Time Series Analysis and Forecasting of Unemployment and Interest Rates

Kuanyi Chen

2025-02-18

```r
# import libraries
library(lattice)
library(ggplot2)
library(foreign)
library(MASS)
library(car)
require(stats)
require(stats4)
library(KernSmooth)
library(fastICA)
library(cluster)
library(leaps)
library(mgcv)
library(rpart)
library(pan)
library(mgcv)
library(DAAG)
library(TTR)
library(tis)
require("datasets")
require(graphics)
library(forecast)
library(xtable)
library(stats)
library(TSA)
library(timeSeries)
library(fBasics)
library(tseries)
library(timsac)
library(TTR)
library(fpp)
library(strucchange)
library(vars)
library(lmtest)
```

I. Introduction

```r
# import data
unemp_rate <- read.csv("unemployment_rate.csv")
int_rate <- read.csv("interest_rate.csv")
```

This project analyzes two key macroeconomic time series: the U.S. monthly unemployment rate and the U.S. 10-year Treasury interest rate. The unemployment rate reflects the percentage of the labor force that is actively seeking employment, serving as a critical indicator of economic health. The 10-year interest rate represents the yield an investor would receive if they held a 10-year government bond until maturity, often serving as a benchmark for long-term borrowing costs and investor sentiment regarding future economic conditions.

The dataset spans a 20-year period, from January 2005 to January 2025, capturing a range of economic cycles, including financial crises, recoveries, and expansionary periods. By fitting appropriate time series models, we aim to analyze the historical trends and forecast the future movements of these economic indicators.

Additionally, we seek to explore potential causal relationships between the two variables. Given that unemployment and interest rates are closely linked to economic conditions, such as inflation, monetary policy, and business cycles, it is possible that changes in one variable may systematically influence the other. Using statistical methods such as Granger causality tests or vector autoregressive (VAR) models, we aim to investigate whether fluctuations in interest rates have predictive power over unemployment rates or vice versa.

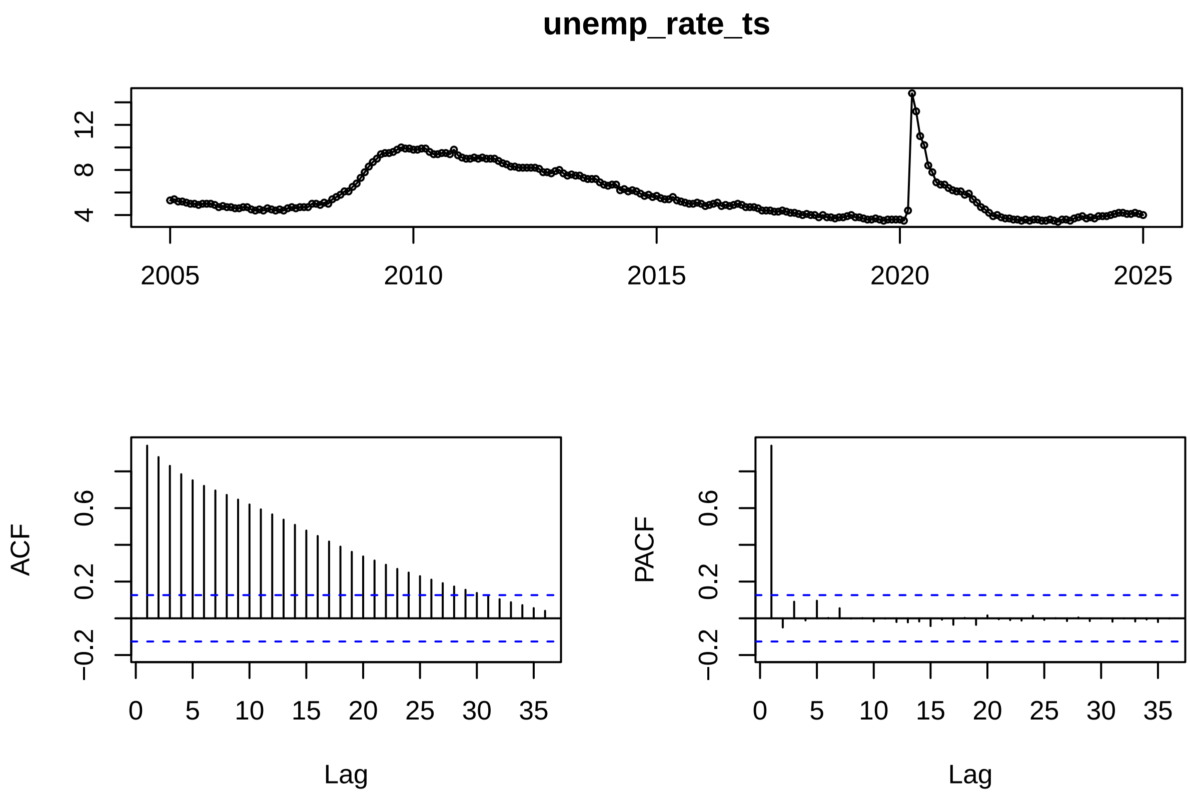
This analysis will provide valuable insights into the dynamics between these two fundamental economic indicators, offering implications for policymakers, investors, and economists interested in understanding macroeconomic trends and forecasting future economic conditions.

II. Results

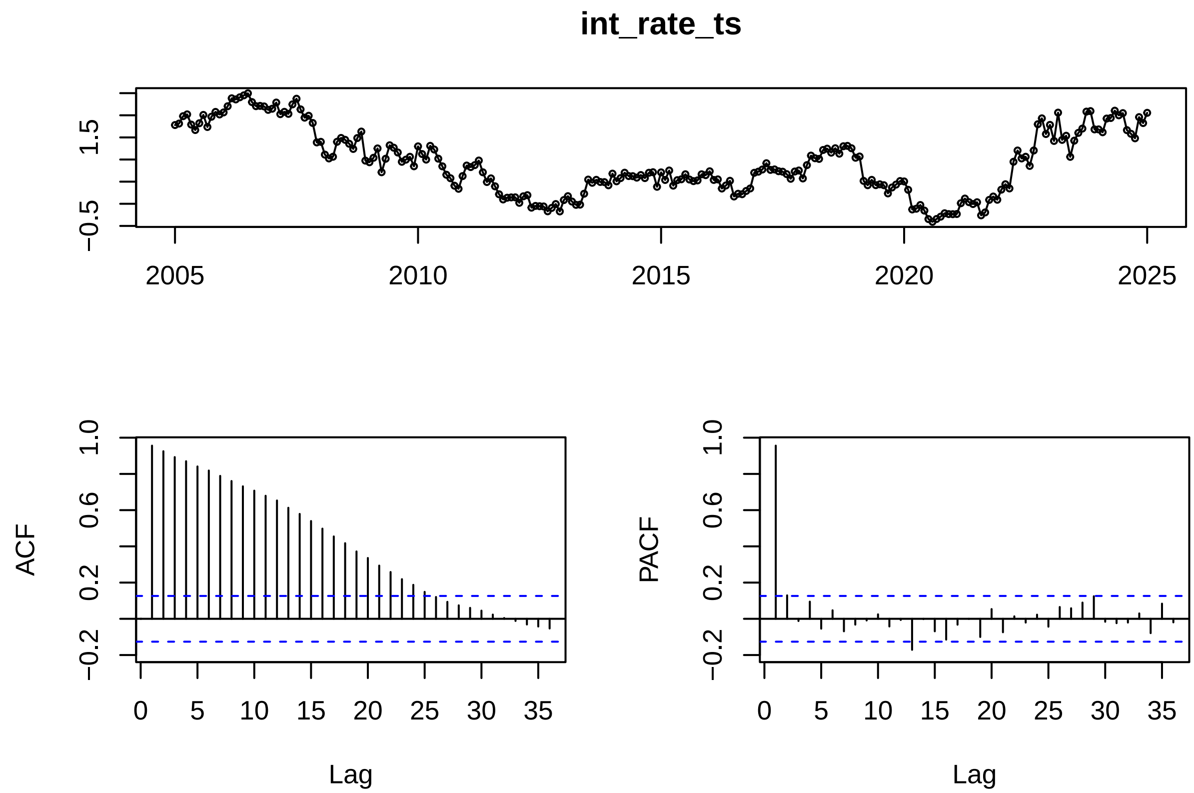
(a)

```r
# turn data into time series object
unemp_rate_ts <- ts(unemp_rate$UNRATE, start = c(2005, 1), frequency = 12)
int_rate_ts <- ts(int_rate$REAINTRATREARAT10Y, start = c(2005, 1), frequency = 12)

# plot the time series
tsdisplay(unemp_rate_ts)
```

