# Time series - Li2Co3 zeroes calculations and statistics

In [ ]:

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
from statsmodels.graphics.tsaplots import plot_acf, plot_pacf
import statsmodels.api as sm
from tabulate import tabulate
from statsmodels.stats.diagnostic import acorr_ljungbox
from statsmodels.tsa.stattools import acf
from google.colab import files
from sklearn.metrics import mean_squared_error
```

# P1. Data Preprocess

In [ ]:

```
# Data from 2017-05-10 to 2024-04-19
li2co3 = pd.read_csv(r'/content/Lithium Carbonate 99%Min China Spot Historical Data (5).csv')
li2co3['Date'] = pd.to_datetime(li2co3['Date'])
# The date order need to be inverted (from early to late)
li2co3 = li2co3.sort_values('Date')
li2co3.set_index('Date', inplace=True)
li2co3 = pd.DataFrame(li2co3["Price"])


li2co3['Price'] = (li2co3['Price'].str.replace(",","").astype(float))
na_count = li2co3['Price'].isna().sum()
print("Number of missing values:", na_count)
if na_count > 0:
  li2co3 = li2co3.dropna(subset=['Price'])

# daily log returns
li2co3['log_ret'] = np.log(li2co3['Price']).diff()
li2co3 = li2co3.dropna(subset=['log_ret'])
```

```
Number of missing values: 0
```

In [ ]:

```
li2co3.head(10)
```

Out[ ]:

| Date | Price | log_ret |
|---|---|---|
| 2017-05-11 | 123000.0 | 0.000000 |
| 2017-05-12 | 123000.0 | 0.000000 |
| 2017-05-15 | 123000.0 | 0.000000 |
| 2017-05-16 | 123000.0 | 0.000000 |
| 2017-05-17 | 123000.0 | 0.000000 |
| 2017-05-18 | 123000.0 | 0.000000 |
| 2017-05-19 | 124000.0 | 0.008097 |
| 2017-05-22 | 124000.0 | 0.000000 |
| 2017-05-23 | 124000.0 | 0.000000 |
| 2017-05-24 | 124000.0 | 0.000000 |

In [ ]:

```
li2co3.tail(10)
```

Out[ ]:

| Date | Price | log_ret |
|---|---|---|
| 2024-04-08 | 109500.0 | 0.027780 |
| 2024-04-09 | 109500.0 | 0.000000 |
| 2024-04-10 | 110500.0 | 0.009091 |
| 2024-04-11 | 109500.0 | -0.009091 |
| 2024-04-12 | 109500.0 | 0.000000 |
| 2024-04-15 | 109500.0 | 0.000000 |
| 2024-04-16 | 109500.0 | 0.000000 |
| 2024-04-17 | 109500.0 | 0.000000 |
| 2024-04-18 | 109500.0 | 0.000000 |
| 2024-04-19 | 109500.0 | 0.000000 |

# P2. Weekly log return series and zero count series

In [ ]:

```
Fridays = pd.DataFrame()
Thursdays = pd.DataFrame()
Wednesdays = pd.DataFrame()
Tuesdays = pd.DataFrame()
Mondays = pd.DataFrame()
```

In [ ]:

```
def weekly_returns(data, chosen_day):
    # chosen_day = ['W-MON', 'W-TUE', 'W-WED', 'W-THU', 'W-FRI']
    weekly_log_return = data.groupby(pd.Grouper(freq=chosen_day))['log_ret'].sum()
    weekly_log_return = weekly_log_return.dropna()
    return weekly_log_return
```

In [ ]:

```
def count_zero(df, chosen_day):
    # chosen_day = ['Monday', 'Tuesday', 'Wednesday', 'Thursday', 'Friday']
    # Create a zero dummy series: 1 if 'log_ret' is zero, 0 otherwise
    data = df.copy()
    data['zero_dummy'] = (data['log_ret'] == 0).astype(int)
    # Compute the sum of past 22 days and past 5 days for zero_dummy
    data['zero_count_22'] = data['zero_dummy'].rolling(window=22).sum()
    data['zero_count_5'] = data['zero_dummy'].rolling(window=5).sum()
    data = data.dropna()
    # Extract chosen day
    data['day_of_week'] = data.index.day_name()
    chosendays_data = data[data['day_of_week'] == chosen_day]

    # Select only the zero count columns and the index for Fridays
    chosendays_data = chosendays_data[['zero_count_22', 'zero_count_5']]
    chosendays_data['zero_count_22'] = chosendays_data['zero_count_22'].astype(int)
    chosendays_data['zero_count_5'] = chosendays_data['zero_count_5'].astype(int)
    return chosendays_data
```

In [ ]:

```
# Friday to Friday
Fridays['Log_Return'] = weekly_returns(li2co3, 'W-FRI')
Fridays['Zero_Count_22'] = count_zero(li2co3, 'Friday')['zero_count_22']
Fridays['Zero_Count_5'] = count_zero(li2co3, 'Friday')['zero_count_5']
Fridays = Fridays.dropna()

# Thursday to Thursday
Thursdays['Log_Return'] = weekly_returns(li2co3, 'W-THU')
Thursdays['Zero_Count_22'] = count_zero(li2co3, 'Thursday')['zero_count_22']
Thursdays['Zero_Count_5'] = count_zero(li2co3, 'Thursday')['zero_count_5']
Thursdays = Thursdays.dropna()

# Wednesday to Wednesday
Wednesdays['Log_Return'] = weekly_returns(li2co3, 'W-WED')
Wednesdays['Zero_Count_22'] = count_zero(li2co3, 'Wednesday')['zero_count_22']
Wednesdays['Zero_Count_5'] = count_zero(li2co3, 'Wednesday')['zero_count_5']
Wednesdays = Wednesdays.dropna()

# Tuesday to Tuesday
Tuesdays['Log_Return'] = weekly_returns(li2co3, 'W-TUE')
Tuesdays['Zero_Count_22'] = count_zero(li2co3, 'Tuesday')['zero_count_22']
Tuesdays['Zero_Count_5'] = count_zero(li2co3, 'Tuesday')['zero_count_5']
Tuesdays = Tuesdays.dropna()

# Monday to Monday
Mondays['Log_Return'] = weekly_returns(li2co3, 'W-MON')
Mondays['Zero_Count_22'] = count_zero(li2co3, 'Monday')['zero_count_22']
Mondays['Zero_Count_5'] = count_zero(li2co3, 'Monday')['zero_count_5']
Mondays = Mondays.dropna()
```

In [ ]:

```
print("Friday to Firday")
print(Fridays.tail(10))
print("Thursday to Thursday")
print(Thursdays.tail(10))
print("Wednesday to Wednesday")
print(Wednesdays.tail(10))
print("Tuesday to Tuesday")
print(Tuesdays.tail(10))
print("Monday to Monday")
print(Mondays.tail(10))
```

```
Friday to Firday
            Log_Return  Zero_Count_22  Zero_Count_5
Date
2024-01-26    0.000000           22.0           5.0
2024-02-02    0.000000           22.0           5.0
2024-02-23   -0.005666           20.0           4.0
2024-03-01    0.092206           16.0           1.0
2024-03-08    0.050516           14.0           3.0
2024-03-15    0.038652           10.0           0.0
2024-03-22    0.028039            8.0           2.0
2024-03-29   -0.037563            7.0           2.0
2024-04-12    0.027780            7.0           2.0
2024-04-19    0.000000           12.0           5.0
Thursday to Thursday
            Log_Return  Zero_Count_22  Zero_Count_5
Date
2024-01-25    0.000000           22.0           5.0
2024-02-01    0.000000           22.0           5.0
2024-02-22   -0.005666           20.0           4.0
2024-02-29    0.071263           17.0           2.0
2024-03-07    0.071459           14.0           2.0
2024-03-14    0.066691           10.0           1.0
2024-03-21    0.000000            8.0           1.0
2024-03-28   -0.037563            6.0           2.0
2024-04-11    0.027780            6.0           1.0
2024-04-18    0.000000           11.0           5.0
Wednesday to Wednesday
            Log_Return  Zero_Count_22  Zero_Count_5
Date
2024-02-07    0.022858           21.0           4.0
2024-02-21   -0.005666           20.0           4.0
2024-02-28    0.039002           18.0           3.0
2024-03-06    0.103720           14.0           1.0
2024-03-13    0.057432           11.0           2.0
2024-03-20    0.000000            8.0           1.0
2024-03-27   -0.028304            5.0           1.0
2024-04-03    0.018958            8.0           3.0
2024-04-10    0.036871            7.0           1.0
2024-04-17   -0.009091           10.0           4.0
Tuesday to Tuesday
            Log_Return  Zero_Count_22  Zero_Count_5
Date
2024-02-06    0.022858           21.0           4.0
2024-02-20   -0.005666           20.0           3.0
2024-02-27    0.016902           19.0           4.0
2024-03-05    0.125820           14.0           0.0
2024-03-12    0.029128           12.0           3.0
2024-03-19    0.028304            8.0           0.0
2024-03-26   -0.018780            6.0           2.0
2024-04-02    0.000000            8.0           3.0
2024-04-09    0.037214            8.0           2.0
2024-04-16    0.000000            9.0           3.0
Monday to Monday
            Log_Return  Zero_Count_22  Zero_Count_5
Date
2024-02-05    0.022858           21.0           4.0
2024-02-19    0.000000           21.0           4.0
2024-02-26   -0.005666           20.0           4.0
2024-03-04    0.122821           15.0           0.0
2024-03-11    0.029705           13.0           3.0
2024-03-18    0.038282            9.0           0.0
2024-03-25    0.009346            7.0           2.0
2024-04-01   -0.028304            8.0           3.0
2024-04-08    0.046737            8.0           2.0
2024-04-15    0.000000            8.0           3.0
```

## P3. Summary statistics

In [ ]:

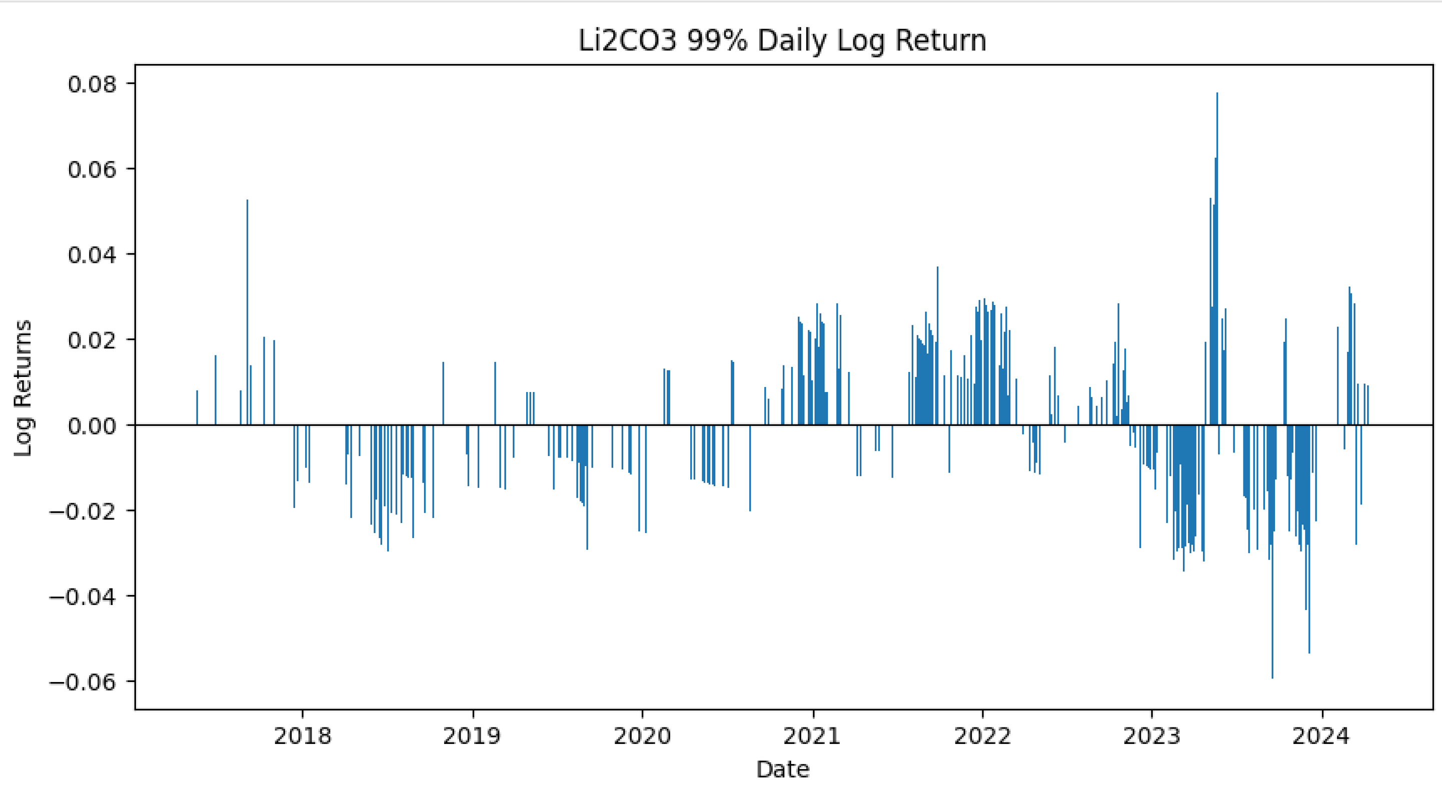
```python
summary_statistics_Fridays = Fridays.describe().loc[['mean', 'min', 'max', 'std']]
summary_statistics_Thursdays = Thursdays.describe().loc[['mean', 'min', 'max', 'std']]
summary_statistics_Wednesdays = Wednesdays.describe().loc[['mean', 'min', 'max', 'std']]
summary_statistics_Tuesdays = Tuesdays.describe().loc[['mean', 'min', 'max', 'std']]
summary_statistics_Mondays = Mondays.describe().loc[['mean', 'min', 'max', 'std']]

print("Friday to Firday")
print(summary_statistics_Fridays)
print("Thursday to Thursday")
print(summary_statistics_Thursdays)
print("Wednesday to Wednesday")
print(summary_statistics_Wednesdays)
print("Tuesday to Tuesday")
print(summary_statistics_Tuesdays)
print("Monday to Monday")
print(summary_statistics_Mondays)
```

```
Friday to Firday
      Log_Return  Zero_Count_22  Zero_Count_5
mean   -0.000397      15.089783      3.421053
min    -0.152469       0.000000      0.000000
max     0.267022      22.000000      5.000000
std     0.044616       5.608998      1.584804
Thursday to Thursday
      Log_Return  Zero_Count_22  Zero_Count_5
mean   -0.000692      15.096386      3.418675
min    -0.172141       0.000000      0.000000
max     0.204501      22.000000      5.000000
std     0.044082       5.556335      1.576646
Wednesday to Wednesday
      Log_Return  Zero_Count_22  Zero_Count_5
mean   -0.000154      15.036145      3.418675
min    -0.190575       0.000000      0.000000
max     0.252326      22.000000      5.000000
std     0.044743       5.529806      1.533911
Tuesday to Tuesday
      Log_Return  Zero_Count_22  Zero_Count_5
mean   -0.000803      15.026946      3.410180
min    -0.163827       0.000000      0.000000
max     0.298955      22.000000      5.000000
std     0.045178       5.532529      1.552684
Monday to Monday
      Log_Return  Zero_Count_22  Zero_Count_5
mean   -0.000810      15.090062      3.403727
min    -0.171980       0.000000      0.000000
max     0.293305      22.000000      5.000000
std     0.045656       5.598839      1.576221
```

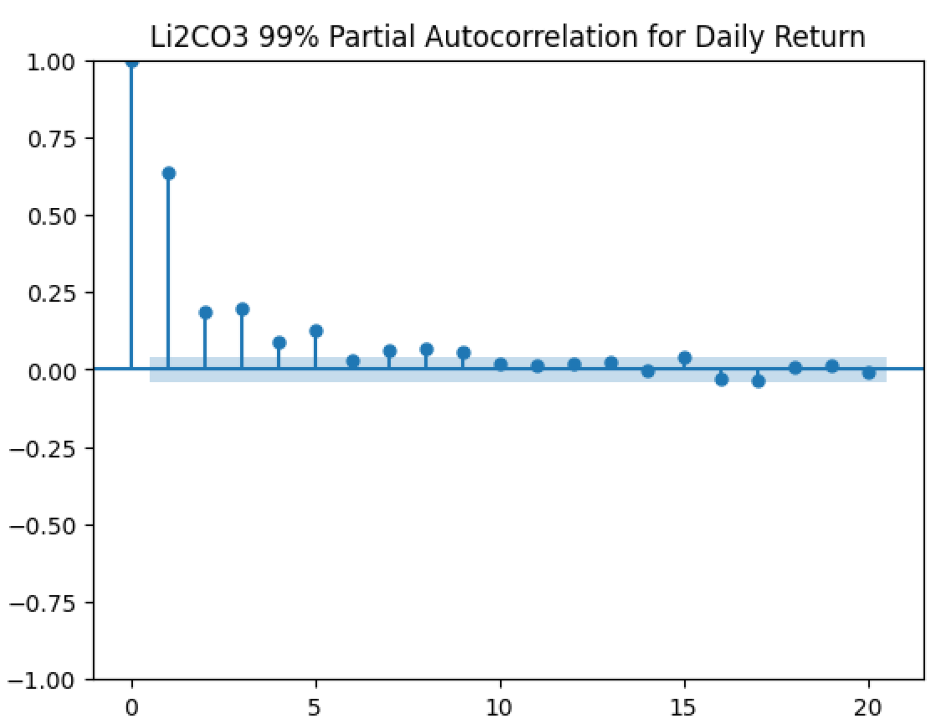
## P4. Autocorrelogram and partial autocorrelogram for daily return

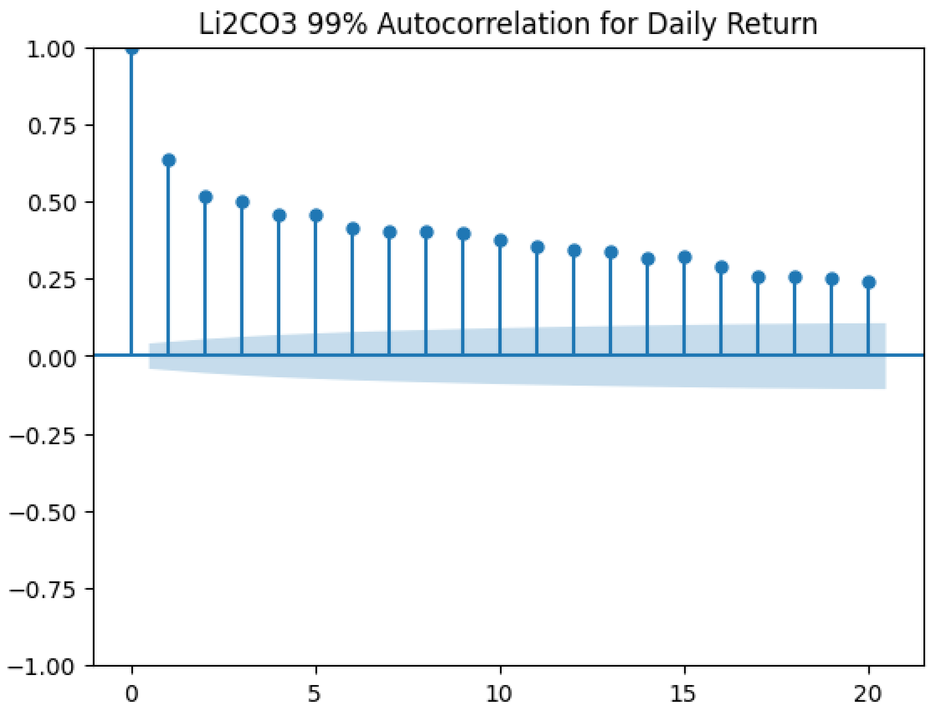
In [ ]:

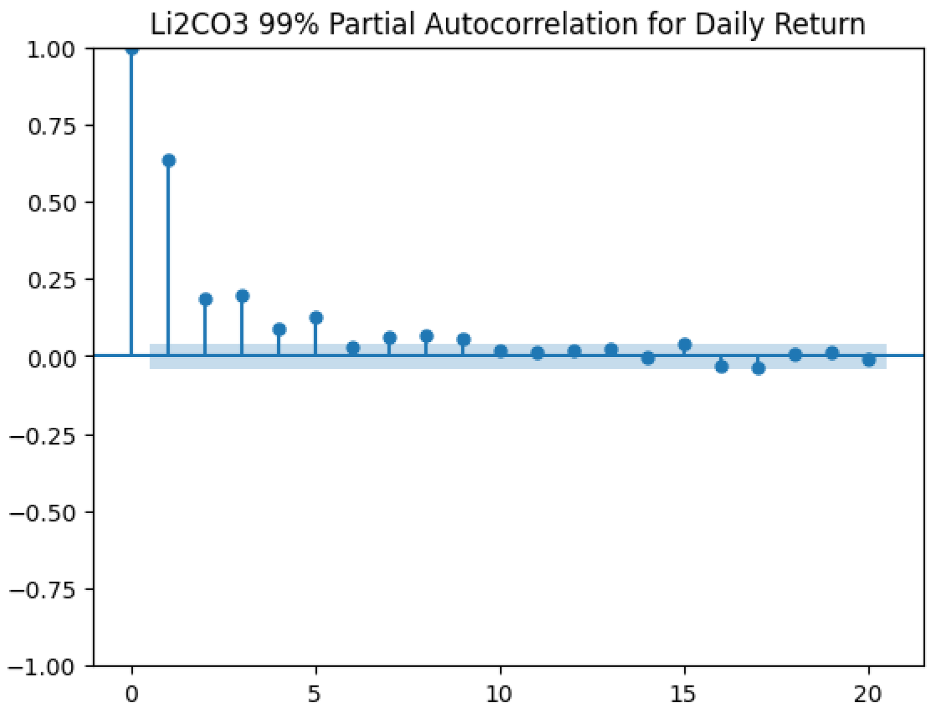
```python
# Daily return
fig, axs = plt.subplots(1, 1, figsize=(10, 5))
axs.bar(li2co3.index, li2co3['log_ret'], width=2.5)
plt.axhline(0, linewidth=0.8, color='k')
plt.xlabel('Date')
plt.ylabel('Log Returns')
plt.title('Li2CO3 99% Daily Log Return')
plt.show()

# Plot autocorrelation and partial autocorrelation
plot_acf(li2co3['log_ret'], lags=20, alpha=0.1, title='Li2CO3 99% Autocorrelation for Daily Return')
plot_pacf(li2co3['log_ret'], lags=20, alpha=0.1, method='ywm', title='Li2CO3 99% Partial Autocorrelation for Daily Return')
```




Out[ ]:



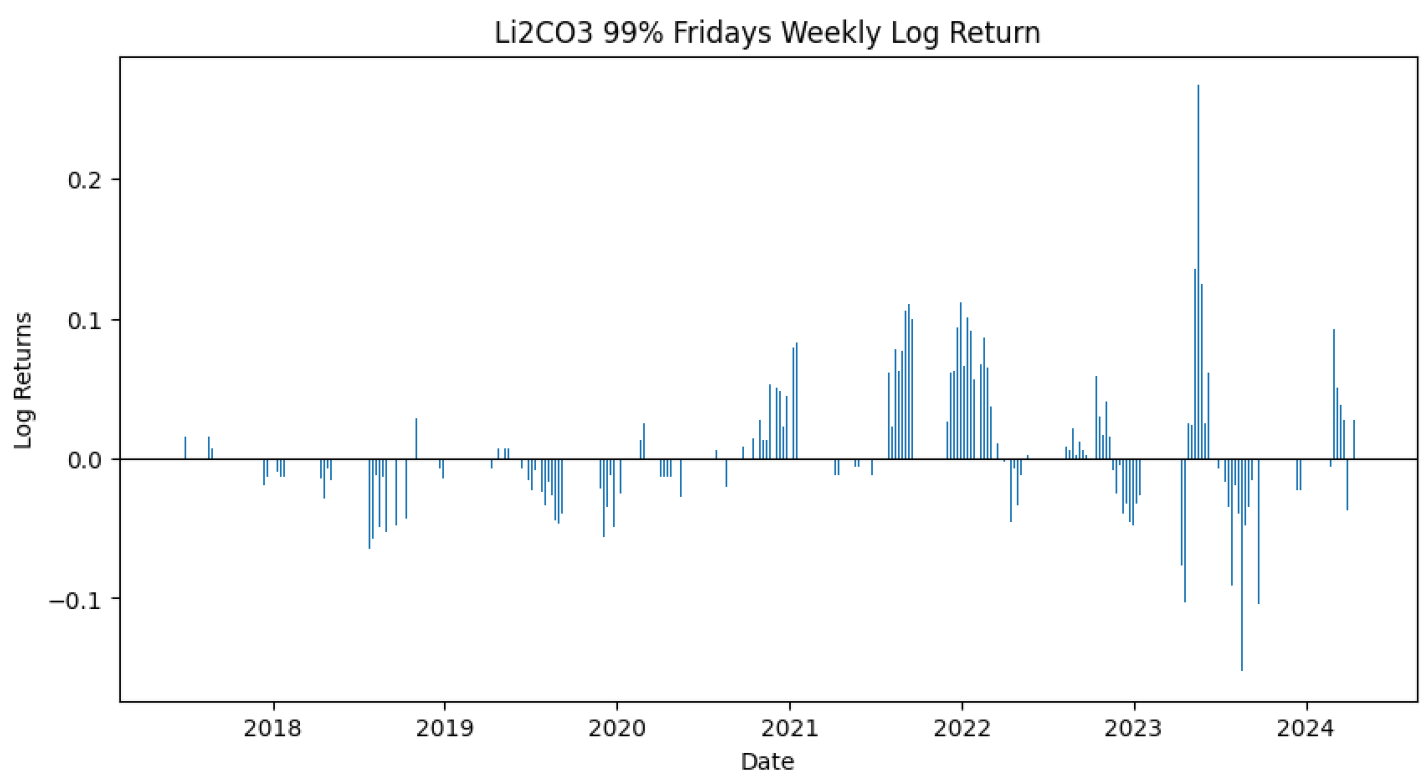


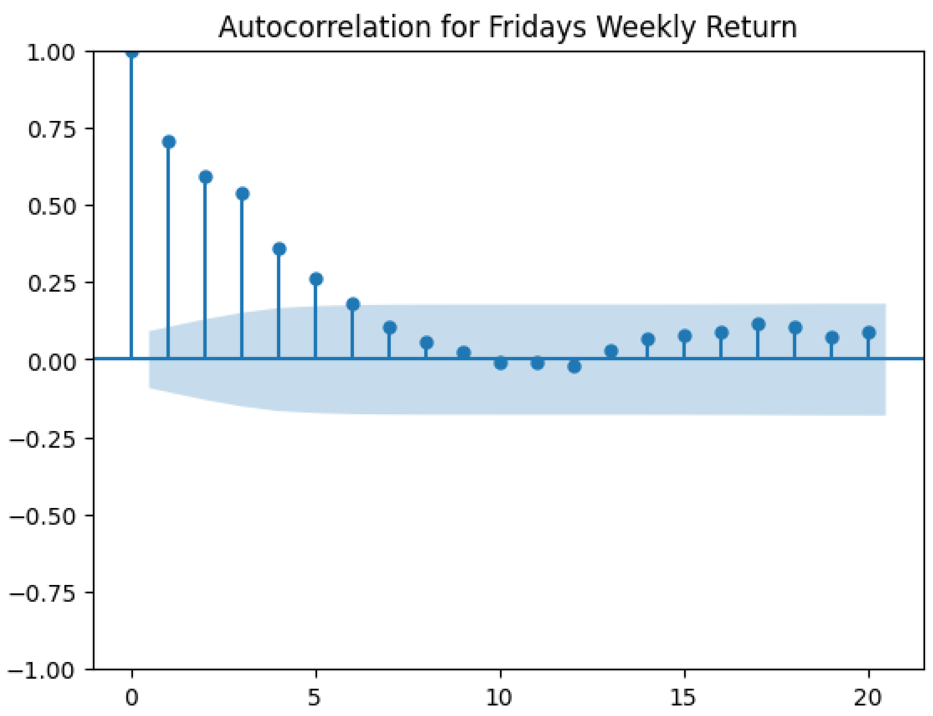



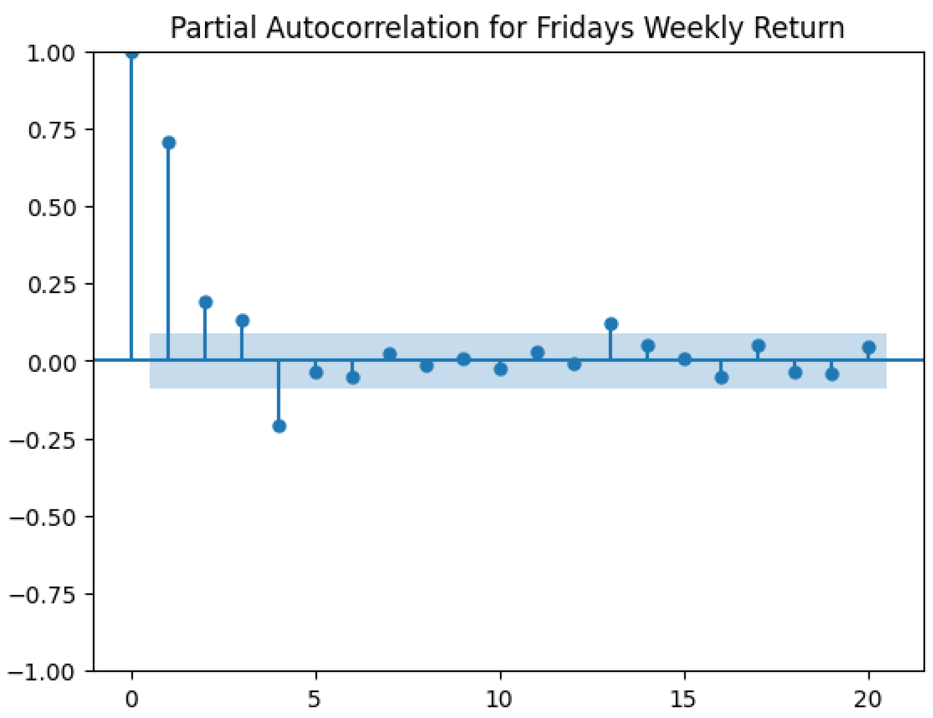

## P5. Autocorrelogram and partial autocorrelogram for weekly return

In [ ]:
```python
def plot_returns_acf_pacf(df, df_name):
    """
    Plots the log returns, autocorrelation, and partial autocorrelation for a given DataFrame.

