

Lab 2

Task 1: Introduction

Monetary policy plays a central role in stabilizing economic activity by influencing interest rates in response to inflationary and output fluctuations. One of the most influential formulations of this relationship is the **Taylor Rule**, which posits that the short-term nominal interest rate should adjust systematically to deviations of inflation from its target and output from potential. Conceptually, the Taylor Rule provides a quantitative guideline for how central banks such as the Federal Reserve set policy to balance price stability and economic growth.

Formally, the baseline Taylor Rule can be expressed as:

$$i_t = r^* + \pi_t + 0.5(\pi_t - \pi^*) + 0.5(\text{Output Gap}_t)$$

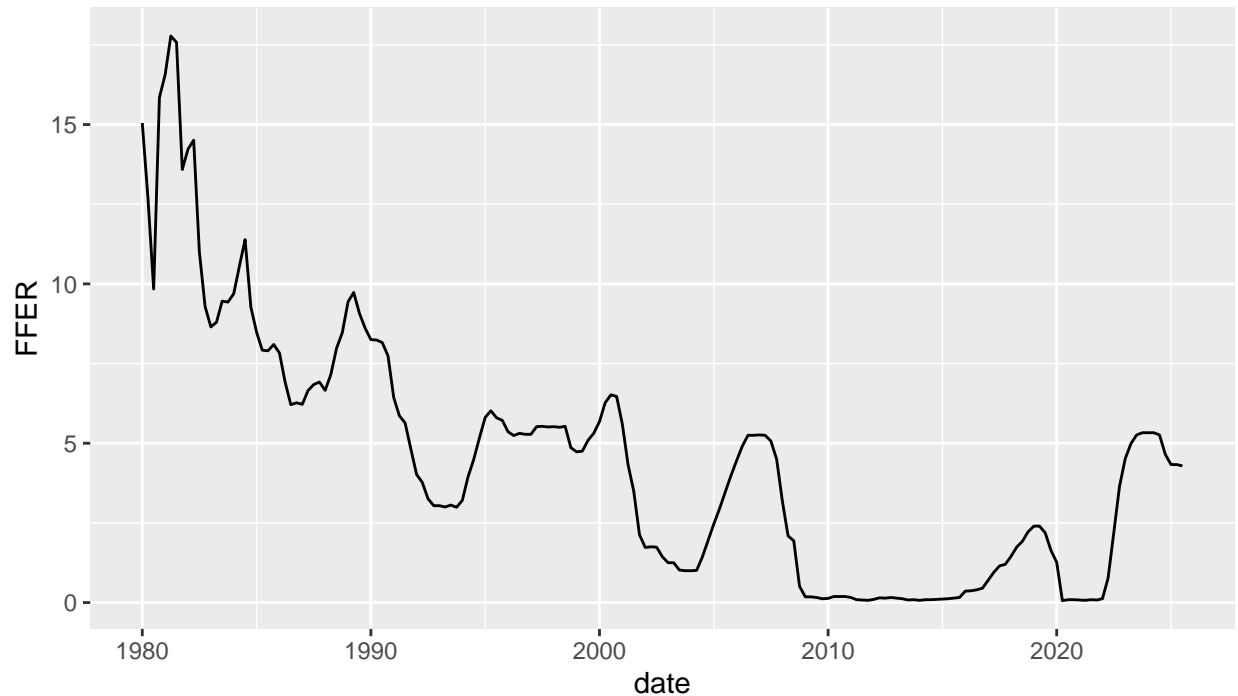
where i_t denotes the nominal policy rate (in our work this is proxied by the federal funds rate), r^* is the equilibrium real interest rate, π_t is the observed inflation rate, π^* is the target inflation rate, and $(y_t - y_t^*)$ represents the output gap. The coefficients 0.5 measure the policy responsiveness to inflation and output deviations, respectively. In practice, this framework serves as both a descriptive model of past monetary policy behavior and a normative benchmark for evaluating central bank actions.

In our analysis, we extend the classical Taylor Rule framework into a sequence of increasingly sophisticated econometric models to evaluate both its explanatory power and dynamic consistency with U.S. monetary policy from 1980 through 2024. We begin by estimating a static **OLS-based Taylor Rule**, which provides a benchmark representation of the policy rate as a function of inflation and output gaps. We then assess the model's residual properties, testing for serial correlation and nonstationarity, and subsequently fit an **ARIMA model** to capture any remaining temporal structure in the residuals. The inclusion of autoregressive and moving average dynamics allows us to model inertia and short-term deviations from the baseline rule.

Building upon this, we incorporate an **Error-Correction Model (ECM)** to evaluate the potential for cointegrating relationships among the interest rate, inflation, and output variables, reflecting the long-run equilibrium behavior implied by the Taylor framework. Finally, we employ a **Vector Autoregression (VAR)** model to capture the full system of interactions among the federal funds rate, inflation gap, and output gap, allowing for feedback effects and joint forecasting. The VAR structure also facilitates the computation of **Impulse Response Functions (IRFs)**, which trace the temporal propagation of policy and macroeconomic shocks across the system.

Together, these modeling stages provide a comprehensive empirical investigation of the Taylor Rule as both a static policy guideline and a dynamic representation of monetary behavior. The progression from OLS to ARIMA, ECM, and VAR frameworks enables us to evaluate how the Federal Reserve's policy reactions have evolved over time, the extent to which they align with theoretical prescriptions, and how well different models capture the interdependence of interest rates, inflation, and output dynamics in the post-1980 monetary regime.

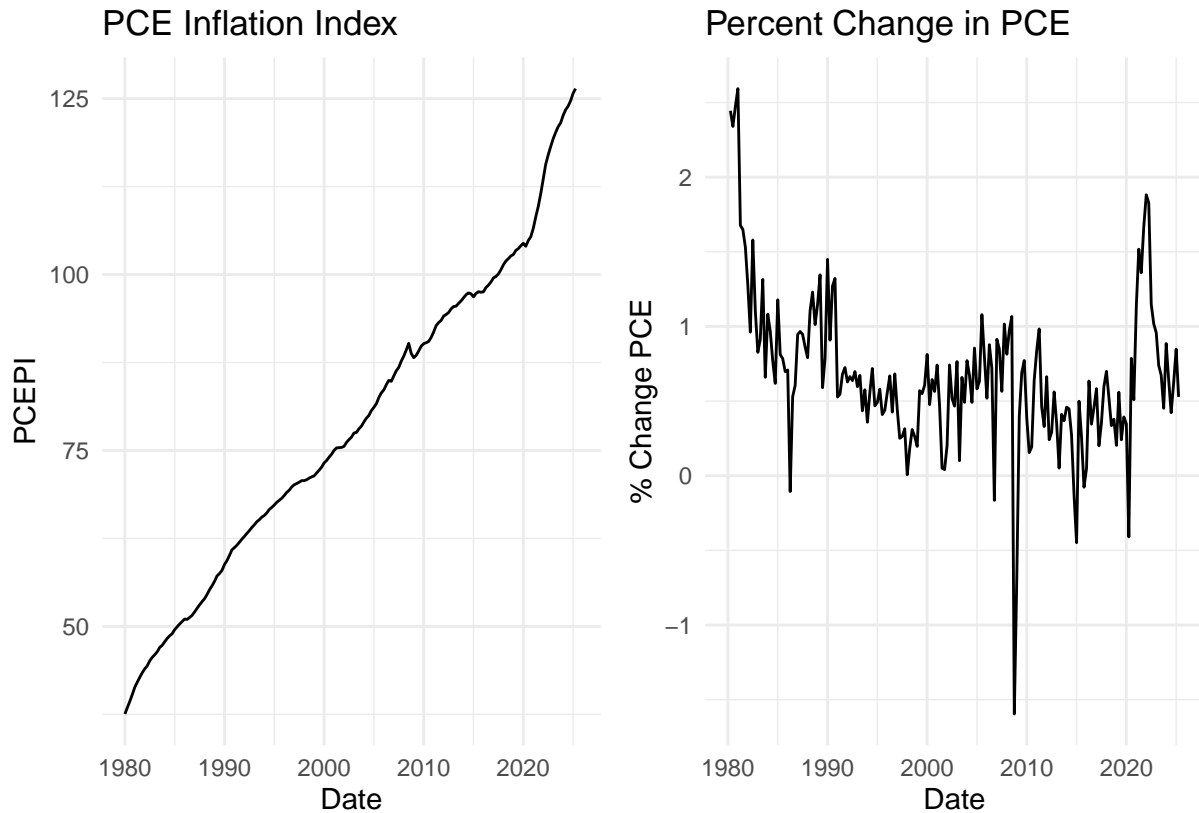
Federal Funds Effective Rate (FFER)



The summary statistics of the federal funds rate series provide a concise quantitative overview of the data used in the analysis. Across 183 quarterly observations spanning 1980 Q1 – 2024 Q4, the average policy rate is approximately 4.42%, with a variance of 15.3 and a standard deviation of 3.91. This level of dispersion highlights the considerable volatility of U.S. monetary policy across distinct economic regimes. The distribution reflects several key macroeconomic episodes visible in the plotted series: the elevated interest rates of the early 1980s aimed at curbing double-digit inflation, the steady declines following the 1990s expansion and the 2001 dot-com correction, the near-zero rates maintained during the 2008–2015 Great Recession recovery, and the renewed monetary tightening in the post-2020 period as inflation re-emerged. Together, these descriptive statistics and historical inflection points frame the context for the econometric modeling that follows, illustrating both the persistence and cyclicity inherent in the policy rate dynamics.

One may note a few key historical events that appear in the graph:

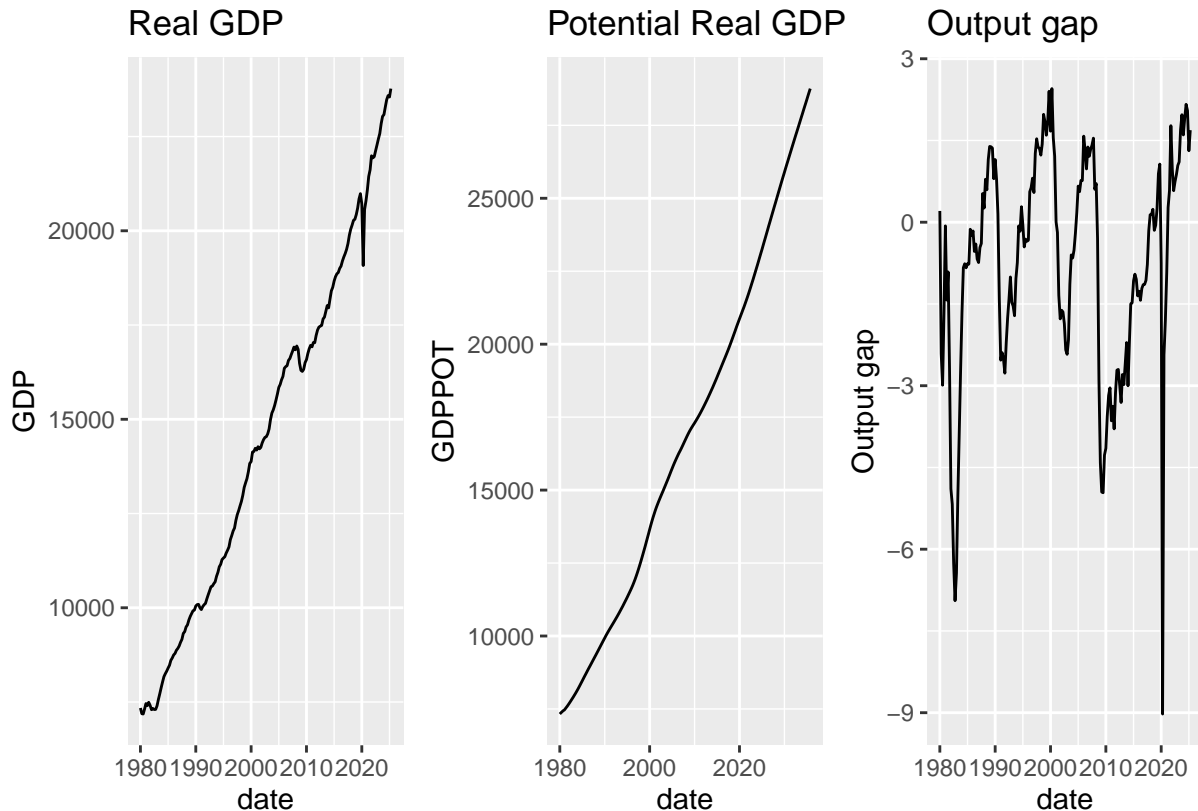
- A near-zero interest rate between 2008 and 2015. This rate was a result of the 2008 Grand Recession
- The 1980s had very high interest rates in order to curb high rates of inflation
- Rates were additionally cut in 2000 in response to the dot-com bubble
- In 2020, rates were cut to near-zero as a result of the pandemic



PCEPI (Inflation) shows events that are related to events that spurred the FFER changes we see in the above graph.