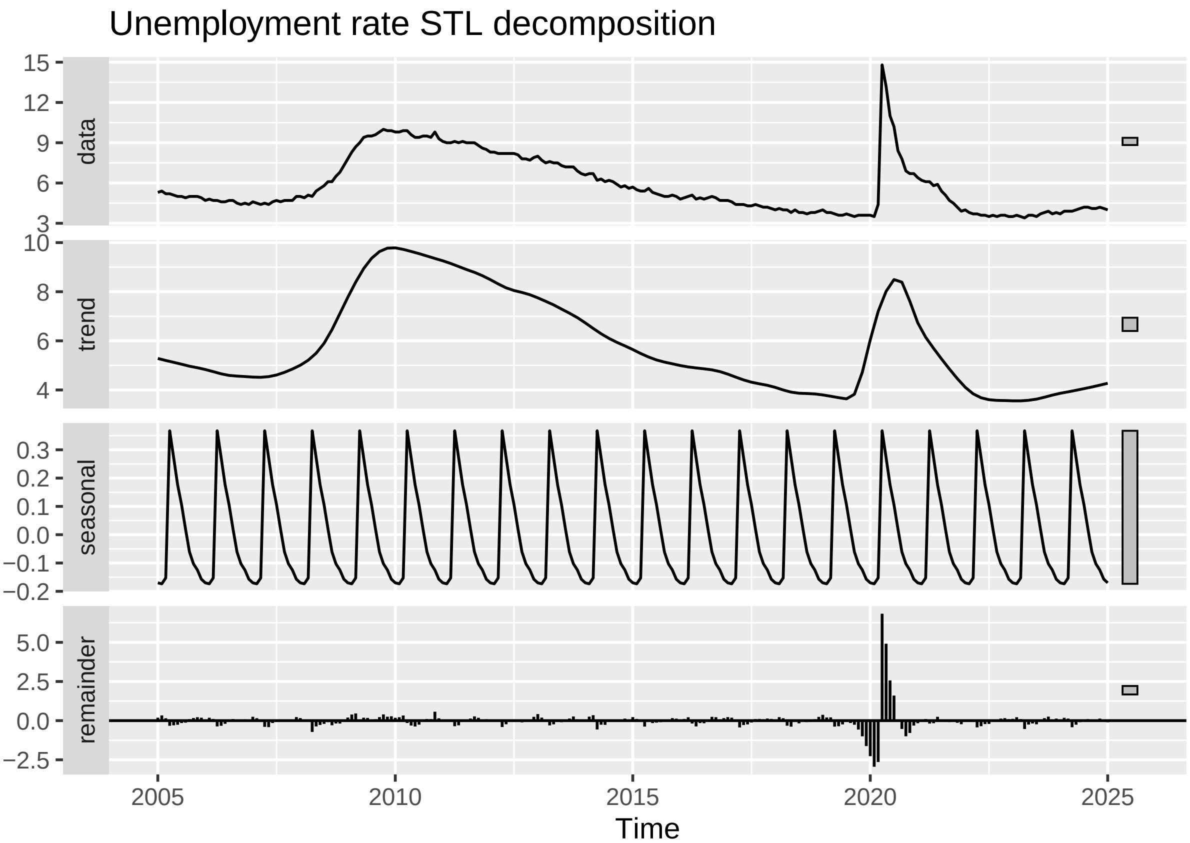



```
tsdisplay(int_rate_ts)
```

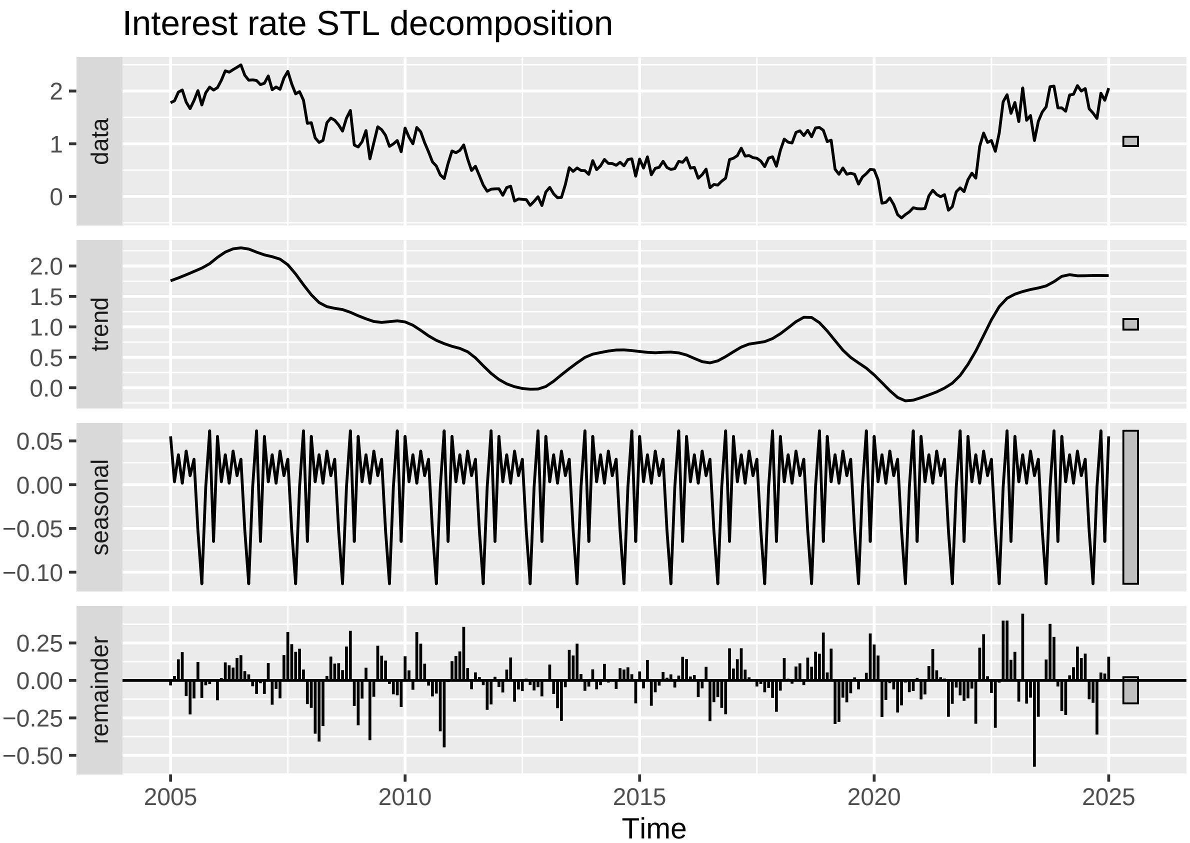


(b)

```r
# plot the stl decompositions
autoplot(stl(unemp_rate_ts, s.window = "periodic")) +
  ggtitle("Unemployment rate STL decomposition")
```




```r
autoplot(stl(int_rate_ts, s.window = "periodic")) +
  ggtitle("Interest rate STL decomposition")
```



Unemployment rate: There seems to be a stochastic trend, as there is not a consistent positive or negative trend. There is a sudden upward spike at 2020, which is during the COVID-19 pandemic. There is a regular seasonal pattern throughout. There is very minimal random effects according to the remainder, except the sudden disruption around the year 2020, as shown by the significant spike.

Interest rate: There also seems to be a stochastic trend, as there is not a consistent positive or negative trend. There is a reccurring seasonal pattern, with more significant impacts certain times during the year, as shown by the more significant spike. There seems to be regular recurring random effects according to the remainder, as shown by the recurring patterns of the spikes.

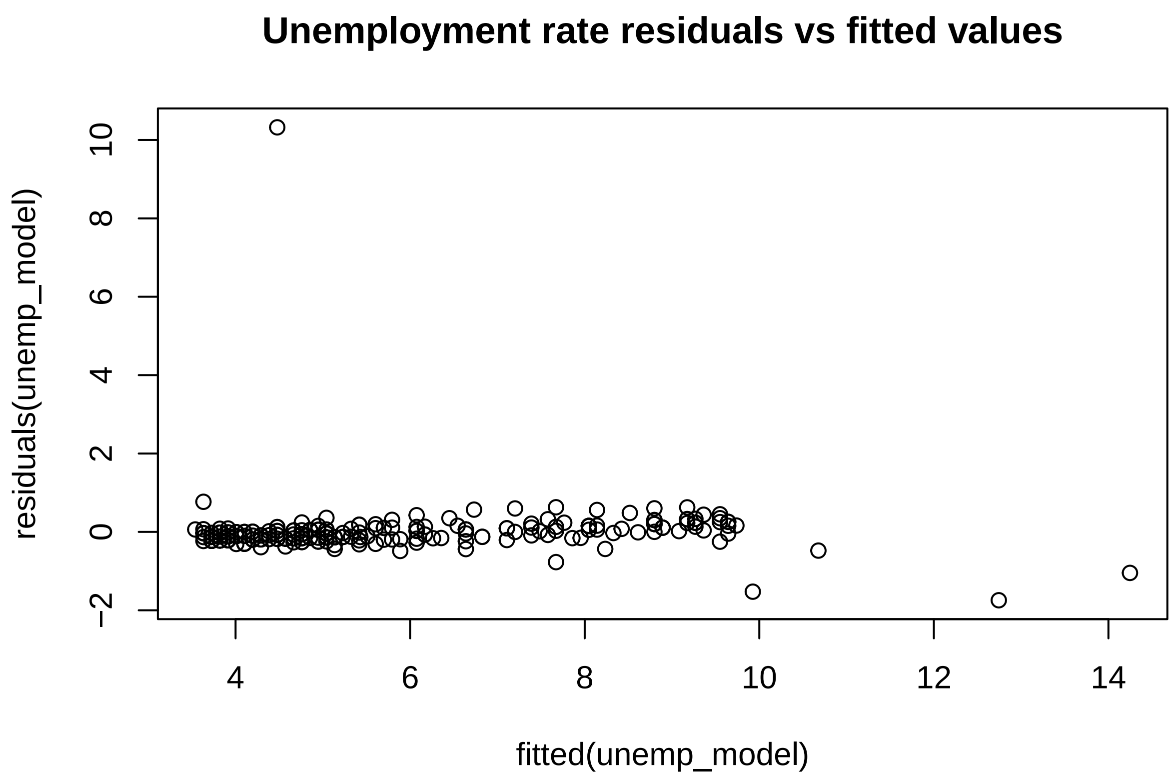
(c)

```r
# model the time series
unemp_model <- Arima(unemp_rate_ts, order = c(1, 0, 0))
int_rate_model <- Arima(int_rate_ts, order = c(1, 0, 0))
```