    Parameters:
    - df: The DataFrame containing the log returns data.
    - df_name: Name of the DataFrame to use in plot titles to indicate the data source.
    """
    data = df.copy()

    # Plotting the log returns
    fig, axs = plt.subplots(1, 1, figsize=(10, 5))
    axs.bar(data.index, data['Log_Return'], width=2.5)
    plt.axhline(0, linewidth=0.8, color='k')
    plt.xlabel('Date')
    plt.ylabel('Log Returns')
    plt.title(f'Li2CO3 99% {df_name} Weekly Log Return')
    plt.show()

    # Plotting the autocorrelation
    fig_acf = plot_acf(data['Log_Return'], lags=20, alpha=0.1, title=f'Autocorrelation for {df_name} Weekly Return')

    # Plotting the partial autocorrelation
    fig_pacf = plot_pacf(data['Log_Return'], lags=20, alpha=0.1, method='ywm', title=f'Partial Autocorrelation for {df_name} Weekly Return')
    plt.show()
```

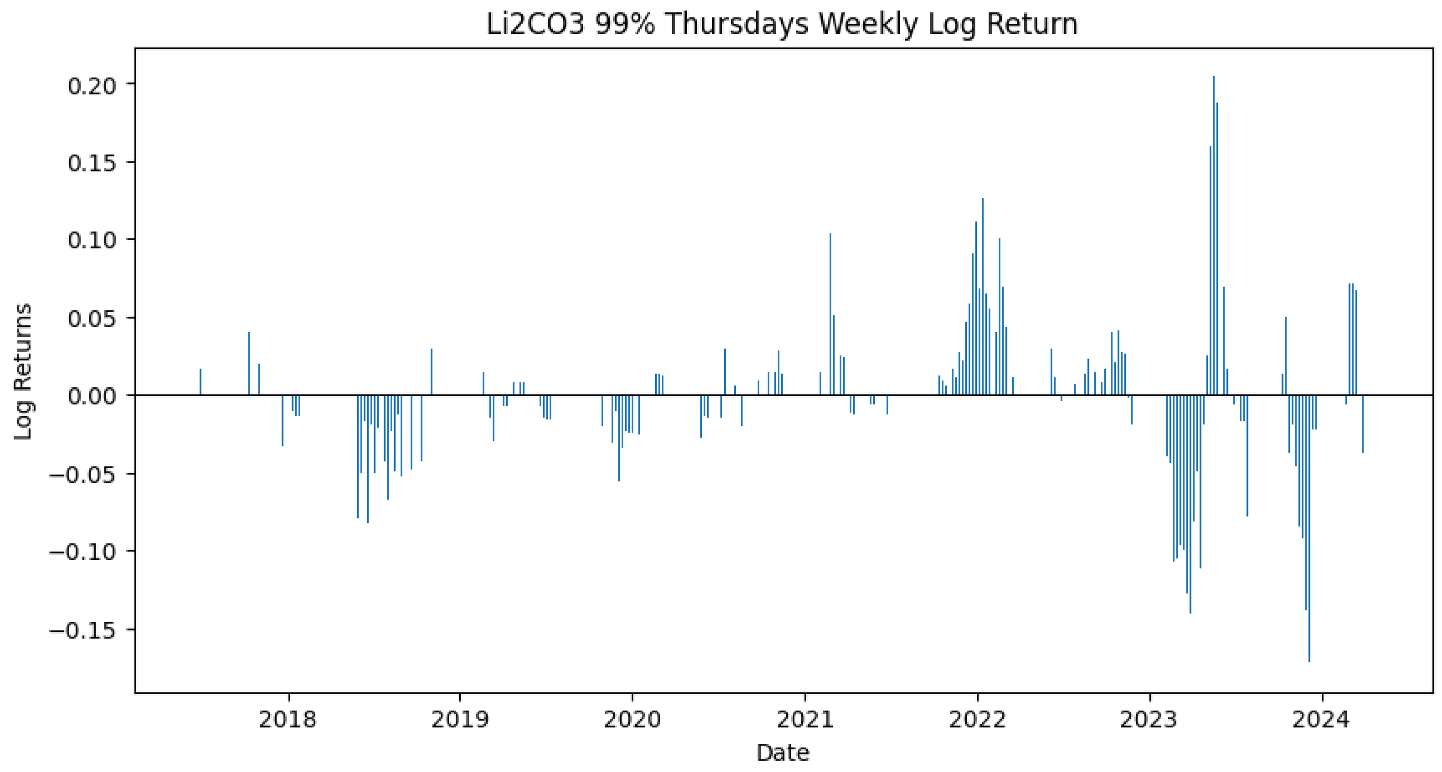
In [ ]:
```python
# Fridays
plot_returns_acf_pacf(Fridays, 'Fridays')
```

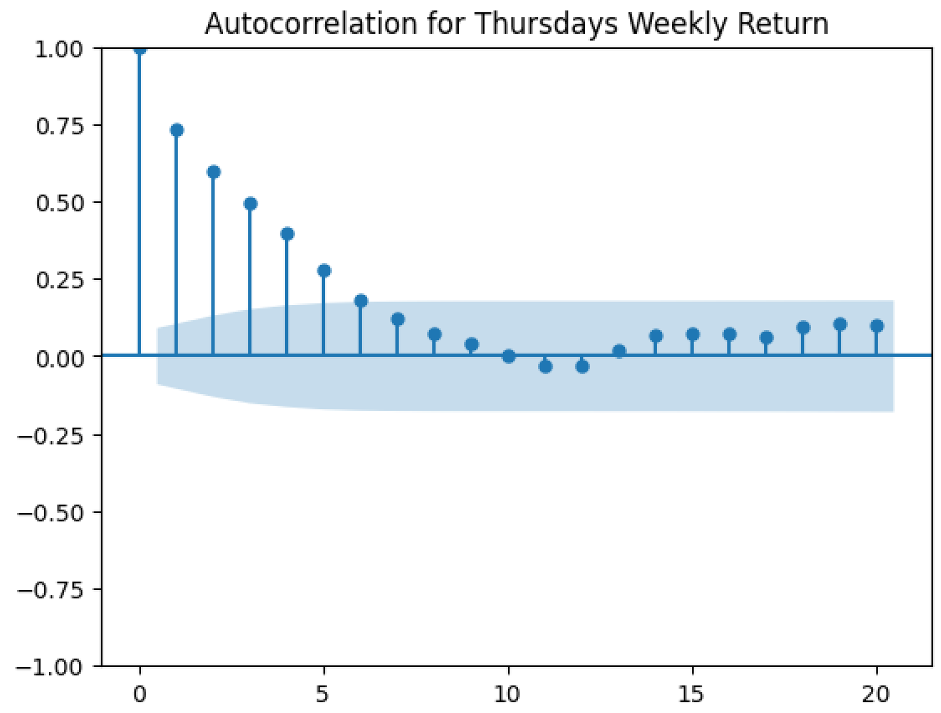


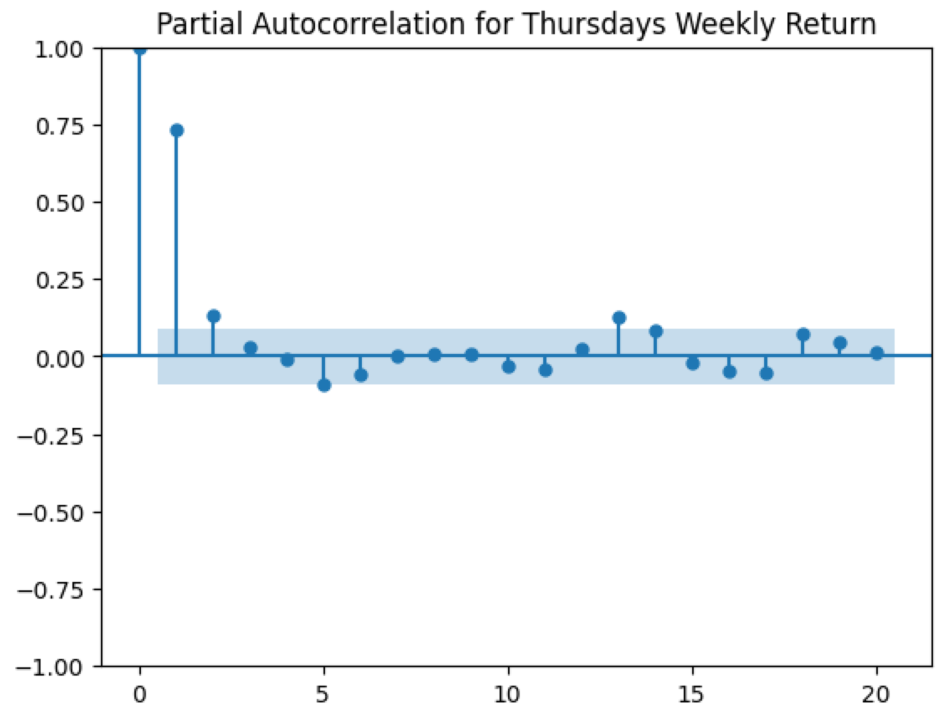



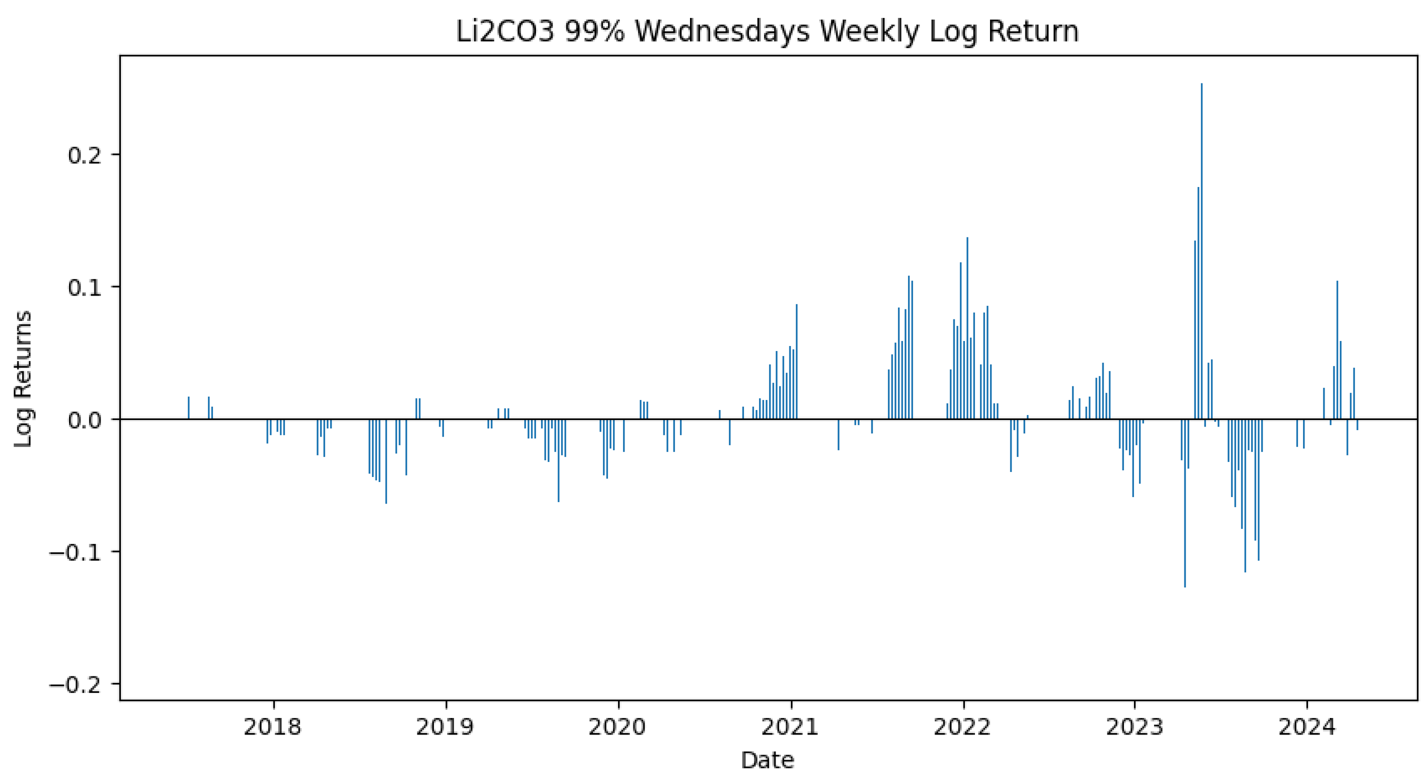

In [ ]:
```python
# Thursdays
plot_returns_acf_pacf(Thursdays, 'Thursdays')
```

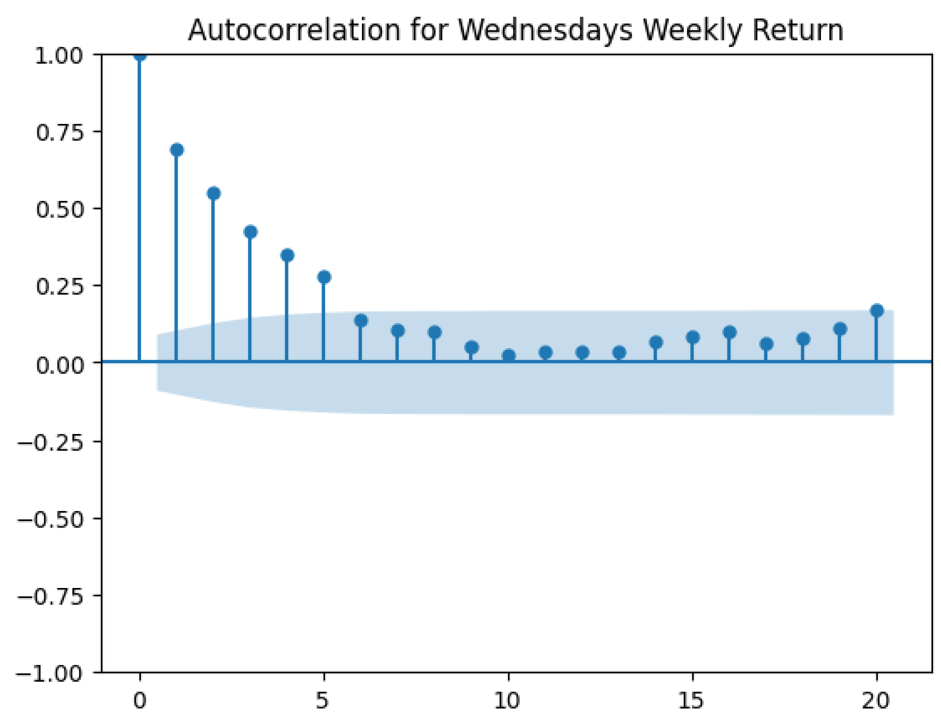


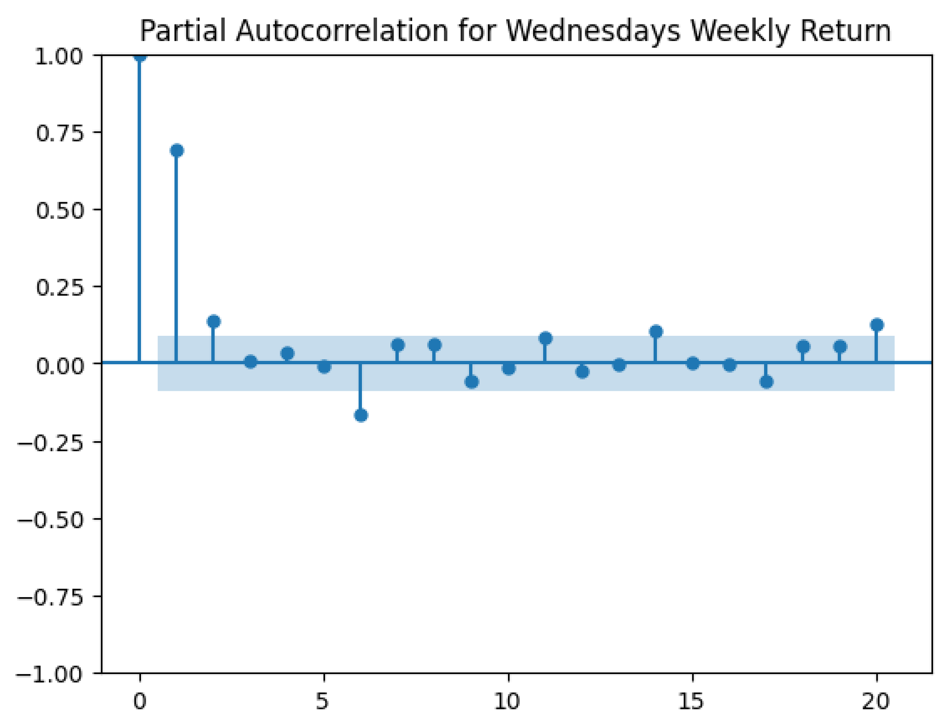



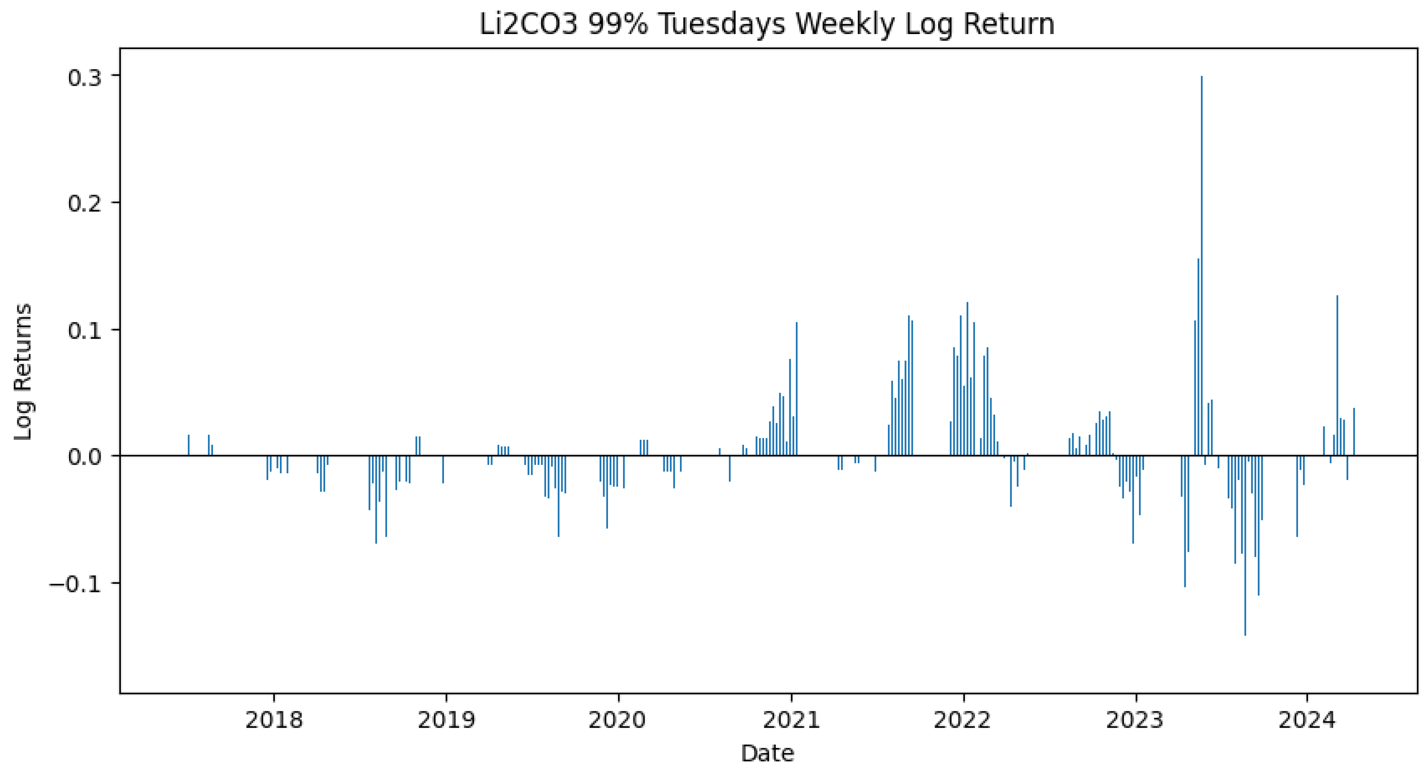

In [ ]:
```
# Wednesdays
plot_returns_acf_pacf(Wednesdays, 'Wednesdays')
```

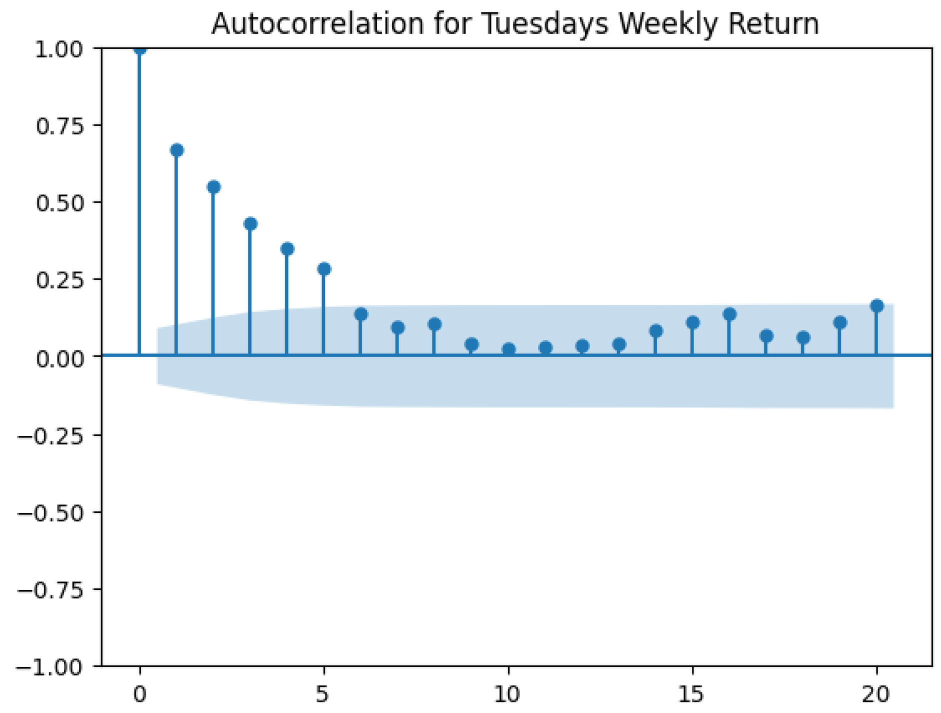


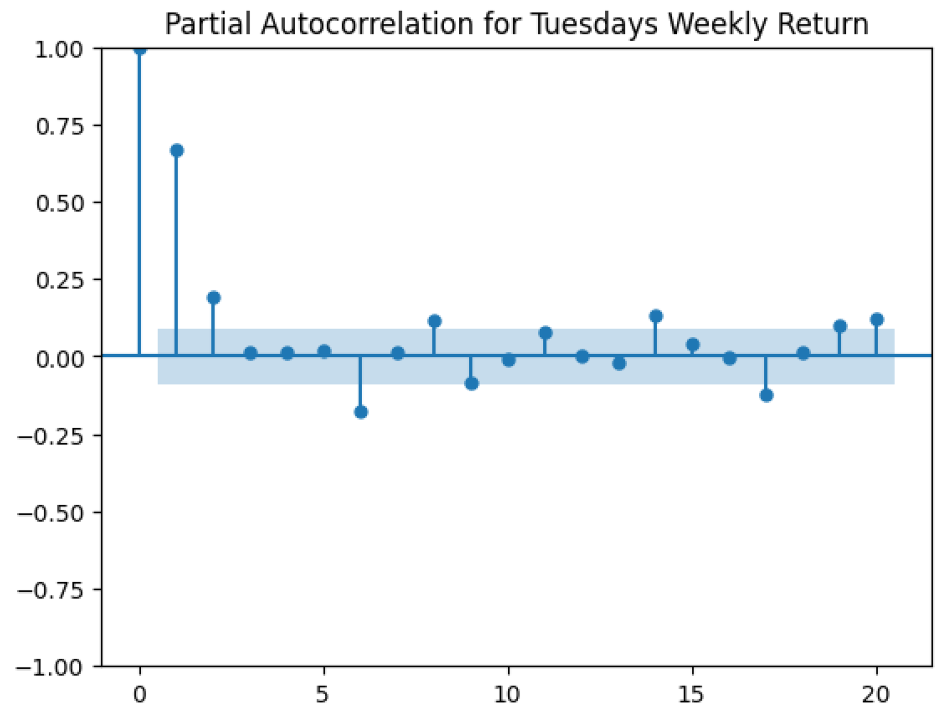



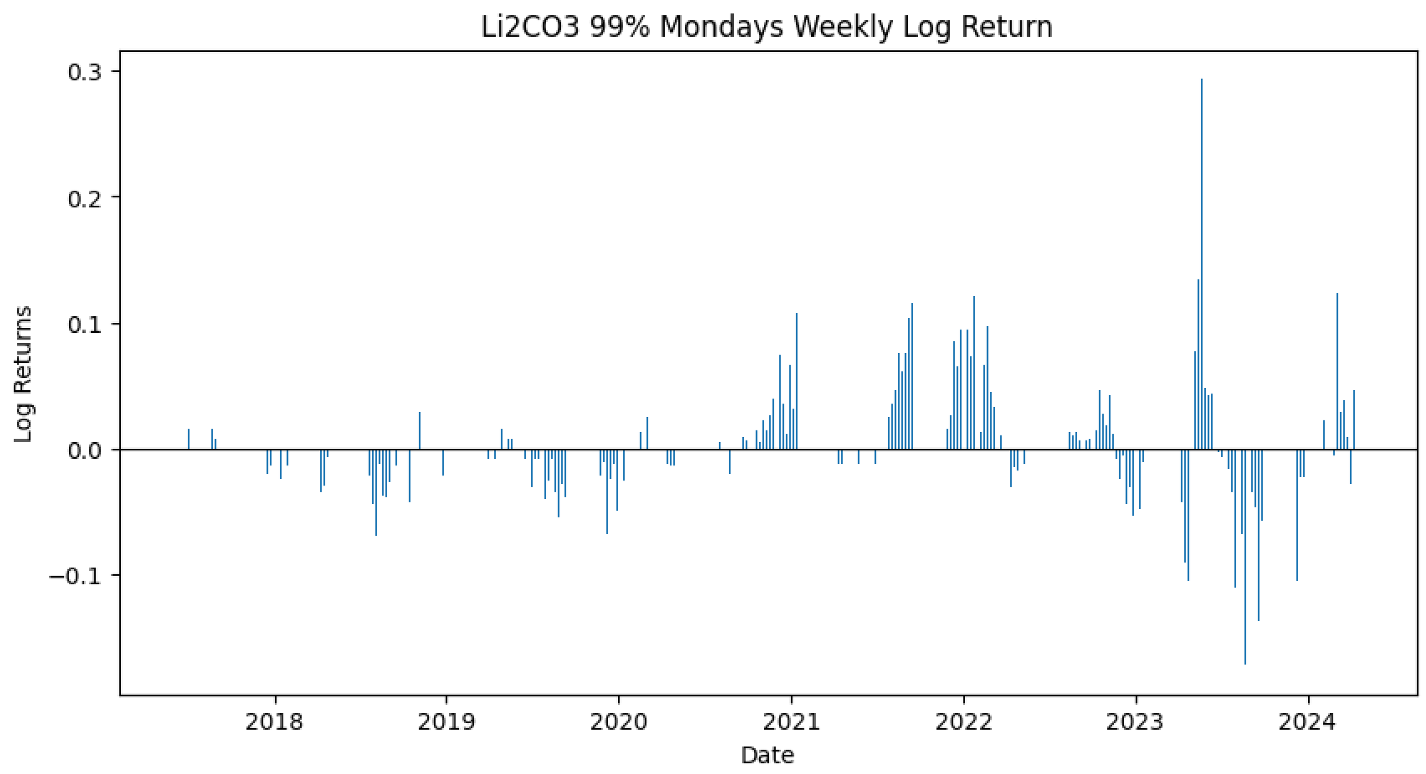

In [ ]:
```
# Tuesdays
plot_returns_acf_pacf(Tuesdays, 'Tuesdays')
```

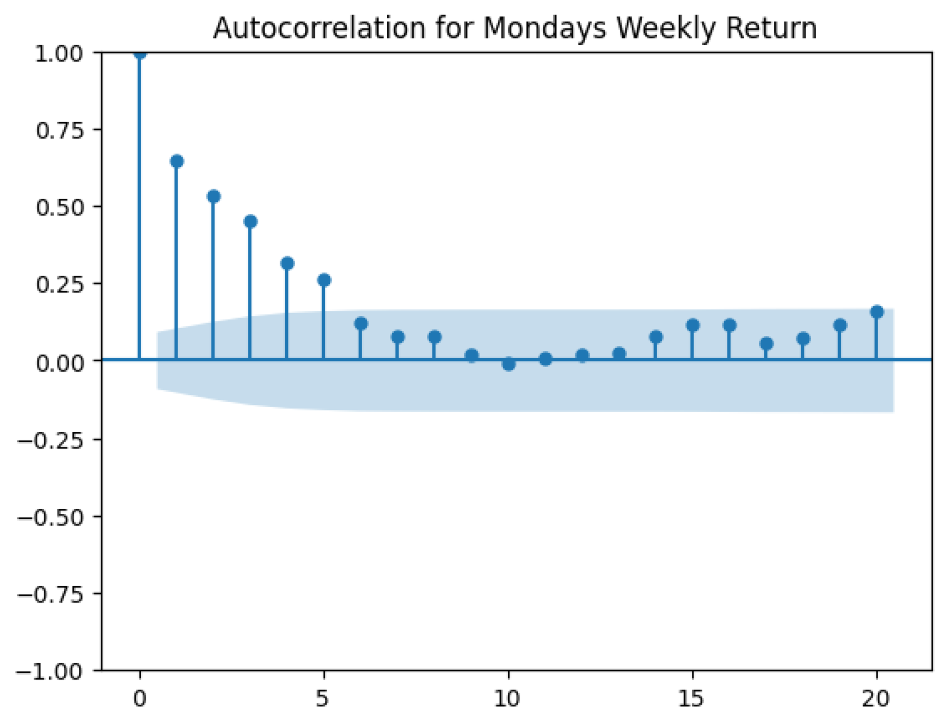


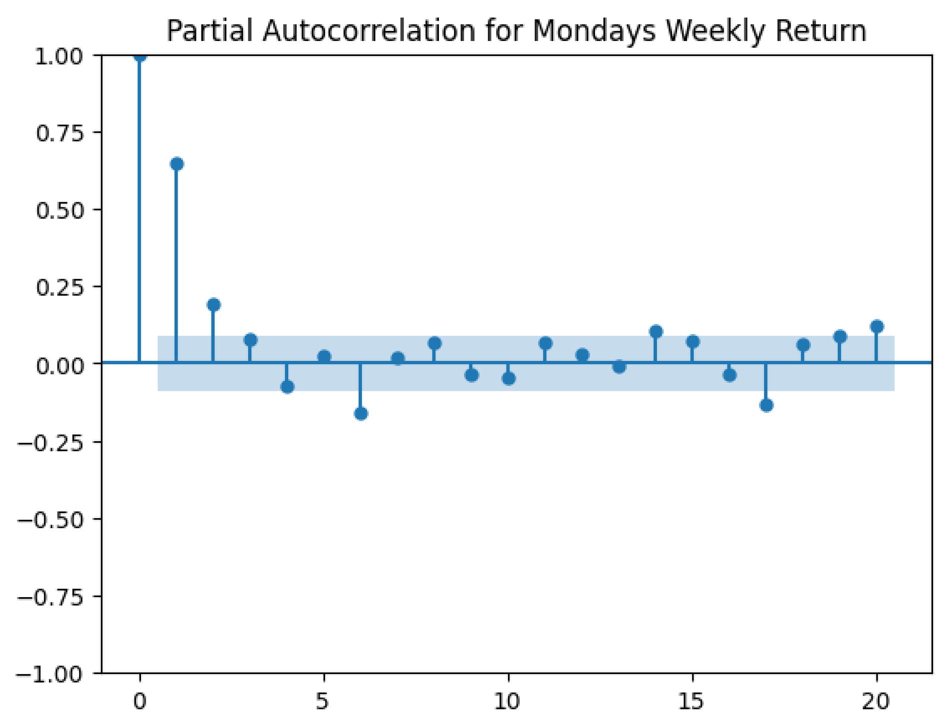