- High inflation characterized the early 1980s, as shown by the peak in the graph
- Very low inflation appears in 2008 during the 2008 financial crisis
- Inflation hit local maxima around 2020 as a result of the pandemic

The Personal Consumption Expenditures Price Index (PCEPI) provides the measure of inflation used in this analysis and serves as the nominal anchor of the Taylor Rule framework. Over the 1980–2024 sample, the index exhibits a mean level of approximately 78.6, with a variance of 501.5 and a standard deviation of 22.4, across 182 quarterly observations. This long-run upward trend reflects the cumulative nature of price growth, while the first-differenced or percentage-change series reveals cyclical fluctuations in inflation dynamics. Periods of particularly high inflation are evident in the early 1980s, consistent with the post-oil-shock stagflation era that motivated aggressive monetary tightening. In contrast, near-zero inflation during the 2008–2015 period reflects deflationary pressures from the global financial crisis. More recently, inflation accelerated sharply around 2020–2022 following the pandemic-induced supply and demand shocks. These fluctuations correspond closely to shifts in the federal funds rate observed earlier, confirming that PCE inflation pressures have historically been a central driver of U.S. monetary policy adjustments.



GDP appears to steadily increase from 1980 to 2020, with a notable dip in 2020 due to the COVID pandemic. The output gap timeseries is affected by similar events as FFER and PCEPI:

- A dip in the early 1980s is seen, which is at the same time as high inflation and spiked interest rates
- A major dip in 2008 exists, which is at the time of the 2008 financial crisis
- A sharp dip in 2020 is seen, corresponding to the time of the COVID pandemic

The level series for real GDP and potential GDP are large-scale aggregates, so their summary moments primarily reflect trend growth rather than short-run variation. Both series have high means ($\approx 1.45 \times 10^4$) and large standard deviations ($\approx 4.75 \times 10^3$), consistent with strong deterministic trend and suggesting that levels are likely **non-stationary** without detrending or differencing. By contrast, the **output gap**, defined as $100 \times (\text{GDPC1} - \text{GDPPOT})/\text{GDPPOT}$, is constructed around zero and exhibits far smaller dispersion; its spikes align with major downturns (early 1980s, 2008-09, 2020). Practically, this implies that GDP and GDPPOT should be modeled in **changes** (or as a **cointegrated** pair), while the output gap can be used directly in regressions because it behaves more like a stationary deviation from trend.

Task 2

Task 2.1: Frequency Alignment (Resampling)

Based on the format of the API call to FRED, data is received at a quarterly cadence, with an average over all months in the quarter calculated and returned in the dataframe. No resampling was done.

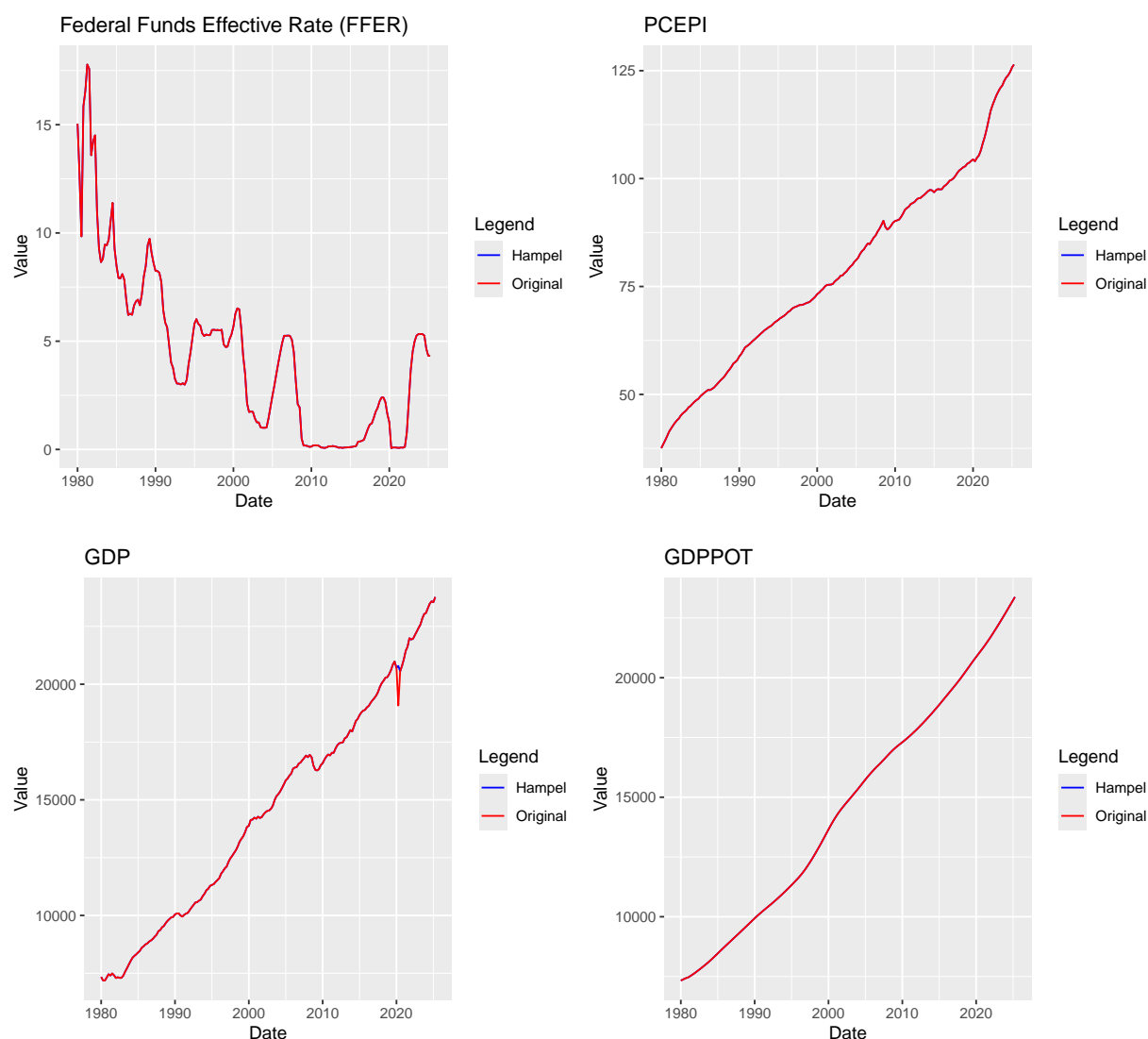
Task 2.2 Outlier Detection & Treatment

Hampel filter

To detect outliers, we've chosen to use the Hampel filter. We prefer this method because based on a visual examination of FFER, PECPI, and GDP data, scores appear to be mostly following either an upward trend or a fluctuating downward trend, with few outliers. The outliers that do appear (i.e. in Real GDP around 2020) are within the data (as opposed to on the upper and lower ends of the data). Thus, this would not be caught by winsorization. The rolling window calculations from a Hampel filter (relying on deviations from a median) would catch these outliers and smooth out the data. Z-scores, as well, is best suited to normally distributed data, which we don't have.

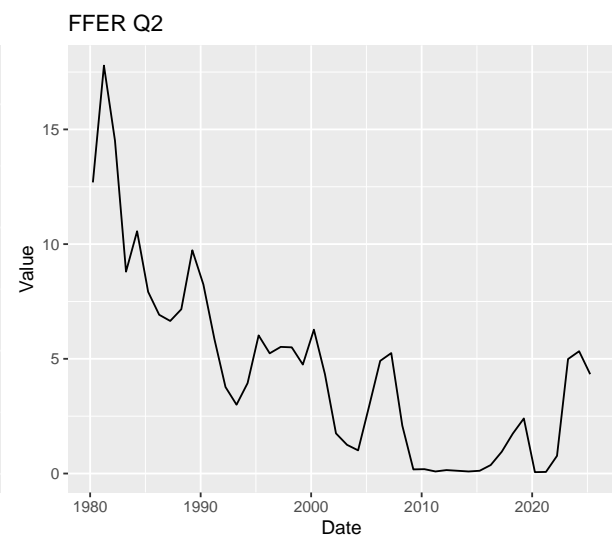
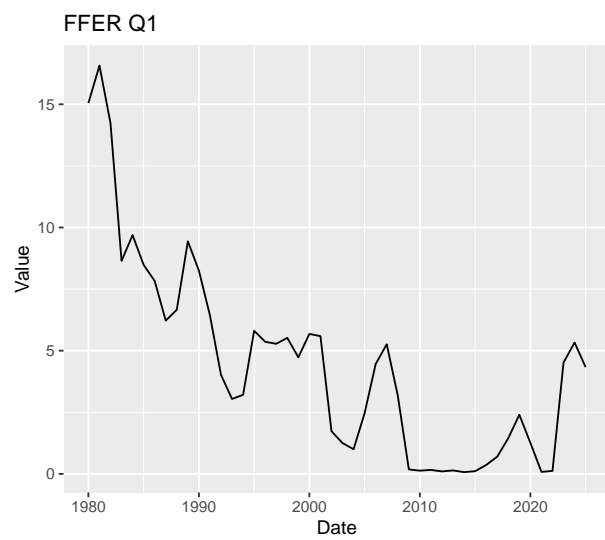
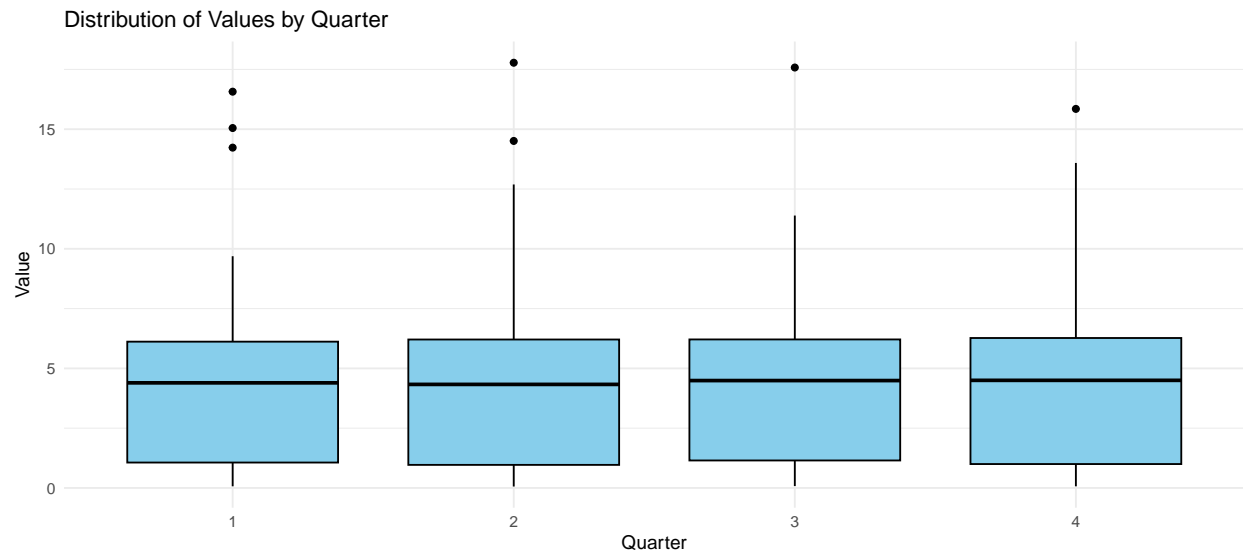
Outlier Detection

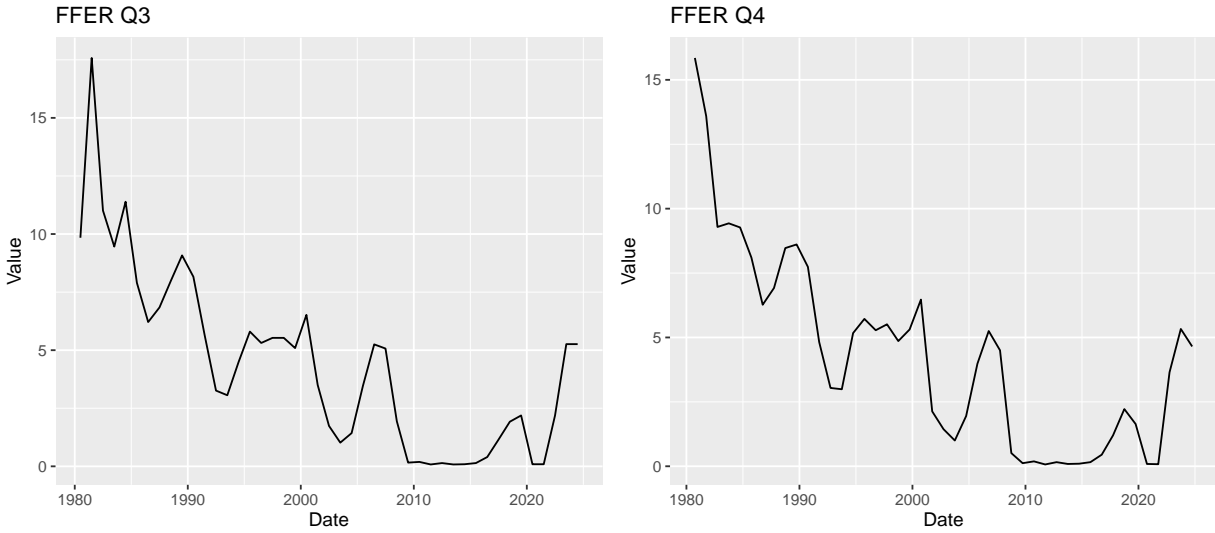
We utilize a Hampel filter for the FFER, PECPI, GDP, GDPPOT, and output gap timeseries. A window size of 5 appeared to smooth out outliers best upon a visual examination of the corrected timeseries data. While, we ran a Hampel filters to detect potential outliers, we retained the original FFER series for estimation because large rate changes reflect actual policy, not data errors.