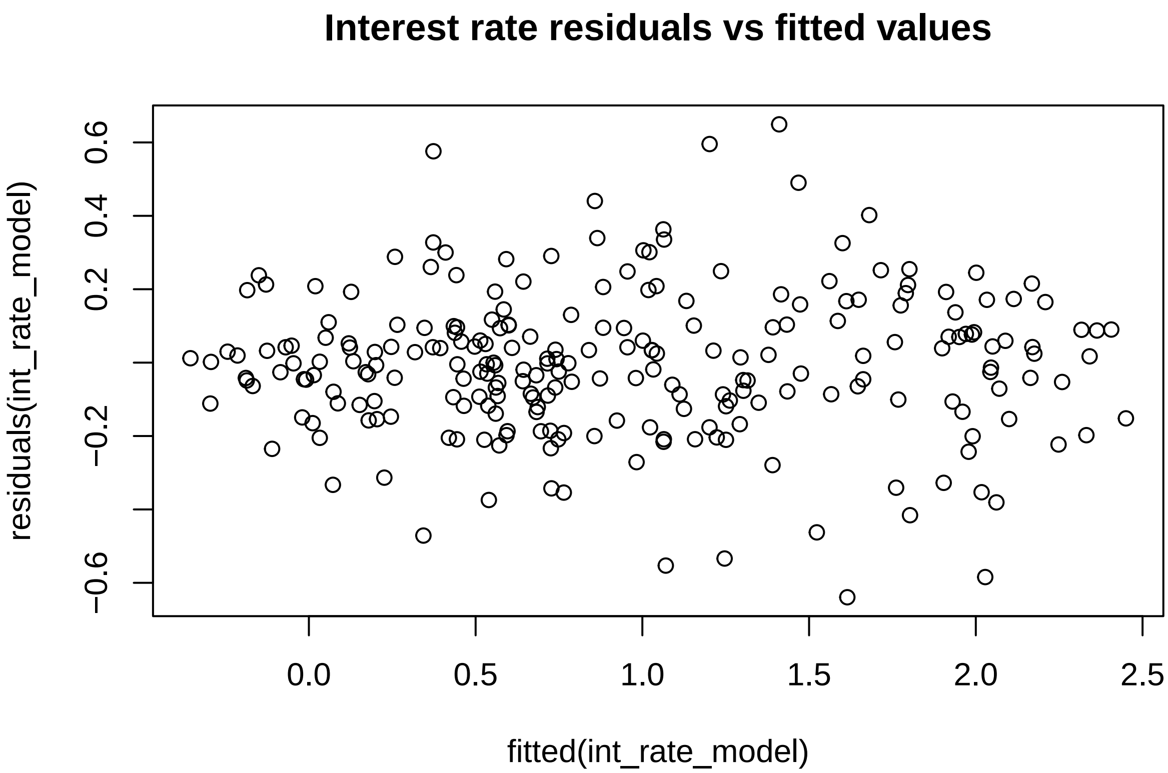
We fit an AR(1) model to both time series. Both their ACF plots showcase a slow gradual decay in the lag significance, indicating a strong time dependence in the data. Both their PACF plots show lag significance till the first lag, followed by a drastic drop in lag significance, so we model the data to the first lag.

(d)

```r
# plot residuals vs fitted values
plot(fitted(unemp_model), residuals(unemp_model),
     main = "Unemployment rate residuals vs fitted values")
```




```r
plot(fitted(int_rate_model), residuals(int_rate_model),
     main = "Interest rate residuals vs fitted values")
```

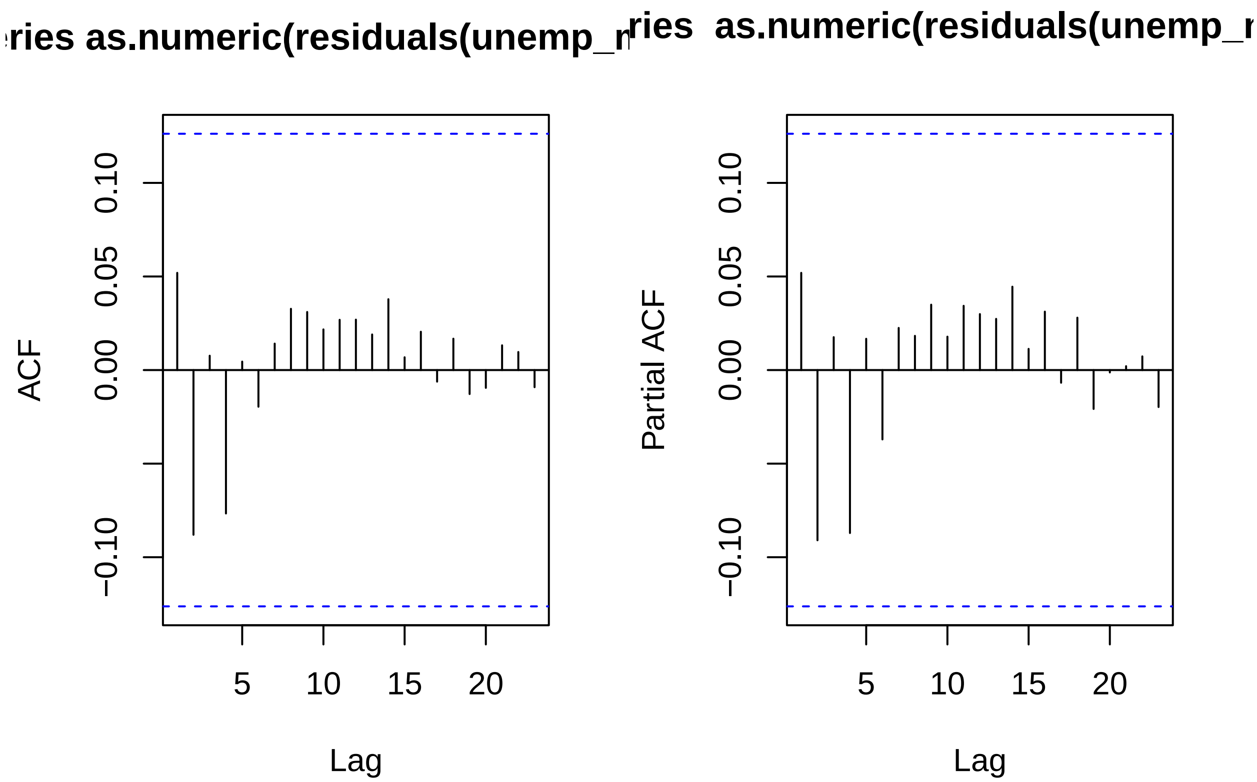


Unemployment rate: The residuals are overall evenly spread around the mean 0, with a slight increase in spread as the fitted values increase. There is one outlier that has a residual over 10, which is the shock at 2020 that is not accounted for by the model.

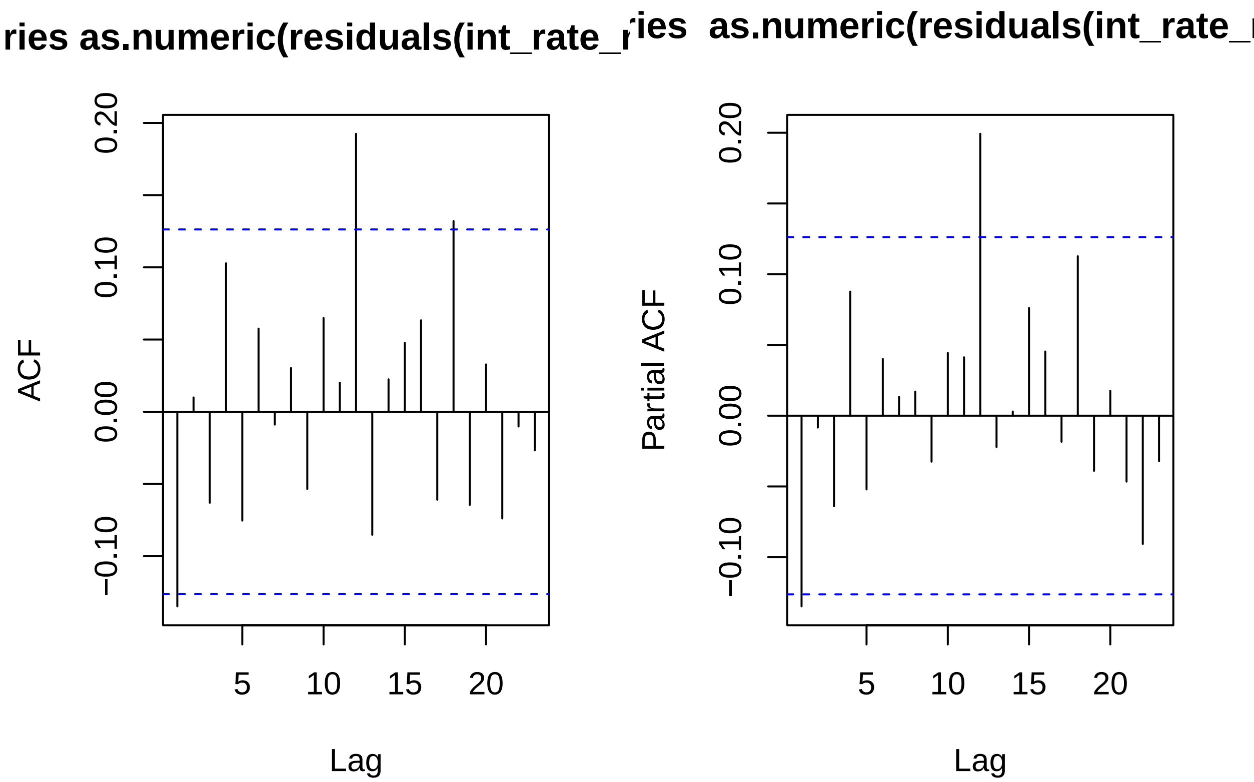
Interest rate: The residuals are overall evenly spread around the mean 0, with a slight increase in the spread as the fitted values increase.