```python
# Mondays
plot_returns_acf_pacf(Mondays, 'Mondays')
```




maxlags = 4 for Fridays

maxlags = 2 for others

## P6. Two AR(2) models for daily return

```python
def estimate_ar2_model(df, lags):
    """
    Estimate a basic AR(2) model for returns.

    Parameters:
    - df: DataFrame containing the log returns series under 'log_ret'.
    - lags: Number of lags to use for HAC standard errors.

    Returns:
    - model: OLS regression results containing the fitted model.
    """
```

```python
    data = df.copy()
    data['log_ret_lag1'] = data['log_ret'].shift(1)
    data['log_ret_lag2'] = data['log_ret'].shift(2)

    data.dropna(inplace=True)

    # Define the variables and add a constant term for the intercept (alpha)
    X = data[['log_ret_lag1', 'log_ret_lag2']]
    X = sm.add_constant(X)  # Adds a constant column to input data set
    Y = data['log_ret']

    # Fit the AR(2) model
    model = sm.OLS(Y, X).fit(cov_type='HAC', cov_kwds={'maxlags': lags})

    return model
```

In [ ]:
```python
# lags=5 acording to the daily pacf
daily_return_ar2_model = estimate_ar2_model(li2co3, lags=5)
```

In [ ]:
```python
print(daily_return_ar2_model.summary())
```

```
                            OLS Regression Results                            
==============================================================================
Dep. Variable:                log_ret   R-squared:                       0.424
Model:                            OLS   Adj. R-squared:                  0.423
Method:                 Least Squares   F-statistic:                     188.0
Date:                Mon, 03 Jun 2024   Prob (F-statistic):           2.62e-74
Time:                        17:03:54   Log-Likelihood:                 5540.0
No. Observations:                1662   AIC:                        -1.107e+04
Df Residuals:                    1659   BIC:                        -1.106e+04
Df Model:                           2                                         
Covariance Type:                  HAC                                         
================================================================================
                   coef    std err          z      P>|z|      [0.025      0.975]
--------------------------------------------------------------------------------
const        -2.075e-05      0.000     -0.104      0.917      -0.000       0.000
log_ret_lag1     0.5158      0.036     14.302      0.000       0.445       0.586
log_ret_lag2     0.1876      0.038      4.933      0.000       0.113       0.262
==============================================================================
Omnibus:                      177.765   Durbin-Watson:                   2.075
Prob(Omnibus):                  0.000   Jarque-Bera (JB):             1463.753
Skew:                          -0.004   Prob(JB):                         0.00
Kurtosis:                       7.598   Cond. No.                         145.
==============================================================================

Notes:
[1] Standard Errors are heteroscedasticity and autocorrelation robust (HAC) using 5 lags and without small sample correction
```

In [ ]:
```python
def estimate_ar2_model_with_zero_dummy(df, lags):
    """
    Estimate an AR(2) model for log returns with modifications to account for zero dummies.

    Parameters:
    - df: DataFrame containing the log returns 'log_ret'.
    - lags: Number of lags to use for HAC standard errors.

    Returns:
    - model: OLS regression results containing the fitted model.
    """

    data = df.copy()
    data['zero_dummy'] = (data['log_ret'] == 0).astype(int)

    # Generate lagged return series
    data['log_ret_lag1'] = data['log_ret'].shift(1)
    data['log_ret_lag2'] = data['log_ret'].shift(2)

    # Generate interaction terms
    data['log_ret_lag1_zero'] = data['log_ret_lag1'] * data['zero_dummy']
    data['log_ret_lag2_zero'] = data['log_ret_lag2'] * data['zero_dummy']

    data.dropna(inplace=True)

    # Define the new model with additional interaction terms
    X = data[['zero_dummy', 'log_ret_lag1', 'log_ret_lag2', 'log_ret_lag1_zero', 'log_ret_lag2_zero']]
    X = sm.add_constant(X)
    Y = data['log_ret']

    # Fit the AR(2) model
    model = sm.OLS(Y, X).fit(cov_type='HAC', cov_kwds={'maxlags': lags})

    return model
```

In [ ]:
```python
# lags=5 acording to the daily pacf
daily_return_ar2_model_with_zero_dummy = estimate_ar2_model_with_zero_dummy(li2co3, lags = 5)
```

In [ ]:
```python
print(daily_return_ar2_model_with_zero_dummy.summary())
```

```
                            OLS Regression Results                            
==============================================================================
Dep. Variable:                log_ret   R-squared:                       0.607
Model:                            OLS   Adj. R-squared:                  0.605
Method:                 Least Squares   F-statistic:                     442.3
Date:                Mon, 03 Jun 2024   Prob (F-statistic):          5.35e-211
Time:                        17:04:00   Log-Likelihood:                 5856.5
No. Observations:                1662   AIC:                        -1.170e+04
Df Residuals:                    1656   BIC:                        -1.167e+04
Df Model:                           5                                         
Covariance Type:                  HAC                                         
=====================================================================================
                        coef    std err          z      P>|z|      [0.025      0.975]
-------------------------------------------------------------------------------------
const                -0.0010      0.001     -1.594      0.111      -0.002       0.000
zero_dummy            0.0010      0.001      1.594      0.111      -0.000       0.002
log_ret_lag1          0.6606      0.045     14.796      0.000       0.573       0.748
log_ret_lag2          0.3622      0.048      7.575      0.000       0.268       0.456
log_ret_lag1_zero    -0.6606      0.045    -14.796      0.000      -0.748      -0.573
log_ret_lag2_zero    -0.3622      0.048     -7.575      0.000      -0.456      -0.268
==============================================================================
Omnibus:                      304.524   Durbin-Watson:                   1.918
Prob(Omnibus):                  0.000   Jarque-Bera (JB):             5065.128
Skew:                           0.339   Prob(JB):                         0.00
Kurtosis:                      11.526   Cond. No.                         413.
==============================================================================

Notes:
[1] Standard Errors are heteroscedasticity and autocorrelation robust (HAC) using 5 lags and without small sample correction
/usr/local/lib/python3.10/dist-packages/statsmodels/base/model.py:1894: ValueWarning: covariance of constraints does not have full rank. The number of constraints is 5, but rank is 3
  warnings.warn('covariance of constraints does not have full '
```

## P7. Four AR(2) models for returns with zero count series (use Fridays as example)

### Model1: Standard AR(2) model

In [ ]:
```python
def estimate_ar2_model(df, lags):
    """
    Estimate a basic AR(2) model for returns.

    Parameters:
    - df: DataFrame containing the log returns series under 'Log_Return'.
    - lags: Number of lags to use for HAC standard errors.

    Returns:
    - model: OLS regression results containing the fitted model.
    """

    data = df.copy()
    data['log_ret_lag1'] = data['Log_Return'].shift(1)
    data['log_ret_lag2'] = data['Log_Return'].shift(2)

    data.dropna(inplace=True)

    # Define the variables and add a constant term for the intercept (alpha)
    X = data[['log_ret_lag1', 'log_ret_lag2']]
    X = sm.add_constant(X)  # Adds a constant column to input data set
    Y = data['Log_Return']

    # Fit the AR(2) model
    model = sm.OLS(Y, X).fit(cov_type='HAC', cov_kwds={'maxlags': lags})

    return model
```

In [ ]:
```python
# lags=4 acording to the weekly pacf
weekly_return_ar2_model = estimate_ar2_model(Fridays, lags=4)

print(weekly_return_ar2_model.summary())
```

```
                            OLS Regression Results                            
==============================================================================
Dep. Variable:             Log_Return   R-squared:                       0.516
Model:                            OLS   Adj. R-squared:                  0.513
Method:                 Least Squares   F-statistic:                     84.51
Date:                Mon, 03 Jun 2024   Prob (F-statistic):           3.69e-30
Time:                        17:04:06   Log-Likelihood:                 658.66
No. Observations:                 321   AIC:                            -1311.
Df Residuals:                     318   BIC:                            -1300.
Df Model:                           2                                         
Covariance Type:                  HAC                                         
================================================================================
                   coef    std err          z      P>|z|      [0.025      0.975]
--------------------------------------------------------------------------------
const        -7.831e-05      0.002     -0.044      0.965      -0.004       0.003
log_ret_lag1     0.5691      0.079      7.160      0.000       0.413       0.725
log_ret_lag2     0.1929      0.088      2.188      0.029       0.020       0.366
==============================================================================
Omnibus:                       73.893   Durbin-Watson:                   2.050
Prob(Omnibus):                  0.000   Jarque-Bera (JB):              539.301
Skew:                           0.713   Prob(JB):                    7.80e-118
Kurtosis:                       9.188   Cond. No.                         41.3
==============================================================================

Notes:
[1] Standard Errors are heteroscedasticity and autocorrelation robust (HAC) using 4 lags and without small sample correction
```

### Model2: AR(2) model using interaction with the weekly zero count series

In [ ]:
```python
def estimate_ar2_model_with_weekly_zero(df, lags):
    """
    Estimate an AR(2) model for weekly log returns, incorporating
    weekly zero count series as interaction effect,
    and allowing for specification of lags for HAC standard errors.

    Parameters:
    - df: DataFrame containing the weekly log returns under 'Log_Return',
        and the weekly zero count series 'Zero_Count_5'.
    - lags: Maximum number of lags to use for HAC standard errors.

    Returns:
    - model: OLS regression results containing the fitted model with HAC standard errors.
    """

    data = df.copy()
    # Generate lagged return series
    data['Log_Return_Lag1'] = data['Log_Return'].shift(1)
    data['Log_Return_Lag2'] = data['Log_Return'].shift(2)

    # Generate interaction terms for lagged returns and 'Zero_Count_5'
    data['Log_Return_Lag1_Zero5'] = data['Log_Return_Lag1'] * data['Zero_Count_5']
    data['Log_Return_Lag2_Zero5'] = data['Log_Return_Lag2'] * data['Zero_Count_5']

    # Drop any rows with NaN values that were created by lagging
    data.dropna(inplace=True)

    # Define the model with additional interaction terms
    X = data[['Zero_Count_5', 'Log_Return_Lag1', 'Log_Return_Lag2',
              'Log_Return_Lag1_Zero5', 'Log_Return_Lag2_Zero5']]
    X = sm.add_constant(X)
    Y = data['Log_Return']

    # Fit the model with HAC standard errors
    model = sm.OLS(Y, X).fit(cov_type='HAC', cov_kwds={'maxlags': lags})

    return model
```

In [ ]:
```python
# lags=4 acording to the weekly pacf
ar2_model_with_weekly_zero = estimate_ar2_model_with_weekly_zero(Fridays, lags=4)

print(ar2_model_with_weekly_zero.summary())
```

```
                            OLS Regression Results
==============================================================================
Dep. Variable:             Log_Return   R-squared:                       0.656
Model:                            OLS   Adj. R-squared:                  0.650
Method:                 Least Squares   F-statistic:                     61.77
Date:                Mon, 03 Jun 2024   Prob (F-statistic):           9.64e-45
Time:                        17:04:12   Log-Likelihood:                 713.29
No. Observations:                 321   AIC:                            -1415.
Df Residuals:                     315   BIC:                            -1392.
Df Model:                           5
Covariance Type:                  HAC
=========================================================================================
                            coef    std err          z      P>|z|      [0.025      0.975]
-----------------------------------------------------------------------------------------
const                    -0.0015      0.007     -0.202      0.840      -0.016       0.013
Zero_Count_5           6.781e-05      0.002      0.041      0.967      -0.003       0.003
Log_Return_Lag1           0.9768      0.188      5.203      0.000       0.609       1.345
Log_Return_Lag2           0.3009      0.179      1.680      0.093      -0.050       0.652
Log_Return_Lag1_Zero5    -0.2143      0.058     -3.664      0.000      -0.329      -0.100
Log_Return_Lag2_Zero5    -0.0485      0.047     -1.022      0.307      -0.141       0.045
==============================================================================
Omnibus:                       88.859   Durbin-Watson:                   1.738
Prob(Omnibus):                  0.000   Jarque-Bera (JB):              614.566
Skew:                           0.938   Prob(JB):                    3.54e-134
Kurtosis:                       9.514   Cond. No.                         340.
==============================================================================

Notes:
[1] Standard Errors are heteroscedasticity and autocorrelation robust (HAC) using 4 lags and without small sample correction
```

### Model3: AR(2) model using interaction with the monthly zero count series

In [ ]:
```python
def estimate_ar2_model_with_monthly_zero(df, lags):
    """
    Estimate an AR(2) model for weekly log returns, incorporating
    monthly zero count series as interaction effect,
    and allowing for specification of lags for HAC standard errors.

    Parameters:
    - df: DataFrame containing the weekly log returns under 'Log_Return',
        and the monthly zero count series 'Zero_Count_22'.
    - lags: Maximum number of lags to use for HAC standard errors.

    Returns:
    - model: OLS regression results containing the fitted model with HAC standard errors.
    """

    data = df.copy()
    # Generate lagged return series
    data['Log_Return_Lag1'] = data['Log_Return'].shift(1)
    data['Log_Return_Lag2'] = data['Log_Return'].shift(2)

    # Generate interaction terms for lagged returns and 'Zero_Count_22'
    data['Log_Return_Lag1_Zero22'] = data['Log_Return_Lag1'] * data['Zero_Count_22']
    data['Log_Return_Lag2_Zero22'] = data['Log_Return_Lag2'] * data['Zero_Count_22']

    # Drop any rows with NaN values that were created by lagging
    data.dropna(inplace=True)

    # Define the model with additional interaction terms
    X = data[['Zero_Count_22', 'Log_Return_Lag1', 'Log_Return_Lag2',
              'Log_Return_Lag1_Zero22', 'Log_Return_Lag2_Zero22']]
    X = sm.add_constant(X)
    Y = data['Log_Return']

    # Fit the model with HAC standard errors
    model = sm.OLS(Y, X).fit(cov_type='HAC', cov_kwds={'maxlags': lags})

    return model
```

In [ ]:
```python
# lags=4 acording to the weekly pacf
ar2_model_with_monthly_zero = estimate_ar2_model_with_monthly_zero(Fridays, lags=4)

print(ar2_model_with_monthly_zero.summary())
```

```
                            OLS Regression Results
==============================================================================
Dep. Variable:             Log_Return   R-squared:                       0.540
Model:                            OLS   Adj. R-squared:                  0.533
Method:                 Least Squares   F-statistic:                     43.84
Date:                Mon, 03 Jun 2024   Prob (F-statistic):           2.99e-34
Time:                        17:04:17   Log-Likelihood:                 666.79
No. Observations:                 321   AIC:                            -1322.
Df Residuals:                     315   BIC:                            -1299.
Df Model:                           5
Covariance Type:                  HAC
==========================================================================================
                             coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------------------
const                      0.0009      0.008      0.122      0.903      -0.014       0.016
Zero_Count_22             -0.0001      0.000     -0.266      0.790      -0.001       0.001
Log_Return_Lag1            0.9183      0.177      5.179      0.000       0.571       1.266
Log_Return_Lag2            0.0360      0.206      0.175      0.861      -0.368       0.440
Log_Return_Lag1_Zero22    -0.0396      0.013     -3.042      0.002      -0.065      -0.014
Log_Return_Lag2_Zero22     0.0109      0.014      0.766      0.444      -0.017       0.039
==============================================================================
Omnibus:                       62.191   Durbin-Watson:                   1.967
Prob(Omnibus):                  0.000   Jarque-Bera (JB):              422.402
Skew:                           0.568   Prob(JB):                     1.89e-92
Kurtosis:                       8.504   Cond. No.                     1.60e+03
==============================================================================

Notes:
[1] Standard Errors are heteroscedasticity and autocorrelation robust (HAC) using 4 lags and without small sample correction
[2] The condition number is large, 1.6e+03. This might indicate that there are
strong multicollinearity or other numerical problems.
```

### Model4: AR(2) model **separately** using interactions with the short-term illiquidity and long-term illiquidity variable

In [ ]:
```python
def estimate_ar2_model_separate_illiquidity(df, lags):
    """
    Estimate a complex AR(2) model for log returns, incorporating
    separate interactions with two types of zero counts.
    The formula incorporates separate interactions for different lags
    with different term-length zero count measures.

    Parameters:
    - df: DataFrame containing the log returns under 'Log_Return',
        two zero count series 'Zero_Count_5' and 'Zero_Count_22'.
    - lags: Maximum number of lags to use for HAC standard errors.

    Returns:
    - model: OLS regression results containing the fitted model with HAC standard errors.
    """

    data = df.copy()
    # Generate lagged return series
    data['Log_Return_Lag1'] = data['Log_Return'].shift(1)
    data['Log_Return_Lag2'] = data['Log_Return'].shift(2)

    # Generate separate interaction terms for lagged returns and zero counts
    data['Log_Return_Lag1_Zero5'] = data['Log_Return_Lag1'] * data['Zero_Count_5']
    data['Log_Return_Lag2_Zero22'] = data['Log_Return_Lag2'] * data['Zero_Count_22']

    # Drop any rows with NaN values that were created by lagging
    data.dropna(inplace=True)

    # Define the model with additional interaction terms
    X = data[['Zero_Count_5', 'Zero_Count_22', 'Log_Return_Lag1', 'Log_Return_Lag2',
              'Log_Return_Lag1_Zero5', 'Log_Return_Lag2_Zero22']]
```

```
    X = sm.add_constant(X)
    Y = data['Log_Return']

    # Fit the model with HAC standard errors
    model = sm.OLS(Y, X).fit(cov_type='HAC', cov_kwds={'maxlags': lags})

    return model
```

In [ ]:
```
# lags=4 acording to the weekly pacf
ar2_model_with_separate_illiquidity = estimate_ar2_model_separate_illiquidity(Fridays, lags=4)

print(ar2_model_with_separate_illiquidity.summary())
```

```
                            OLS Regression Results
==============================================================================
Dep. Variable:             Log_Return   R-squared:                       0.665
Model:                            OLS   Adj. R-squared:                  0.659
Method:                 Least Squares   F-statistic:                     89.29
Date:                Mon, 03 Jun 2024   Prob (F-statistic):           6.64e-65
Time:                        17:04:21   Log-Likelihood:                 717.72
No. Observations:                 321   AIC:                            -1421.
Df Residuals:                     314   BIC:                            -1395.
Df Model:                           6
Covariance Type:                  HAC
==========================================================================================
                             coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------------------
const                     -0.0006      0.006     -0.089      0.929      -0.013       0.012
Zero_Count_5               0.0010      0.002      0.410      0.682      -0.004       0.006
Zero_Count_22             -0.0003      0.001     -0.447      0.655      -0.001       0.001
Log_Return_Lag1            1.2167      0.127      9.596      0.000       0.968       1.465
Log_Return_Lag2           -0.1038      0.133     -0.782      0.434      -0.364       0.156
Log_Return_Lag1_Zero5     -0.2886      0.036     -8.052      0.000      -0.359      -0.218
Log_Return_Lag2_Zero22     0.0286      0.011      2.618      0.009       0.007       0.050
==============================================================================
Omnibus:                       87.066   Durbin-Watson:                   1.952
Prob(Omnibus):                  0.000   Jarque-Bera (JB):              525.799
Skew:                           0.960   Prob(JB):                    6.67e-115
Kurtosis:                       8.969   Cond. No.                     1.27e+03
==============================================================================

Notes:
[1] Standard Errors are heteroscedasticity and autocorrelation robust (HAC) using 4 lags and without small sample correction
[2] The condition number is large, 1.27e+03. This might indicate that there are
strong multicollinearity or other numerical problems.
```

## Comparsion among last three models

In [ ]:
```
model1_results = ar2_model_with_weekly_zero
model2_results = ar2_model_with_monthly_zero
model3_results = ar2_model_with_separate_illiquidity

# Create a DataFrame to summarize the fit statistics
summary_stats = pd.DataFrame({
    'Model': ['Model with Weekly Zero', 'Model with Monthly Zero', 'Model with Separate Illiquidity'],
    'R-squared': [model1_results.rsquared, model2_results.rsquared, model3_results.rsquared],
    'Adj. R-squared': [model1_results.rsquared_adj, model2_results.rsquared_adj, model3_results.rsquared_adj],
    'AIC': [model1_results.aic, model2_results.aic, model3_results.aic],
    'BIC': [model1_results.bic, model2_results.bic, model3_results.bic],
    'F-statistic': [model1_results.fvalue, model2_results.fvalue, model3_results.fvalue]
})

# Print the summary statistics using tabulate
print(tabulate(summary_stats, headers='keys', tablefmt='pretty', showindex=False))
```

