Series." *SAP HANA Predictive Analysis Library (PAL)*, help.sap.com/docs/ SAP_HANA_PLATFORM/2cfbc5cf2bc14f028cfbe2a2bba60a50/b9972576368640da9831d73a9d749c3b.html. Accessed 8 Apr. 2024.

Radtke, Tim. "Minimize Regret." *Minimize Regret - Rediscovering Bayesian Structural Time Series*, minimizeregret.com/post/2020/06/07/rediscovering-bayesian-structural-time-series/. Accessed 9 Apr. 2024.

Yabe, Taka. "Pystan - Causal Inference Using Bayesian Structural Time Series." *Takahiro Yabe*, 21 Feb. 2021, www.takayabe.net/post/pystan-bsts. Accessed 9 Apr. 2024.

Appendix D - Estimation of Hyperparameters for BSTS Model Using the Empirical Bayes Model

Before we can specify the BSTS model for the US Unemployment rates, we need to estimate the hyperparameters that will inform the prior distributions. To do this, we will use an Empirical Bayes approach to estimate these hyper parameters from the data itself.

The Empirical Bayes method is used in scenarios where we have little prior information about the hyper parameters, and involves using the data to estimate values that will then inform our prior beliefs. This is particularly useful in predictive modeling, as it allows us to use data-driven estimates while still maintaining a Bayesian framework.

Defining the Hyperpriors for the Noise Variance

We fit an Auto-Regressive Integrated Moving Average (ARIMA) model to the unemployment rate time series to capture its temporal patterns. The ARIMA model's residuals provide an empirical basis for estimating the noise variance in the BSTS model.

```
arima_fit <- auto.arima(ts_unrate)
summary(arima_fit)
```

```
## Series: ts_unrate
## ARIMA(1,1,1)
##
## Coefficients:
##           ar1     ma1
##       -0.7763  0.8323
## s.e.   0.0973  0.0847
##
## sigma^2 = 0.1747:  log likelihood = -497.94
## AIC=1001.88   AICc=1001.9   BIC=1016.33
##
## Training set error measures:
##                        ME      RMSE       MAE        MPE    MAPE      MASE
## Training set 0.0005455781 0.4172323 0.1682336 -0.1161398 2.90668 0.1768403
##                    ACF1
## Training set -0.0126775
```

```
resid_var <- var(arima_fit$residuals)

# We use a relatively small 2 to remain less informative
alpha_y <- 2
beta_y <- (alpha_y + 1) * resid_var

sigma_sqr_y_prior <- list(shape = alpha_y, scale = beta_y)
sigma_sqr_y_prior
```

```
## $shape
## [1] 2
##
## $scale
## [1] 0.5228194
```

Defining the Hyperpriors for the Trend and Seasonal Components

We decompose the time series of the US unemployment rates data into its constituent element and evaluate the variances of these components.

```r
decomp <- stl(ts_unrate, s.window = "periodic")

trend_var <- var(decomp$time.series[, "trend"])
seasonal_var <- var(decomp$time.series[, "seasonal"])

# We use a relatively small 2 to remain less informative
alpha_mu <- 2    # For the trend component
alpha_tau <- 2  # For the seasonal component

beta_mu <- (alpha_mu + 1) * trend_var
beta_tau <- (alpha_tau + 1) * seasonal_var

sigma_sqr_mu_prior <- list(shape = alpha_mu, scale = beta_mu)
sigma_sqr_mu_prior
```

```
## $shape
## [1] 2
##
## $scale
## [1] 7.724585
```

```r
sigma_sqr_tau_prior <- list(shape = alpha_tau, scale = beta_tau)
sigma_sqr_tau_prior
```

```
## $shape
## [1] 2
##
## $scale
## [1] 0.007432969
```

Appendix [] - []

[To be added]