(e)

```
# plot ACF and PACF of residuals
par(mfrow = c(1, 2))
acf(as.numeric(residuals(unemp_model)))
pacf(as.numeric(residuals(unemp_model)))
```




```r
par(mfrow = c(1, 2))
acf(as.numeric(residuals(int_rate_model)))
pacf(as.numeric(residuals(int_rate_model)))
```

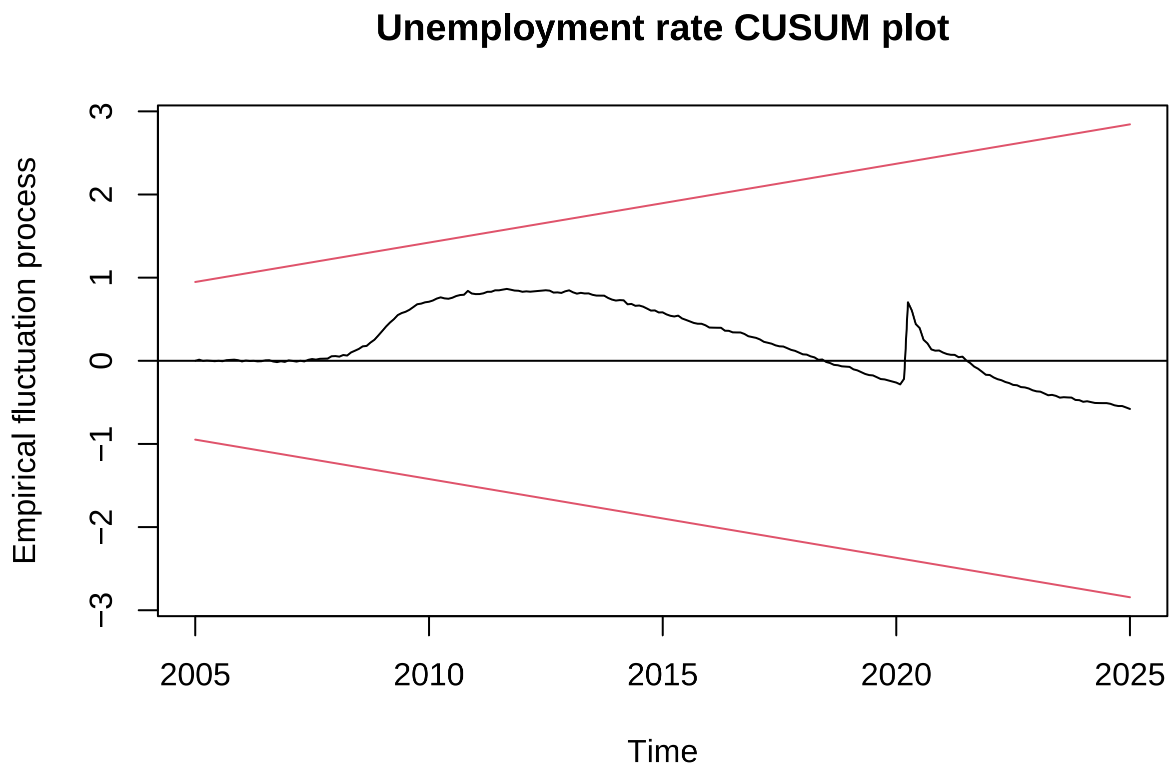


Unemployment rate: In both the ACF and PACF plots, the residuals are insignificant at all the lags, suggesting they are random (white noise) and the model is a good fit.

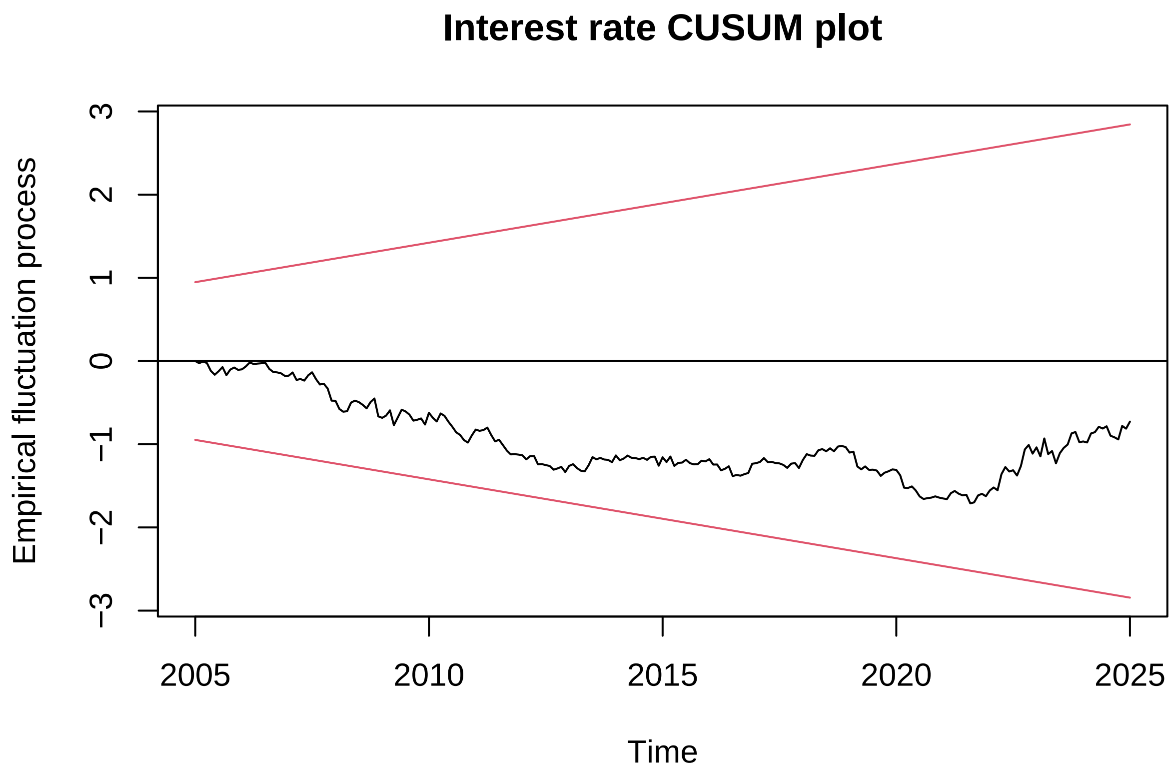
Interest rate: In both the ACF and PACF plots, the residuals are significant at lag 1 and lag 12, indicating there may be monthly seasonality. This means we should improve the model by incorporating a seasonal AR, seasonal MA, or both.

(f)

```r
# plot the CUSUM
plot(efp(residuals(unemp_model) ~ 1, type = "Rec-CUSUM"),
     main = "Unemployment rate CUSUM plot")
```




```r
plot(efp(residuals(int_rate_model) ~ 1, type = "Rec-CUSUM"),
     main = "Interest rate CUSUM plot")
```



Unemployment rate: The line in the plot indicates the process starts out being stable and close to the target value, then increases in the mean, then starts to decline, with a sudden increase shock in 2020. There isn't a consistent upward or downward trend in the unemployment rate.

Interest rate: The line in the plot indicates the process has a gradual decrease in the mean, with seasonality within the years. The decreasing trend turns into an increasing trend after around 2021. This shows interest rate overall has a gradual decreasing trend throughout the years observed.

(g)

```r
# summary of the unemployment rate model
summary(unemp_model)
```

```
## Series: unemp_rate_ts
## ARIMA(1,0,0) with non-zero mean
##
## Coefficients:
##          ar1    mean
##       0.9393  5.6785
## s.e.  0.0210  0.7195
##
## sigma^2 = 0.5204:  log likelihood = -263.32
## AIC=532.64   AICc=532.74   BIC=543.1
##
## Training set error measures:
```

```
##                        ME      RMSE       MAE        MPE     MAPE      MASE
## Training set 0.002511394 0.7183792 0.2265287 -0.9613369 3.496293 0.1900882
##                     ACF1
## Training set 0.05195086
```

The error statistics of the unemployment rate model are all within reasonable values, indicating a good fit of the model.

```r
# ljung-box test on unemployment rate model residuals
Box.test(residuals(unemp_model))
```

```
##
##  Box-Pierce test
##
## data:  residuals(unemp_model)
## X-squared = 0.65043, df = 1, p-value = 0.42
```

The p-value of the Ljung-box test on the unemployment rate model residuals is not significant at the 5% significance level, so we fail to reject the null hypothesis and conclude that the residuals are independently distributed with no autocorrelation. This shows the model is a good fit.

```r
# summary of the interest rate model
summary(int_rate_model)
```

```
## Series: int_rate_ts
## ARIMA(1,0,0) with non-zero mean
##
## Coefficients:
##          ar1    mean
##       0.9662  1.1291
## s.e.  0.0162  0.3466
##
## sigma^2 = 0.03903:  log likelihood = 48.52
## AIC=-91.03   AICc=-90.93   BIC=-80.58
##
## Training set error measures:
##                         ME      RMSE       MAE      MPE     MAPE      MASE
## Training set -0.004760914 0.1967378 0.1471113 10.01313 51.14577 0.3346705
##                    ACF1
## Training set -0.134757
```

Some of the error statistics of the interest rate model, such as MAPE, are high, indicating the model may not be a good fit, and may need justifying or replacing.