```
+---------------------------------+--------------------+--------------------+---------------------+---------------------+--------------------+
|              Model              |     R-squared      |   Adj. R-squared   |         AIC         |         BIC         |    F-statistic     |
+---------------------------------+--------------------+--------------------+---------------------+---------------------+--------------------+
|     Model with Weekly Zero      | 0.6555624644924585 | 0.6500952020240849 | -1414.5814877806774 | -1391.9528410418973 | 61.767219648707986 |
|     Model with Monthly Zero     | 0.5398221716814222 | 0.5325177617081115 | -1321.587709665845  | -1298.959062927065  | 43.83645709472929  |
| Model with Separate Illiquidity | 0.6649370501762542 | 0.6585345734280297 | -1421.4392679155076 | -1395.0391800535974 | 89.28868785518402  |
+---------------------------------+--------------------+--------------------+---------------------+---------------------+--------------------+
```

## Interpretation

**R-squared** & **Adj. R-squared**: The proportion of the variance in the dependent variable that is predictable from the independent variables. The higher these values, the better the model explains the variability of the response variable. From the table, the *Model with Separate Illiquidity* has both the highest R-squared and the highest Adjusted R-squared, which means it is the best model that explains the highest proportion of variance in the weekly log return series among the last three models.

**AIC** & **BIC**: Both criteria help in model selection where lower values generally indicate a better model. AIC shows the goodness of fit with a penalty for the number of parameters to avoid overfitting, and BIC considers a larger penalty for models with more parameters. The *Model with Separate Illiquidity* has both the lowest AIC and the lowest BIC, which means it is the best model from a complexity-fit trade-off perspective.

**F-statistic**: Indicates the overall significance of the regression AR(2) model. The higher the F-statistic, the more significant the model is. The *Model with Separate Illiquidity* has the highest F-statistic, which means it is statistically the most significant model in terms of the contribution of the explanatory variables used in the model.

## Conclusion

*Model with Separate Illiquidity* has advantages on all three aspects: explaining the variance in the weekly log return; balance between model complexity and fit; and also, the prediction ability, because the statistical significance of this model's explanatory variables is highest, making it potentially more reliable when making predictions or inferences.

Besides, *Model with Weekly Zero* is the second best model, where its variance explaining ability and complexity-fit balance are all very close to the *Model with Separate Illiquidity*.

The conclusion is just for Friday to Friday return series, it might be different for other days.

# P8: Rescale zero count series to zero fraction series

In [ ]:
```
# Rescale the zero counts to be fractions of their respective periods
Fridays['Zero_Fraction_5'] = Fridays['Zero_Count_5'] / 5
Fridays['Zero_Fraction_22'] = Fridays['Zero_Count_22'] / 22

print(Fridays[['Zero_Count_5', 'Zero_Fraction_5', 'Zero_Count_22', 'Zero_Fraction_22']].head())
# Double check the range
print(Fridays[['Zero_Count_5', 'Zero_Fraction_5', 'Zero_Count_22', 'Zero_Fraction_22']].describe())
```

```
            Zero_Count_5  Zero_Fraction_5  Zero_Count_22  Zero_Fraction_22
Date
2017-06-16           5.0              1.0           21.0          0.954545
2017-06-23           5.0              1.0           22.0          1.000000
2017-06-30           4.0              0.8           21.0          0.954545
2017-07-07           5.0              1.0           21.0          0.954545
2017-07-14           5.0              1.0           21.0          0.954545
       Zero_Count_5  Zero_Fraction_5  Zero_Count_22  Zero_Fraction_22
count    323.000000       323.000000     323.000000        323.000000
mean       3.421053         0.684211      15.089783          0.685899
std        1.584804         0.316961       5.608998          0.254954
min        0.000000         0.000000       0.000000          0.000000
25%        2.000000         0.400000      11.500000          0.522727
50%        4.000000         0.800000      17.000000          0.772727
75%        5.000000         1.000000      20.000000          0.909091
max        5.000000         1.000000      22.000000          1.000000
```

In [ ]:
```
Thursdays['Zero_Fraction_5'] = Thursdays['Zero_Count_5'] / 5
Thursdays['Zero_Fraction_22'] = Thursdays['Zero_Count_22'] / 22
print(Thursdays[['Zero_Count_5', 'Zero_Fraction_5', 'Zero_Count_22', 'Zero_Fraction_22']].describe())

Wednesdays['Zero_Fraction_5'] = Wednesdays['Zero_Count_5'] / 5
Wednesdays['Zero_Fraction_22'] = Wednesdays['Zero_Count_22'] / 22
print(Wednesdays[['Zero_Count_5', 'Zero_Fraction_5', 'Zero_Count_22', 'Zero_Fraction_22']].describe())

Tuesdays['Zero_Fraction_5'] = Tuesdays['Zero_Count_5'] / 5
Tuesdays['Zero_Fraction_22'] = Tuesdays['Zero_Count_22'] / 22
print(Tuesdays[['Zero_Count_5', 'Zero_Fraction_5', 'Zero_Count_22', 'Zero_Fraction_22']].describe())

Mondays['Zero_Fraction_5'] = Mondays['Zero_Count_5'] / 5
Mondays['Zero_Fraction_22'] = Mondays['Zero_Count_22'] / 22
print(Mondays[['Zero_Count_5', 'Zero_Fraction_5', 'Zero_Count_22', 'Zero_Fraction_22']].describe())
```

```
       Zero_Count_5  Zero_Fraction_5  Zero_Count_22  Zero_Fraction_22
count    332.000000       332.000000     332.000000        332.000000
mean       3.418675         0.683735      15.096386          0.686199
std        1.576646         0.315329       5.556335          0.252561
min        0.000000         0.000000       0.000000          0.000000
25%        2.000000         0.400000      11.000000          0.500000
50%        4.000000         0.800000      17.000000          0.772727
75%        5.000000         1.000000      20.000000          0.909091
max        5.000000         1.000000      22.000000          1.000000
       Zero_Count_5  Zero_Fraction_5  Zero_Count_22  Zero_Fraction_22
count    332.000000       332.000000     332.000000        332.000000
mean       3.418675         0.683735      15.036145          0.683461
std        1.533911         0.306782       5.529806          0.251355
min        0.000000         0.000000       0.000000          0.000000
25%        2.000000         0.400000      11.000000          0.500000
50%        4.000000         0.800000      17.000000          0.772727
75%        5.000000         1.000000      20.000000          0.909091
max        5.000000         1.000000      22.000000          1.000000
       Zero_Count_5  Zero_Fraction_5  Zero_Count_22  Zero_Fraction_22
count    334.000000       334.000000     334.000000        334.000000
mean       3.410180         0.682036      15.026946          0.683043
std        1.552684         0.310537       5.532529          0.251479
min        0.000000         0.000000       0.000000          0.000000
25%        2.000000         0.400000      11.000000          0.500000
50%        4.000000         0.800000      17.000000          0.772727
75%        5.000000         1.000000      20.000000          0.909091
max        5.000000         1.000000      22.000000          1.000000
       Zero_Count_5  Zero_Fraction_5  Zero_Count_22  Zero_Fraction_22
count    322.000000       322.000000     322.000000        322.000000
mean       3.403727         0.680745      15.090062          0.685912
std        1.576221         0.315244       5.598839          0.254493
min        0.000000         0.000000       0.000000          0.000000
25%        2.250000         0.450000      11.000000          0.500000
50%        4.000000         0.800000      17.000000          0.772727
75%        5.000000         1.000000      20.000000          0.909091
max        5.000000         1.000000      22.000000          1.000000
```

In [ ]:
```
# Fridays
csv_filename = 'Fridays_return_zeros_data.csv'
Fridays.to_csv(csv_filename, index=True, header=True)
files.download(csv_filename)

# Thursdays
csv_filename = 'Thursdays_return_zeros_data.csv'
Thursdays.to_csv(csv_filename, index=True, header=True)
```

```
files.download(csv_filename)

# Wednesdays
csv_filename = 'Wednesdays_return_zeros_data.csv'
Wednesdays.to_csv(csv_filename, index=True, header=True)
files.download(csv_filename)

# Tuesdays
csv_filename = 'Tuesdays_return_zeros_data.csv'
Tuesdays.to_csv(csv_filename, index=True, header=True)
files.download(csv_filename)

# Mondays
csv_filename = 'Mondays_return_zeros_data.csv'
Mondays.to_csv(csv_filename, index=True, header=True)
files.download(csv_filename)
```

## P9: Ten new models with rescaled zero fraction series

### M1: Constant

In [ ]:
```
# lags have no meaning, just for uniforming the parameter structure with AR models to perpare for the model summary function, add lags will not affect the result
def estimate_model_constant(df, lags):
    data = df.copy()
    X = sm.add_constant(pd.Series(1, index=data.index))
    Y = data['Log_Return']
    model = sm.OLS(Y, X).fit()
    return model
```

### M2: Constant + z5(t-1)

In [ ]:
```
# lags have no meaning, just for uniforming the parameter structure with AR models to perpare for the model summary function, add lags will not affect the result
def estimate_model_constant_z5(df, lags):
    data = df.copy()
    data['Z5_Lag1'] = data['Zero_Fraction_5'].shift(1)
    data.dropna(inplace=True)
    X = sm.add_constant(data['Z5_Lag1'])  # Adding a constant
    Y = data['Log_Return']
    model = sm.OLS(Y, X).fit()
    return model
```

### M3: Constant + z22(t-1)

In [ ]:
```
# lags have no meaning, just for uniforming the parameter structure with AR models to perpare for the model summary function, add lags will not affect the result
def estimate_model_constant_z22(df, lags):
    data = df.copy()
    data['Z22_Lag1'] = data['Zero_Fraction_22'].shift(1)
    data.dropna(inplace=True)
    X = sm.add_constant(data['Z22_Lag1'])
    Y = data['Log_Return']
    model = sm.OLS(Y, X).fit()
    return model
```

### M4: AR(1)

In [ ]:
```
def estimate_model_ar1(df, lags):
    data = df.copy()
    data['Log_Return_Lag1'] = data['Log_Return'].shift(1)
    data.dropna(inplace=True)
    X = sm.add_constant(data['Log_Return_Lag1'])
    Y = data['Log_Return']
    model = sm.OLS(Y, X).fit(cov_type='HAC', cov_kwds={'maxlags': lags})
    return model
```

### M5: AR(2)

In [ ]:
```
def estimate_model_ar2(df, lags):
    data = df.copy()
    data['Log_Return_Lag1'] = data['Log_Return'].shift(1)
    data['Log_Return_Lag2'] = data['Log_Return'].shift(2)
    data.dropna(inplace=True)
    X = sm.add_constant(data[['Log_Return_Lag1', 'Log_Return_Lag2']])
    Y = data['Log_Return']
    model = sm.OLS(Y, X).fit(cov_type='HAC', cov_kwds={'maxlags': lags})
    return model
```

### M6: AR(1) with z5(t-1) interaction

In [ ]:
```
def estimate_model_ar1_z5_interaction(df, lags):
    data = df.copy()
    data['Log_Return_Lag1'] = data['Log_Return'].shift(1)
    data['Z5_Lag1'] = data['Zero_Fraction_5'].shift(1)
    data.dropna(inplace=True)
    data['Log_Return_Lag1_Zero5'] = data['Log_Return_Lag1'] * data['Z5_Lag1']
    X = sm.add_constant(data[['Z5_Lag1', 'Log_Return_Lag1', 'Log_Return_Lag1_Zero5']])
    Y = data['Log_Return']
    model = sm.OLS(Y, X).fit(cov_type='HAC', cov_kwds={'maxlags': lags})
    return model
```

### M7: AR(1) with z22(t-1) interaction

In [ ]:
```
def estimate_model_ar1_z22_interaction(df, lags):
    data = df.copy()
    data['Log_Return_Lag1'] = data['Log_Return'].shift(1)
    data['Z22_Lag1'] = data['Zero_Fraction_22'].shift(1)
    data.dropna(inplace=True)
    data['Log_Return_Lag1_Zero22'] = data['Log_Return_Lag1'] * data['Z22_Lag1']
    X = sm.add_constant(data[['Z22_Lag1', 'Log_Return_Lag1', 'Log_Return_Lag1_Zero22']])
    Y = data['Log_Return']
    model = sm.OLS(Y, X).fit(cov_type='HAC', cov_kwds={'maxlags': lags})
    return model
```

### M8: AR(2) with z5(t-1) interaction

In [ ]:
```
def estimate_model_ar2_z5_interaction(df, lags):
    """
    Estimate an AR(2) model for weekly log returns, incorporating
    weekly zero fraction series as interaction effect,
    and allowing for specification of lags for HAC standard errors.

    Parameters:
    - df: DataFrame containing the weekly log returns under 'Log_Return',
      and the weekly zero fraction series 'Zero_Fraction_5'.
    - lags: Maximum number of lags to use for HAC standard errors.

    Returns:
    - model: OLS regression results containing the fitted model with HAC standard errors.
    """

    data = df.copy()
    data['Log_Return_Lag1'] = data['Log_Return'].shift(1)
    data['Log_Return_Lag2'] = data['Log_Return'].shift(2)
    data['Z5_Lag1'] = data['Zero_Fraction_5'].shift(1)
    data.dropna(inplace=True)
    data['Log_Return_Lag1_Zero5'] = data['Log_Return_Lag1'] * data['Z5_Lag1']
    data['Log_Return_Lag2_Zero5'] = data['Log_Return_Lag2'] * data['Z5_Lag1']
    X = data[['Z5_Lag1', 'Log_Return_Lag1', 'Log_Return_Lag2',
              'Log_Return_Lag1_Zero5', 'Log_Return_Lag2_Zero5']]
    X = sm.add_constant(X)
    Y = data['Log_Return']
    model = sm.OLS(Y, X).fit(cov_type='HAC', cov_kwds={'maxlags': lags})
    return model
```

### M9: AR(2) with z22(t-1) interaction

In [ ]:
```python
def estimate_model_ar2_z22_interaction(df, lags):
    """
    Estimate an AR(2) model for weekly log returns, incorporating
    monthly zero fraction series as interaction effect,
    and allowing for specification of lags for HAC standard errors.

    Parameters:
    - df: DataFrame containing the weekly log returns under 'Log_Return',
        and the monthly zero fraction series 'Zero_Fraction_22'.
    - lags: Maximum number of lags to use for HAC standard errors.

    Returns:
    - model: OLS regression results containing the fitted model with HAC standard errors.
    """

    data = df.copy()
    data['Log_Return_Lag1'] = data['Log_Return'].shift(1)
    data['Log_Return_Lag2'] = data['Log_Return'].shift(2)
    data['Z22_Lag1'] = data['Zero_Fraction_22'].shift(1)
    data.dropna(inplace=True)
    data['Log_Return_Lag1_Zero22'] = data['Log_Return_Lag1'] * data['Z22_Lag1']
    data['Log_Return_Lag2_Zero22'] = data['Log_Return_Lag2'] * data['Z22_Lag1']
    X = data[['Z22_Lag1', 'Log_Return_Lag1', 'Log_Return_Lag2',
              'Log_Return_Lag1_Zero22', 'Log_Return_Lag2_Zero22']]
    X = sm.add_constant(X)
    Y = data['Log_Return']
    model = sm.OLS(Y, X).fit(cov_type='HAC', cov_kwds={'maxlags': lags})
    return model
```

## M10: AR(2) with z5(t-1) interaction for first lag and z22(t-1) for second lag

In [ ]:
```python
def estimate_model_ar2_z5_z22_separate_interaction(df, lags):
    """
    Estimate an AR(2) model for weekly log returns, incorporating
    weekly and monthly zero fraction series as interaction effect,
    and allowing for specification of lags for HAC standard errors.

    Parameters:
    - df: DataFrame containing the weekly log returns under 'Log_Return',
        and two zero fraction series 'Zero_Fraction_22' and 'Zero_Fraction_5'.
    - lags: Maximum number of lags to use for HAC standard errors.

    Returns:
    - model: OLS regression results containing the fitted model with HAC standard errors.
    """
    data = df.copy()
    data['Log_Return_Lag1'] = data['Log_Return'].shift(1)
    data['Log_Return_Lag2'] = data['Log_Return'].shift(2)
    data['Z5_Lag1'] = data['Zero_Fraction_5'].shift(1)
    data['Z22_Lag1'] = data['Zero_Fraction_22'].shift(1)
    data.dropna(inplace=True)
    data['Log_Return_Lag1_Zero5'] = data['Log_Return_Lag1'] * data['Z5_Lag1']
    data['Log_Return_Lag2_Zero22'] = data['Log_Return_Lag2'] * data['Z22_Lag1']
    X = data[['Z5_Lag1', 'Z22_Lag1', 'Log_Return_Lag1', 'Log_Return_Lag2',
              'Log_Return_Lag1_Zero5', 'Log_Return_Lag2_Zero22']]
    X = sm.add_constant(X)
    Y = data['Log_Return']
    model = sm.OLS(Y, X).fit(cov_type='HAC', cov_kwds={'maxlags': lags})
    return model
```

In [ ]:
```python
# Ten Model functions mapped to their descriptive names
models = [('Model with Only Constant', estimate_model_constant),
      ('Model with Weekly Zero', estimate_model_constant_z5),
      ('Model with Monthly Zero', estimate_model_constant_z22),
      ('AR(1)', estimate_model_ar1),
      ('AR(2)', estimate_model_ar2),
      ('AR(1) with Weekly Zero Interaction', estimate_model_ar1_z5_interaction),
      ('AR(1) with Monthly Zero Interaction', estimate_model_ar1_z22_interaction),
      ('AR(2) with Weekly Zero Interaction', estimate_model_ar2_z5_interaction),
      ('AR(2) with Monthly Zero Interaction', estimate_model_ar2_z22_interaction),
      ('AR(2) with Separate Weekly and Monthly Zero Interactions', estimate_model_ar2_z5_z22_separate_interaction)]
```

In [ ]:
```python
model_features = {
    'Model with Weekly Zero': ['const', 'Z5_Lag1'],
    'Model with Monthly Zero': ['const', 'Z22_Lag1'],
    'AR(1)': ['const', 'Log_Return_Lag1'],
    'AR(2)': ['const', 'Log_Return_Lag1', 'Log_Return_Lag2'],
    'AR(1) with Weekly Zero Interaction': ['const', 'Z5_Lag1', 'Log_Return_Lag1', 'Log_Return_Lag1_Zero5'],
    'AR(1) with Monthly Zero Interaction': ['const', 'Z22_Lag1', 'Log_Return_Lag1', 'Log_Return_Lag1_Zero22'],
    'AR(2) with Weekly Zero Interaction': ['const', 'Z5_Lag1', 'Log_Return_Lag1', 'Log_Return_Lag2', 'Log_Return_Lag1_Zero5', 'Log_Return_Lag2_Zero5'],
    'AR(2) with Monthly Zero Interaction': ['const', 'Z22_Lag1', 'Log_Return_Lag1', 'Log_Return_Lag2', 'Log_Return_Lag1_Zero22', 'Log_Return_Lag2_Zero22'],
    'AR(2) with Separate Weekly and Monthly Zero Interactions': ['const', 'Z5_Lag1', 'Z22_Lag1', 'Log_Return_Lag1', 'Log_Return_Lag2', 'Log_Return_Lag1_Zero5', 'Log_Return_Lag2_Zero22']
}
```

## P10: Models Summary & Comparsion (Use Fridays as example with maxlag=4)

In [ ]:
```python
def autocorrelations_residuals(model):
    residuals = model.resid
    # Compute autocorrelations for the first two lags
    autocorr = acf(residuals, nlags=2, fft=True)
    # Perform Ljung-Box test for three lags
    ljungbox_results = acorr_ljungbox(residuals, lags=[3], return_df=True)
    results = {
        'Rho_1': autocorr[1],
        'Rho_2': autocorr[2],
        'Ljung-Box Test Statistic': ljungbox_results['lb_stat'].iloc[0],
        'Ljung-Box P-value': ljungbox_results['lb_pvalue'].iloc[0]
    }
    return results
```

In [ ]:
```python
def models_comparison(df, lags, df_name):
    """
    Analyze and compare multiple regression models on a given DataFrame.
    This function fits nine different models, each incorporating
    different aspects of zero fraction interactions and lags, to a
    specified DataFrame. It evaluates each model's performance using
    Adjusted R-squared, AIC, BIC, and conducts residual diagnostics
    using autocorrelations and the Ljung-Box test.

    Parameters:
    - df: The input DataFrame containing the time series data. The
        DataFrame should include columns for log returns and zero fractions.
    - lags: Maximum number of lags to use for HAC standard errors.
    - df_name: A string that specifies the name of the DataFrame,
          used to name the output CSV file.

    Returns:
    - model_results_dict: A dictionary containing the fitted models.

    Outputs:
    - Prints a summary table directly to the console, displaying
      key model statistics and parameters' results.
    - Prints detailed summary statistics for each model, providing
      comparable insights into the model fit and residuals.
    - Saves a CSV file named '{df_name}_model_comparison_results.csv'
      containing the detailed summary comparsion results.

    Example of usage:
    Assuming 'Fridays' is a DataFrame suitable for the analysis:
    fridays_models = models_comparison(Fridays, lags=4, df_name='Fridays')
    """

    data = df.copy()
    model_results_dict = {}
    results_list = []

    for name, model_func in models:
        model_fit = model_func(df=data, lags=lags)
        model_results_dict[name] = model_fit
        analysis_results = autocorrelations_residuals(model_fit)

        # Prepare a dictionary for each model's results
        model_info = {
            'Model Name': name,
            'Adj R^2': model_fit.rsquared_adj,
            'AIC': model_fit.aic,
            'BIC': model_fit.bic,
            'Rho_1': analysis_results['Rho_1'],
            'Rho_2': analysis_results['Rho_2'],
            'Ljung-Box Test Statistic': analysis_results['Ljung-Box Test Statistic'],
            'Ljung-Box P-value': analysis_results['Ljung-Box P-value']
        }

        results_list.append(model_info)

    results_df = pd.DataFrame(results_list)

    # Save the DataFrame to a CSV file with a dynamic name based on df_name
    # files.download is just for google colab, if u use other enviroments, plz change
    csv_filename = f'{df_name}_model_comparison_results.csv'
    results_df.to_csv(csv_filename, index=True, header=True)
    files.download(csv_filename)

    print(tabulate(results_df, headers='keys', tablefmt='pretty', showindex=False))

    # Print parameter summaries for each model, separated by two blank lines
    for model_name, model_result in model_results_dict.items():
        print(f"Parameters for Model: {model_name}:")
        print(model_result.summary())
        print('\n' * 2)
    return model_results_dict
```

In [ ]:
```python
fridays_models = models_comparison(Fridays, lags=4, df_name='Fridays')
```

| Model Name | Adj R^2 | AIC | BIC | Rho_1 | Rho_2 | Ljung-Box Test Statistic | Ljung-Box P-value |
|---|---|---|---|---|---|---|---|
| Model with Only Constant | 2.220446049250313e-16 | -1091.2087686271934 | -1087.4311163039708 | 0.7051491115846917 | 0.5939695382896838 | 372.71304911867526 | 1.7996068972050845e-80 |
| Model with Weekly Zero | 0.0006396222829437859 | -1086.0365555375033 | -1078.4874524464144 | 0.7033701906196563 | 0.5908909175502778 | 370.0266260053934 | 6.870124334073737e-80 |
| Model with Monthly Zero | 0.00043030332805416016 | -1085.9691187576136 | -1078.4200156665247 | 0.704146496376675 | 0.5930101123396581 | 370.92228693755567 | 4.395380344949907e-80 |
| AR(1) | 0.4956643744483322 | -1306.2438167084683 | -1298.6947136173794 | -0.1356912954248646 | 0.02417237892107988 | 29.881012620618524 | 1.461922715402241e-06 |
| AR(2) | 0.5128527872241293 | -1311.318225508016 | -1300.003902138626 | -0.025673654543057242 | -0.05067886504282071 | 17.42342656969924 | 0.0005782616530108395 |
| AR(1) with Weekly Zero Interaction | 0.49294967265322254 | -1302.5340415943133 | -1287.4358354121357 | -0.13479051264275915 | 0.02251678901729949 | 29.763726221081924 | 1.5473588401116546e-06 |
| AR(1) with Monthly Zero Interaction | 0.49628199580928833 | -1304.6572025805049 | -1289.5589963983273 | -0.10700004709102476 | 0.026896555881336548 | 29.414772333817737 | 1.832154994375916e-06 |
| AR(2) with Weekly Zero Interaction | 0.5127447891107964 | -1308.2897461017178 | -1285.6610993629376 | -0.02327658656985997 | -0.06344992042876285 | 18.971899628236205 | 0.00027708098141720073 |
| AR(2) with Monthly Zero Interaction | 0.5194093779465099 | -1312.7106295895924 | -1290.0819828508122 | -0.035235679222936384 | -0.08749058938374132 | 23.101765389640835 | 3.8457884576106816e-05 |
| AR(2) with Separate Weekly and Monthly Zero Interactions | 0.5131122848344245 | -1307.552609400545 | -1281.1525215386348 | -0.0364309638420215 | -0.08550912709705655 | 20.599383574776333 | 0.00012749525510854525 |

Parameters for Model: Model with Only Constant:

```
                            OLS Regression Results
==============================================================================
Dep. Variable:             Log_Return   R-squared:                       0.000
Model:                            OLS   Adj. R-squared:                  0.000
Method:                 Least Squares   F-statistic:                       nan
Date:                Mon, 03 Jun 2024   Prob (F-statistic):                nan
Time:                        17:59:04   Log-Likelihood:                 546.60
No. Observations:                 323   AIC:                            -1091.
Df Residuals:                     322   BIC:                            -1087.
Df Model:                           0
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
0             -0.0004      0.002     -0.160      0.873      -0.005       0.004
==============================================================================
Omnibus:                       59.871   Durbin-Watson:                   0.590
Prob(Omnibus):                  0.000   Jarque-Bera (JB):              426.209
Skew:                           0.509   Prob(JB):                     2.82e-93
Kurtosis:                       8.535   Cond. No.                         1.00
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
```

Parameters for Model: Model with Weekly Zero:

```
                            OLS Regression Results
==============================================================================
Dep. Variable:             Log_Return   R-squared:                       0.004
Model:                            OLS   Adj. R-squared:                  0.001
Method:                 Least Squares   F-statistic:                     1.205
Date:                Mon, 03 Jun 2024   Prob (F-statistic):              0.273
Time:                        17:59:04   Log-Likelihood:                 545.02
No. Observations:                 322   AIC:                            -1086.
Df Residuals:                     320   BIC:                            -1078.
Df Model:                           1
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const          0.0055      0.006      0.929      0.354      -0.006       0.017
Z5_Lag1       -0.0086      0.008     -1.098      0.273      -0.024       0.007
==============================================================================
Omnibus:                       51.319   Durbin-Watson:                   0.593
Prob(Omnibus):                  0.000   Jarque-Bera (JB):              400.989
Skew:                           0.315   Prob(JB):                     8.44e-88
Kurtosis:                       8.431   Cond. No.                         4.74
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
```

Parameters for Model: Model with Monthly Zero:

```
                            OLS Regression Results
==============================================================================
Dep. Variable:             Log_Return   R-squared:                       0.004
Model:                            OLS   Adj. R-squared:                  0.000
Method:                 Least Squares   F-statistic:                     1.138
Date:                Mon, 03 Jun 2024   Prob (F-statistic):              0.287
Time:                        17:59:04   Log-Likelihood:                 544.98
No. Observations:                 322   AIC:                            -1086.
Df Residuals:                     320   BIC:                            -1078.
Df Model:                           1
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const          0.0068      0.007      0.945      0.346      -0.007       0.021
Z22_Lag1      -0.0104      0.010     -1.067      0.287      -0.030       0.009
==============================================================================
Omnibus:                       50.717   Durbin-Watson:                   0.592
Prob(Omnibus):                  0.000   Jarque-Bera (JB):              395.353
Skew:                           0.303   Prob(JB):                     1.41e-86
Kurtosis:                       8.394   Cond. No.                         5.86
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
```

Parameters for Model: AR(1):

```
                            OLS Regression Results
==============================================================================
Dep. Variable:             Log_Return   R-squared:                       0.497
Model:                            OLS   Adj. R-squared:                  0.496
Method:                 Least Squares   F-statistic:                     156.3
Date:                Mon, 03 Jun 2024   Prob (F-statistic):           1.77e-29
Time:                        17:59:04   Log-Likelihood:                 655.12
No. Observations:                 322   AIC:                            -1306.
Df Residuals:                     320   BIC:                            -1299.
Df Model:                           1
Covariance Type:                  HAC
===================================================================================
                      coef    std err          z      P>|z|      [0.025      0.975]
-----------------------------------------------------------------------------------
const              -0.0001      0.002     -0.066      0.947      -0.004       0.003
Log_Return_Lag1     0.7051      0.056     12.503      0.000       0.595       0.816
==============================================================================
Omnibus:                       64.236   Durbin-Watson:                   2.270
Prob(Omnibus):                  0.000   Jarque-Bera (JB):              356.332
Skew:                           0.671   Prob(JB):                     4.20e-78
Kurtosis:                       7.976   Cond. No.                         22.4
==============================================================================

Notes:
[1] Standard Errors are heteroscedasticity and autocorrelation robust (HAC) using 4 lags and without small sample correction
```

Parameters for Model: AR(2):