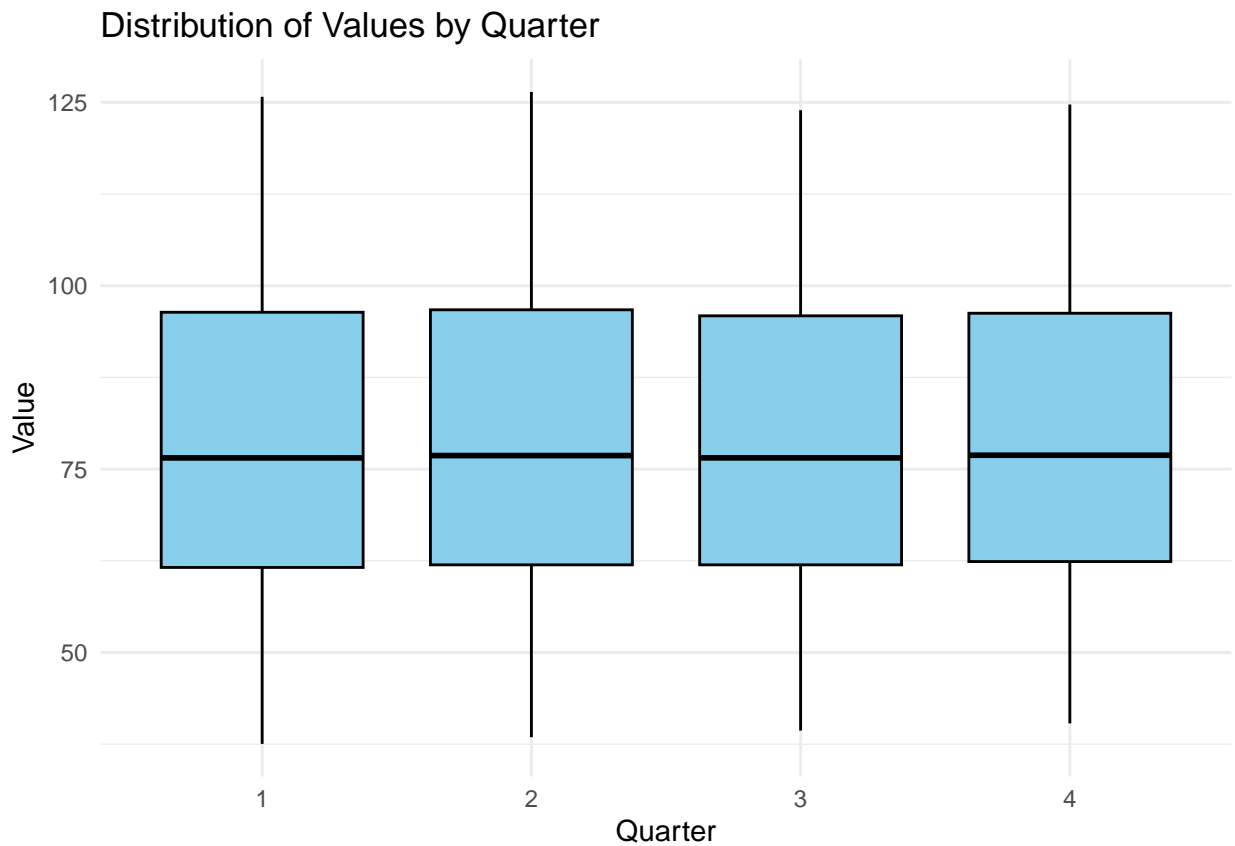
Task 2.3: Seasonal Adjustment

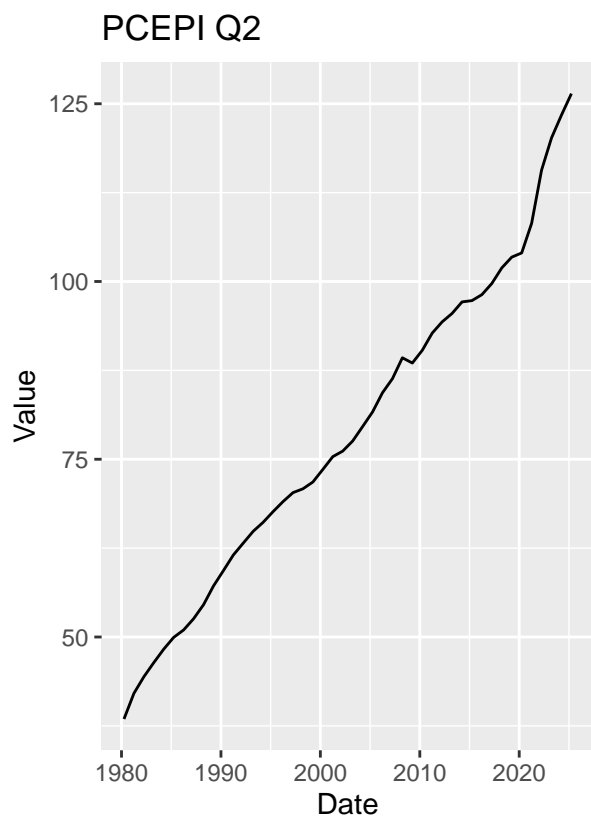
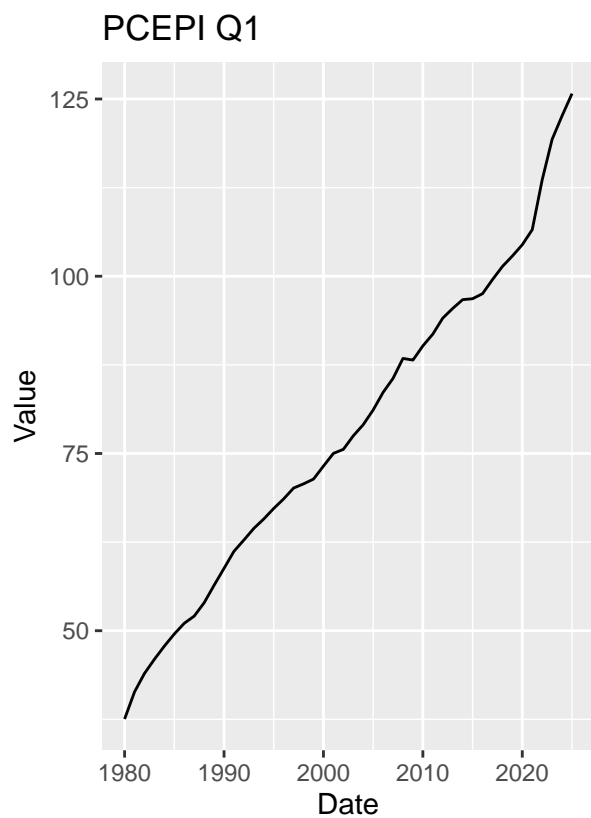
Detection

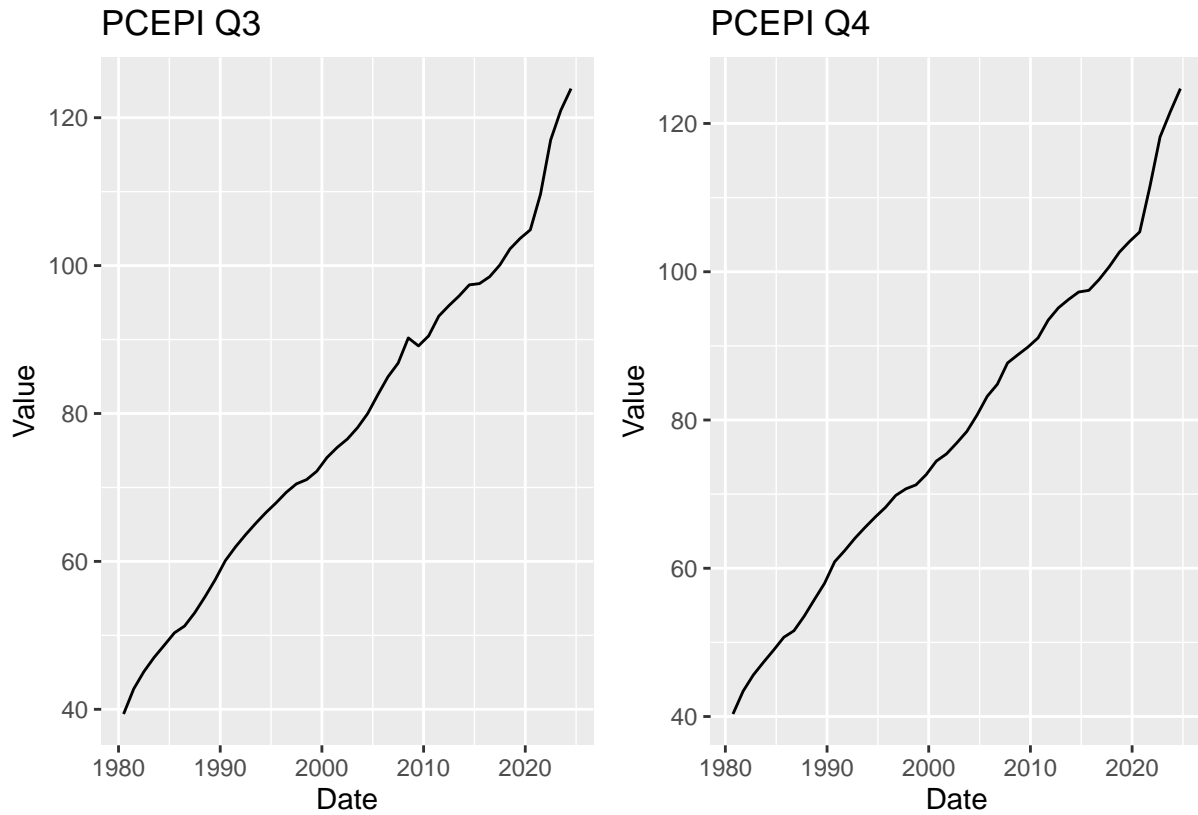




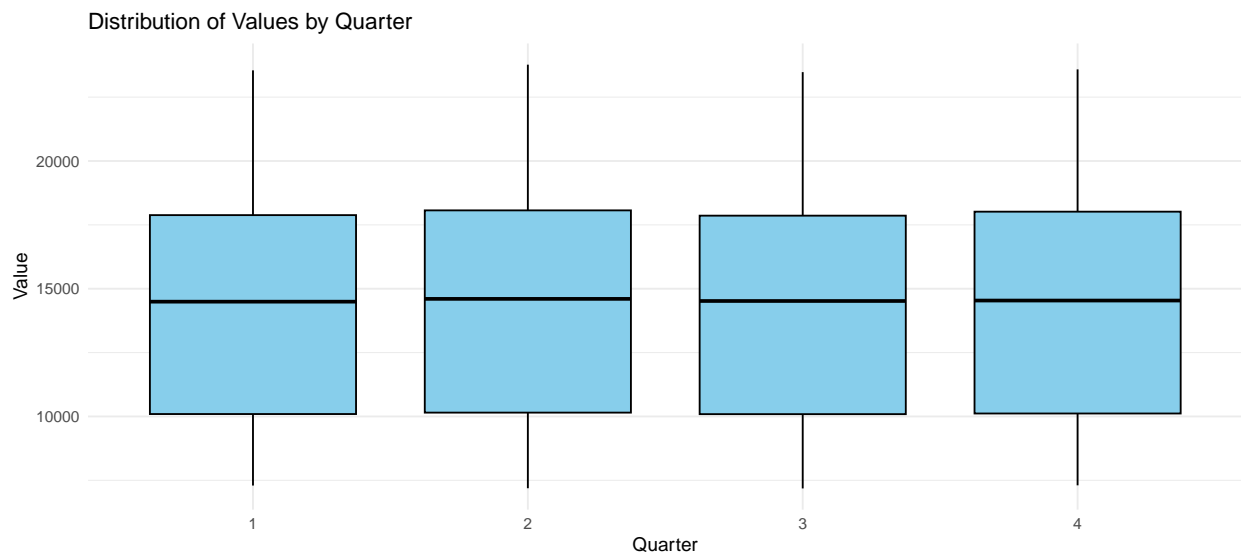
The QS test for seasonality applied to the Federal Funds Effective Rate (FFER) produced a test statistic of 0 with a p-value of 1. This result provides no evidence of seasonality in the FFER series. In other words, the Federal Funds Rate does not exhibit systematic quarterly or annual seasonal fluctuations. This is consistent with the institutional nature of monetary policy decisions—interest rate adjustments occur in response to macroeconomic conditions rather than regular seasonal cycles. Consequently, no seasonal adjustment or differencing is necessary for the FFER series before modeling.