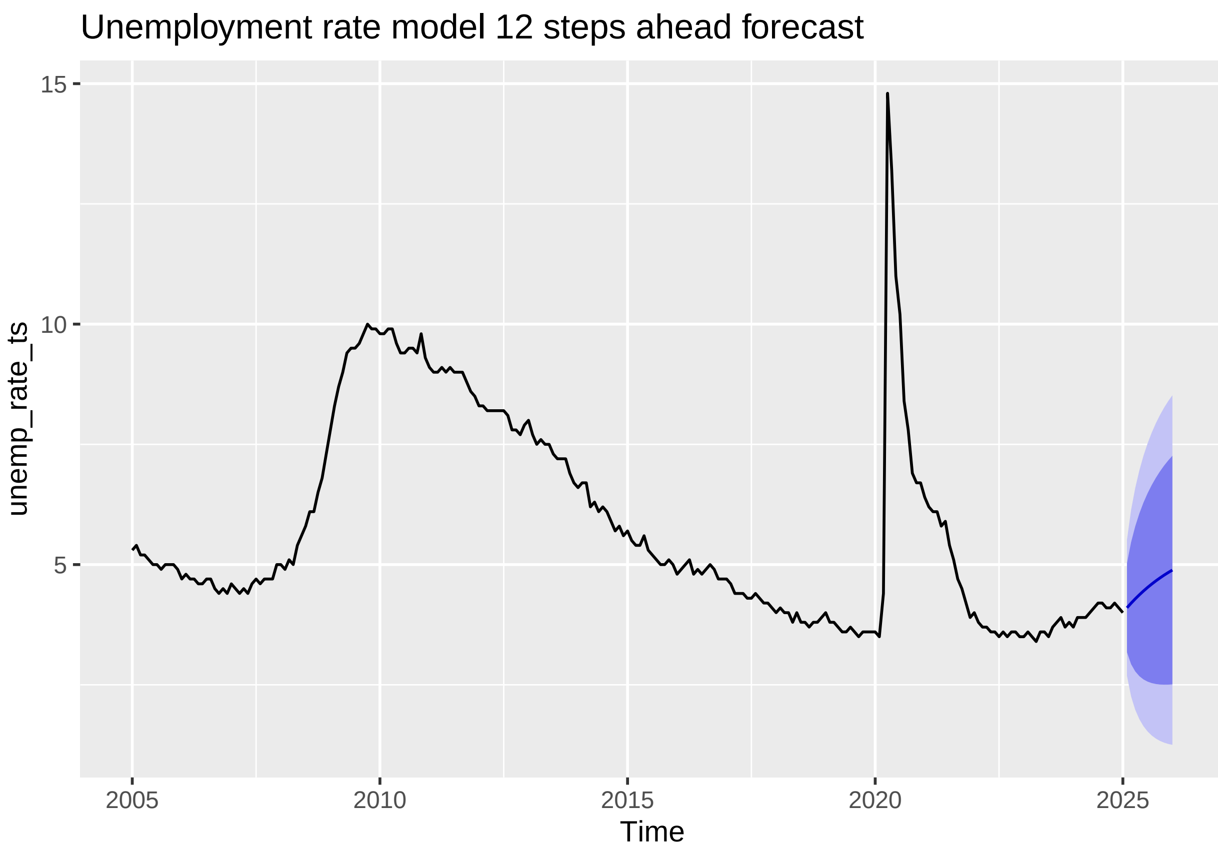
```r
# ljung-box test on interest rate model residuals
Box.test(residuals(int_rate_model))
```

```
##
##  Box-Pierce test
##
## data:  residuals(int_rate_model)
## X-squared = 4.3764, df = 1, p-value = 0.03644
```

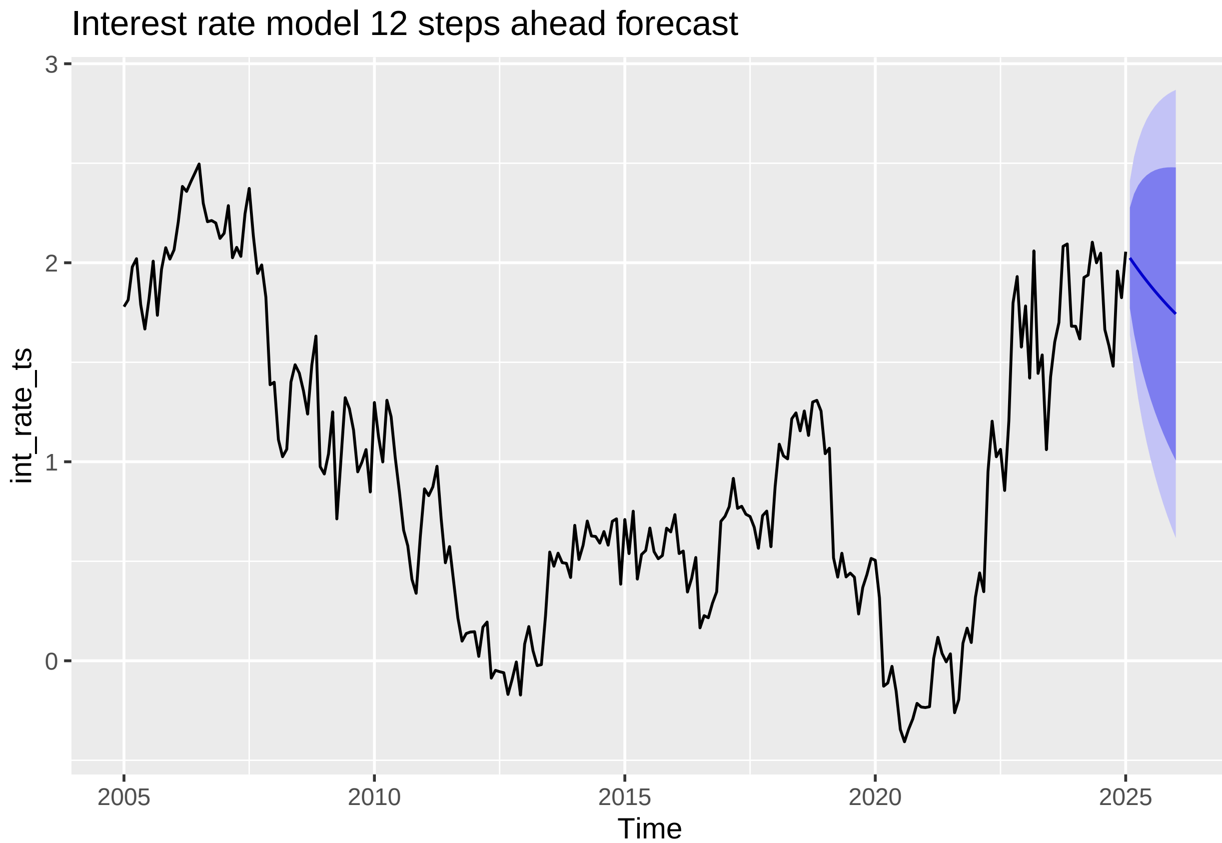
The p-value of the Ljung-box test on the interest rate model residuals is significant at the 5% significance level, so we reject the null hypothesis and conclude that the residuals exhibit autocorrelation. So we need to fit a new model that is better fitting.

(h)

```r
# plot the 12 steps ahead forecasts
autoplot(forecast(unemp_model, h = 12)) +
  ggtitle("Unemployment rate model 12 steps ahead forecast")
```



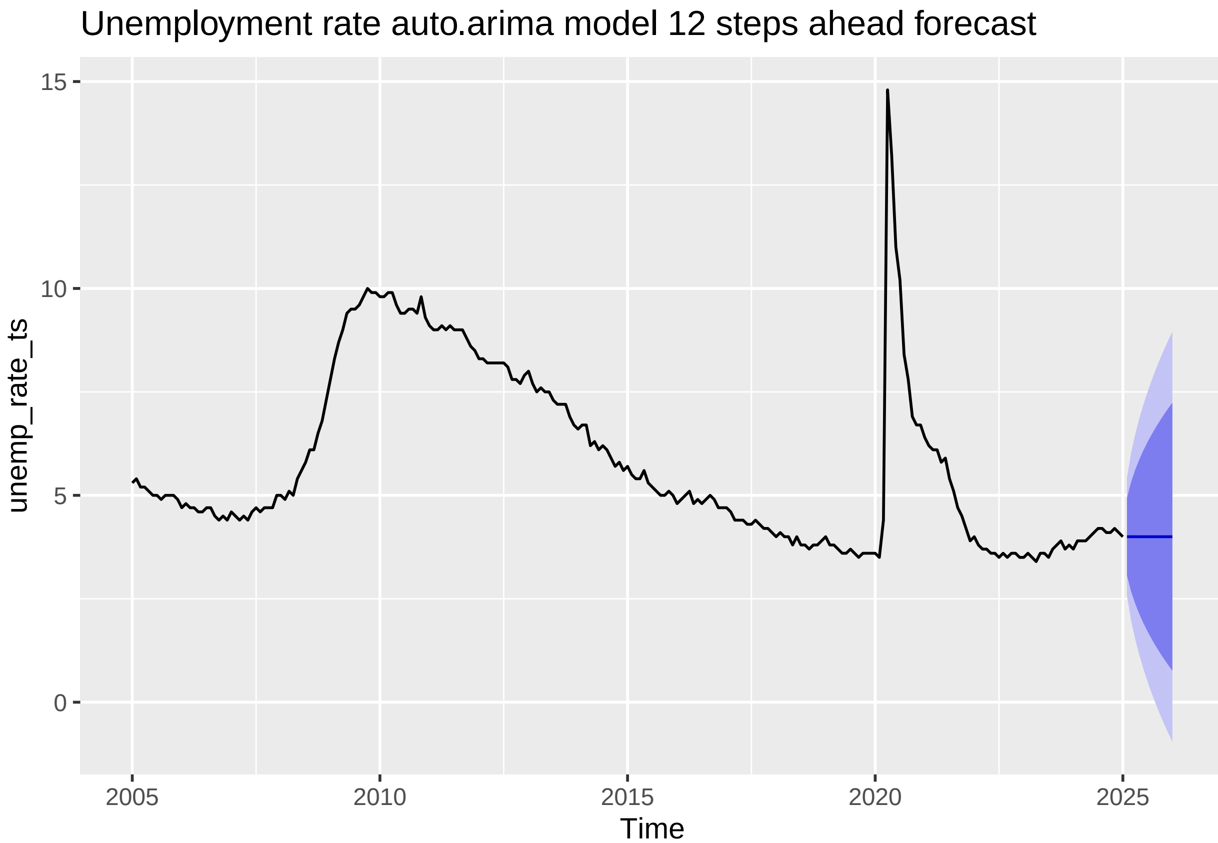

```r
autoplot(forecast(int_rate_model, h = 12)) +
  ggtitle("Interest rate model 12 steps ahead forecast")
```