```
                            OLS Regression Results
==============================================================================
Dep. Variable:             Log_Return   R-squared:                       0.516
Model:                            OLS   Adj. R-squared:                  0.513
Method:                 Least Squares   F-statistic:                     84.51
Date:                Mon, 03 Jun 2024   Prob (F-statistic):           3.69e-30
Time:                        17:59:04   Log-Likelihood:                 658.66
No. Observations:                 321   AIC:                            -1311.
Df Residuals:                     318   BIC:                            -1300.
Df Model:                           2
Covariance Type:                  HAC
===================================================================================
                      coef    std err          z      P>|z|      [0.025      0.975]
-----------------------------------------------------------------------------------
const           -7.831e-05      0.002     -0.044      0.965      -0.004       0.003
Log_Return_Lag1     0.5691      0.079      7.160      0.000       0.413       0.725
Log_Return_Lag2     0.1929      0.088      2.188      0.029       0.020       0.366
==============================================================================
Omnibus:                       73.893   Durbin-Watson:                   2.050
Prob(Omnibus):                  0.000   Jarque-Bera (JB):              539.301
Skew:                           0.713   Prob(JB):                    7.80e-118
Kurtosis:                       9.188   Cond. No.                         41.3
==============================================================================

Notes:
[1] Standard Errors are heteroscedasticity and autocorrelation robust (HAC) using 4 lags and without small sample correction
```

Parameters for Model: AR(1) with Weekly Zero Interaction:

```
                            OLS Regression Results
==============================================================================
Dep. Variable:             Log_Return   R-squared:                       0.498
Model:                            OLS   Adj. R-squared:                  0.493
Method:                 Least Squares   F-statistic:                     52.12
Date:                Mon, 03 Jun 2024   Prob (F-statistic):           2.00e-27
Time:                        17:59:04   Log-Likelihood:                 655.27
No. Observations:                 322   AIC:                            -1303.
Df Residuals:                     318   BIC:                            -1287.
Df Model:                           3
Covariance Type:                  HAC
=========================================================================================
                            coef    std err          z      P>|z|      [0.025      0.975]
-----------------------------------------------------------------------------------------
const                     0.0019      0.007      0.286      0.775      -0.011       0.015
Z5_Lag1                  -0.0029      0.008     -0.365      0.715      -0.019       0.013
Log_Return_Lag1           0.6988      0.083      8.449      0.000       0.537       0.861
Log_Return_Lag1_Zero5     0.0263      0.276      0.096      0.924      -0.514       0.567
==============================================================================
Omnibus:                       60.877   Durbin-Watson:                   2.268
Prob(Omnibus):                  0.000   Jarque-Bera (JB):              341.475
Skew:                           0.618   Prob(JB):                     7.07e-75
Kurtosis:                       7.891   Cond. No.                         132.
==============================================================================

Notes:
[1] Standard Errors are heteroscedasticity and autocorrelation robust (HAC) using 4 lags and without small sample correction
```

Parameters for Model: AR(1) with Monthly Zero Interaction:

```
                            OLS Regression Results
==============================================================================
Dep. Variable:             Log_Return   R-squared:                       0.501
Model:                            OLS   Adj. R-squared:                  0.496
Method:                 Least Squares   F-statistic:                     55.86
Date:                Mon, 03 Jun 2024   Prob (F-statistic):           4.97e-29
Time:                        17:59:04   Log-Likelihood:                 656.33
No. Observations:                 322   AIC:                            -1305.
Df Residuals:                     318   BIC:                            -1290.
Df Model:                           3
Covariance Type:                  HAC
========================================================================================
                           coef    std err          z      P>|z|      [0.025      0.975]
----------------------------------------------------------------------------------------
const                    0.0007      0.007      0.101      0.920      -0.014       0.015
Z22_Lag1                -0.0016      0.009     -0.185      0.853      -0.019       0.016
Log_Return_Lag1          0.7979      0.087      9.149      0.000       0.627       0.969
Log_Return_Lag1_Zero22  -0.2807      0.184     -1.525      0.127      -0.642       0.080
==============================================================================
Omnibus:                       61.436   Durbin-Watson:                   2.213
Prob(Omnibus):                  0.000   Jarque-Bera (JB):              348.038
Skew:                           0.623   Prob(JB):                     2.66e-76
Kurtosis:                       7.939   Cond. No.                         135.
==============================================================================

Notes:
[1] Standard Errors are heteroscedasticity and autocorrelation robust (HAC) using 4 lags and without small sample correction


Parameters for Model: AR(2) with Weekly Zero Interaction:
                            OLS Regression Results
==============================================================================
Dep. Variable:             Log_Return   R-squared:                       0.520
Model:                            OLS   Adj. R-squared:                  0.513
Method:                 Least Squares   F-statistic:                     33.76
Date:                Mon, 03 Jun 2024   Prob (F-statistic):           1.40e-27
Time:                        17:59:04   Log-Likelihood:                 660.14
No. Observations:                 321   AIC:                            -1308.
Df Residuals:                     315   BIC:                            -1286.
Df Model:                           5
Covariance Type:                  HAC
=======================================================================================
                          coef    std err          z      P>|z|      [0.025      0.975]
---------------------------------------------------------------------------------------
const                   0.0016      0.007      0.235      0.814      -0.012       0.015
Z5_Lag1                -0.0024      0.008     -0.305      0.760      -0.018       0.013
Log_Return_Lag1         0.6817      0.131      5.185      0.000       0.424       0.939
Log_Return_Lag2         0.0430      0.197      0.218      0.828      -0.344       0.429
Log_Return_Lag1_Zero5  -0.2924      0.287     -1.021      0.307      -0.854       0.269
Log_Return_Lag2_Zero5   0.3010      0.256      1.176      0.240      -0.201       0.803
==============================================================================
Omnibus:                       65.819   Durbin-Watson:                   2.045
Prob(Omnibus):                  0.000   Jarque-Bera (JB):              467.017
Skew:                           0.606   Prob(JB):                    3.88e-102
Kurtosis:                       8.783   Cond. No.                         199.
==============================================================================

Notes:
[1] Standard Errors are heteroscedasticity and autocorrelation robust (HAC) using 4 lags and without small sample correction


Parameters for Model: AR(2) with Monthly Zero Interaction:
                            OLS Regression Results
==============================================================================
Dep. Variable:             Log_Return   R-squared:                       0.527
Model:                            OLS   Adj. R-squared:                  0.519
Method:                 Least Squares   F-statistic:                     35.94
Date:                Mon, 03 Jun 2024   Prob (F-statistic):           4.50e-29
Time:                        17:59:04   Log-Likelihood:                 662.36
No. Observations:                 321   AIC:                            -1313.
Df Residuals:                     315   BIC:                            -1290.
Df Model:                           5
Covariance Type:                  HAC
========================================================================================
                           coef    std err          z      P>|z|      [0.025      0.975]
----------------------------------------------------------------------------------------
const                    0.0008      0.008      0.105      0.916      -0.014       0.016
Z22_Lag1                -0.0011      0.009     -0.119      0.905      -0.019       0.017
Log_Return_Lag1          0.8167      0.156      5.240      0.000       0.511       1.122
Log_Return_Lag2         -0.0901      0.195     -0.463      0.644      -0.472       0.292
Log_Return_Lag1_Zero22  -0.5191      0.233     -2.227      0.026      -0.976      -0.062
Log_Return_Lag2_Zero22   0.6352      0.321      1.980      0.048       0.006       1.264
==============================================================================
Omnibus:                       63.066   Durbin-Watson:                   2.069
Prob(Omnibus):                  0.000   Jarque-Bera (JB):              406.659
Skew:                           0.601   Prob(JB):                     4.96e-89
Kurtosis:                       8.381   Cond. No.                         239.
==============================================================================

Notes:
[1] Standard Errors are heteroscedasticity and autocorrelation robust (HAC) using 4 lags and without small sample correction


Parameters for Model: AR(2) with Separate Weekly and Monthly Zero Interactions:
                            OLS Regression Results
==============================================================================
Dep. Variable:             Log_Return   R-squared:                       0.522
Model:                            OLS   Adj. R-squared:                  0.513
Method:                 Least Squares   F-statistic:                     29.40
Date:                Mon, 03 Jun 2024   Prob (F-statistic):           6.58e-28
Time:                        17:59:04   Log-Likelihood:                 660.78
No. Observations:                 321   AIC:                            -1308.
Df Residuals:                     314   BIC:                            -1281.
Df Model:                           6
Covariance Type:                  HAC
========================================================================================
                           coef    std err          z      P>|z|      [0.025      0.975]
----------------------------------------------------------------------------------------
const                    0.0008      0.008      0.097      0.923      -0.015       0.016
Z5_Lag1                 -0.0038      0.010     -0.372      0.710      -0.024       0.016
Z22_Lag1                 0.0028      0.012      0.223      0.824      -0.022       0.027
Log_Return_Lag1          0.6352      0.100      6.377      0.000       0.440       0.830
Log_Return_Lag2          0.0540      0.156      0.346      0.730      -0.252       0.360
Log_Return_Lag1_Zero5   -0.2334      0.276     -0.845      0.398      -0.775       0.308
Log_Return_Lag2_Zero22   0.3948      0.255      1.546      0.122      -0.106       0.895
==============================================================================
Omnibus:                       71.379   Durbin-Watson:                   2.072
Prob(Omnibus):                  0.000   Jarque-Bera (JB):              511.792
Skew:                           0.684   Prob(JB):                    7.34e-112
Kurtosis:                       9.033   Cond. No.                         204.
==============================================================================

Notes:
[1] Standard Errors are heteroscedasticity and autocorrelation robust (HAC) using 4 lags and without small sample correction
```

In [ ]:

```python
# Example for extracting a stored model
print(fridays_models['AR(1) with Monthly Zero Interaction'].summary())
```

```
                            OLS Regression Results
==============================================================================
Dep. Variable:             Log_Return   R-squared:                       0.501
Model:                            OLS   Adj. R-squared:                  0.496
Method:                 Least Squares   F-statistic:                     55.86
Date:                Mon, 03 Jun 2024   Prob (F-statistic):           4.97e-29
Time:                        17:59:57   Log-Likelihood:                 656.33
No. Observations:                 322   AIC:                            -1305.
Df Residuals:                     318   BIC:                            -1290.
Df Model:                           3
Covariance Type:                  HAC
========================================================================================
                           coef    std err          z      P>|z|      [0.025      0.975]
----------------------------------------------------------------------------------------
const                    0.0007      0.007      0.101      0.920      -0.014       0.015
Z22_Lag1                -0.0016      0.009     -0.185      0.853      -0.019       0.016
Log_Return_Lag1          0.7979      0.087      9.149      0.000       0.627       0.969
Log_Return_Lag1_Zero22  -0.2807      0.184     -1.525      0.127      -0.642       0.080
==============================================================================
Omnibus:                       61.436   Durbin-Watson:                   2.213
Prob(Omnibus):                  0.000   Jarque-Bera (JB):              348.038
Skew:                           0.623   Prob(JB):                     2.66e-76
Kurtosis:                       7.939   Cond. No.                         135.
==============================================================================

Notes:
[1] Standard Errors are heteroscedasticity and autocorrelation robust (HAC) using 4 lags and without small sample correction
```

## Evaluating Model Comparisons

We consider several factors such as the $R^2_{adj}$, $AIC$, $BIC$, autocorrelations ($Rho_1$ and $Rho_2$), and results from the Ljung-Box test. We break down each model's performance based on these criteria and determine the best model based on the evidence.

### Evaluating Model Performance

1. **$R^2_{adj}$:** Higher values indicate that the model explains a higher proportion of the variance in the log return, adjusted for the number of predictors. The AR(2) models with interactions, especially the model **AR(2) with Monthly Zero Interaction**, which has the highest value $(0.519)$, is the winner in this criteria. the model **AR(2) with Separate Weekly and Monthly Zero Interactions** is the second best model in this criteria with second highest value $(0.513)$

2. **$AIC$ and $BIC$:** Lower values of $AIC$ and $BIC$ indicate a better model fit, adjusted for the number of parameters. The model **AR(2) with Monthly Zero Interaction** has lowest $AIC$, and the model **AR(2)** has lowest $BIC$, suggesting they provide strong balances of model fit and complexity. The model **AR(2) with Separate Weekly and Monthly Zero Interactions** also has closely low $AIC$ and $BIC$ to these two winners.

3. **Autocorrelations ($Rho_1$ and $Rho_2$):** Ideally, these should be close to zero, indicating that the residuals from the model do not exhibit autocorrelation. The AR(1) and AR(2) models generally show very low autocorrelation. Specifically, the $Rho_1$ of model **AR(2) with Weekly Zero Interaction** is closest to zero, where the $Rho_1$ of model **AR(2) with Separate Weekly and Monthly Zero Interactions** is also close to zero. the $Rho_2$ of model **AR(2) with Weekly Zero Interaction** is closest to zero, where the $Rho_2$ of model **AR(2) with Separate Weekly and Monthly Zero Interactions** is still close enough to zero. But both differences to first closest model are tiny, so **AR(2) with Separate Weekly and Monthly Zero Interactions** can still be acceptable in this criteria.

4. **Ljung-Box Test:** This test checks for autocorrelation in the residuals at multiple lag lengths. A high p-value (close to 1) indicates that there is little evidence to reject the null hypothesis of no serial autocorrelations among residuals. The model **AR(2) with Separate Weekly and Monthly Zero Interactions** shows the third highest p-value (around $0.000127$), but is still too low.

### Conclusion

The model **AR(2) with Separate Weekly and Monthly Zero Interactions** and model **AR(2) with Monthly Zero Interaction** both balance complexity with performance effectively and manages residuals better than simpler models or those considering fewer interaction terms. They should be chosen for further forecasting assuming these results hold consistently across different dataframes (Eg. Wednesdays). These models' ability to handle interactions between different periods of zero fractions provides a comprehensive understanding of the factors influencing log returns, which can be especially valuable in further forecasting where such dynamics are often significant.

## *Robustness work for P10*

In [ ]:
```
# Thursdays
thursdays_models = models_comparison(Thursdays, lags=2, df_name='Thursdays')
```

| Model Name | Adj R^2 | AIC | BIC | Rho_1 | Rho_2 | Ljung-Box Test Statistic | Ljung-Box P-value |
|---|---|---|---|---|---|---|---|
| Model with Only Constant | 3.3306690738754696e-16 | -1129.6314046367324 | -1125.826269667816 | 0.7345946179183339 | 0.5997278534251737 | 383.393733423569 | 8.749828086232154e-83 |
| Model with Weekly Zero | -0.0015243351787677106 | -1123.7250132381773 | -1116.1207764874232 | 0.7327797490558713 | 0.5983789793575569 | 380.61905623055236 | 3.490952743252467e-82 |
| Model with Monthly Zero | -9.826455511507959e-05 | -1124.1966600474789 | -1116.5924232967247 | 0.733883153652476 | 0.5989658994510134 | 381.73385775671676 | 2.002164075177208e-82 |
| AR(1) | 0.5382307541461196 | -1379.9895675709663 | -1372.3853308202122 | -0.09593281484946242 | 0.04771025742450191 | 4.922570199686606 | 0.17755553347138409 |
| AR(2) | 0.5447343740533538 | -1378.5015778982604 | -1367.1042999348788 | -0.0038355974340777074 | -0.019796150845353307 | 0.5311140933125282 | 0.9120044970665508 |
| AR(1) with Weekly Zero Interaction | 0.5369816932146954 | -1377.1137380806786 | -1361.9052645791703 | -0.08889427167496415 | 0.05465040262697347 | 4.679437982052594 | 0.19683247834698503 |
| AR(1) with Monthly Zero Interaction | 0.5361682469072273 | -1376.5327362367798 | -1361.3242627352715 | -0.08368647486013457 | 0.04993412616090531 | 4.309916299168381 | 0.22988499813851404 |
| AR(2) with Weekly Zero Interaction | 0.5623701361413097 | -1388.5805650781188 | -1365.7860091513555 | -0.011922386047237972 | 0.017260370733125974 | 0.19937132616917788 | 0.9776907233741919 |
| AR(2) with Monthly Zero Interaction | 0.5503206131290315 | -1379.617315688911 | -1356.8227597621478 | -0.02411845582948623 | -0.04455039034174976 | 1.388723554182887 | 0.708180108850241 |
| AR(2) with Separate Weekly and Monthly Zero Interactions | 0.5505979703054391 | -1378.84101231318 | -1352.2473637319563 | -0.017609125547069608 | -0.04974084196507659 | 1.1814110684277064 | 0.7574660705088695 |

```
Parameters for Model: Model with Only Constant:
                            OLS Regression Results
==============================================================================
Dep. Variable:             Log_Return   R-squared:                       0.000
Model:                            OLS   Adj. R-squared:                  0.000
Method:                 Least Squares   F-statistic:                       nan
Date:                Mon, 03 Jun 2024   Prob (F-statistic):                nan
Time:                        18:00:03   Log-Likelihood:                 565.82
No. Observations:                 332   AIC:                            -1130.
Df Residuals:                     331   BIC:                            -1126.
Df Model:                           0
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
0             -0.0007      0.002     -0.286      0.775      -0.005       0.004
==============================================================================
Omnibus:                       44.240   Durbin-Watson:                   0.531
Prob(Omnibus):                  0.000   Jarque-Bera (JB):              237.047
Skew:                           0.349   Prob(JB):                     3.36e-52
Kurtosis:                       7.080   Cond. No.                         1.00
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.


Parameters for Model: Model with Weekly Zero:
                            OLS Regression Results
==============================================================================
Dep. Variable:             Log_Return   R-squared:                       0.002
Model:                            OLS   Adj. R-squared:                 -0.002
Method:                 Least Squares   F-statistic:                    0.4977
Date:                Mon, 03 Jun 2024   Prob (F-statistic):              0.481
Time:                        18:00:03   Log-Likelihood:                 563.86
No. Observations:                 331   AIC:                            -1124.
Df Residuals:                     329   BIC:                            -1116.
Df Model:                           1
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const          0.0030      0.006      0.521      0.603      -0.008       0.014
Z5_Lag1       -0.0054      0.008     -0.706      0.481      -0.021       0.010
==============================================================================
Omnibus:                       39.663   Durbin-Watson:                   0.534
Prob(Omnibus):                  0.000   Jarque-Bera (JB):              225.323
Skew:                           0.217   Prob(JB):                     1.18e-49
Kurtosis:                       7.019   Cond. No.                         4.76
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.


Parameters for Model: Model with Monthly Zero:
                            OLS Regression Results
==============================================================================
Dep. Variable:             Log_Return   R-squared:                       0.003
Model:                            OLS   Adj. R-squared:                 -0.000
Method:                 Least Squares   F-statistic:                    0.9676
Date:                Mon, 03 Jun 2024   Prob (F-statistic):              0.326
Time:                        18:00:03   Log-Likelihood:                 564.10
No. Observations:                 331   AIC:                            -1124.
Df Residuals:                     329   BIC:                            -1117.
Df Model:                           1
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const          0.0058      0.007      0.825      0.410      -0.008       0.020
Z22_Lag1      -0.0095      0.010     -0.984      0.326      -0.028       0.009
==============================================================================
Omnibus:                       38.218   Durbin-Watson:                   0.532
Prob(Omnibus):                  0.000   Jarque-Bera (JB):              220.889
Skew:                           0.158   Prob(JB):                     1.08e-48
Kurtosis:                       6.989   Cond. No.                         5.91
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.


Parameters for Model: AR(1):
                            OLS Regression Results
==============================================================================
Dep. Variable:             Log_Return   R-squared:                       0.540
Model:                            OLS   Adj. R-squared:                  0.538
Method:                 Least Squares   F-statistic:                     135.6
Date:                Mon, 03 Jun 2024   Prob (F-statistic):           1.78e-26
Time:                        18:00:03   Log-Likelihood:                 691.99
No. Observations:                 331   AIC:                            -1380.
Df Residuals:                     329   BIC:                            -1372.
Df Model:                           1
Covariance Type:                  HAC
==================================================================================
                     coef    std err          z      P>|z|      [0.025      0.975]
----------------------------------------------------------------------------------
const             -0.0002      0.002     -0.117      0.907      -0.003       0.003
Log_Return_Lag1    0.7346      0.063     11.645      0.000       0.611       0.858
==============================================================================
Omnibus:                       40.177   Durbin-Watson:                   2.190
Prob(Omnibus):                  0.000   Jarque-Bera (JB):              240.627
Skew:                           0.191   Prob(JB):                     5.60e-53
Kurtosis:                       7.159   Cond. No.                         22.7
==============================================================================

Notes:
[1] Standard Errors are heteroscedasticity and autocorrelation robust (HAC) using 2 lags and without small sample correction


Parameters for Model: AR(2):
                            OLS Regression Results
==============================================================================
Dep. Variable:             Log_Return   R-squared:                       0.548
Model:                            OLS   Adj. R-squared:                  0.545
Method:                 Least Squares   F-statistic:                     69.37
Date:                Mon, 03 Jun 2024   Prob (F-statistic):           7.71e-26
Time:                        18:00:03   Log-Likelihood:                 692.25
No. Observations:                 330   AIC:                            -1379.
Df Residuals:                     327   BIC:                            -1367.
Df Model:                           2
Covariance Type:                  HAC
===================================================================================
                      coef    std err          z      P>|z|      [0.025      0.975]
-----------------------------------------------------------------------------------
const              -0.0001      0.002     -0.092      0.926      -0.003       0.003
Log_Return_Lag1     0.6384      0.094      6.784      0.000       0.454       0.823
Log_Return_Lag2     0.1309      0.084      1.566      0.117      -0.033       0.295
==============================================================================
Omnibus:                       44.652   Durbin-Watson:                   2.006
Prob(Omnibus):                  0.000   Jarque-Bera (JB):              316.589
Skew:                           0.192   Prob(JB):                     1.79e-69
Kurtosis:                       7.783   Cond. No.                         44.0
==============================================================================

Notes:
[1] Standard Errors are heteroscedasticity and autocorrelation robust (HAC) using 2 lags and without small sample correction


Parameters for Model: AR(1) with Weekly Zero Interaction:
                            OLS Regression Results
==============================================================================
Dep. Variable:             Log_Return   R-squared:                       0.541
Model:                            OLS   Adj. R-squared:                  0.537
Method:                 Least Squares   F-statistic:                     49.46
Date:                Mon, 03 Jun 2024   Prob (F-statistic):           2.21e-26
Time:                        18:00:03   Log-Likelihood:                 692.56
No. Observations:                 331   AIC:                            -1377.
Df Residuals:                     327   BIC:                            -1362.
Df Model:                           3
Covariance Type:                  HAC
===========================================================================================
                              coef    std err          z      P>|z|      [0.025      0.975]
-------------------------------------------------------------------------------------------
const                       0.0008      0.006      0.134      0.894      -0.010       0.012
Z5_Lag1                    -0.0016      0.007     -0.240      0.811      -0.015       0.012
Log_Return_Lag1             0.7751      0.088      8.823      0.000       0.603       0.947
Log_Return_Lag1_Zero5      -0.1879      0.239     -0.788      0.431      -0.655       0.280
==============================================================================
Omnibus:                       40.701   Durbin-Watson:                   2.176
Prob(Omnibus):                  0.000   Jarque-Bera (JB):              256.558
Skew:                           0.164   Prob(JB):                     1.95e-56
Kurtosis:                       7.301   Cond. No.                         142.
==============================================================================

Notes:
[1] Standard Errors are heteroscedasticity and autocorrelation robust (HAC) using 2 lags and without small sample correction


Parameters for Model: AR(1) with Monthly Zero Interaction:
```

```
                            OLS Regression Results
==============================================================================
Dep. Variable:             Log_Return   R-squared:                       0.540
Model:                            OLS   Adj. R-squared:                  0.536
Method:                 Least Squares   F-statistic:                     50.58
Date:                Mon, 03 Jun 2024   Prob (F-statistic):           6.99e-27
Time:                        18:00:03   Log-Likelihood:                 692.27
No. Observations:                 331   AIC:                            -1377.
Df Residuals:                     327   BIC:                            -1361.
Df Model:                           3
Covariance Type:                  HAC
==========================================================================================
                             coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------------------
const                      0.0008      0.007      0.113      0.910      -0.013       0.015
Z22_Lag1                  -0.0016      0.008     -0.186      0.852      -0.018       0.015
Log_Return_Lag1            0.7743      0.114      6.778      0.000       0.550       0.998
Log_Return_Lag1_Zero22    -0.1197      0.224     -0.534      0.593      -0.559       0.320
==============================================================================
Omnibus:                       40.456   Durbin-Watson:                   2.166
Prob(Omnibus):                  0.000   Jarque-Bera (JB):              256.931
Skew:                           0.147   Prob(JB):                     1.61e-56
Kurtosis:                       7.306   Cond. No.                         136.
==============================================================================

Notes:
[1] Standard Errors are heteroscedasticity and autocorrelation robust (HAC) using 2 lags and without small sample correction


Parameters for Model: AR(2) with Weekly Zero Interaction:
                            OLS Regression Results
==============================================================================
Dep. Variable:             Log_Return   R-squared:                       0.569
Model:                            OLS   Adj. R-squared:                  0.562
Method:                 Least Squares   F-statistic:                     35.24
Date:                Mon, 03 Jun 2024   Prob (F-statistic):           9.34e-29
Time:                        18:00:03   Log-Likelihood:                 700.29
No. Observations:                 330   AIC:                            -1389.
Df Residuals:                     324   BIC:                            -1366.
Df Model:                           5
Covariance Type:                  HAC
=========================================================================================
                            coef    std err          z      P>|z|      [0.025      0.975]
-----------------------------------------------------------------------------------------
const                     0.0013      0.006      0.232      0.816      -0.010       0.012
Z5_Lag1                  -0.0020      0.006     -0.313      0.755      -0.015       0.011
Log_Return_Lag1           0.9350      0.153      6.116      0.000       0.635       1.235
Log_Return_Lag2          -0.2224      0.184     -1.211      0.226      -0.582       0.137
Log_Return_Lag1_Zero5    -0.6471      0.290     -2.228      0.026      -1.216      -0.078
Log_Return_Lag2_Zero5     0.6378      0.228      2.791      0.005       0.190       1.086
==============================================================================
Omnibus:                       41.071   Durbin-Watson:                   2.020
Prob(Omnibus):                  0.000   Jarque-Bera (JB):              259.103
Skew:                           0.178   Prob(JB):                     5.45e-57
Kurtosis:                       7.326   Cond. No.                         202.
==============================================================================

Notes:
[1] Standard Errors are heteroscedasticity and autocorrelation robust (HAC) using 2 lags and without small sample correction


Parameters for Model: AR(2) with Monthly Zero Interaction:
                            OLS Regression Results
==============================================================================
Dep. Variable:             Log_Return   R-squared:                       0.557
Model:                            OLS   Adj. R-squared:                  0.550
Method:                 Least Squares   F-statistic:                     35.01
Date:                Mon, 03 Jun 2024   Prob (F-statistic):           1.35e-28
Time:                        18:00:03   Log-Likelihood:                 695.81
No. Observations:                 330   AIC:                            -1380.
Df Residuals:                     324   BIC:                            -1357.
Df Model:                           5
Covariance Type:                  HAC
==========================================================================================
                             coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------------------
const                      0.0008      0.007      0.118      0.906      -0.013       0.015
Z22_Lag1                  -0.0011      0.009     -0.124      0.901      -0.018       0.016
Log_Return_Lag1            0.8544      0.225      3.800      0.000       0.414       1.295
Log_Return_Lag2           -0.1476      0.205     -0.721      0.471      -0.549       0.253
Log_Return_Lag1_Zero22    -0.4278      0.341     -1.254      0.210      -1.096       0.241
Log_Return_Lag2_Zero22     0.6352      0.332      1.911      0.056      -0.016       1.287
==============================================================================
Omnibus:                       42.840   Durbin-Watson:                   2.046
Prob(Omnibus):                  0.000   Jarque-Bera (JB):              275.611
Skew:                           0.215   Prob(JB):                     1.42e-60
Kurtosis:                       7.456   Cond. No.                         255.
==============================================================================

Notes:
[1] Standard Errors are heteroscedasticity and autocorrelation robust (HAC) using 2 lags and without small sample correction