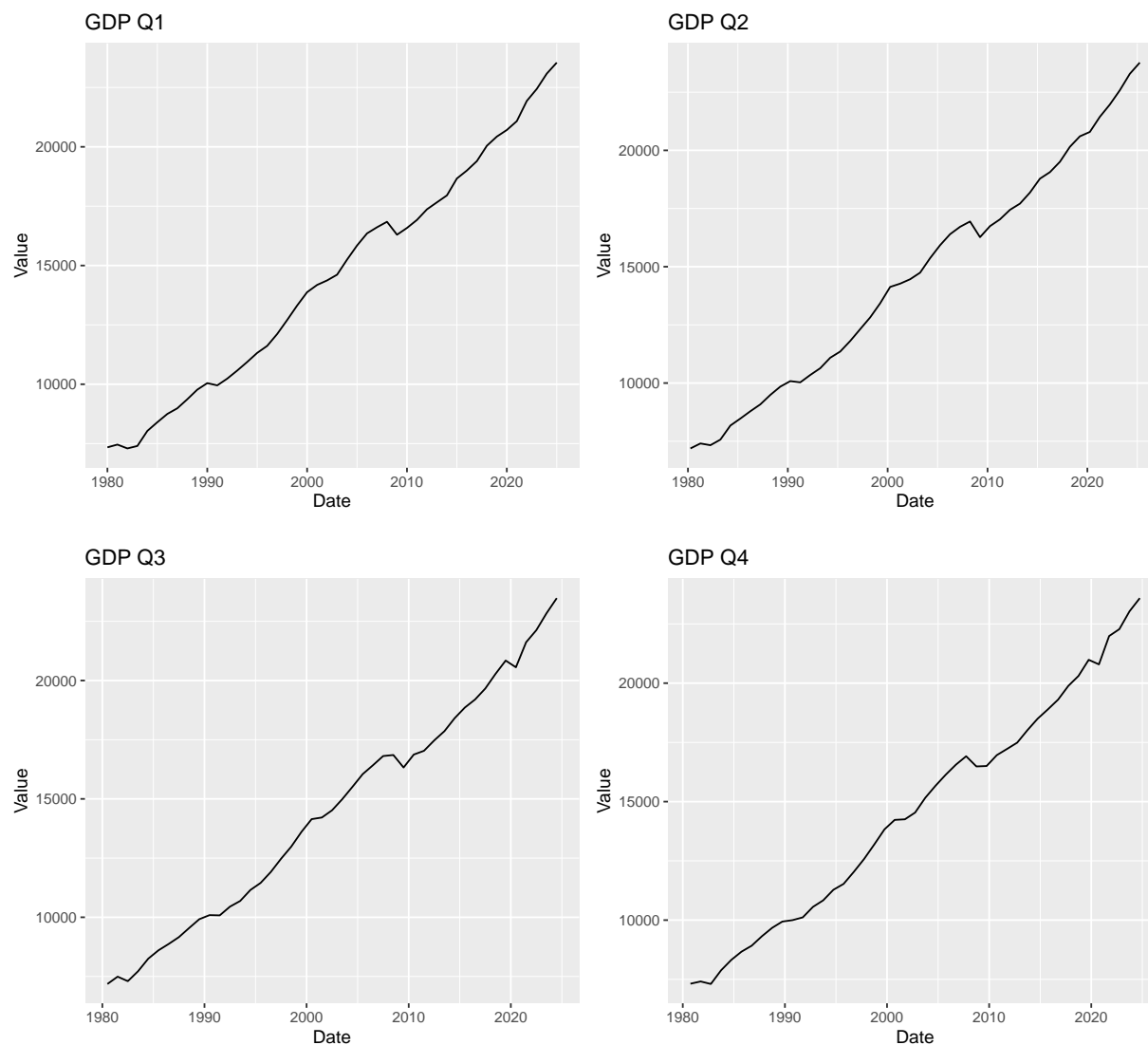






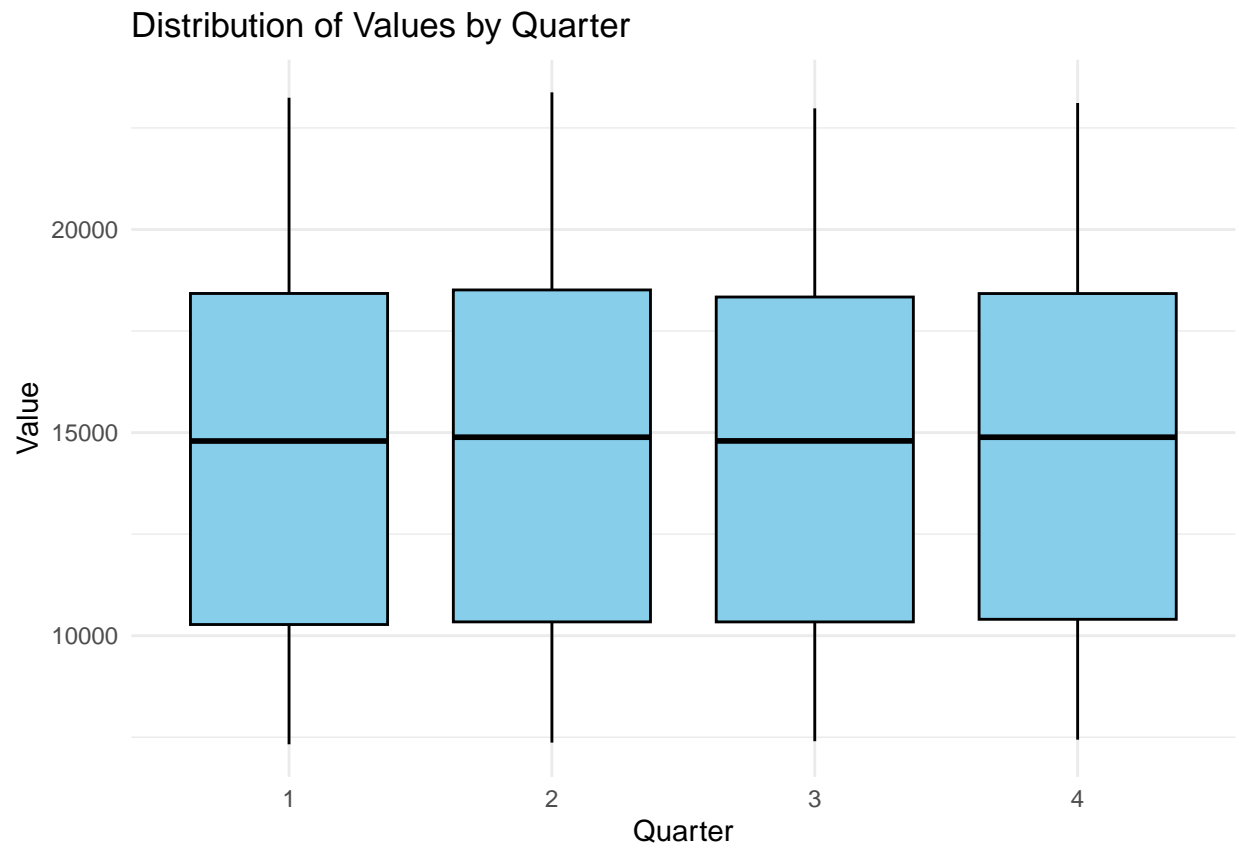
The QS test in the quarterly PCEPI inflation series produced a statistic of 16.15 with a p-value of 0.00031, indicating statistically significant seasonal effects. Therefore, the raw series exhibits recurring within-year variation, justifying the use of seasonal differencing or seasonal adjustment before fitting ARIMA-type models.

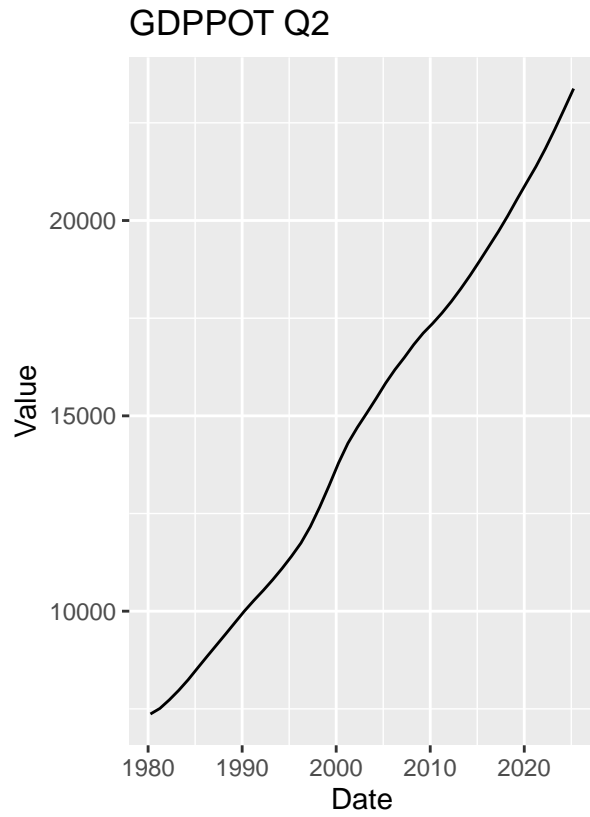
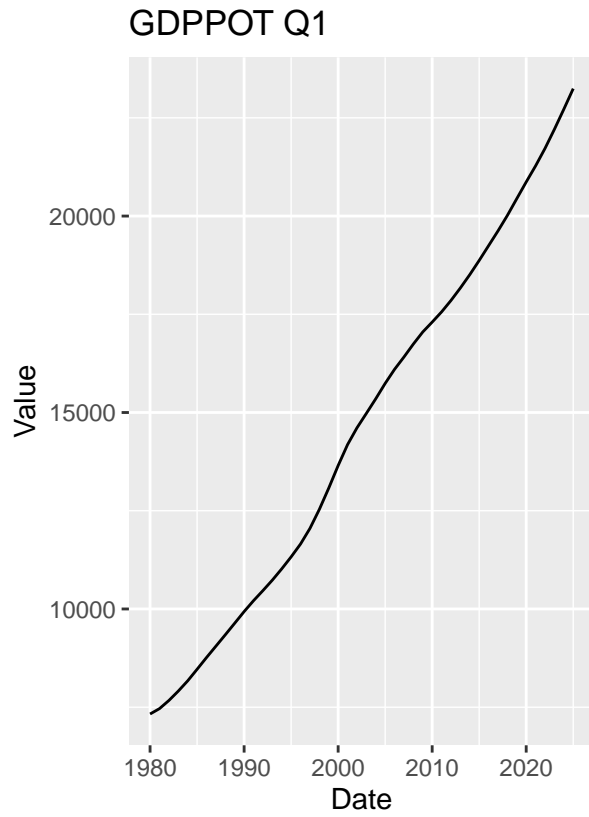