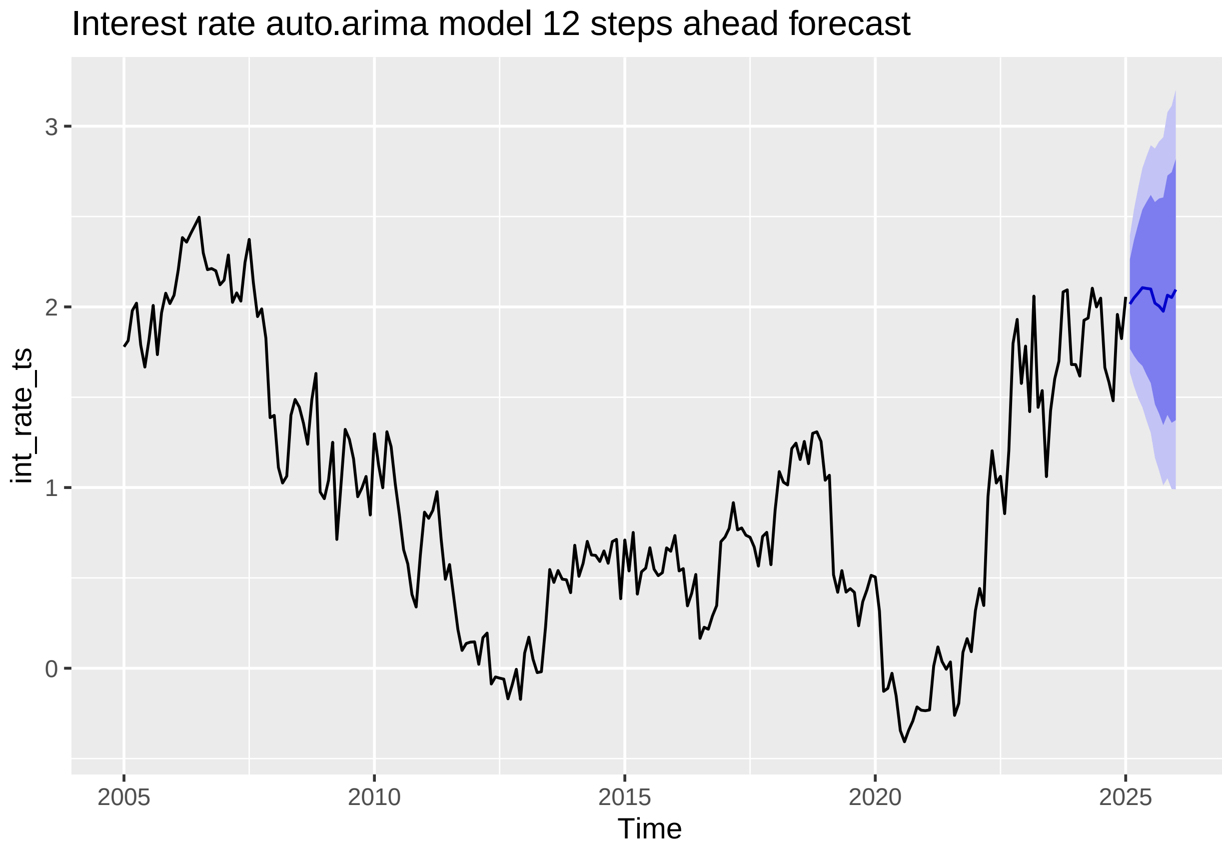
(i)

```
# model the time series with auto.arima
unemp_auto_arima <- auto.arima(unemp_rate_ts)
int_rate_auto_arima <- auto.arima(int_rate_ts)

# forecast the auto.arima models
autoplot(forecast(unemp_auto_arima, h = 12)) +
  ggtitle("Unemployment rate auto.arima model 12 steps ahead forecast")
```




```r
autoplot(forecast(int_rate_auto_arima, h = 12)) +
  ggtitle("Interest rate auto.arima model 12 steps ahead forecast")
```




```
# create a function that calculates MAPE
MAPE <- function(actual, forecast) {
  mean(abs((actual - forecast) / actual))
}

## calculate MAPE for unemployment rate

# divide original time series data into training and testing data
unemp_train_data <- ts(unemp_rate_ts[1:(length(unemp_rate_ts) - 12)],
                       start = c(2005, 1), frequency = 12)
unemp_test_data <- ts(unemp_rate_ts[(length(unemp_rate_ts) - 11):length(unemp_rate_ts)],
                      start = c(2024, 2), frequency = 12)

# fit to training data, forecast and compare with testing data
unemp_model_mape <- Arima(unemp_train_data, order = c(1, 0, 0))
cat("MAPE for unemployment rate AR(1):",
    MAPE(unemp_test_data, forecast(unemp_model_mape, h = 12)$mean))
```

```
## MAPE for unemployment rate AR(1): 0.07765713
```

```
unemp_auto_arima_mape <- auto.arima(unemp_train_data)
cat("MAPE for unemployment rate auto.arima:",
    MAPE(unemp_test_data, forecast(unemp_auto_arima_mape, h = 12)$mean))
```

```
## MAPE for unemployment rate auto.arima: 0.08760274
```

```r
# calculate MAPE for interest rate

# divide original time series data into training and testing data
int_rate_train_data <- ts(int_rate_ts[1:(length(int_rate_ts) - 12)],
                          start = c(2005, 1), frequency = 12)
int_rate_test_data <- ts(int_rate_ts[(length(int_rate_ts) - 11):length(int_rate_ts)],
                         start = c(2024, 2), frequency = 12)

# fit to training data, forecast and compare with testing data
int_rate_model_mape <- Arima(int_rate_train_data, order = c(1, 0, 0))
cat("MAPE for interest rate AR(1):",
    MAPE(int_rate_test_data, forecast(int_rate_model_mape, h = 12)$mean))
```

```
## MAPE for interest rate AR(1): 0.1579509
```

```r
int_rate_auto_arima_mape <- auto.arima(int_rate_train_data)
cat("MAPE for interest rate auto.arima:",
    MAPE(int_rate_test_data, forecast(int_rate_auto_arima_mape, h = 12)$mean))
```

```
## MAPE for interest rate auto.arima: 0.1352711
```

Unemployment rate: The AR(1) model performs better than the auto.arima model in terms of MAPE.

Interest rate: The auto.arima model performs better than the AR(1) model in terms of MAPE.

Overall, between all the models of the two variables, the AR(1) model for unemployment rate performs the best, as it has the lowest MAPE out of all.

(j)

```r
## combine forecasts for unemployment rate using linear regression

# create data frame of forecasts and actual data
unemp_rate_forecast_df <- data.frame(fc1 = forecast(unemp_model_mape, h = 12)$mean,
                                     fc2 = forecast(unemp_auto_arima_mape, h = 12)$mean,
                                     actual = unemp_test_data)

# fit a linear model to combine forecasts
unemp_comb_model <- lm(actual ~ fc1 + fc2, data = unemp_rate_forecast_df)

# forecast using combined model
unemp_comb_fc <- predict(unemp_comb_model, newdata = unemp_rate_forecast_df)

# calculate MAPE for combined forecasts
cat("MAPE for unemployment rate combined:", MAPE(unemp_test_data, unemp_comb_fc))
```

```
## MAPE for unemployment rate combined: 0.0166471
```

```r
## combine forecasts for interest rate using linear regression

# create data frame of forecasts and actual data
int_rate_forecast_df <- data.frame(fc1 = forecast(int_rate_model_mape, h = 12)$mean,
                                   fc2 = forecast(int_rate_auto_arima_mape, h = 12)$mean,
                                   actual = int_rate_test_data)

# fit a linear model to combine forecasts
int_comb_model <- lm(actual ~ fc1 + fc2, data = int_rate_forecast_df)

# forecast using combined model
int_comb_fc <- predict(int_comb_model, newdata = int_rate_forecast_df)

# calculate MAPE for combined forecasts
cat("MAPE for interest rate combined:", MAPE(int_rate_test_data, int_comb_fc))
```

```
## MAPE for interest rate combined: 0.09660713
```

For both unemployment rate and interest rate, the combined models performed better forecasts than both the AR(1) and auto.arima models, in terms of lowest MAPE.

(k)

```r
## fit VAR(p) model to the data

# combine the two time series into a data frame
y <- cbind(unemp_rate_ts, int_rate_ts)
y_tot <- data.frame(y)

VARselect(y_tot, lag.max = 10)
```

```
## $selection
## AIC(n)  HQ(n)  SC(n) FPE(n)
##      2      1      1      2
##
## $criteria
##                  1           2           3           4           5           6
## AIC(n) -3.82966026 -3.85297604 -3.83069357 -3.80714539 -3.79245559 -3.76540544
## HQ(n)  -3.79359671 -3.79287012 -3.74654528 -3.69895473 -3.66022256 -3.60913005
## SC(n)  -3.74024681 -3.70395363 -3.62206219 -3.53890505 -3.46460628 -3.37794717
## FPE(n)  0.02171706  0.02121679  0.02169537  0.02221325  0.02254343  0.02316374
##                  7           8           9          10
## AIC(n) -3.73579756 -3.70321039 -3.67081091 -3.64160494
## HQ(n)  -3.55547979 -3.49885026 -3.44240841 -3.38916007
## SC(n)  -3.28873032 -3.19653419 -3.10452575 -3.01571081
## FPE(n)  0.02386287  0.02465742  0.02547479  0.02623664
```

```r
y_model <- VAR(y_tot, p = 2)
```

We selected `p = 2` over `p = 1` based on both statistical and diagnostic considerations. Although SC and HQ favored p = 1 due to their stronger penalty on complexity, AIC and FPE suggested p = 2, indicating a better predictive fit. Since the goal is to capture dynamic relationships effectively, we prioritized the criteria that emphasize forecast accuracy.