Parameters for Model: AR(2) with Separate Weekly and Monthly Zero Interactions:
                            OLS Regression Results
==============================================================================
Dep. Variable:             Log_Return   R-squared:                       0.559
Model:                            OLS   Adj. R-squared:                  0.551
Method:                 Least Squares   F-statistic:                     29.20
Date:                Mon, 03 Jun 2024   Prob (F-statistic):           6.84e-28
Time:                        18:00:03   Log-Likelihood:                 696.42
No. Observations:                 330   AIC:                            -1379.
Df Residuals:                     323   BIC:                            -1352.
Df Model:                           6
Covariance Type:                  HAC
==========================================================================================
                             coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------------------
const                      0.0006      0.007      0.084      0.933      -0.013       0.015
Z5_Lag1                   -0.0018      0.009     -0.203      0.839      -0.019       0.015
Z22_Lag1                   0.0009      0.011      0.083      0.934      -0.021       0.023
Log_Return_Lag1            0.7685      0.135      5.701      0.000       0.504       1.033
Log_Return_Lag2           -0.0692      0.153     -0.453      0.650      -0.369       0.230
Log_Return_Lag1_Zero5     -0.4289      0.318     -1.348      0.178      -1.052       0.194
Log_Return_Lag2_Zero22     0.5299      0.309      1.717      0.086      -0.075       1.135
==============================================================================
Omnibus:                       45.677   Durbin-Watson:                   2.033
Prob(Omnibus):                  0.000   Jarque-Bera (JB):              299.370
Skew:                           0.274   Prob(JB):                     9.83e-66
Kurtosis:                       7.634   Cond. No.                         222.
==============================================================================

Notes:
[1] Standard Errors are heteroscedasticity and autocorrelation robust (HAC) using 2 lags and without small sample correction
```

In [ ]:

```
# Wednesdays
wednesdays_models = models_comparison(Wednesdays, lags=2, df_name='Wednesdays')
```

| Model Name | Adj R^2 | AIC | BIC | Rho_1 | Rho_2 | Ljung-Box Test Statistic | Ljung-Box P-value |
|---|---|---|---|---|---|---|---|
| Model with Only Constant | 3.3306690738754696e-16 | -1119.750613118897 | -1115.9454781499805 | 0.6901856806523485 | 0.5499285557504089 | 322.735302435179 | 1.1929501201673566e-69 |
| Model with Weekly Zero | 0.0004368016252594975 | -1114.5225316713363 | -1106.9182949205822 | 0.6866726736508274 | 0.5443736775616459 | 316.5934385593608 | 2.5477960643187162e-68 |
| Model with Monthly Zero | 0.0038472895343881763 | -1115.6538275555095 | -1108.0495908047553 | 0.6880407000350561 | 0.5475280649274151 | 319.63337470919527 | 5.599033872173402e-69 |
| AR(1) | 0.4748224489524412 | -1327.54816838933 | -1319.9439316385758 | -0.09646255565409773 | 0.07713618543293972 | 5.192412746328644 | 0.1582378980095613 |
| AR(2) | 0.4835561612904895 | -1327.0716217640604 | -1315.6743438006788 | -0.0010928707204485918 | -0.010150063874339348 | 0.14188400659053077 | 0.9863758870967192 |
| AR(1) with Weekly Zero Interaction | 0.4752940524181547 | -1325.8638350962042 | -1310.655361594696 | -0.10196049924721672 | 0.06908142068319112 | 5.132818557273332 | 0.16232589315664808 |
| AR(1) with Monthly Zero Interaction | 0.47373272031542635 | -1324.880363219294 | -1309.6718897177857 | -0.10622987687680226 | 0.07417725211380555 | 5.692722017173288 | 0.12755548532932812 |
| AR(2) with Weekly Zero Interaction | 0.48444840983854753 | -1324.6837446898273 | -1301.889188763064 | -0.012274501768083616 | -0.0012947042746150601 | 0.22864941205736708 | 0.9728369303156228 |
| AR(2) with Monthly Zero Interaction | 0.4834123944965004 | -1324.0212658708258 | -1301.2267099440626 | -0.010727682128786092 | -0.027283820424752706 | 0.47393767643844487 | 0.9245787273403366 |
| AR(2) with Separate Weekly and Monthly Zero Interactions | 0.48211485340423454 | -1322.213519760148 | -1295.6198711789243 | -0.011206934511727055 | -0.024723289355335673 | 0.4017086456565259 | 0.9398893028756881 |

```
Parameters for Model: Model with Only Constant:
                            OLS Regression Results
==============================================================================
Dep. Variable:             Log_Return   R-squared:                       0.000
Model:                            OLS   Adj. R-squared:                  0.000
Method:                 Least Squares   F-statistic:                       nan
Date:                Mon, 03 Jun 2024   Prob (F-statistic):                nan
Time:                        18:01:09   Log-Likelihood:                 560.88
No. Observations:                 332   AIC:                            -1120.
Df Residuals:                     331   BIC:                            -1116.
Df Model:                           0
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
0             -0.0002      0.002     -0.063      0.950      -0.005       0.005
==============================================================================
Omnibus:                       58.640   Durbin-Watson:                   0.620
Prob(Omnibus):                  0.000   Jarque-Bera (JB):              427.180
Skew:                           0.460   Prob(JB):                     1.73e-93
Kurtosis:                       8.480   Cond. No.                         1.00
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
```

```
Parameters for Model: Model with Weekly Zero:
                            OLS Regression Results
==============================================================================
Dep. Variable:             Log_Return   R-squared:                       0.003
Model:                            OLS   Adj. R-squared:                  0.000
Method:                 Least Squares   F-statistic:                     1.144
Date:                Mon, 03 Jun 2024   Prob (F-statistic):              0.286
Time:                        18:01:10   Log-Likelihood:                 559.26
No. Observations:                 331   AIC:                            -1115.
Df Residuals:                     329   BIC:                            -1107.
Df Model:                           1
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const          0.0057      0.006      0.950      0.343      -0.006       0.018
Z5_Lag1       -0.0086      0.008     -1.070      0.286      -0.024       0.007
==============================================================================
Omnibus:                       50.746   Durbin-Watson:                   0.626
Prob(Omnibus):                  0.000   Jarque-Bera (JB):              408.354
Skew:                           0.261   Prob(JB):                     2.12e-89
Kurtosis:                       8.416   Cond. No.                         4.88
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
```

```
Parameters for Model: Model with Monthly Zero:
                            OLS Regression Results
==============================================================================
Dep. Variable:             Log_Return   R-squared:                       0.007
Model:                            OLS   Adj. R-squared:                  0.004
Method:                 Least Squares   F-statistic:                     2.275
Date:                Mon, 03 Jun 2024   Prob (F-statistic):              0.132
Time:                        18:01:10   Log-Likelihood:                 559.83
No. Observations:                 331   AIC:                            -1116.
Df Residuals:                     329   BIC:                            -1108.
Df Model:                           1
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const          0.0099      0.007      1.394      0.164      -0.004       0.024
Z22_Lag1      -0.0148      0.010     -1.508      0.132      -0.034       0.004
==============================================================================
Omnibus:                       48.086   Durbin-Watson:                   0.624
Prob(Omnibus):                  0.000   Jarque-Bera (JB):              395.855
Skew:                           0.167   Prob(JB):                     1.10e-86
Kurtosis:                       8.347   Cond. No.                         5.93
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
```

```
Parameters for Model: AR(1):
                            OLS Regression Results
==============================================================================
Dep. Variable:             Log_Return   R-squared:                       0.476
Model:                            OLS   Adj. R-squared:                  0.475
Method:                 Least Squares   F-statistic:                     90.18
Date:                Mon, 03 Jun 2024   Prob (F-statistic):           4.63e-19
Time:                        18:01:10   Log-Likelihood:                 665.77
No. Observations:                 331   AIC:                            -1328.
Df Residuals:                     329   BIC:                            -1320.
Df Model:                           1
Covariance Type:                  HAC
==================================================================================
                     coef    std err          z      P>|z|      [0.025      0.975]
----------------------------------------------------------------------------------
const          -6.678e-05      0.002     -0.039      0.969      -0.003       0.003
Log_Return_Lag1    0.6903      0.073      9.496      0.000       0.548       0.833
==============================================================================
Omnibus:                       50.930   Durbin-Watson:                   2.189
Prob(Omnibus):                  0.000   Jarque-Bera (JB):              497.491
Skew:                           0.072   Prob(JB):                    9.36e-109
Kurtosis:                       9.004   Cond. No.                         22.4
==============================================================================

Notes:
[1] Standard Errors are heteroscedasticity and autocorrelation robust (HAC) using 2 lags and without small sample correction
```

```
Parameters for Model: AR(2):
                            OLS Regression Results
==============================================================================
Dep. Variable:             Log_Return   R-squared:                       0.487
Model:                            OLS   Adj. R-squared:                  0.484
Method:                 Least Squares   F-statistic:                     49.99
Date:                Mon, 03 Jun 2024   Prob (F-statistic):           1.14e-19
Time:                        18:01:10   Log-Likelihood:                 666.54
No. Observations:                 330   AIC:                            -1327.
Df Residuals:                     327   BIC:                            -1316.
Df Model:                           2
Covariance Type:                  HAC
==================================================================================
                     coef    std err          z      P>|z|      [0.025      0.975]
----------------------------------------------------------------------------------
const          -4.575e-05      0.002     -0.026      0.979      -0.003       0.003
Log_Return_Lag1    0.5932      0.103      5.740      0.000       0.391       0.796
Log_Return_Lag2    0.1405      0.094      1.503      0.133      -0.043       0.324
==============================================================================
Omnibus:                       58.291   Durbin-Watson:                   1.999
Prob(Omnibus):                  0.000   Jarque-Bera (JB):              635.630
Skew:                           0.244   Prob(JB):                    9.43e-139
Kurtosis:                       9.782   Cond. No.                         40.2
==============================================================================

Notes:
[1] Standard Errors are heteroscedasticity and autocorrelation robust (HAC) using 2 lags and without small sample correction
```

```
Parameters for Model: AR(1) with Weekly Zero Interaction:
                            OLS Regression Results
==============================================================================
Dep. Variable:             Log_Return   R-squared:                       0.480
Model:                            OLS   Adj. R-squared:                  0.475
Method:                 Least Squares   F-statistic:                     32.58
Date:                Mon, 03 Jun 2024   Prob (F-statistic):           1.89e-18
Time:                        18:01:10   Log-Likelihood:                 666.93
No. Observations:                 331   AIC:                            -1326.
Df Residuals:                     327   BIC:                            -1311.
Df Model:                           3
Covariance Type:                  HAC
=========================================================================================
                            coef    std err          z      P>|z|      [0.025      0.975]
-----------------------------------------------------------------------------------------
const                     0.0024      0.007      0.358      0.720      -0.011       0.016
Z5_Lag1                  -0.0033      0.008     -0.420      0.674      -0.019       0.012
Log_Return_Lag1           0.6328      0.097      6.535      0.000       0.443       0.823
Log_Return_Lag1_Zero5     0.2675      0.235      1.138      0.255      -0.193       0.728
==============================================================================
Omnibus:                       51.218   Durbin-Watson:                   2.200
Prob(Omnibus):                  0.000   Jarque-Bera (JB):              510.488
Skew:                          -0.044   Prob(JB):                    1.41e-111
Kurtosis:                       9.083   Cond. No.                         135.
==============================================================================

Notes:
[1] Standard Errors are heteroscedasticity and autocorrelation robust (HAC) using 2 lags and without small sample correction
```

Parameters for Model: AR(1) with Monthly Zero Interaction:

```
                            OLS Regression Results
==============================================================================
Dep. Variable:             Log_Return   R-squared:                       0.479
Model:                            OLS   Adj. R-squared:                  0.474
Method:                 Least Squares   F-statistic:                     36.95
Date:                Mon, 03 Jun 2024   Prob (F-statistic):           1.38e-20
Time:                        18:01:10   Log-Likelihood:                 666.44
No. Observations:                 331   AIC:                            -1325.
Df Residuals:                     327   BIC:                            -1310.
Df Model:                           3
Covariance Type:                  HAC
==========================================================================================
                             coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------------------
const                      0.0032      0.008      0.390      0.696      -0.013       0.019
Z22_Lag1                  -0.0045      0.010     -0.453      0.651      -0.024       0.015
Log_Return_Lag1            0.6262      0.129      4.845      0.000       0.373       0.880
Log_Return_Lag1_Zero22     0.1863      0.254      0.733      0.464      -0.312       0.685
==============================================================================
Omnibus:                       49.419   Durbin-Watson:                   2.209
Prob(Omnibus):                  0.000   Jarque-Bera (JB):              464.173
Skew:                          -0.021   Prob(JB):                    1.61e-101
Kurtosis:                       8.801   Cond. No.                         135.
==============================================================================

Notes:
[1] Standard Errors are heteroscedasticity and autocorrelation robust (HAC) using 2 lags and without small sample correction



Parameters for Model: AR(2) with Weekly Zero Interaction:
                            OLS Regression Results
==============================================================================
Dep. Variable:             Log_Return   R-squared:                       0.492
Model:                            OLS   Adj. R-squared:                  0.484
Method:                 Least Squares   F-statistic:                     23.15
Date:                Mon, 03 Jun 2024   Prob (F-statistic):           6.97e-20
Time:                        18:01:10   Log-Likelihood:                 668.34
No. Observations:                 330   AIC:                            -1325.
Df Residuals:                     324   BIC:                            -1302.
Df Model:                           5
Covariance Type:                  HAC
=========================================================================================
                            coef    std err          z      P>|z|      [0.025      0.975]
-----------------------------------------------------------------------------------------
const                     0.0022      0.007      0.306      0.760      -0.012       0.016
Z5_Lag1                  -0.0028      0.008     -0.330      0.741      -0.019       0.014
Log_Return_Lag1           0.6362      0.206      3.087      0.002       0.232       1.040
Log_Return_Lag2          -0.0038      0.265     -0.014      0.989      -0.524       0.517
Log_Return_Lag1_Zero5     0.0890      0.313      0.284      0.776      -0.524       0.702
Log_Return_Lag2_Zero5     0.2501      0.330      0.757      0.449      -0.397       0.897
==============================================================================
Omnibus:                       54.835   Durbin-Watson:                   2.021
Prob(Omnibus):                  0.000   Jarque-Bera (JB):              610.773
Skew:                           0.096   Prob(JB):                    2.36e-133
Kurtosis:                       9.662   Cond. No.                         215.
==============================================================================

Notes:
[1] Standard Errors are heteroscedasticity and autocorrelation robust (HAC) using 2 lags and without small sample correction



Parameters for Model: AR(2) with Monthly Zero Interaction:
                            OLS Regression Results
==============================================================================
Dep. Variable:             Log_Return   R-squared:                       0.491
Model:                            OLS   Adj. R-squared:                  0.483
Method:                 Least Squares   F-statistic:                     26.89
Date:                Mon, 03 Jun 2024   Prob (F-statistic):           9.72e-23
Time:                        18:01:10   Log-Likelihood:                 668.01
No. Observations:                 330   AIC:                            -1324.
Df Residuals:                     324   BIC:                            -1301.
Df Model:                           5
Covariance Type:                  HAC
==========================================================================================
                             coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------------------
const                      0.0030      0.009      0.351      0.725      -0.014       0.020
Z22_Lag1                  -0.0039      0.010     -0.378      0.706      -0.024       0.016
Log_Return_Lag1            0.5414      0.218      2.484      0.013       0.114       0.969
Log_Return_Lag2            0.0784      0.217      0.361      0.718      -0.347       0.504
Log_Return_Lag1_Zero22     0.1605      0.358      0.449      0.654      -0.540       0.861
Log_Return_Lag2_Zero22     0.1943      0.360      0.540      0.589      -0.511       0.900
==============================================================================
Omnibus:                       54.167   Durbin-Watson:                   2.018
Prob(Omnibus):                  0.000   Jarque-Bera (JB):              565.428
Skew:                           0.152   Prob(JB):                    1.66e-123
Kurtosis:                       9.405   Cond. No.                         233.
==============================================================================

Notes:
[1] Standard Errors are heteroscedasticity and autocorrelation robust (HAC) using 2 lags and without small sample correction



Parameters for Model: AR(2) with Separate Weekly and Monthly Zero Interactions:
                            OLS Regression Results
==============================================================================
Dep. Variable:             Log_Return   R-squared:                       0.492
Model:                            OLS   Adj. R-squared:                  0.482
Method:                 Least Squares   F-statistic:                     22.47
Date:                Mon, 03 Jun 2024   Prob (F-statistic):           4.03e-22
Time:                        18:01:10   Log-Likelihood:                 668.11
No. Observations:                 330   AIC:                            -1322.
Df Residuals:                     323   BIC:                            -1296.
Df Model:                           6
Covariance Type:                  HAC
==========================================================================================
                             coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------------------
const                      0.0029      0.009      0.329      0.742      -0.014       0.020
Z5_Lag1                   -0.0007      0.009     -0.077      0.939      -0.019       0.017
Z22_Lag1                  -0.0030      0.012     -0.245      0.807      -0.027       0.021
Log_Return_Lag1            0.5666      0.136      4.181      0.000       0.301       0.832
Log_Return_Lag2            0.0601      0.174      0.346      0.729      -0.280       0.401
Log_Return_Lag1_Zero5      0.1688      0.263      0.642      0.521      -0.347       0.684
Log_Return_Lag2_Zero22     0.2207      0.292      0.755      0.450      -0.352       0.793
==============================================================================
Omnibus:                       54.368   Durbin-Watson:                   2.019
Prob(Omnibus):                  0.000   Jarque-Bera (JB):              593.480
Skew:                           0.102   Prob(JB):                    1.34e-129
Kurtosis:                       9.567   Cond. No.                         220.
==============================================================================

Notes:
[1] Standard Errors are heteroscedasticity and autocorrelation robust (HAC) using 2 lags and without small sample correction
```

In [ ]:

```python
# Tuesdays
tuesdays_models = models_comparison(Tuesdays, lags=2, df_name='Tuesdays')
```

| Model Name | Adj R^2 | AIC | BIC | Rho_1 | Rho_2 | Ljung-Box Test Statistic | Ljung-Box P-value |
|---|---|---|---|---|---|---|---|
| Model with Only Constant | 4.440892098500626e-16 | -1120.0482104914004 | -1116.2370694984236 | 0.6662548458224645 | 0.5498091033843614 | 314.5186833119641 | 7.165941752616387e-68 |
| Model with Weekly Zero | -0.0023361244588984142 | -1113.9181033525988 | -1106.3018183726379 | 0.6638863509966867 | 0.5484560552409659 | 311.52807915742983 | 3.1813629512928694e-67 |
| Model with Monthly Zero | -0.0013050677167152358 | -1114.2608213270487 | -1106.6445363470878 | 0.6649057542551791 | 0.5489401719970008 | 312.5577682727266 | 1.9042811551799388e-67 |
| AR(1) | 0.44221628864425344 | -1309.095199471638 | -1301.478914491677 | -0.12692869813086013 | 0.1139968538880168 | 10.397345606056946 | 0.015473677649263303 |
| AR(2) | 0.4608973245266801 | -1314.4740652539508 | -1303.0586603472013 | -0.002327121756879495 | -0.010146721029427166 | 0.04215333467183039 | 0.9977270952194889 |
| AR(1) with Weekly Zero Interaction | 0.45924533809348944 | -1317.438248004738 | -1302.205678044816 | -0.108377596894946 | 0.11199084546835066 | 8.534682158407868 | 0.03616208959814599 |
| AR(1) with Monthly Zero Interaction | 0.4430502109444814 | -1307.6116162864646 | -1292.3790463265427 | -0.13970115048928833 | 0.10707789666812177 | 10.992685410854715 | 0.011765494153635555 |
| AR(2) with Weekly Zero Interaction | 0.474663823179042 | -1320.1033679704096 | -1297.2725581569107 | 0.013898330908768757 | -0.002095090567061654 | 0.1023064175528134 | 0.9915592155562123 |
| AR(2) with Monthly Zero Interaction | 0.46367873783225144 | -1313.2326423223885 | -1290.4018325088896 | -0.009458205968811311 | -0.031045780700168047 | 0.39459295827242463 | 0.9413571844141486 |
| AR(2) with Separate Weekly and Monthly Zero Interactions | 0.4720451434617029 | -1317.4725056977572 | -1290.8365609153418 | 0.008171376609158726 | 0.0021736982956762512 | 0.0645649730698349 | 0.9957202565371582 |

```
Parameters for Model: Model with Only Constant:
                            OLS Regression Results
==============================================================================
Dep. Variable:             Log_Return   R-squared:                       0.000
Model:                            OLS   Adj. R-squared:                  0.000
Method:                 Least Squares   F-statistic:                       nan
Date:                Mon, 03 Jun 2024   Prob (F-statistic):                nan
Time:                        18:01:27   Log-Likelihood:                 561.02
No. Observations:                 334   AIC:                            -1120.
Df Residuals:                     333   BIC:                            -1116.
Df Model:                           0
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
0             -0.0008      0.002     -0.325      0.746      -0.006       0.004
==============================================================================
Omnibus:                       85.736   Durbin-Watson:                   0.667
Prob(Omnibus):                  0.000   Jarque-Bera (JB):              812.466
Skew:                           0.749   Prob(JB):                    3.76e-177
Kurtosis:                      10.493   Cond. No.                         1.00
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.


Parameters for Model: Model with Weekly Zero:
                            OLS Regression Results
==============================================================================
Dep. Variable:             Log_Return   R-squared:                       0.001
Model:                            OLS   Adj. R-squared:                 -0.002
Method:                 Least Squares   F-statistic:                    0.2262
Date:                Mon, 03 Jun 2024   Prob (F-statistic):              0.635
Time:                        18:01:27   Log-Likelihood:                 558.96
No. Observations:                 333   AIC:                            -1114.
Df Residuals:                     331   BIC:                            -1106.
Df Model:                           1
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const          0.0018      0.006      0.299      0.765      -0.010       0.014
Z5_Lag1       -0.0038      0.008     -0.476      0.635      -0.020       0.012
==============================================================================
Omnibus:                       79.774   Durbin-Watson:                   0.672
Prob(Omnibus):                  0.000   Jarque-Bera (JB):              779.334
Skew:                           0.658   Prob(JB):                    5.88e-170
Kurtosis:                      10.378   Cond. No.                         4.82
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.


Parameters for Model: Model with Monthly Zero:
                            OLS Regression Results
==============================================================================
Dep. Variable:             Log_Return   R-squared:                       0.002
Model:                            OLS   Adj. R-squared:                 -0.001
Method:                 Least Squares   F-statistic:                    0.5673
Date:                Mon, 03 Jun 2024   Prob (F-statistic):              0.452
Time:                        18:01:27   Log-Likelihood:                 559.13
No. Observations:                 333   AIC:                            -1114.
Df Residuals:                     331   BIC:                            -1107.
Df Model:                           1
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const          0.0043      0.007      0.595      0.552      -0.010       0.018
Z22_Lag1      -0.0074      0.010     -0.753      0.452      -0.027       0.012
==============================================================================
Omnibus:                       76.288   Durbin-Watson:                   0.670
Prob(Omnibus):                  0.000   Jarque-Bera (JB):              759.179
Skew:                           0.601   Prob(JB):                    1.40e-165
Kurtosis:                      10.299   Cond. No.                         5.93
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.


Parameters for Model: AR(1):
                            OLS Regression Results
==============================================================================
Dep. Variable:             Log_Return   R-squared:                       0.444
Model:                            OLS   Adj. R-squared:                  0.442
Method:                 Least Squares   F-statistic:                     66.37
Date:                Mon, 03 Jun 2024   Prob (F-statistic):           7.73e-15
Time:                        18:01:27   Log-Likelihood:                 656.55
No. Observations:                 333   AIC:                            -1309.
Df Residuals:                     331   BIC:                            -1301.
Df Model:                           1
Covariance Type:                  HAC
==================================================================================
                     coef    std err          z      P>|z|      [0.025      0.975]
----------------------------------------------------------------------------------
const             -0.0003      0.002     -0.152      0.879      -0.004       0.003
Log_Return_Lag1    0.6663      0.082      8.147      0.000       0.506       0.827
==============================================================================
Omnibus:                       68.206   Durbin-Watson:                   2.252
Prob(Omnibus):                  0.000   Jarque-Bera (JB):             1086.804
Skew:                           0.214   Prob(JB):                    1.01e-236
Kurtosis:                      11.840   Cond. No.                         22.1
==============================================================================

Notes:
[1] Standard Errors are heteroscedasticity and autocorrelation robust (HAC) using 2 lags and without small sample correction


Parameters for Model: AR(2):
                            OLS Regression Results
==============================================================================
Dep. Variable:             Log_Return   R-squared:                       0.464
Model:                            OLS   Adj. R-squared:                  0.461
Method:                 Least Squares   F-statistic:                     42.85
Date:                Mon, 03 Jun 2024   Prob (F-statistic):           2.89e-17
Time:                        18:01:27   Log-Likelihood:                 660.24
No. Observations:                 332   AIC:                            -1314.
Df Residuals:                     329   BIC:                            -1303.
Df Model:                           2
Covariance Type:                  HAC
==================================================================================
                     coef    std err          z      P>|z|      [0.025      0.975]
----------------------------------------------------------------------------------
const             -0.0002      0.002     -0.109      0.913      -0.004       0.003
Log_Return_Lag1    0.5389      0.098      5.493      0.000       0.347       0.731
Log_Return_Lag2    0.1912      0.076      2.518      0.012       0.042       0.340
==============================================================================
Omnibus:                       79.052   Durbin-Watson:                   2.004
Prob(Omnibus):                  0.000   Jarque-Bera (JB):             1219.009
Skew:                           0.465   Prob(JB):                    1.97e-265
Kurtosis:                      12.341   Cond. No.                         38.3
==============================================================================

Notes:
[1] Standard Errors are heteroscedasticity and autocorrelation robust (HAC) using 2 lags and without small sample correction


Parameters for Model: AR(1) with Weekly Zero Interaction:
                            OLS Regression Results
==============================================================================
Dep. Variable:             Log_Return   R-squared:                       0.464
Model:                            OLS   Adj. R-squared:                  0.459
Method:                 Least Squares   F-statistic:                     46.91
Date:                Mon, 03 Jun 2024   Prob (F-statistic):           2.90e-25
Time:                        18:01:27   Log-Likelihood:                 662.72
No. Observations:                 333   AIC:                            -1317.
Df Residuals:                     329   BIC:                            -1302.
Df Model:                           3
Covariance Type:                  HAC
==========================================================================================
                             coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------------------
const                     -0.0002      0.007     -0.024      0.981      -0.014       0.014
Z5_Lag1                    0.0010      0.008      0.120      0.905      -0.015       0.017
Log_Return_Lag1            0.5509      0.124      4.435      0.000       0.307       0.794
Log_Return_Lag1_Zero5      0.6625      0.370      1.791      0.073      -0.062       1.387
==============================================================================
Omnibus:                       75.294   Durbin-Watson:                   2.214
Prob(Omnibus):                  0.000   Jarque-Bera (JB):              791.103
Skew:                           0.565   Prob(JB):                    1.64e-172
Kurtosis:                      10.466   Cond. No.                         129.
==============================================================================

Notes:
[1] Standard Errors are heteroscedasticity and autocorrelation robust (HAC) using 2 lags and without small sample correction


Parameters for Model: AR(1) with Monthly Zero Interaction:
```

```
                            OLS Regression Results
==============================================================================
Dep. Variable:             Log_Return   R-squared:                       0.448
Model:                            OLS   Adj. R-squared:                  0.443
Method:                 Least Squares   F-statistic:                     43.65
Date:                Mon, 03 Jun 2024   Prob (F-statistic):           9.05e-24
Time:                        18:01:27   Log-Likelihood:                 657.81
No. Observations:                 333   AIC:                            -1308.
Df Residuals:                     329   BIC:                            -1292.
Df Model:                           3
Covariance Type:                  HAC
==========================================================================================
                             coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------------------
const                      0.0004      0.008      0.052      0.958      -0.016       0.017
Z22_Lag1                  -0.0006      0.010     -0.059      0.953      -0.020       0.019
Log_Return_Lag1            0.5669      0.153      3.715      0.000       0.268       0.866
Log_Return_Lag1_Zero22     0.3067      0.281      1.091      0.275      -0.244       0.857
==============================================================================
Omnibus:                       69.384   Durbin-Watson:                   2.278
Prob(Omnibus):                  0.000   Jarque-Bera (JB):              957.605
Skew:                           0.339   Prob(JB):                    1.14e-208
Kurtosis:                      11.280   Cond. No.                         135.
==============================================================================