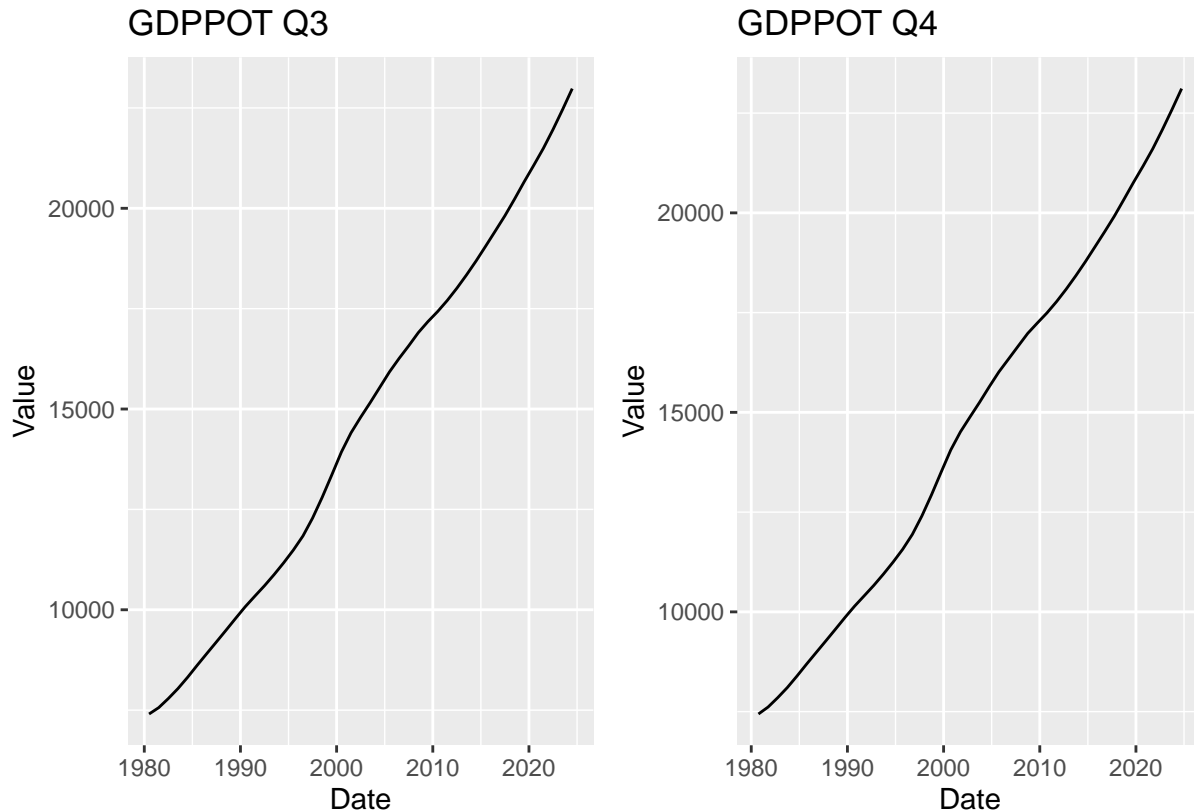


Here, our QS seasonality test for the Real GDP series yielded a test statistic of 0 and a p-value of 1, indicating no detectable seasonal pattern in the data. This result suggests that GDP, as reported in seasonally adjusted form by the Bureau of Economic Analysis (BEA), already has its intra-year seasonal components removed. Therefore, no additional seasonal adjustment or seasonal differencing is necessary prior to modeling.

This is consistent with standard macroeconomic data practices: most official GDP series are *seasonally adjusted annualized rates (SAAR)*, meaning any regular quarterly effects have already been filtered out by the data provider.







The QS test for the Potential GDP (GDPPOT) series produced a test statistic of 204.06 with a p-value of 0, providing overwhelming evidence of seasonality. This result indicates that the GDPPOT series, as currently specified, contains recurring within-year fluctuations that are not seasonally adjusted.

In practice, this is somewhat expected: the *potential output* series is an estimated smooth trend of what the economy could produce at full capacity, but depending on how it was constructed or interpolated, the quarterly data may still exhibit minor cyclical or deterministic seasonal behavior.

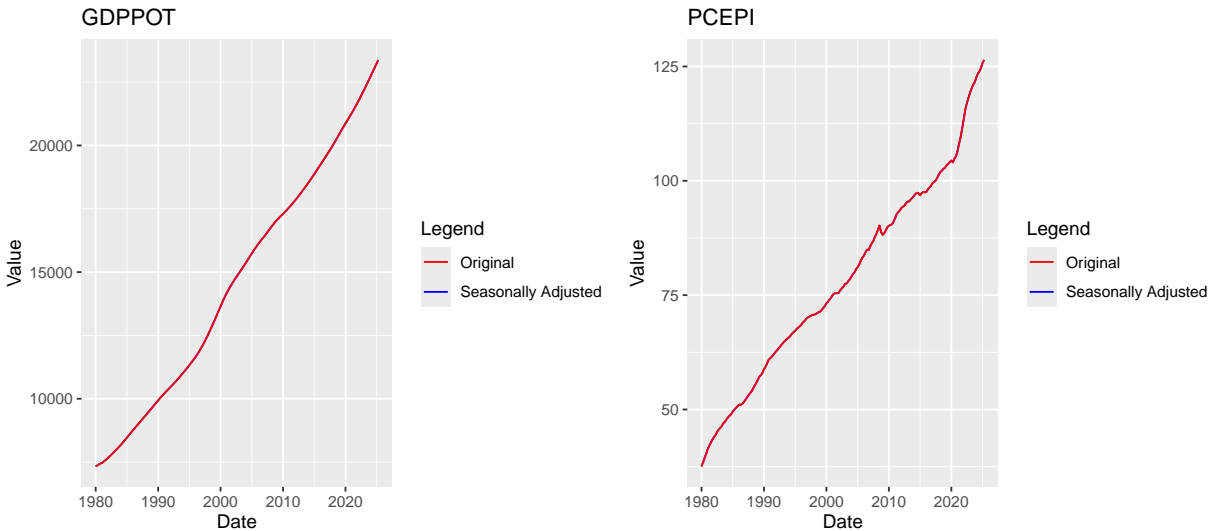
Therefore, to ensure consistency with other macroeconomic variables in the model (such as Real GDP and FFER), it may be appropriate to apply **seasonal differencing** or a **low-pass filter** (e.g., HP filter) to the GDPPOT series before computing the output gap or including it in time series models.

Seasonal Adjustment via X-13ARIMA-SEATS

This method is used because it allows for timeseries that can be decomposed additively or multiplicatively, allowing for maximum flexibility. The trend, seasonal component, and irregular component are all estimated and the algorithm uses centered moving averages to remove the seasonal component.

Source: <https://en.wikipedia.org/wiki/X-13ARIMA-SEATS>

```
## Don't know how to automatically pick scale for object of type <ts>. Defaulting
## to continuous.
## Don't know how to automatically pick scale for object of type <ts>. Defaulting
## to continuous.
```



Task 3

```
vars <- names(select(infl_data_tsbl_train, where(is.numeric)))

results <- data.frame(
  variable      = vars,
  stationary_3 = NA,
  stationary_2 = NA
)

for (i in seq_along(vars)) {
  x <- infl_data_tsbl_train[[vars[i]]]
  res <- aTSA::adf.test(x, output = FALSE)

  # Evaluate if evidence of Trend and Drift
  p3 <- min(res$type3[, "p.value"], na.rm = TRUE)
  results$stationary_3[i] <- p3 < 0.05

  # Evaluate for Drift only
  p2 <- min(res$type2[, "p.value"], na.rm = TRUE)
  results$stationary_2[i] <- p2 < 0.05
}
```

```
## Warning in summary.lm(m1): essentially perfect fit: summary may be unreliable
## Warning in summary.lm(m1): essentially perfect fit: summary may be unreliable
## Warning in summary.lm(m1): essentially perfect fit: summary may be unreliable
## Warning in summary.lm(m1): essentially perfect fit: summary may be unreliable
## Warning in summary.lm(m1): essentially perfect fit: summary may be unreliable
## Warning in summary.lm(m1): essentially perfect fit: summary may be unreliable

## Warning in summary.lm(m2): essentially perfect fit: summary may be unreliable
## Warning in summary.lm(m2): essentially perfect fit: summary may be unreliable

## Warning in summary.lm(m3): essentially perfect fit: summary may be unreliable
## Warning in summary.lm(m3): essentially perfect fit: summary may be unreliable
```

```
## Warning in summary.lm(m1): essentially perfect fit: summary may be unreliable
## Warning in summary.lm(m1): essentially perfect fit: summary may be unreliable
```

```
as_tibble(results)
```

```
## # A tibble: 9 x 3
##   variable      stationary_3 stationary_2
##   <chr>          <lgl>          <lgl>
## 1 FEDFUNDS_hampel TRUE           TRUE
## 2 PCEPI_s_adj    FALSE          FALSE
## 3 pct_chg_pce    TRUE           TRUE
## 4 GDPC1_hampel   FALSE          FALSE
## 5 GDPPOT_s_adj   TRUE           FALSE
## 6 output_gap     TRUE           TRUE
## 7 pct_chg_pce_s_adj TRUE          TRUE
## 8 inflation_gap  TRUE           TRUE
## 9 date           FALSE          FALSE
```

Discussion:

Using the General to Specific approach, we first evaluate for drift and trend terms for each variable in the dataset. For this test, only the Seasonally Adjusted PCEPI and the GDPC with hampel filter variables fail to reject the null hypothesis of stationarity. From there, we evaluate all variable for drift only. This time, the seasonally adjusted GDPPOT variable also fails to reject the null, in addition to the aforementioned variables.

These test results suggest that the PCEPI and GDPC variables should be differenced before use in modeling, to avoid spurious regression. The GDPPOT variable is trend-stationary because it does not contain a unit-root but it does contain a deterministic trend. In order to capture the explanatory power of the variable, GDPPOT_s_adj should not be differenced.

```
infl_data_tsbl_train <- infl_data_tsbl_train %>%
  mutate(
    d_FEDFUNDS_hampel = difference(FEDFUNDS_hampel),
    d_PCEPI_s_adj = difference(PCEPI_s_adj),
    d_GDPC1_hampel = difference(GDPC1_hampel),
    d_GDPPOT_s_adj = difference(GDPPOT_s_adj)
  )
```

```
adf.test(infl_data_tsbl_train$d_PCEPI_s_adj)
```

```
## Augmented Dickey-Fuller Test
## alternative: stationary
##
## Type 1: no drift no trend
##      lag      ADF p.value
## [1,]  0 -3.713  0.0100
## [2,]  1 -2.815  0.0100
## [3,]  2 -1.896  0.0581
## [4,]  3 -1.825  0.0686
## [5,]  4 -0.944  0.3408
## Type 2: with drift no trend
##      lag      ADF p.value
## [1,]  0 -6.32  0.0100
## [2,]  1 -5.11  0.0100
## [3,]  2 -3.78  0.0100
## [4,]  3 -3.89  0.0100
## [5,]  4 -2.63  0.0926
```

```
## Type 3: with drift and trend
##      lag    ADF p.value
## [1,]   0 -6.38  0.0100
## [2,]   1 -5.19  0.0100
## [3,]   2 -3.87  0.0174
## [4,]   3 -3.96  0.0130
## [5,]   4 -2.67  0.2968
## ----
## Note: in fact, p.value = 0.01 means p.value <= 0.01
adf.test(infl_data_tsbl_train$d_GDPC1_hampel)
```

```
## Augmented Dickey-Fuller Test
## alternative: stationary
##
## Type 1: no drift no trend
##      lag    ADF p.value
## [1,]   0 -6.39   0.01
## [2,]   1 -3.90   0.01
## [3,]   2 -3.61   0.01
## [4,]   3 -3.27   0.01
## [5,]   4 -2.76   0.01
## Type 2: with drift no trend
##      lag    ADF p.value
## [1,]   0 -9.81   0.01
## [2,]   1 -6.57   0.01
## [3,]   2 -6.35   0.01
## [4,]   3 -6.09   0.01
## [5,]   4 -5.86   0.01
## Type 3: with drift and trend
##      lag    ADF p.value
## [1,]   0 -9.91   0.01
## [2,]   1 -6.64   0.01
## [3,]   2 -6.50   0.01
## [4,]   3 -6.31   0.01
## [5,]   4 -6.04   0.01
## ----
## Note: in fact, p.value = 0.01 means p.value <= 0.01
```

Discussion

After differencing the two variables identified by the ADF test above, they no longer show trend or drift. Notably, there are some lags for both variables that result in a failure to reject the null hypothesis. Therefore care must be taken when determining which lags to use when modeling with these variables to ensure they remain stationary.

```
za_results <- map_dfr(vars, function(v) {
  x <- infl_data_tsbl_train[[v]]
  test <- ur.za(x, model = "both")
  result <- summary(test)
  tibble(
    variable = v,
    model = test@model[1],
    sig_thresh = result@cval[2],
    test_res = result@teststat,
```

```

    stry_w_brk = result@teststat < result@cval[2]
  )
})

za_results

## # A tibble: 9 x 5
##   variable      model sig_thresh test_res stry_w_brk
##   <chr>         <chr>      <dbl>    <dbl> <lgl>
## 1 FEDFUNDS_hampel both      -5.08   -4.66 FALSE
## 2 PCEPI_s_adj    both      -5.08   -4.18 FALSE
## 3 pct_chg_pce    both      -5.08   -7.78 TRUE
## 4 GDPC1_hampel   both      -5.08   -4.42 FALSE
## 5 GDPPOT_s_adj   both      -5.08   -5.03 FALSE
## 6 output_gap     both      -5.08   -4.23 FALSE
## 7 pct_chg_pce_s_adj both      -5.08   -7.78 TRUE
## 8 inflation_gap  both      -5.08   -7.78 TRUE
## 9 date          both      -5.08  -11.5 TRUE

```

Discussion:

In comparison to the ADF test, the Zivot Andrews test identifies more variables that are not stationary when accounting for structural breaks in the data. The FEDFUNDS_hampel, PCEPI_s_adj, GDPC1_hampel, GDPPOT_s_adj and output_gap variables fail to reject the null hypothesis of stationarity with a structural break.