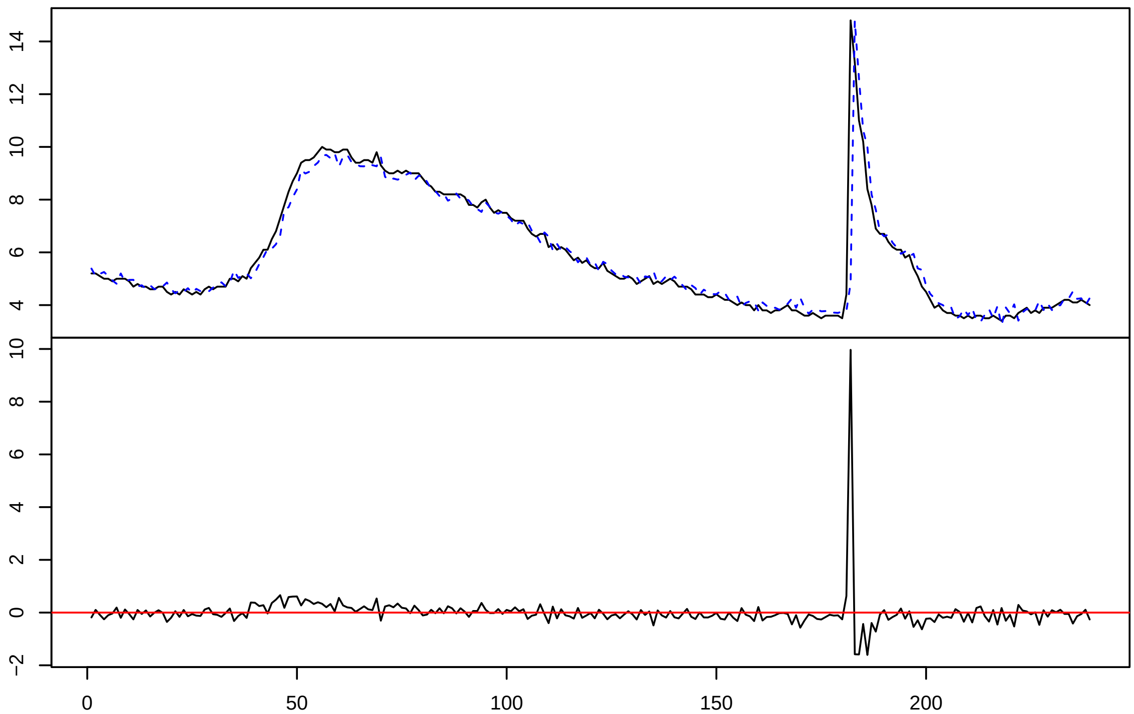
```r
# summary of the VAR model
summary(y_model)
```

```
##
## VAR Estimation Results:
## =========================
## Endogenous variables: unemp_rate_ts, int_rate_ts
## Deterministic variables: const
## Sample size: 239
## Log Likelihood: -203.274
## Roots of the characteristic polynomial:
## 0.9602 0.9382 0.06667 0.06667
## Call:
## VAR(y = y_tot, p = 2)
##
##
## Estimation results for equation unemp_rate_ts:
## ==============================================
## unemp_rate_ts = unemp_rate_ts.l1 + int_rate_ts.l1 + unemp_rate_ts.l2 + int_rate_ts.l2 + const
##
##                  Estimate Std. Error t value Pr(>|t|)
## unemp_rate_ts.l1  0.98813    0.06439  15.346  < 2e-16 ***
## int_rate_ts.l1   -0.63087    0.23775  -2.654  0.00851 **
## unemp_rate_ts.l2 -0.05475    0.06482  -0.845  0.39916
## int_rate_ts.l2    0.60604    0.23625   2.565  0.01093 *
## const             0.40535    0.18311   2.214  0.02781 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##
## Residual standard error: 0.7172 on 234 degrees of freedom
## Multiple R-Squared: 0.8902,  Adjusted R-squared: 0.8884
## F-statistic: 474.5 on 4 and 234 DF,  p-value: < 2.2e-16
##
##
## Estimation results for equation int_rate_ts:
## ============================================
## int_rate_ts = unemp_rate_ts.l1 + int_rate_ts.l1 + unemp_rate_ts.l2 + int_rate_ts.l2 + const
##
##                   Estimate Std. Error t value Pr(>|t|)
## unemp_rate_ts.l1  0.006243   0.017544   0.356   0.7223
## int_rate_ts.l1    0.814450   0.064779  12.573   <2e-16 ***
## unemp_rate_ts.l2 -0.019264   0.017661  -1.091   0.2765
## int_rate_ts.l2    0.140090   0.064370   2.176   0.0305 *
## const             0.119164   0.049891   2.388   0.0177 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##
## Residual standard error: 0.1954 on 234 degrees of freedom
## Multiple R-Squared: 0.9305,  Adjusted R-squared: 0.9293
```

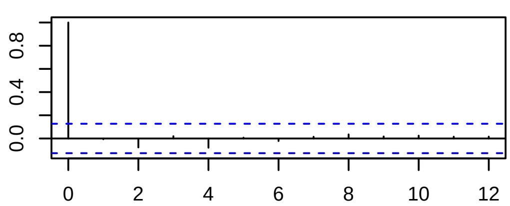
```
## F-statistic: 783.2 on 4 and 234 DF,  p-value: < 2.2e-16
##
##
##
## Covariance matrix of residuals:
##              unemp_rate_ts int_rate_ts
## unemp_rate_ts      0.514431   -0.007137
## int_rate_ts       -0.007137    0.038192
##
## Correlation matrix of residuals:
##              unemp_rate_ts int_rate_ts
## unemp_rate_ts       1.00000    -0.05092
## int_rate_ts        -0.05092     1.00000
```

```r
# plot the VAR model
quartz()
plot(y_model)
```

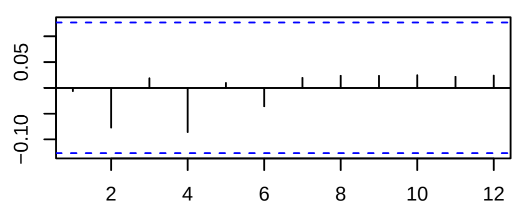
Diagram of fit and residuals for unemp_rate_ts

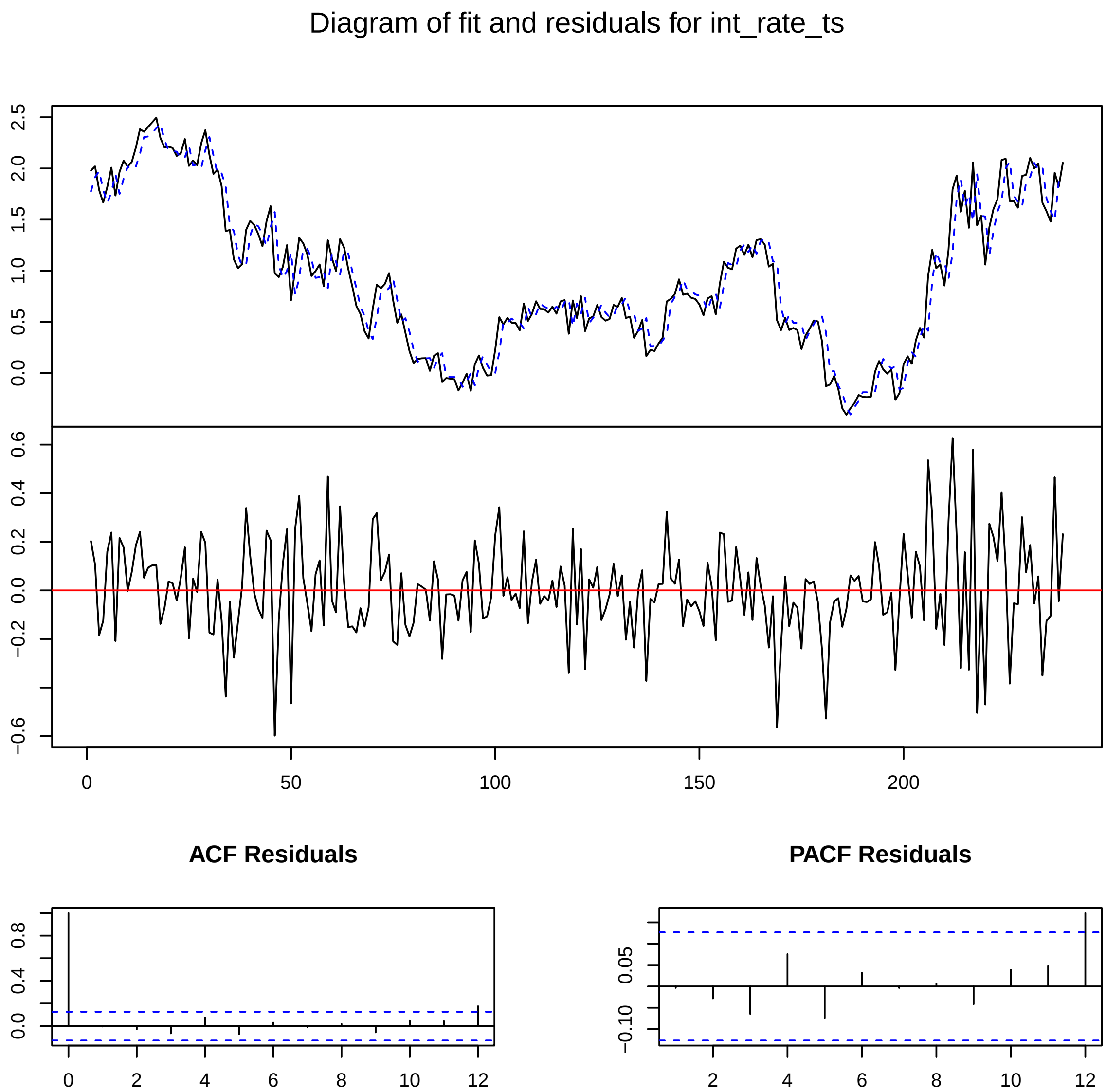


ACF Residuals



PACF Residuals





According to the plot, both unemployment rate and interest rate are well-fitted by the model.