Notes:
[1] Standard Errors are heteroscedasticity and autocorrelation robust (HAC) using 2 lags and without small sample correction


Parameters for Model: AR(2) with Weekly Zero Interaction:
                            OLS Regression Results
==============================================================================
Dep. Variable:             Log_Return   R-squared:                       0.483
Model:                            OLS   Adj. R-squared:                  0.475
Method:                 Least Squares   F-statistic:                     29.69
Date:                Mon, 03 Jun 2024   Prob (F-statistic):           7.72e-25
Time:                        18:01:27   Log-Likelihood:                 666.05
No. Observations:                 332   AIC:                            -1320.
Df Residuals:                     326   BIC:                            -1297.
Df Model:                           5
Covariance Type:                  HAC
=========================================================================================
                            coef    std err          z      P>|z|      [0.025      0.975]
-----------------------------------------------------------------------------------------
const                 -7.368e-05      0.007     -0.011      0.991      -0.013       0.013
Z5_Lag1                   0.0009      0.008      0.109      0.913      -0.015       0.016
Log_Return_Lag1           0.3833      0.197      1.950      0.051      -0.002       0.769
Log_Return_Lag2           0.2670      0.223      1.198      0.231      -0.170       0.704
Log_Return_Lag1_Zero5     0.7203      0.415      1.736      0.083      -0.093       1.534
Log_Return_Lag2_Zero5    -0.1585      0.277     -0.573      0.567      -0.701       0.384
==============================================================================
Omnibus:                       94.572   Durbin-Watson:                   1.970
Prob(Omnibus):                  0.000   Jarque-Bera (JB):              965.767
Skew:                           0.847   Prob(JB):                    1.93e-210
Kurtosis:                      11.182   Cond. No.                         192.
==============================================================================

Notes:
[1] Standard Errors are heteroscedasticity and autocorrelation robust (HAC) using 2 lags and without small sample correction


Parameters for Model: AR(2) with Monthly Zero Interaction:
                            OLS Regression Results
==============================================================================
Dep. Variable:             Log_Return   R-squared:                       0.472
Model:                            OLS   Adj. R-squared:                  0.464
Method:                 Least Squares   F-statistic:                     31.91
Date:                Mon, 03 Jun 2024   Prob (F-statistic):           1.91e-26
Time:                        18:01:27   Log-Likelihood:                 662.62
No. Observations:                 332   AIC:                            -1313.
Df Residuals:                     326   BIC:                            -1290.
Df Model:                           5
Covariance Type:                  HAC
==========================================================================================
                             coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------------------
const                      0.0003      0.009      0.030      0.976      -0.017       0.017
Z22_Lag1               -2.857e-05      0.011     -0.003      0.998      -0.021       0.021
Log_Return_Lag1            0.4130      0.202      2.046      0.041       0.017       0.809
Log_Return_Lag2            0.1794      0.169      1.064      0.287      -0.151       0.510
Log_Return_Lag1_Zero22     0.3672      0.339      1.084      0.278      -0.297       1.031
Log_Return_Lag2_Zero22     0.0781      0.296      0.264      0.792      -0.502       0.659
==============================================================================
Omnibus:                       86.370   Durbin-Watson:                   2.018
Prob(Omnibus):                  0.000   Jarque-Bera (JB):             1083.975
Skew:                           0.661   Prob(JB):                    4.15e-236
Kurtosis:                      11.753   Cond. No.                         233.
==============================================================================

Notes:
[1] Standard Errors are heteroscedasticity and autocorrelation robust (HAC) using 2 lags and without small sample correction


Parameters for Model: AR(2) with Separate Weekly and Monthly Zero Interactions:
                            OLS Regression Results
==============================================================================
Dep. Variable:             Log_Return   R-squared:                       0.482
Model:                            OLS   Adj. R-squared:                  0.472
Method:                 Least Squares   F-statistic:                     26.88
Date:                Mon, 03 Jun 2024   Prob (F-statistic):           5.56e-26
Time:                        18:01:27   Log-Likelihood:                 665.74
No. Observations:                 332   AIC:                            -1317.
Df Residuals:                     325   BIC:                            -1291.
Df Model:                           6
Covariance Type:                  HAC
==========================================================================================
                             coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------------------
const                      0.0006      0.009      0.068      0.946      -0.017       0.018
Z5_Lag1                    0.0032      0.009      0.362      0.718      -0.014       0.021
Z22_Lag1                  -0.0033      0.013     -0.259      0.795      -0.028       0.022
Log_Return_Lag1            0.4456      0.144      3.093      0.002       0.163       0.728
Log_Return_Lag2            0.1631      0.147      1.113      0.266      -0.124       0.450
Log_Return_Lag1_Zero5      0.5968      0.416      1.435      0.151      -0.218       1.412
Log_Return_Lag2_Zero22     0.0405      0.282      0.144      0.886      -0.512       0.593
==============================================================================
Omnibus:                       89.247   Durbin-Watson:                   1.982
Prob(Omnibus):                  0.000   Jarque-Bera (JB):              946.813
Skew:                           0.762   Prob(JB):                    2.52e-206
Kurtosis:                      11.131   Cond. No.                         211.
==============================================================================

Notes:
[1] Standard Errors are heteroscedasticity and autocorrelation robust (HAC) using 2 lags and without small sample correction
```

In [ ]:

```
# Mondays
mondays_models = models_comparison(Mondays, lags=2, df_name='Mondays')
```

| Model Name | Adj R^2 | AIC | BIC | Rho_1 | Rho_2 | Ljung-Box Test Statistic | Ljung-Box P-value |
|---|---|---|---|---|---|---|---|
| Model with Only Constant | 1.1102230246251565e-16 | -1072.9949067216508 | -1069.2203551761063 | 0.6485785132468169 | 0.5326556440054085 | 295.15763921355324 | 1.1111279373154317e-63 |
| Model with Weekly Zero | -0.0024909154714884973 | -1066.8643895536563 | -1059.3215073073964 | 0.6465267666177132 | 0.5308985745446347 | 292.6986400399263 | 3.783768250056632e-63 |
| Model with Monthly Zero | -0.001776650965297888 | -1067.0931802800376 | -1059.5502980337776 | 0.6479011354989274 | 0.5323797096947639 | 293.89808110230047 | 2.081379342285421e-63 |
| AR(1) | 0.4188387831233582 | -1241.8783715514755 | -1234.3354893052156 | -0.125428185903783 | 0.07747961535798127 | 14.154821894854246 | 0.0027018115793091355 |
| AR(2) | 0.43892848337180257 | -1247.2675692410312 | -1235.96260625365 | -0.015422930053294053 | -0.034779787506850623 | 3.305239746236195 | 0.3469140351821625 |
| AR(1) with Weekly Zero Interaction | 0.432410091283264 | -1247.482165984658 | -1232.396401492138 | -0.11802157243230886 | 0.07510237002044547 | 11.99037976299498 | 0.0074161881314606945 |
| AR(1) with Monthly Zero Interaction | 0.41539348518678243 | -1237.9998804249653 | -1222.9141159324454 | -0.12051640776192568 | 0.07904285025745658 | 13.82845963979641 | 0.0031481955107964615 |
| AR(2) with Weekly Zero Interaction | 0.44619190652084795 | -1248.4800303332695 | -1225.870104358507 | -0.020981953443048807 | -0.022541411224164705 | 2.6736435186448326 | 0.4447248998754597 |
| AR(2) with Monthly Zero Interaction | 0.441039554381889 | -1245.5166746862674 | -1222.9067487115049 | -0.033731545683699175 | -0.05758742097668787 | 4.6694945741200895 | 0.19766092418126427 |
| AR(2) with Separate Weekly and Monthly Zero Interactions | 0.44682382954652655 | -1247.86610950278 | -1221.4878625322237 | -0.022023988031967683 | -0.03511626269765109 | 3.1317283942301386 | 0.37175778240574625 |

```
Parameters for Model: Model with Only Constant:
                            OLS Regression Results
==============================================================================
Dep. Variable:             Log_Return   R-squared:                       0.000
Model:                            OLS   Adj. R-squared:                  0.000
Method:                 Least Squares   F-statistic:                       nan
Date:                Mon, 03 Jun 2024   Prob (F-statistic):                nan
Time:                        18:01:46   Log-Likelihood:                 537.50
No. Observations:                 322   AIC:                            -1073.
Df Residuals:                     321   BIC:                            -1069.
Df Model:                           0
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
0             -0.0008      0.003     -0.318      0.750      -0.006       0.004
==============================================================================
Omnibus:                       72.298   Durbin-Watson:                   0.703
Prob(Omnibus):                  0.000   Jarque-Bera (JB):              687.606
Skew:                           0.589   Prob(JB):                    4.88e-150
Kurtosis:                      10.061   Cond. No.                         1.00
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.


Parameters for Model: Model with Weekly Zero:
                            OLS Regression Results
==============================================================================
Dep. Variable:             Log_Return   R-squared:                       0.001
Model:                            OLS   Adj. R-squared:                 -0.002
Method:                 Least Squares   F-statistic:                    0.2049
Date:                Mon, 03 Jun 2024   Prob (F-statistic):              0.651
Time:                        18:01:46   Log-Likelihood:                 535.43
No. Observations:                 321   AIC:                            -1067.
Df Residuals:                     319   BIC:                            -1059.
Df Model:                           1
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const          0.0017      0.006      0.277      0.782      -0.010       0.014
Z5_Lag1       -0.0037      0.008     -0.453      0.651      -0.020       0.012
==============================================================================
Omnibus:                       67.859   Durbin-Watson:                   0.707
Prob(Omnibus):                  0.000   Jarque-Bera (JB):              666.207
Skew:                           0.510   Prob(JB):                    2.16e-145
Kurtosis:                       9.984   Cond. No.                         4.75
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.


Parameters for Model: Model with Monthly Zero:
                            OLS Regression Results
==============================================================================
Dep. Variable:             Log_Return   R-squared:                       0.001
Model:                            OLS   Adj. R-squared:                 -0.002
Method:                 Least Squares   F-statistic:                    0.4325
Date:                Mon, 03 Jun 2024   Prob (F-statistic):              0.511
Time:                        18:01:46   Log-Likelihood:                 535.55
No. Observations:                 321   AIC:                            -1067.
Df Residuals:                     319   BIC:                            -1060.
Df Model:                           1
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const          0.0037      0.007      0.507      0.613      -0.011       0.018
Z22_Lag1      -0.0066      0.010     -0.658      0.511      -0.026       0.013
==============================================================================
Omnibus:                       65.549   Durbin-Watson:                   0.704
Prob(Omnibus):                  0.000   Jarque-Bera (JB):              652.238
Skew:                           0.466   Prob(JB):                    2.34e-142
Kurtosis:                       9.921   Cond. No.                         5.88
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.


Parameters for Model: AR(1):
                            OLS Regression Results
==============================================================================
Dep. Variable:             Log_Return   R-squared:                       0.421
Model:                            OLS   Adj. R-squared:                  0.419
Method:                 Least Squares   F-statistic:                     91.99
Date:                Mon, 03 Jun 2024   Prob (F-statistic):           2.62e-19
Time:                        18:01:46   Log-Likelihood:                 622.94
No. Observations:                 321   AIC:                            -1242.
Df Residuals:                     319   BIC:                            -1234.
Df Model:                           1
Covariance Type:                  HAC
===================================================================================
                      coef    std err          z      P>|z|      [0.025      0.975]
-----------------------------------------------------------------------------------
const              -0.0003      0.002     -0.156      0.876      -0.004       0.003
Log_Return_Lag1     0.6486      0.068      9.591      0.000       0.516       0.781
==============================================================================
Omnibus:                       81.877   Durbin-Watson:                   2.249
Prob(Omnibus):                  0.000   Jarque-Bera (JB):              641.468
Skew:                           0.800   Prob(JB):                    5.09e-140
Kurtosis:                       9.738   Cond. No.                         21.9
==============================================================================

Notes:
[1] Standard Errors are heteroscedasticity and autocorrelation robust (HAC) using 2 lags and without small sample correction


Parameters for Model: AR(2):
                            OLS Regression Results
==============================================================================
Dep. Variable:             Log_Return   R-squared:                       0.442
Model:                            OLS   Adj. R-squared:                  0.439
Method:                 Least Squares   F-statistic:                     53.35
Date:                Mon, 03 Jun 2024   Prob (F-statistic):           1.07e-20
Time:                        18:01:46   Log-Likelihood:                 626.63
No. Observations:                 320   AIC:                            -1247.
Df Residuals:                     317   BIC:                            -1236.
Df Model:                           2
Covariance Type:                  HAC
===================================================================================
                      coef    std err          z      P>|z|      [0.025      0.975]
-----------------------------------------------------------------------------------
const              -0.0002      0.002     -0.109      0.913      -0.004       0.003
Log_Return_Lag1     0.5224      0.077      6.792      0.000       0.372       0.673
Log_Return_Lag2     0.1945      0.070      2.791      0.005       0.058       0.331
==============================================================================
Omnibus:                       95.002   Durbin-Watson:                   2.030
Prob(Omnibus):                  0.000   Jarque-Bera (JB):              788.916
Skew:                           0.963   Prob(JB):                    4.89e-172
Kurtosis:                      10.447   Cond. No.                         37.0
==============================================================================

Notes:
[1] Standard Errors are heteroscedasticity and autocorrelation robust (HAC) using 2 lags and without small sample correction


Parameters for Model: AR(1) with Weekly Zero Interaction:
                            OLS Regression Results
==============================================================================
Dep. Variable:             Log_Return   R-squared:                       0.438
Model:                            OLS   Adj. R-squared:                  0.432
Method:                 Least Squares   F-statistic:                     37.17
Date:                Mon, 03 Jun 2024   Prob (F-statistic):           1.31e-20
Time:                        18:01:46   Log-Likelihood:                 627.74
No. Observations:                 321   AIC:                            -1247.
Df Residuals:                     317   BIC:                            -1232.
Df Model:                           3
Covariance Type:                  HAC
=========================================================================================
                            coef    std err          z      P>|z|      [0.025      0.975]
-----------------------------------------------------------------------------------------
const                  4.618e-05      0.007      0.006      0.995      -0.014       0.014
Z5_Lag1                   0.0004      0.009      0.047      0.962      -0.017       0.017
Log_Return_Lag1           0.5515      0.098      5.619      0.000       0.359       0.744
Log_Return_Lag1_Zero5     0.5985      0.358      1.673      0.094      -0.103       1.300
==============================================================================
Omnibus:                       84.021   Durbin-Watson:                   2.232
Prob(Omnibus):                  0.000   Jarque-Bera (JB):              557.364
Skew:                           0.887   Prob(JB):                    9.33e-122
Kurtosis:                       9.207   Cond. No.                         124.
==============================================================================

Notes:
[1] Standard Errors are heteroscedasticity and autocorrelation robust (HAC) using 2 lags and without small sample correction


Parameters for Model: AR(1) with Monthly Zero Interaction:
```

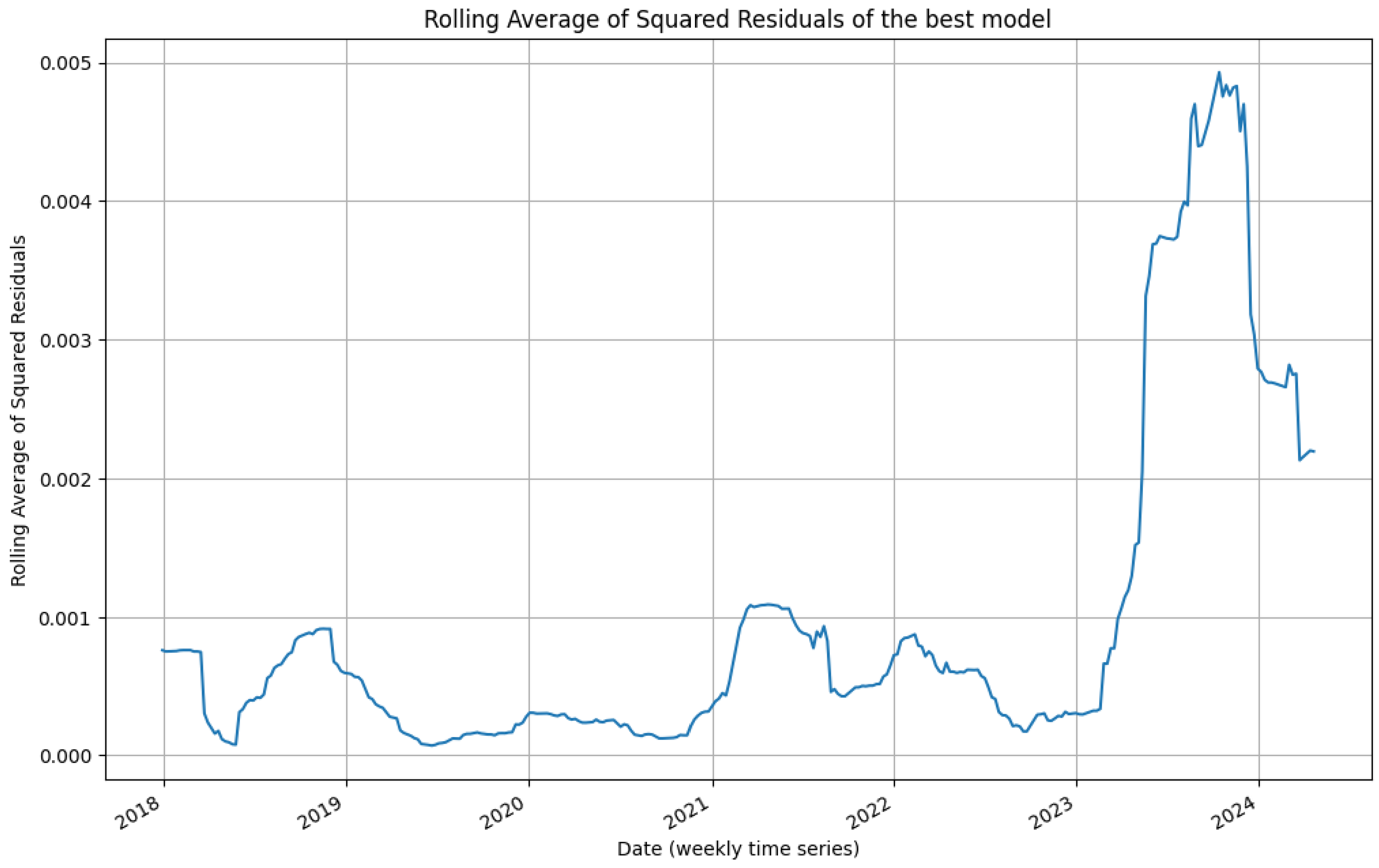
```
                            OLS Regression Results
==============================================================================
Dep. Variable:             Log_Return   R-squared:                       0.421
Model:                            OLS   Adj. R-squared:                  0.415
Method:                 Least Squares   F-statistic:                     32.88
Date:                Mon, 03 Jun 2024   Prob (F-statistic):           1.57e-18
Time:                        18:01:46   Log-Likelihood:                 623.00
No. Observations:                 321   AIC:                            -1238.
Df Residuals:                     317   BIC:                            -1223.
Df Model:                           3
Covariance Type:                  HAC
==========================================================================================
                             coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------------------
const                      0.0005      0.008      0.055      0.956      -0.016       0.017
Z22_Lag1                  -0.0012      0.010     -0.116      0.908      -0.021       0.018
Log_Return_Lag1            0.6688      0.128      5.236      0.000       0.418       0.919
Log_Return_Lag1_Zero22    -0.0614      0.269     -0.228      0.820      -0.590       0.467
==============================================================================
Omnibus:                       79.704   Durbin-Watson:                   2.239
Prob(Omnibus):                  0.000   Jarque-Bera (JB):              640.105
Skew:                           0.761   Prob(JB):                    1.01e-139
Kurtosis:                       9.748   Cond. No.                         131.
==============================================================================

Notes:
[1] Standard Errors are heteroscedasticity and autocorrelation robust (HAC) using 2 lags and without small sample correction


Parameters for Model: AR(2) with Weekly Zero Interaction:
                            OLS Regression Results
==============================================================================
Dep. Variable:             Log_Return   R-squared:                       0.455
Model:                            OLS   Adj. R-squared:                  0.446
Method:                 Least Squares   F-statistic:                     23.57
Date:                Mon, 03 Jun 2024   Prob (F-statistic):           4.11e-20
Time:                        18:01:46   Log-Likelihood:                 630.24
No. Observations:                 320   AIC:                            -1248.
Df Residuals:                     314   BIC:                            -1226.
Df Model:                           5
Covariance Type:                  HAC
=========================================================================================
                            coef    std err          z      P>|z|      [0.025      0.975]
-----------------------------------------------------------------------------------------
const                    -0.0002      0.007     -0.027      0.979      -0.014       0.014
Z5_Lag1                   0.0008      0.008      0.092      0.927      -0.015       0.017
Log_Return_Lag1           0.4724      0.122      3.868      0.000       0.233       0.712
Log_Return_Lag2           0.1412      0.176      0.801      0.423      -0.204       0.487
Log_Return_Lag1_Zero5     0.4751      0.346      1.373      0.170      -0.203       1.153
Log_Return_Lag2_Zero5     0.0586      0.225      0.261      0.794      -0.382       0.499
==============================================================================
Omnibus:                       96.832   Durbin-Watson:                   2.040
Prob(Omnibus):                  0.000   Jarque-Bera (JB):              691.602
Skew:                           1.039   Prob(JB):                    6.62e-151
Kurtosis:                       9.896   Cond. No.                         169.
==============================================================================

Notes:
[1] Standard Errors are heteroscedasticity and autocorrelation robust (HAC) using 2 lags and without small sample correction


Parameters for Model: AR(2) with Monthly Zero Interaction:
                            OLS Regression Results
==============================================================================
Dep. Variable:             Log_Return   R-squared:                       0.450
Model:                            OLS   Adj. R-squared:                  0.441
Method:                 Least Squares   F-statistic:                     25.24
Date:                Mon, 03 Jun 2024   Prob (F-statistic):           2.18e-21
Time:                        18:01:46   Log-Likelihood:                 628.76
No. Observations:                 320   AIC:                            -1246.
Df Residuals:                     314   BIC:                            -1223.
Df Model:                           5
Covariance Type:                  HAC
==========================================================================================
                             coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------------------
const                      0.0007      0.009      0.084      0.933      -0.016       0.018
Z22_Lag1                  -0.0009      0.010     -0.087      0.931      -0.021       0.020
Log_Return_Lag1            0.6231      0.182      3.433      0.001       0.267       0.979
Log_Return_Lag2           -0.0047      0.162     -0.029      0.977      -0.323       0.313
Log_Return_Lag1_Zero22    -0.1798      0.309     -0.582      0.560      -0.785       0.425
Log_Return_Lag2_Zero22     0.4939      0.277      1.783      0.075      -0.049       1.037
==============================================================================
Omnibus:                       88.673   Durbin-Watson:                   2.066
Prob(Omnibus):                  0.000   Jarque-Bera (JB):              768.471
Skew:                           0.862   Prob(JB):                    1.35e-167
Kurtosis:                      10.394   Cond. No.                         218.
==============================================================================

Notes:
[1] Standard Errors are heteroscedasticity and autocorrelation robust (HAC) using 2 lags and without small sample correction


Parameters for Model: AR(2) with Separate Weekly and Monthly Zero Interactions:
                            OLS Regression Results
==============================================================================
Dep. Variable:             Log_Return   R-squared:                       0.457
Model:                            OLS   Adj. R-squared:                  0.447
Method:                 Least Squares   F-statistic:                     22.41
Date:                Mon, 03 Jun 2024   Prob (F-statistic):           5.90e-22
Time:                        18:01:46   Log-Likelihood:                 630.93
No. Observations:                 320   AIC:                            -1248.
Df Residuals:                     313   BIC:                            -1221.
Df Model:                           6
Covariance Type:                  HAC
==========================================================================================
                             coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------------------
const                      0.0009      0.009      0.098      0.922      -0.017       0.019
Z5_Lag1                    0.0027      0.010      0.274      0.784      -0.016       0.022
Z22_Lag1                  -0.0033      0.013     -0.252      0.801      -0.029       0.022
Log_Return_Lag1            0.4799      0.101      4.762      0.000       0.282       0.677
Log_Return_Lag2            0.0802      0.135      0.594      0.553      -0.184       0.345
Log_Return_Lag1_Zero5      0.4399      0.367      1.199      0.230      -0.279       1.159
Log_Return_Lag2_Zero22     0.2526      0.231      1.094      0.274      -0.200       0.705
==============================================================================
Omnibus:                       94.877   Durbin-Watson:                   2.042
Prob(Omnibus):                  0.000   Jarque-Bera (JB):              684.064
Skew:                           1.009   Prob(JB):                    2.87e-149
Kurtosis:                       9.872   Cond. No.                         184.
==============================================================================

Notes:
[1] Standard Errors are heteroscedasticity and autocorrelation robust (HAC) using 2 lags and without small sample correction
```

Not like Fridays, the mix model **AR(2) with Separate Weekly and Monthly Zero Interactions** shows high enough Ljung-Box p-values ($0.757, 0.940, 0.996, 0.372$) based on these four dataframes, supports there not exist serial correlations of residuals.

## P11: Plot rolling average of SQUARED residuals of the best model (Use fridays_models as the example)

In [83]:
```python
def plot_rolling_average_of_squared_residuals(model_results):
    """
    Plots the rolling average of squared residuals for a given model.

    Parameters:
    - model_results (RegressionResultsWrapper): The fitted model result.