This means that the additional variables identified over the ADF only appear non-stationary because of a structural break. Once that break is included, the variables are found to be non-stationary. They should be differenced to reduce issues with model inference due to non-stationary residuals or spurious regression.

Task 4

```

taylor_mod <- infl_data_tsbl_train %>%
  model(tlsm = TSLM(FEDFUNDS_hampel ~ inflation_gap + output_gap))
report(taylor_mod)

## Series: FEDFUNDS_hampel
## Model: TSLM
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -9.33831 -2.50848 -0.01132  2.21617  9.51659
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    9.8247     0.6965  14.106 < 2e-16 ***
## inflation_gap 406.7634    48.3740   8.409 1.71e-14 ***
## output_gap     0.1176     0.1357   0.866  0.388
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.328 on 168 degrees of freedom
## Multiple R-squared:  0.2999, Adjusted R-squared:  0.2915
## F-statistic: 35.98 on 2 and 168 DF, p-value: 9.8768e-14

```

```
taylor_mod %>%
  gg_tsresiduals()
```

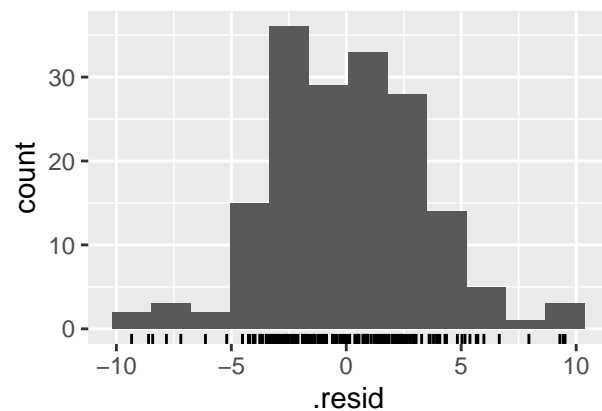
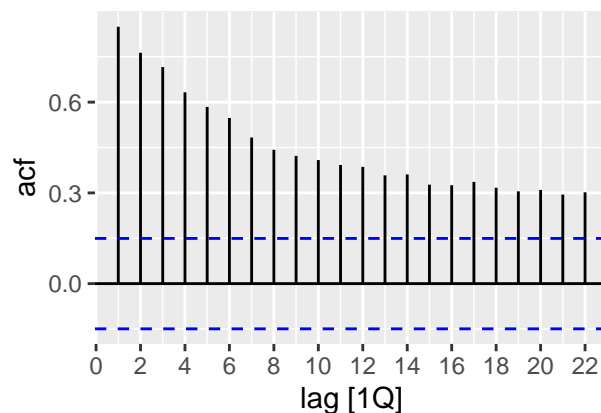
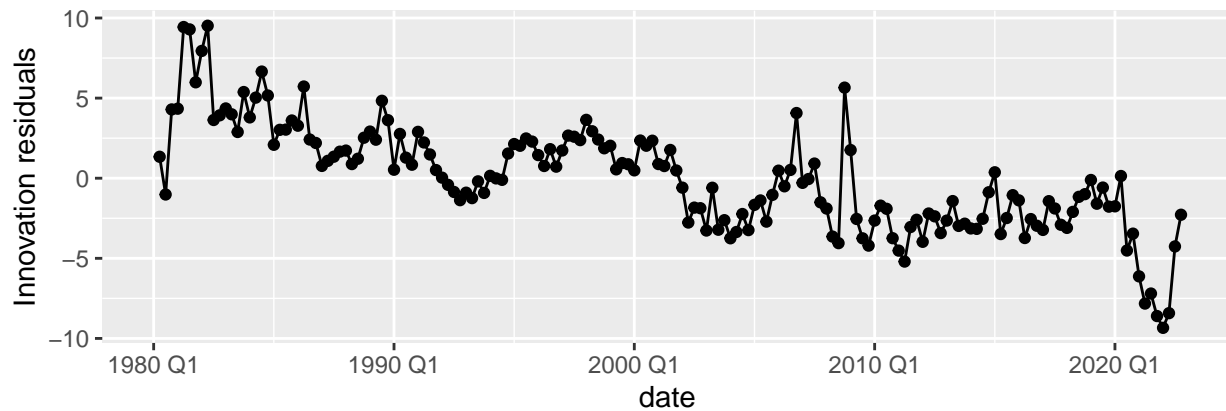
```
## Warning: `gg_tsresiduals()` was deprecated in feasts 0.4.2.
## i Please use `ggtime::gg_tsresiduals()` instead.
## This warning is displayed once every 8 hours.
## Call `lifecycle::last_lifecycle_warnings()` to see where this warning was
## generated.

## Warning: Removed 1 row containing missing values or values outside the scale range
## (`geom_line()`).

## Warning: Removed 1 row containing missing values or values outside the scale range
## (`geom_point()`).

## Warning: Removed 1 row containing non-finite outside the scale range
## (`stat_bin()`).

## Warning: Removed 1 row containing missing values or values outside the scale range
## (`geom_rug()`).
```



Discussion:

Previous analysis on the variables indicated that they should be differenced before use in regression. However doing so results in a different interpretation of the Taylor Rule. The non-stationary variables were used so that discussion could revolve around the models adherence to economic theory.

With this model, the inflation gap estimate is significant with a value of 406.8. This variable is reported in hundredths, so the correct interpretation about a 4% increase for every 1% inflation point above the inflation

target. This does match economic theory, in that the Fed Fund rate increases as inflation increases in an effort to “cool off” the economy. This estimate is slightly high in comparison to real targets of 2%. In general, this means that in this sample, the funds rate reacts much more strongly to inflation than what would be considered the “rule of thumb”. This may be due to the several economic shocks that occurred through the training time periods.

The Autocorrelation plots for this model shows that the residuals are significant across all lags. This shows that there is additional information that is not captured with the model, as the residuals are not white noise.

```
accuracy(taylor_mod)$RMSE
```

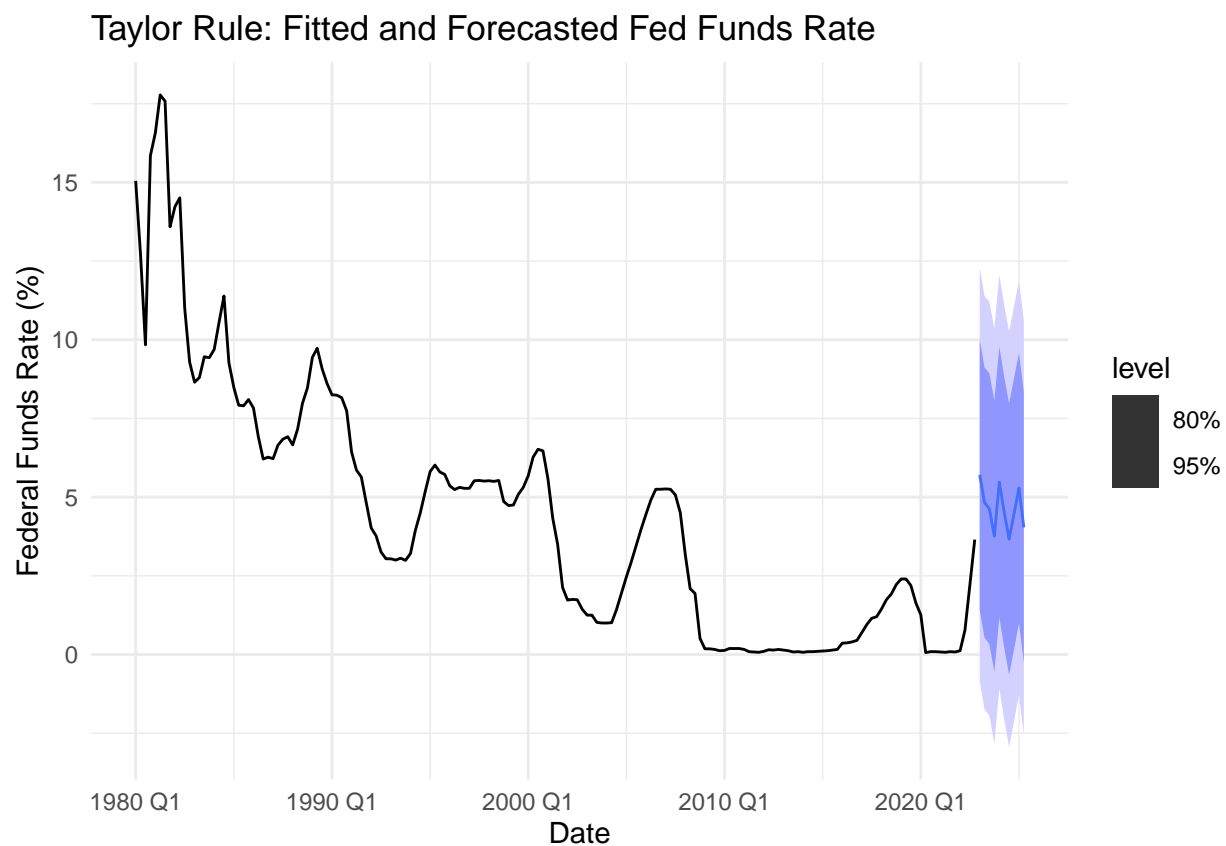
```
## [1] 3.29877
```

```
taylor_fc <- forecast(taylor_mod, new_data = infl_data_tsbl_test)
```

```
accuracy(taylor_fc, infl_data_tsbl_test)$RMSE
```

```
## [1] 0.9260019
```

```
autoplot(taylor_fc, infl_data_tsbl_train) +  
  labs(  
    title = "Taylor Rule: Fitted and Forecasted Fed Funds Rate",  
    x = "Date",  
    y = "Federal Funds Rate (%)")  
  ) +  
  theme_minimal()
```



Discussion:

The Root Mean Squared Error value for the training data is 3.29. For the test data, the value is 0.92. So for the training data, the model is generally within 3.3 percentage points of the actual data, a pretty large spread overall. However this tightens to within 1 percentage point for the test data.

This difference may be attributed to the relative calm in the market during the test period compared to the training period. For example, the training period contains large abnormalities, like the Dot-Com bubble, housing crisis, and COVID, all within nearly 20 years of the end of the training data. The test data has no such unrest, and is therefore able to more “accurate” in describing the fed rate response to inflation. For these reasons, it is likely improper to suggest that this is an optimized model for future forecasting.

Task 5

5.1: Cointegration Test (5 Points)

```
## Tom Commented this code after creating the model in task 4
# taylor_mod <- lm(FEDFUNDS_hampel ~ inflation_gap + output_gap,
#                 data = infl_data_tsbl_train)
# summary(taylor_mod)
```

Anthony comment - theres an error here. `taylor_mod` above is a MABLE not a numeric vector. Need the

```
# taylor_resid <- resid(taylor_mod)
# adf.test(taylor_resid)
# corrected code below -
```

```
taylor_resid_tsbl <- residuals(taylor_mod)
```

```
taylor_resid <- taylor_resid_tsbl %>%
  filter(.model == "tlsm") %>%
  pull(.resid)
```

```
adf.test(taylor_resid)
```

```
## Augmented Dickey-Fuller Test
## alternative: stationary
##
## Type 1: no drift no trend
##      lag    ADF p.value
## [1,]   0 -3.67  0.0100
## [2,]   1 -2.94  0.0100
## [3,]   2 -2.35  0.0204
## [4,]   3 -2.51  0.0133
## [5,]   4 -2.43  0.0171
## Type 2: with drift no trend
##      lag    ADF p.value
## [1,]   0 -3.66  0.0100
## [2,]   1 -2.93  0.0463
## [3,]   2 -2.33  0.1981
## [4,]   3 -2.50  0.1348
## [5,]   4 -2.39  0.1765
## Type 3: with drift and trend
##      lag    ADF p.value
## [1,]   0 -6.39   0.01
```

```
## [2,] 1 -5.85 0.01
## [3,] 2 -4.58 0.01
## [4,] 3 -4.93 0.01
## [5,] 4 -4.19 0.01
## ----
## Note: in fact, p.value = 0.01 means p.value <= 0.01
```

Since the p value is 0.01024 and less than 0.05, we reject the null hypothesis that the series has a unit root. This means that the residuals are stationary and there is a long-run equilibrium relationship.