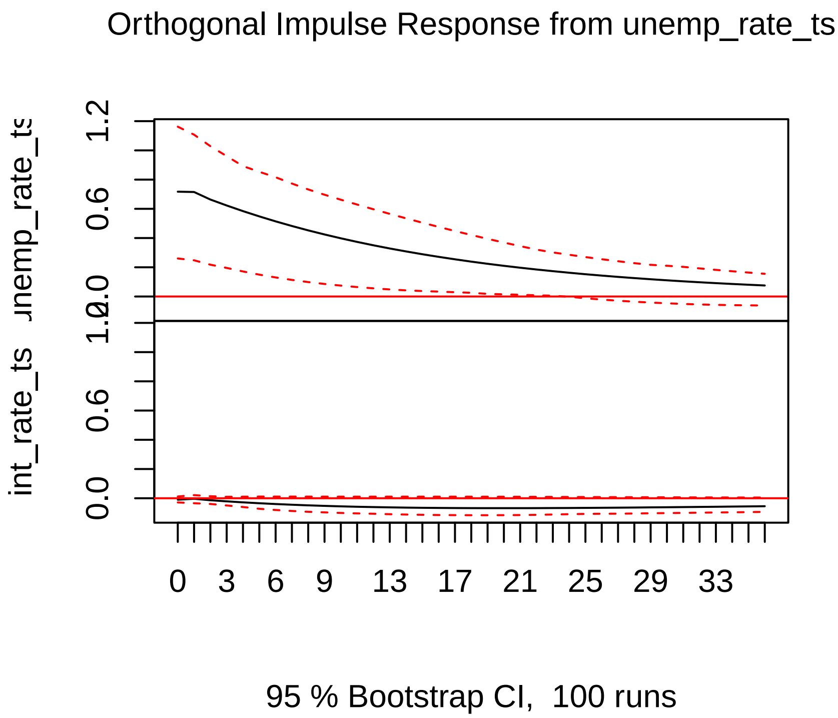
According to the significance of the lagged coefficients, it seems that interest rate is significant in Granger-causing unemployment rate. We can verify this later in the Granger-causality test.

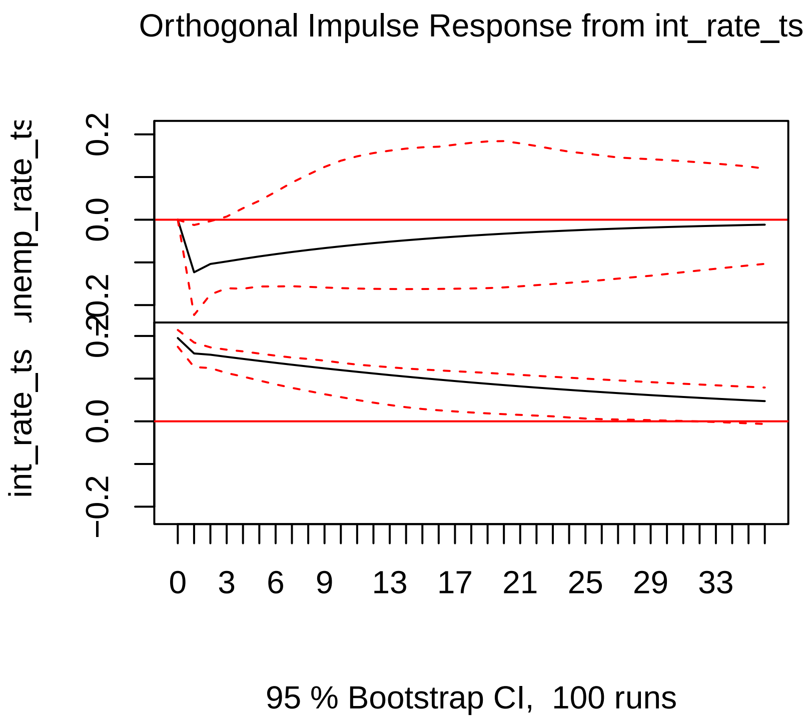
Unemployment rate: The residuals are not significant at the lags in both the ACF and PACF, so they are random (white noise).

Interest rate: Most of the residuals are not significant at the lags in both the ACF and PACF, so they are random (white noise). There is one spike at lag 12, which signals there may be a presence of autocorrelation or seasonality.

(1)

```r
# plot the impulse response functions
par(mfrow = c(1, 2))
plot(irf(y_model, n.ahead = 36))
```





Effect of unemployment rate on unemployment rate: Initially a positive effect, then decays gradually.

Effect of unemployment rate on interest rate: Has little, almost negligible, effect the whole time.

Effect of interest rate on unemployment rate: Initially no effect, then a negative effect, then gradually smaller effects.

Effect of interest rate on interest rate: Initially a positive effect, then decays gradually.

(m)

```r
# perform granger-causality test

# does interest rate granger-cause unemployment rate
grangertest(unemp_rate_ts ~ int_rate_ts, order = 2)
```

```
## Granger causality test
##
## Model 1: unemp_rate_ts ~ Lags(unemp_rate_ts, 1:2) + Lags(int_rate_ts, 1:2)
## Model 2: unemp_rate_ts ~ Lags(unemp_rate_ts, 1:2)
##   Res.Df Df     F  Pr(>F)
## 1    234
## 2    236 -2 3.525 0.03103 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

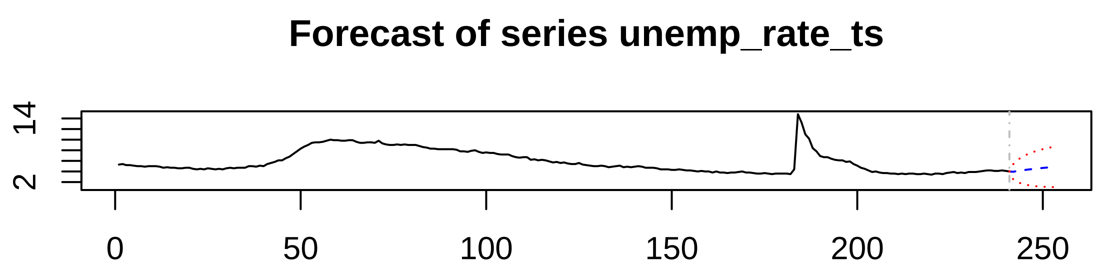
```r
# does unemplotment rate granger-cause interest rate
grangertest(int_rate_ts ~ unemp_rate_ts, order = 2)
```

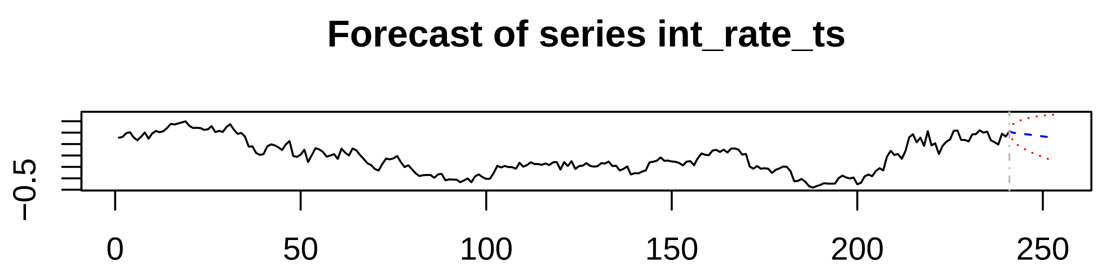
```
## Granger causality test
##
## Model 1: int_rate_ts ~ Lags(int_rate_ts, 1:2) + Lags(unemp_rate_ts, 1:2)
## Model 2: int_rate_ts ~ Lags(int_rate_ts, 1:2)
##   Res.Df Df     F Pr(>F)
## 1    234
## 2    236 -2 2.208 0.1122
```

According to the Granger test at the 5% significance level, interest rate Granger-causes unemployment rate, while unemployment rate does not Granger-cause interest rate.

(n)

```r
# forecast and plot 12 steps ahead with VAR model
var_predict <- predict(object = y_model, n.ahead = 12)
plot(var_predict)
```





The forecast using VAR considers the relationship between the two time series, unemployment rate and interest rate, capturing how the changes and volatility in one variable affects another. The forecasts with other models considers the time series independently, so does not capture the dynamic between the variables.

III. Conclusions and Future Work

Conclusions

This study analyzed the relationship between the U.S. monthly unemployment rate and the 10-year Treasury interest rate over a 20-year period (2005–2025) using various time series models, including ARIMA, auto.ARIMA, and Vector Autoregression (VAR). The key findings are summarized as follows:

1. **Time Series Characteristics**
   - Both the unemployment rate and interest rate exhibited stochastic trends with seasonal components.
   - The unemployment rate displayed significant shocks, particularly around 2020 due to the COVID-19 pandemic.
   - The interest rate showed a general declining trend, with a reversal around 2021.
2. **Modeling Performance**
   - The AR(1) model provided the best fit for unemployment rate forecasting, with the lowest Mean Absolute Percentage Error (MAPE).
   - For the interest rate, the auto.ARIMA model outperformed the AR(1) model, suggesting that a more flexible model was required to capture its dynamics.
   - A combined forecasting approach using linear regression further improved the prediction accuracy for both variables, achieving the lowest MAPE.
3. **Dynamic Relationships and Causality**
   - The VAR(2) model was selected based on AIC and FPE criteria, allowing for a more accurate representation of the dynamic interactions.
   - The Granger causality test indicated that the interest rate Granger-causes the unemployment rate, but not vice versa. This suggests that changes in interest rates may have predictive power over unemployment fluctuations, likely due to monetary policy effects.
   - Impulse response analysis showed that shocks in interest rates initially had a positive effect on unemployment, which then decayed over time. In contrast, changes in unemployment had negligible long-term effects on interest rates.
4. **Forecasting Insights**
   - The VAR model incorporated the interdependence between variables, offering a more comprehensive forecasting approach.
   - Forecasts suggested that while unemployment is expected to remain stable, interest rates may continue fluctuating based on economic conditions.

Future Work

While this study provides valuable insights, several areas warrant further exploration:

1. **Enhancing Model Performance**
   - The current ARIMA and VAR models could be refined by incorporating external economic indicators such as inflation rates, GDP growth, and labor market participation to improve predictive accuracy.

2. **Addressing Seasonality and Structural Breaks**

    - The presence of seasonal patterns and structural changes (e.g., financial crises, policy shifts) suggests the need for models that can adapt dynamically.

3. **Policy Implications and Real-World Applications**

    - Further research could explore how central bank policies influence these dynamics, particularly under different monetary policy regimes.

In conclusion, this analysis highlights the interactions between unemployment and interest rates, demonstrating the importance of using robust time series methodologies. Future research should focus on refining models, incorporating additional economic factors, and exploring alternative forecasting techniques to further enhance predictive accuracy and economic insights.

IV. References

The data are from the Federal Reserve Economic Data (FRED).

Unemployment rate: https://fred.stlouisfed.org/series/UNRATE

Interest rate: https://fred.stlouisfed.org/series/REAINTRATREARAT10Y