5.2: Estimating the ECM via OLS (5 Points)

```
# full residuals to add back to tsibble
long_run_model <- lm(FEDFUNDS_hampel ~ inflation_gap + output_gap,
                     data = infl_data_tsbl)
long_run_resid <- c(NA, resid(long_run_model))

# add ECM variables
infl_data_tsbl_ecm <- infl_data_tsbl %>%
  mutate(
    d_it = FEDFUNDS_hampel - lag(FEDFUNDS_hampel),
    d_infl_gap = inflation_gap - lag(inflation_gap),
    d_output_gap = output_gap - lag(output_gap),
    lag_resid = lag(long_run_resid)
  )

infl_data_tsbl_ecm_train <- infl_data_tsbl_ecm %>%
  filter(date <= yearquarter("2022 Q4"))
infl_data_tsbl_ecm_test <- infl_data_tsbl_ecm %>%
  filter(date > yearquarter("2022 Q4"))

ecm_model <- lm(d_it ~ d_infl_gap + d_output_gap + lag_resid,
               data = infl_data_tsbl_ecm)
summary(ecm_model)

##
## Call:
## lm(formula = d_it ~ d_infl_gap + d_output_gap + lag_resid, data = infl_data_tsbl_ecm)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.6049 -0.3062 -0.0009  0.2609  5.4345
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  -0.05479    0.05493  -0.997 0.319938
## d_infl_gap    12.63397    14.01487   0.901 0.368571
## d_output_gap   0.42169     0.08784   4.801 3.36e-06 ***
## lag_resid    -0.06119     0.01750  -3.497 0.000596 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.7361 on 176 degrees of freedom
## (2 observations deleted due to missingness)
## Multiple R-squared:  0.1956, Adjusted R-squared:  0.1819
```

```
## F-statistic: 14.27 on 3 and 176 DF, p-value: 2.3e-08
```

5.3: Compare ECM Forecasts (5 Points)

```
ecm_model_train <- lm(d_it ~ d_infl_gap + d_output_gap + lag_resid,
                      data = infl_data_tsbl_ecm_train)
summary(ecm_model_train)

##
## Call:
## lm(formula = d_it ~ d_infl_gap + d_output_gap + lag_resid, data = infl_data_tsbl_ecm_train)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.6112 -0.3227  0.0011  0.2753  5.4414
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  -0.06031    0.05787  -1.042  0.29884
## d_infl_gap    12.70562   14.53533   0.874  0.38332
## d_output_gap   0.42150    0.09111   4.627 7.45e-06 ***
## lag_resid     -0.05995    0.01798  -3.334  0.00105 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.7539 on 166 degrees of freedom
## (2 observations deleted due to missingness)
## Multiple R-squared:  0.1932, Adjusted R-squared:  0.1786
## F-statistic: 13.25 on 3 and 166 DF, p-value: 8.524e-08

# forecast d_it
ecm_fc <- forecast(ecm_model_train, newdata = as.data.frame(infl_data_tsbl_ecm_test))

d_hat <- ecm_fc$mean

# Convert to level forecasts
i_hat <- numeric(length(d_hat))
# Start with last value of train data and recursively add
i_hat[1] <- tail(infl_data_tsbl_ecm_train$FEDFUNDS_hampel, 1) + d_hat[1]
for (t in 2:length(d_hat)) {
  i_hat[t] <- i_hat[t - 1] + d_hat[t]
}

# We dont show the in-sample performance
ecm_rmse <- sqrt(mean((infl_data_tsbl_ecm_test$FEDFUNDS_hampel - i_hat)^2, na.rm = TRUE))

ecm_rmse

## [1] 1.272986

# Anthony comment - this is not working because taylor_mod is a table not numeric
# ols_fc <- forecast(taylor_mod, newdata = infl_data_tsbl_test)
# ols_pred <- ols_fc$mean
# ols_rmse <- sqrt(mean((infl_data_tsbl_test$FEDFUNDS_hampel - ols_pred)^2, na.rm = TRUE))
# ols_rmse
```

```
ols_lm <- lm(FEDFUNDS_hampel ~ inflation_gap + output_gap, data = infl_data_tsbl_train)

ols_fc <- forecast(ols_lm, newdata = as.data.frame(infl_data_tsbl_test))

ols_pred <- as.numeric(ols_fc$mean)

ols_rmse <- sqrt(mean(
  (infl_data_tsbl_test$FEDFUNDS_hampel - ols_pred)^2,
  na.rm = TRUE
))

ols_rmse

## [1] 0.9260019
```

The ECM model performs slightly worse with a higher RMSE.

Task 6: ARIMA Modeling of Taylor Rule Errors

To build on work completed above, we provide additional stationarity tests on the residuals. Here, applied three complementary unit-root diagnostics to the Taylor Rule residuals, denoted $\{\hat{\varepsilon}_t\}$. The **KPSS** test (null H_0 : stationarity) indicates that the series is stationary in levels once a deterministic trend is allowed. Specifically, the level-stationarity version yields $p \geq 0.10$, and the trend-stationarity version also yields $p \geq 0.10$, while the “with drift, no trend” variant rejects with $p \leq 0.01$. Taken together, these results suggest that $\hat{\varepsilon}_t$ is **trend-stationary** rather than containing a stochastic trend: deviations are mean-reverting around a deterministic component. Formally, we fail to reject stationarity for $\hat{\varepsilon}_t = \mu + \tau t + u_t$ with u_t covariance-stationary.

Allowing for an endogenous structural change, the **Zivot–Andrews** test rejects the unit-root null once a single break in intercept and trend is permitted. The test statistic is -7.09 , which is more negative than the 1% critical value (-5.57), implying stationarity conditional on one regime shift. The estimated break occurs early in the sample (position ≈ 9), consistent with a Volcker-era policy shift [1]. Economically, this means the Taylor Rule relationship is stable within regimes but parameters likely changed once, so that $\hat{\varepsilon}_t$ behaves as a stationary process after accounting for that break.

The **Phillips–Perron** test, which shares the ADF null H_0 unit root but is robust to serial correlation and heteroskedasticity, also rejects nonstationarity decisively ($Z\text{-tau} = -6.41$, beyond the 1% critical value). This corroborates that the residual process does not harbor a stochastic trend once inflation and output gaps are included on the right-hand side.

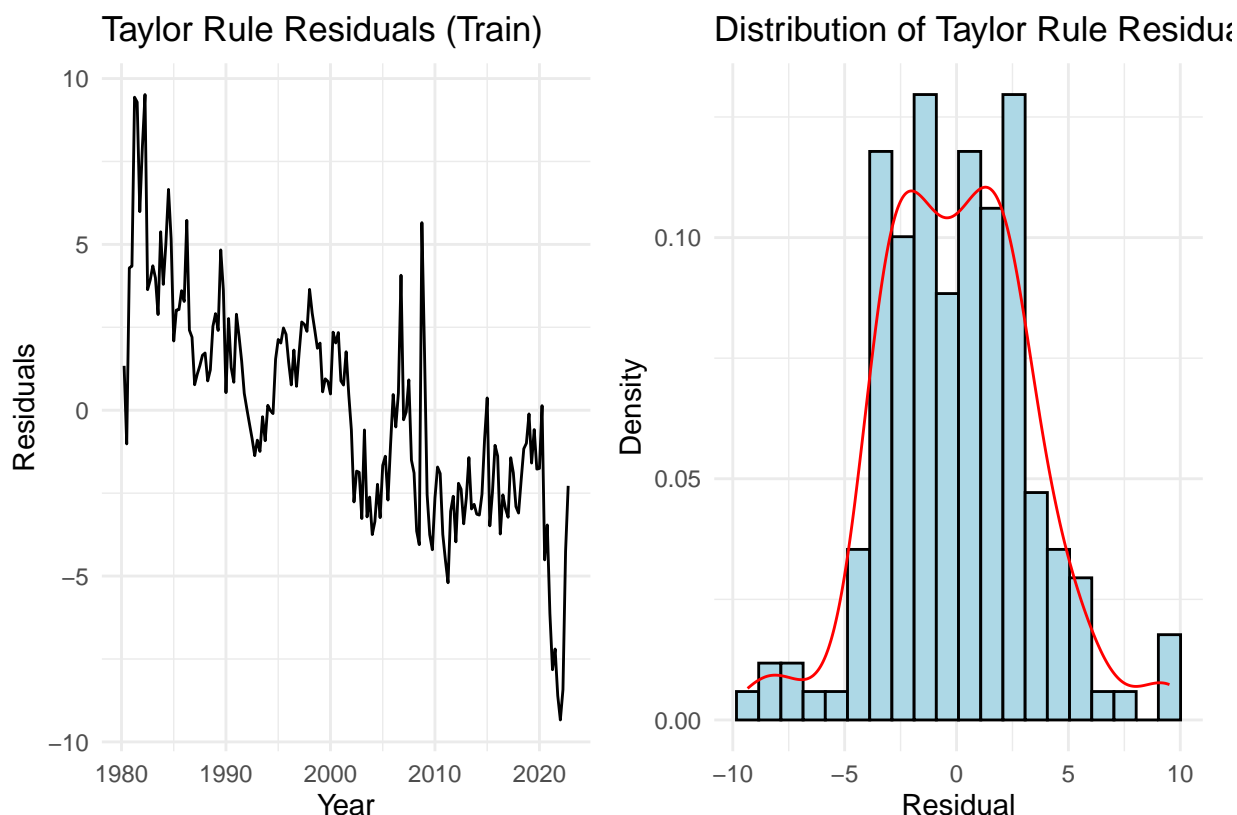
Overall, the three diagnostics are mutually consistent: $\hat{\varepsilon}_t$ is best characterized as **stationary**, possibly around a deterministic trend and with evidence of a **single structural break**. This justifies modeling the unexplained component with a low-order ARIMA on the residuals. However, we take the route often taken in practice, where differencing once (ARIMA with $d = 1$) can act as a convenient device to remove the deterministic drift and any slow-moving break-induced level shift, yielding a stationary innovation process for forecasting. Consequently, the combined forecast we report later follows

$$\hat{i}_t^{\text{combined}} = \hat{i}_t^{\text{Taylor}} + \hat{\varepsilon}_t^{\text{ARIMA}},$$

where $\hat{i}_t^{\text{Taylor}}$ is the OLS Taylor Rule fit and $\hat{\varepsilon}_t^{\text{ARIMA}}$ is the ARIMA forecast of the residual process that is stationary within regimes.

```
## Warning: The dot-dot notation (``..density..`) was deprecated in ggplot2 3.4.0.
## i Please use `after_stat(density)` instead.
## This warning is displayed once every 8 hours.
```

```
## Call `lifecycle::last_lifecycle_warnings()` to see where this warning was
## generated.
```



We begin by examining the residual series $\hat{\varepsilon}_t$ from the OLS Taylor Rule model.

The time-series plot of residuals shows long memory and a clear downward drift until the early 2000s, followed by a period of volatility spikes (notably during the 2008 financial crisis and the COVID-19 period). This persistence suggests that the residuals are not white noise and may exhibit autocorrelation or a weak unit root, despite the test statistics on the non-differenced residuals above.

The autocorrelation and partial autocorrelation plots (ACF and PACF) reinforce this observation.

The **ACF** decays slowly, indicating strong persistence in shocks, while the **PACF** shows a sharp cutoff after lag 1. This pattern is consistent with an autoregressive process of order one, AR(1), or potentially an ARIMA(1,1,0) structure with a single difference needed to achieve stationarity.

Such behavior implies that the shocks to the Taylor Rule equation tend to persist over multiple quarters, rather than dissipating immediately.

After differencing the residuals, the time series of $\Delta\hat{\varepsilon}_t$ becomes mean-reverting with stable variance and no clear trend. Visually, this differenced series appears stationary, suggesting that one level of differencing ($d = 1$) is sufficient to remove the long-run component. Formally, both the **Augmented Dickey–Fuller (ADF)** and **KPSS** tests confirm this:

the ADF strongly rejects the null of a unit root (all $p \leq 0.01$), while the KPSS fails to reject the null of stationarity ($p \geq 0.10$). Together, these imply that the differenced residuals are stationary.

To formally capture this structure, we fit several ARIMA models. The AIC and BIC comparison indicates that higher-order mixed models outperform pure AR or MA specifications.

Specifically, **ARIMA(2,1,3)**, chosen by auto-ARIMA, achieves the lowest AIC ($AIC = 669.37$), outperforming all ARIMA models tested.

The estimated parameters show significant AR and MA terms, confirming that both short-run momentum and shock effects are present.

In summary, the residual diagnostics show that:

1. The raw residuals $\hat{\varepsilon}_t$ are non-stationary, consistent with persistent policy shocks.
2. First differencing yields stationary residuals $\Delta\hat{\varepsilon}_t$.
3. The optimal model based on information criteria is **ARIMA(2,1,3)**, which balances autoregressive persistence with moving-average smoothing.

This result justifies the use of the combined model:

$$\hat{i}_t^{\text{combined}} = \hat{i}_t^{\text{Taylor}} + \hat{\varepsilon}_t^{\text{ARIMA}},$$

where the ARIMA component refines the Taylor Rule forecast by modeling serial correlation in the policy deviations. The combined model therefore captures both the systematic and dynamic elements of the Federal Reserve’s interest rate decisions.

The table below compares the RMSE for the three models across the training and test sets:

Model	Train RMSE	Test RMSE
Taylor Rule (OLS)	3.30	0.93
ECM	0.96	1.27
Taylor–ARIMA Combined	1.66	4.00

The **Taylor Rule (OLS)** model provides a relatively loose in-sample fit (RMSE ≈ 3.3) but performs best on the **test set** (RMSE ≈ 0.93). This suggests that while the OLS model does not fully capture short-term fluctuations during the estimation period, it generalizes reasonably well for the more stable post-2022 period, when macroeconomic volatility was subdued.

The **Error Correction Model (ECM)** achieves the lowest training RMSE (≈ 0.96), implying that differencing and inclusion of lagged disequilibrium terms successfully capture most short-run adjustments within the training sample. However, its slightly higher out-of-sample RMSE (≈ 1.27) indicates mild overfitting—expected since the ECM optimizes within-sample dynamic correction rather than long-horizon prediction.

By contrast, the **Taylor–ARIMA Combined** model shows improved in-sample residual structure (RMSE ≈ 1.66) but a larger out-of-sample RMSE (≈ 4.00). This outcome arises because the ARIMA component (ARIMA(2, 1, 3)) is tuned to capture serial correlation in historical policy deviations ($\hat{\varepsilon}_t$), which were highly volatile during the 1980–2020 training period.

When applied to the test period—marked by comparatively stable rates, the ARIMA extrapolation amplifies noise, yielding over-dispersion in forecasts. Which allows us to surmise that while ARIMA enhances the *in-sample* dynamics by filtering autocorrelation, it reduces *out-of-sample robustness* under regime stability.

In economic terms, this suggests that the **Taylor Rule itself remains structurally sound** for recent policy behavior, and that much of the autocorrelation observed historically reflects **temporary shocks or regime-specific persistence**. The combined model’s poorer forecast performance emphasizes the importance of *model parsimony* when the policy environment changes: the additional ARIMA layer, while statistically valid, can misinterpret macroeconomic calm as signal.

The final combined forecast equation is:

$$\hat{i}_t^{\text{combined}} = \hat{i}_t^{\text{Taylor}} + \hat{\varepsilon}_t^{\text{ARIMA}},$$

where $\hat{i}_t^{\text{Taylor}}$ represents the structural component implied by inflation and output gaps, and $\hat{\epsilon}_t^{\text{ARIMA}}$ captures the dynamic residual correction term. Overall, the findings highlight that post-2020 policy behavior is best captured by the **canonical Taylor Rule** rather than by more complex dynamic extensions.

Task 7: VAR Modeling

We treat $\{i_t, (\pi_t - \pi^*), \text{OutputGap}_t\}$ as a joint system and estimate a VAR on **levels**.

7.1 Estimate, Forecast, and Granger Causality

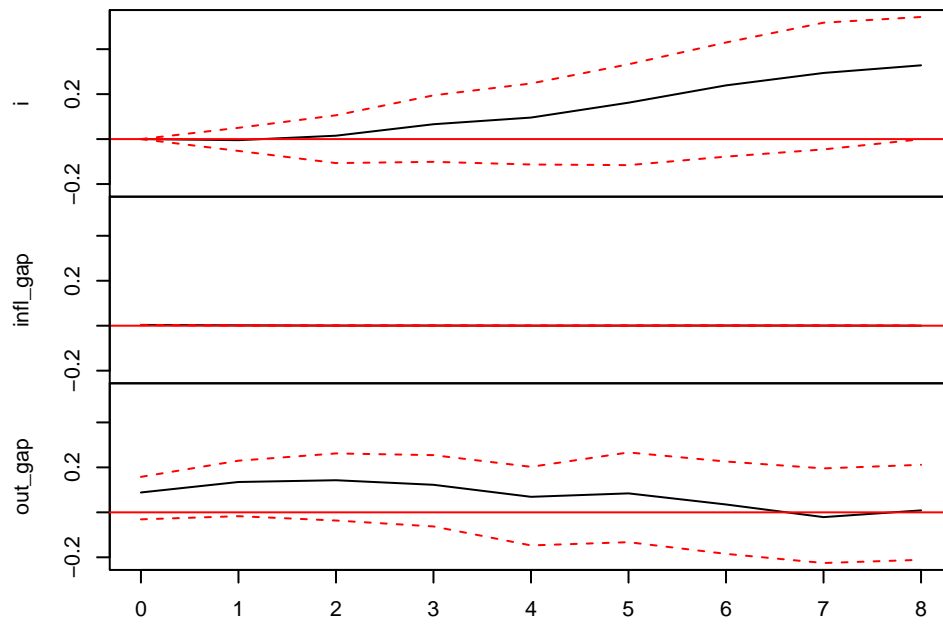
A Vector Autoregression (VAR) was estimated over the system $\{i_t, (\pi_t - \pi^*), \text{OutputGap}_t\}$ using quarterly data from 1980 Q1 to 2022 Q4. The optimal lag length, selected using the Akaike Information Criterion, was eight, implying two years of dynamic feedback among the federal funds rate, inflation gap, and output gap. The resulting model exhibited high explanatory power with $R_i^2 = 0.986$, $R_{\pi\text{gap}}^2 = 0.44$, and $R_{\text{out gap}}^2 = 0.93$, showing that past lags explain most variation in both the policy rate and the output gap, while the inflation gap remains more stochastic. All characteristic roots lie within the unit circle, confirming that the system is dynamically stable and suitable for forecasting.

The model’s predictive accuracy was assessed through in-sample and out-of-sample RMSEs. In the training data, RMSE values were 0.388 for i_t , 0.0032 for $(\pi_t - \pi^*)$, and 0.479 for the output gap. The corresponding test RMSEs were 0.761, 0.0035, and 1.728, respectively. The model captures the dynamics of the federal funds rate with relatively low error, while the inflation gap is predicted with remarkable precision and the output gap displays greater volatility in the test set, particularly in the post-pandemic period. Compared with the Taylor Rule (OLS) and Error-Correction (ECM) specifications, the VAR provides a richer multivariate framework at the cost of added complexity. Its predictive accuracy for i_t sits between the ECM and Taylor–ARIMA combined models, indicating a balanced trade-off between fit and generalization.

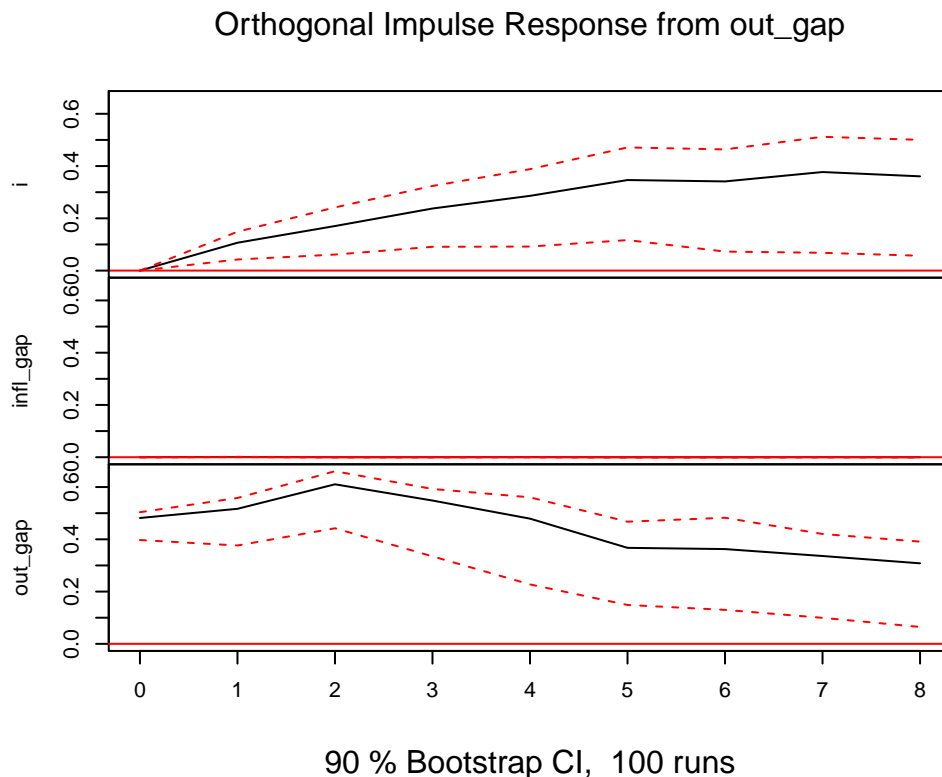
Causality testing further illuminates the system’s internal dynamics. The Granger causality test, with the null hypothesis that inflation and output gaps do not Granger-cause i_t , yielded $F(16, 414) = 1.4004$ and $p = 0.137$, leading to a failure to reject the null. This suggests that lagged information about inflation and real activity does not significantly improve forecasts of the policy rate once its own history is included. However, the instantaneous causality test produced $\chi^2(2) = 20.48$ with $p < 0.001$, implying strong contemporaneous correlations among $\{i_t, (\pi_t - \pi^*), \text{OutputGap}_t\}$. These results support the interpretation that monetary policy actions are largely contemporaneous with prevailing macroeconomic conditions, consistent with a responsive but forward-looking central bank reaction function.

Which indicates strong contemporaneous interaction among $\{i_t, (\pi_t - \pi^*), \text{OutputGap}_t\}$, consistent with the idea that the Fed adjusts rates within the same quarter in response to inflation and output signals. Together, these results imply that policy moves are largely contemporaneous with macroeconomic conditions rather than driven by lagged information.

Orthogonal Impulse Response from infl_gap



90 % Bootstrap CI, 100 runs



```
## integer(0)
```

Next, Impulse Response Functions were generated from the fitted VAR to trace the system's dynamic adjustment to one-time shocks. The IRFs describe how each variable in $\{i_t, (\pi_t - \pi^*), \text{OutputGap}_t\}$ evolves following a temporary disturbance in another variable. The responses were computed over eight quarters with 90% bootstrapped confidence intervals based on 100 replications.

A positive shock to the inflation gap ($\pi_t - \pi^*$) induces an upward response in the federal funds rate i_t over subsequent quarters, consistent with a tightening of monetary policy in reaction to higher inflation. The increase in i_t is gradual during the first year and stabilizes thereafter, indicating a sustained but controlled policy adjustment. The output gap rises slightly in the short run, suggesting a temporary expansion before the contractionary impact of higher interest rates dampens activity. The inflation gap itself displays mild mean reversion, implying that inflation expectations remain anchored despite the initial shock.

When the system experiences a positive disturbance in the output gap, the interest rate again rises for approximately two years, reflecting the Federal Reserve's response to overheating conditions in the real economy. Inflation reacts modestly and within the confidence interval, indicating weak inflationary spillovers from output movements. The output gap gradually converges back to equilibrium within six to eight quarters, demonstrating a stable adjustment path and the self-correcting nature of the system.

Together, the impulse response and causality analyses reveal a coherent monetary-policy transmission mechanism. Inflation and output shocks generate policy-rate responses that are persistent yet mean-reverting, indicating the stabilizing role of monetary policy. The presence of strong instantaneous causality but weak Granger causality suggests that rate adjustments occur in response to current economic information rather than lagged data. Overall, the VAR and IRF evidence reinforce the central principle of the Taylor Rule in a multivariate dynamic setting: the federal funds rate systematically responds to deviations in inflation and real activity to guide the economy back toward equilibrium.

Relative to the Taylor-ARIMA combined model from Task 6, the VAR framework provides a more structural

and interactive representation of monetary policy behavior. While the ARIMA extension improved short-run forecast precision by modeling residual autocorrelation in the Taylor Rule, it remained a univariate correction focused solely on i_t . The VAR, by contrast, endogenizes inflation and the output gap, allowing policy reactions and macroeconomic feedback to emerge jointly within a unified system. The ARIMA-based approach excels at capturing serial dependence but lacks cross-variable dynamics; the VAR overcomes that limitation by explicitly linking shocks across equations. Although the Taylor-ARIMA model achieved a lower test RMSE, the VAR reveals the economic mechanism underlying those patterns: interest-rate adjustments are immediate and persistent responses to contemporaneous inflation and output fluctuations. Thus, the VAR complements the previous models by translating statistical fit into a richer narrative of policy interaction and macroeconomic stabilization.

Appendix