    Outputs:
    - A plot showing the rolling average of squared residuals.
    """
    residuals = model_results.resid
    squared_residuals = residuals ** 2
    rolling_average = squared_residuals.rolling(window=26).mean()

    plt.figure(figsize=(12, 8))
    rolling_average.plot(title='Rolling Average of Squared Residuals of the best model')
    plt.xlabel('Date (weekly time series)')
    plt.ylabel('Rolling Average of Squared Residuals')
    plt.grid(True)
    plt.show()
```

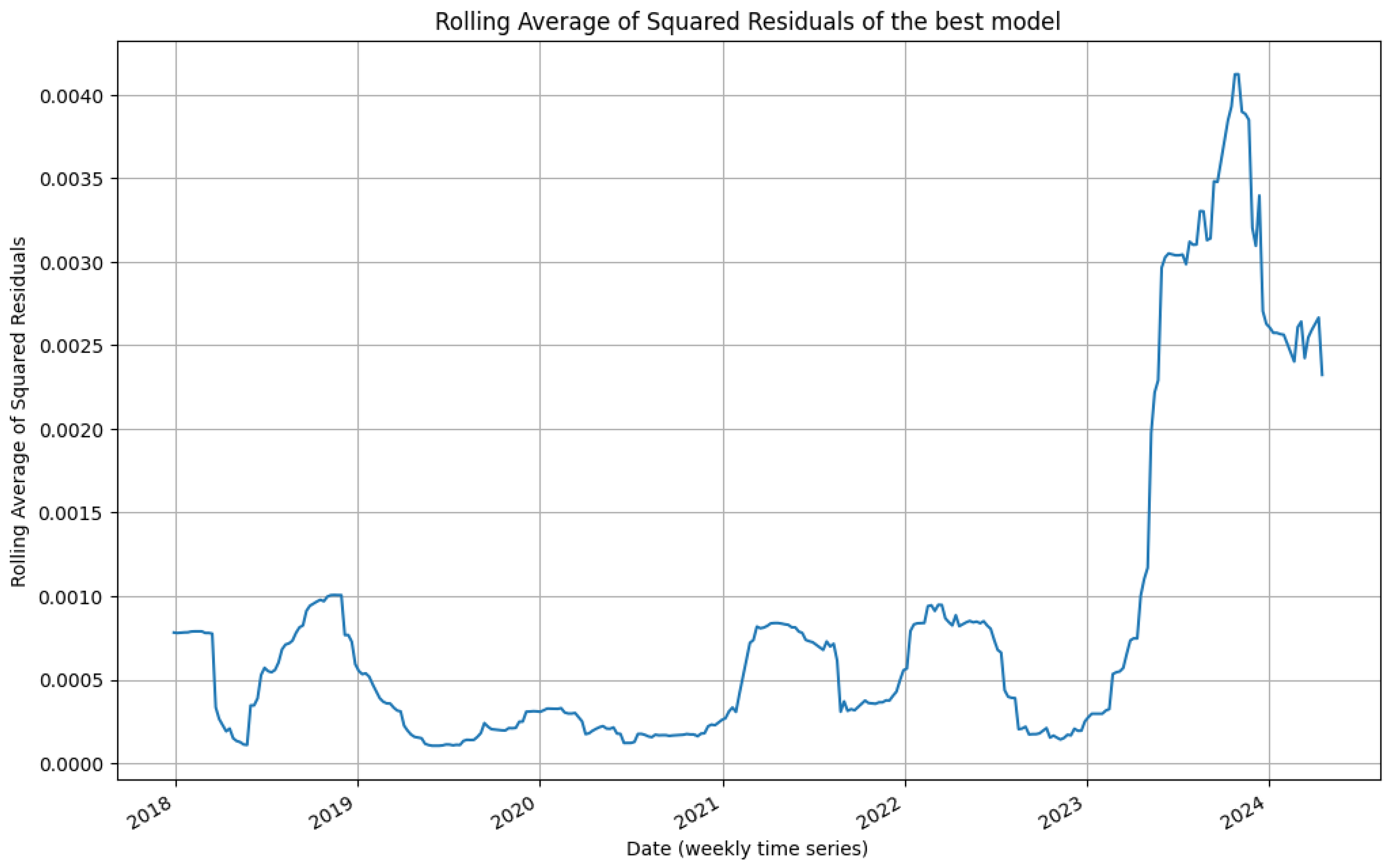
In [84]:
```python
plot_rolling_average_of_squared_residuals(model_results=fridays_models['AR(2) with Separate Weekly and Monthly Zero Interactions'])
```



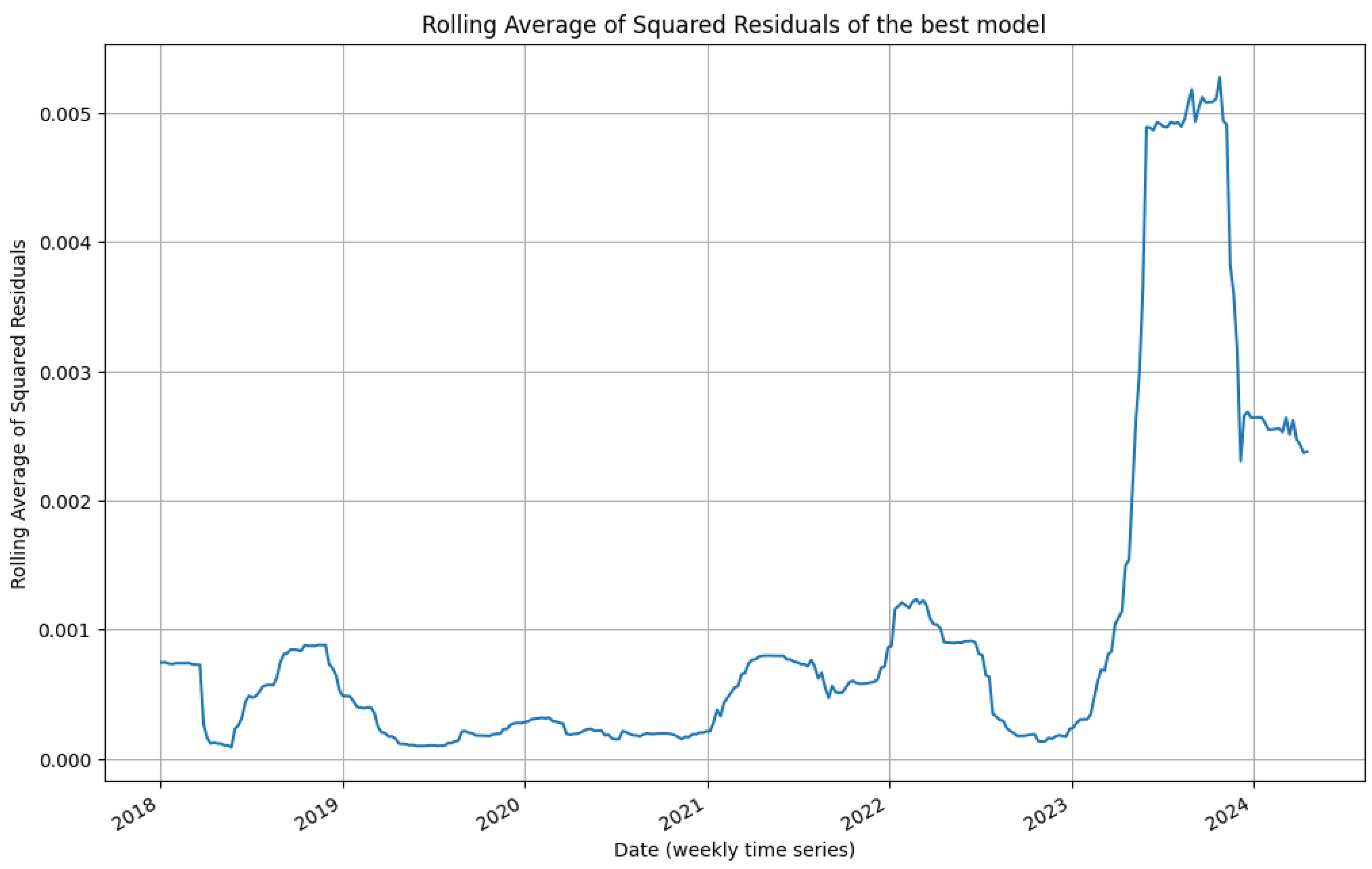

The rolling average follows a totally increasing trend, this indicates a rising volatility, but there also exists a sharp decreasing near the end of year 2023.

## *Robustness work for P11*

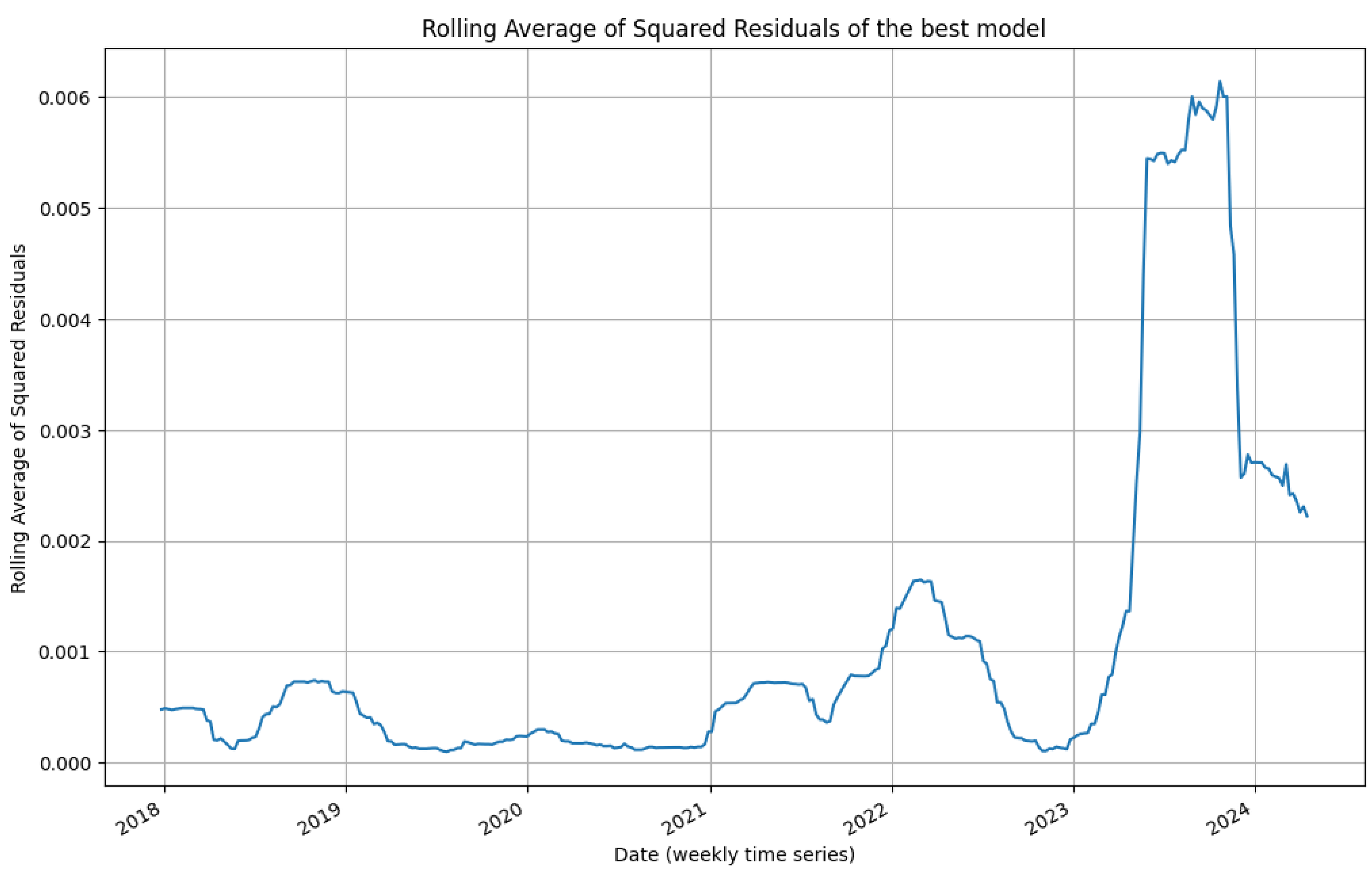
In [85]:
```
# Thursdays
plot_rolling_average_of_squared_residuals(model_results=thursdays_models['AR(2) with Separate Weekly and Monthly Zero Interactions'])
```



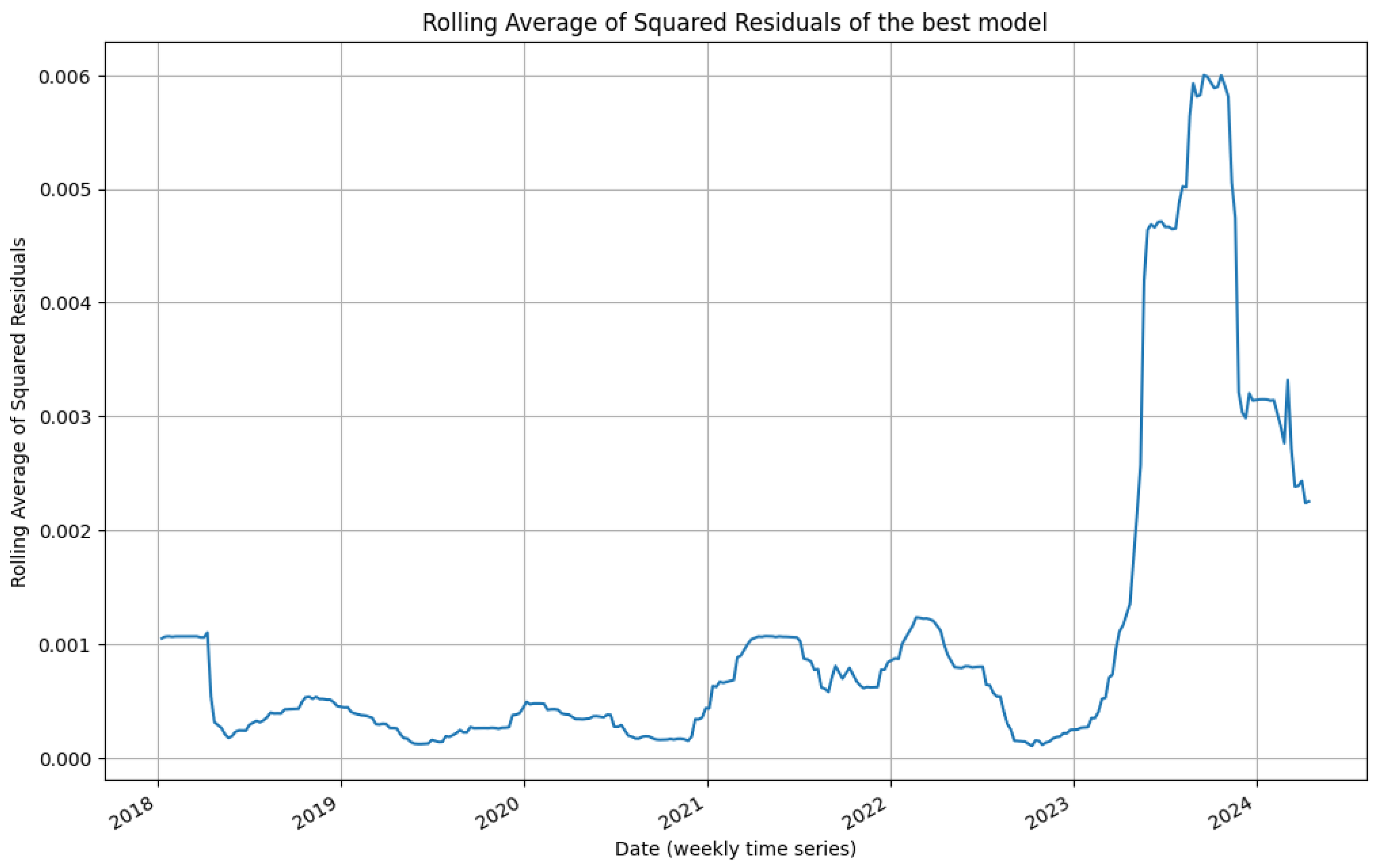

In [86]:
```
# Wednesdays
plot_rolling_average_of_squared_residuals(model_results=wednesdays_models['AR(2) with Separate Weekly and Monthly Zero Interactions'])
```




In [87]:
```
# Tuesdays
plot_rolling_average_of_squared_residuals(model_results=tuesdays_models['AR(2) with Separate Weekly and Monthly Zero Interactions'])
```




In [88]:
```
# Mondays
plot_rolling_average_of_squared_residuals(model_results=mondays_models['AR(2) with Separate Weekly and Monthly Zero Interactions'])
```




## P12: Perform Out-of-Sample Forecasting and Compute RMSE (Use Fridays as an example)

In [107…
```
def prepare_input_data(df):
    """
    Prepares input data for out-of-sample predictions by creating necessary lagged
    variables and interactions based on the specified models. The function safely
    operates on a copy of the input DataFrame to avoid modifying the original data.

    Parameters:
    - df: Original DataFrame containing the columns ['Log_Return',
        'Zero_Fraction_5', 'Zero_Fraction_22'].

    Returns:
    - input_data: A new DataFrame with original and derived columns
            necessary for making predictions with all ten models.
    """
    data = df.copy()

    input_data = pd.DataFrame(index=data.index)
    input_data['Log_Return'] = data['Log_Return']
    input_data['Zero_Fraction_5'] = data['Zero_Fraction_5']
    input_data['Zero_Fraction_22'] = data['Zero_Fraction_22']

    # Create lagged features for Log_Return
    input_data['Log_Return_Lag1'] = input_data['Log_Return'].shift(1)
    input_data['Log_Return_Lag2'] = input_data['Log_Return'].shift(2)
    input_data['Z5_Lag1'] = input_data['Zero_Fraction_5'].shift(1)
    input_data['Z22_Lag1'] = input_data['Zero_Fraction_22'].shift(1)

    # Create interactions for lagged Log_Return with Zero_Fractions
    input_data['Log_Return_Lag1_Zero5'] = input_data['Log_Return_Lag1'] * input_data['Z5_Lag1']
    input_data['Log_Return_Lag1_Zero22'] = input_data['Log_Return_Lag1'] * input_data['Z22_Lag1']
    input_data['Log_Return_Lag2_Zero5'] = input_data['Log_Return_Lag2']  * input_data['Z5_Lag1']
    input_data['Log_Return_Lag2_Zero22'] = input_data['Log_Return_Lag2']  * input_data['Z22_Lag1']

    input_data.dropna(inplace=True)

    # Add constant term for regression
    input_data = sm.add_constant(input_data)

    return input_data
```

In [108…
```
def calculate_rmse(errors):
    """
    Calculate the root mean squared error for a list of errors.
    """
    mse = np.mean([e**2 for e in errors])
    return np.sqrt(mse)
```

In [109…
```
def forecast_and_compute_RMSE(df, lags, df_name):
    """
    Performs out-of-sample forecasting using a list of predefined
    models on an expanding window basis. Starting with the first
    52 observations, this function forecasts the 53rd observation
    and continues expanding the sample one observation at a time,
    re-estimating models and forecasting the next observation.
    The forecast errors are calculated for each model and used to
    compute the Root Mean Squared Error (RMSE) for each model's
    predictions. The results are then ranked by RMSE and saved to
    a csv file.

    Parameters:
    - df: A DataFrame containing the time series data, specifically
        a column 'Log_Return' which is used for the forecasting.
    - lags: Maximum number of lags to use for HAC standard errors.
    - df_name: A string that specifies the name of the DataFrame,
          used to name the output CSV file.

    Returns:
    - rmse_df: a DataFrame containing the model names and their
           corresponding RMSE, sorted by RMSE.

    Outputs:
    - Prints a sorted table of the RMSE results for each model.
    - Saves a CSV file named '{df_name}_forecast_rmse_results.csv'
      containing the RMSE results for each model based on input
      dataframe.
    """
    data = df.copy()
    perpared_data = prepare_input_data(data)
    forecast_errors = {name: [] for name, _ in models}

    # Start forecasting from the 53rd observation, using the first 52 observations initially
    for i in range(52, len(perpared_data)):
        current_sample = perpared_data[['Log_Return', 'Zero_Fraction_5',        'Zero_Fraction_22']].iloc[:i]

        for name, model_func in models:

            if name == 'Model with Only Constant':  # Special case for constant model
                forecast = current_sample['Log_Return'].mean()
            else:
                features = model_features[name]
                model = model_func(current_sample, lags)
                input_data = perpared_data.iloc[i][features]
                forecast = model.predict(input_data)

            # Actual return for the next time point (i)
            actual_return = perpared_data['Log_Return'].iloc[i]
            forecast_error = actual_return - forecast
            forecast_errors[name].append(forecast_error)

    # Compute RMSE for each model
    rmses = {name: calculate_rmse(errors) for name, errors in forecast_errors.items()}
    rmse_df = pd.DataFrame(list(rmses.items()), columns=['Model', 'RMSE'])

    # Sort by RMSE
    rmse_df.sort_values(by='RMSE', inplace=True)

    # Print the RMSE table
    print(tabulate(rmse_df, headers='keys', tablefmt='pretty', showindex=False))

    # Save the RMSE results to CSV
    # files.download is just for google colab, if u use other enviroments, plz change
    csv_filename = f'{df_name}_forecast_rmse_results.csv'
    rmse_df.to_csv(csv_filename, index=True, header=True)
    files.download(csv_filename)

    return rmse_df
```

In [110…
```
fridays_rmse = forecast_and_compute_RMSE(df=Fridays, lags=4, df_name='Fridays')
```

```
+---------------------------------------------------------+-----------------------+
|                          Model                          |         RMSE          |
+---------------------------------------------------------+-----------------------+
|                          AR(2)                          | 0.03319323376216553   |
|                          AR(1)                          | 0.03347683497575105   |
|              AR(2) with Weekly Zero Interaction         | 0.03405335419043286   |
|              AR(1) with Weekly Zero Interaction         | 0.034065014359989744  |
|             AR(1) with Monthly Zero Interaction         | 0.03411353705615823   |
|             AR(2) with Monthly Zero Interaction         | 0.03424400124657069   |
| AR(2) with Separate Weekly and Monthly Zero Interactions | 0.03428830145726344   |
|                 Model with Only Constant                | 0.047601851591042715  |
|                  Model with Weekly Zero                 | 0.048056696113039886  |
|                 Model with Monthly Zero                 | 0.04812928878062246   |
+---------------------------------------------------------+-----------------------+
```

The model **AR(2)** wins in $RMSE$ criteria with lowest RMSE. This mix model has acceptably low RMSE (very close to the winner), which signifies high accuracy in predictive performance. It means that the mix model's predictions are, on average, close to the actual observed weekly log return.

## *Robustness work for P12*

In [111…
```
# Thursdays
thursdays_rmse = forecast_and_compute_RMSE(df=Thursdays, lags=2, df_name='Thursdays')
```

```
+---------------------------------------------------------+------------------------+
|                          Model                          |          RMSE          |
+---------------------------------------------------------+------------------------+
|                          AR(1)                          | 0.031228836428829586 |
|                          AR(2)                          | 0.03131017690997134  |
|            AR(2) with Weekly Zero Interaction            | 0.03131377832533065  |
|            AR(1) with Weekly Zero Interaction            | 0.031661200730862635 |
|           AR(1) with Monthly Zero Interaction            | 0.03195213191312714  |
|           AR(2) with Monthly Zero Interaction            | 0.03205756605217784  |
| AR(2) with Separate Weekly and Monthly Zero Interactions | 0.03208375760971013  |
|                 Model with Only Constant                 | 0.04684102151231067  |
|                  Model with Weekly Zero                  | 0.04728509441637568  |
|                 Model with Monthly Zero                  | 0.04735288128098435  |
+---------------------------------------------------------+------------------------+
```

In [112…

```python
# Wednesdays
wednesdays_rmse = forecast_and_compute_RMSE(df=Wednesdays, lags=2, df_name='Wednesdays')
```

```
+---------------------------------------------------------+------------------------+
|                          Model                          |          RMSE          |
+---------------------------------------------------------+------------------------+
|                          AR(1)                          |  0.0344754974330957  |
|                          AR(2)                          | 0.034532832194297196 |
|            AR(1) with Weekly Zero Interaction            | 0.03495969979970849  |
|           AR(1) with Monthly Zero Interaction            | 0.03531270390755053  |
|            AR(2) with Weekly Zero Interaction            | 0.035383770340930286 |
|           AR(2) with Monthly Zero Interaction            | 0.03584111892682286  |
| AR(2) with Separate Weekly and Monthly Zero Interactions | 0.03587648812686814  |
|                 Model with Only Constant                 | 0.047647999920785465 |
|                 Model with Monthly Zero                  | 0.048032922633657125 |
|                  Model with Weekly Zero                  | 0.04804314037456729  |
+---------------------------------------------------------+------------------------+
```

In [113…

```python
# Tuesdays
tuesdays_rmse = forecast_and_compute_RMSE(df=Tuesdays, lags=2, df_name='Tuesdays')
```

```
+---------------------------------------------------------+------------------------+
|                          Model                          |          RMSE          |
+---------------------------------------------------------+------------------------+
|                          AR(2)                          |  0.0363193179205231  |
|                          AR(1)                          | 0.03668898286049331  |
|            AR(1) with Weekly Zero Interaction            | 0.03706180633461377  |
|            AR(2) with Weekly Zero Interaction            | 0.03721136921820577  |
| AR(2) with Separate Weekly and Monthly Zero Interactions | 0.037419351322153495 |
|           AR(1) with Monthly Zero Interaction            | 0.03750723478198474  |
|           AR(2) with Monthly Zero Interaction            | 0.03768697952584409  |
|                 Model with Only Constant                 | 0.048320938820529416 |
|                  Model with Weekly Zero                  | 0.04879089668801973  |
|                 Model with Monthly Zero                  | 0.04882150180841967  |
+---------------------------------------------------------+------------------------+
```

In [114…

```python
# Mondays
mondays_rmse = forecast_and_compute_RMSE(df=Mondays, lags=2, df_name='Mondays')
```

```
+---------------------------------------------------------+------------------------+
|                          Model                          |          RMSE          |
+---------------------------------------------------------+------------------------+
|                          AR(2)                          | 0.03668184362733399  |
|                          AR(1)                          | 0.037050763630292645 |
|            AR(1) with Weekly Zero Interaction            | 0.03739738770736273  |
|            AR(2) with Weekly Zero Interaction            | 0.03754236275505407  |
| AR(2) with Separate Weekly and Monthly Zero Interactions | 0.03761768044852644  |
|           AR(2) with Monthly Zero Interaction            | 0.03776959641579725  |
|           AR(1) with Monthly Zero Interaction            |  0.0379527480142736  |
|                 Model with Only Constant                 | 0.04851899399219849  |
|                 Model with Monthly Zero                  | 0.04911067061442834  |
|                  Model with Weekly Zero                  | 0.049170200104615945 |
+---------------------------------------------------------+------------------------+
```

The winners in $RMSE$ criteria across different dataframes are **AR(2)** or **AR(1)** models, but the mix model also has accpetably low RMSE, which is very close to these two winners.

# Source of the weirdly lowh Ljung Box p-values for the Fridays series

In [115…

```python
model_test = fridays_models['AR(2) with Separate Weekly and Monthly Zero Interactions']
residuals = model_test.resid
ljung_box_results = acorr_ljungbox(residuals, lags=[3], return_df=True)

print("Ljung-Box test results for the first three lags:")
print(ljung_box_results)
```

```
Ljung-Box test results for the first three lags:
     lb_stat  lb_pvalue
3  20.599384   0.000127
```

large 'lb_stat' and low 'lb_pvalue'

### Compute the autocorrelations of the residuals at the first three lags

In [116…

```python
# Fridays
model_test = fridays_models['AR(2) with Separate Weekly and Monthly Zero Interactions']
residuals = model_test.resid
autocorrs = acf(residuals, nlags=3, fft=True)
print("Autocorrelations at the first three lags:", autocorrs[1:])
print("Fridays' third autocorrelation:", autocorrs[3])
```

```
Autocorrelations at the first three lags: [-0.03643096 -0.08550913  0.2336051 ]
Fridays' third autocorrelation: 0.23360509797963613
```

In [117…

```python
# Mondays
model_test = mondays_models['AR(2) with Separate Weekly and Monthly Zero Interactions']
residuals = model_test.resid
autocorrs = acf(residuals, nlags=3, fft=True)
print("Autocorrelations at the first three lags:", autocorrs[1:])
print("Mondays' third autocorrelation:", autocorrs[3])
```

```
Autocorrelations at the first three lags: [-0.02202399 -0.03511626  0.08901346]
Mondays' third autocorrelation: 0.08901345731509394
```

In [118…

```python
# Tuesdays
model_test = tuesdays_models['AR(2) with Separate Weekly and Monthly Zero Interactions']
residuals = model_test.resid
autocorrs = acf(residuals, nlags=3, fft=True)
print("Autocorrelations at the first three lags:", autocorrs[1:])
print("Tuesdays' third autocorrelation:", autocorrs[3])
```

```
Autocorrelations at the first three lags: [ 0.00817138  0.0021737  -0.01097647]
Tuesdays' third autocorrelation: -0.010976472633772105
```

In [119…

```python
# Wednesdays
model_test = wednesdays_models['AR(2) with Separate Weekly and Monthly Zero Interactions']
residuals = model_test.resid
autocorrs = acf(residuals, nlags=3, fft=True)
print("Autocorrelations at the first three lags:", autocorrs[1:])
print("Wednesdays' third autocorrelation:", autocorrs[3])
```

```
Autocorrelations at the first three lags: [-0.01120693 -0.02472329 -0.02155822]
Wednesdays' third autocorrelation: -0.02155821724733596
```

In [120…

```python
# Thursdays
model_test = thursdays_models['AR(2) with Separate Weekly and Monthly Zero Interactions']
residuals = model_test.resid
autocorrs = acf(residuals, nlags=3, fft=True)
print("Autocorrelations at the first three lags:", autocorrs[1:])
print("Thursdays' third autocorrelation:", autocorrs[3])
```

```
Autocorrelations at the first three lags: [-0.01760913 -0.04974084  0.02741009]
Thursdays' third autocorrelation: 0.027410086612236765
```

Comparing Fridays and other four data, it is obivous that the reason for the Ljung-Box test rejecting the null hypothesis of no autocorrelation for models based on Fridays is the unusually higher third autocorrelation $(0.234)$ than other four data $(0.089, -0.011, -0.022, 0.027)$, where first two autocorrelations have no significant difference between these five data.

# P13: Forecast Graph

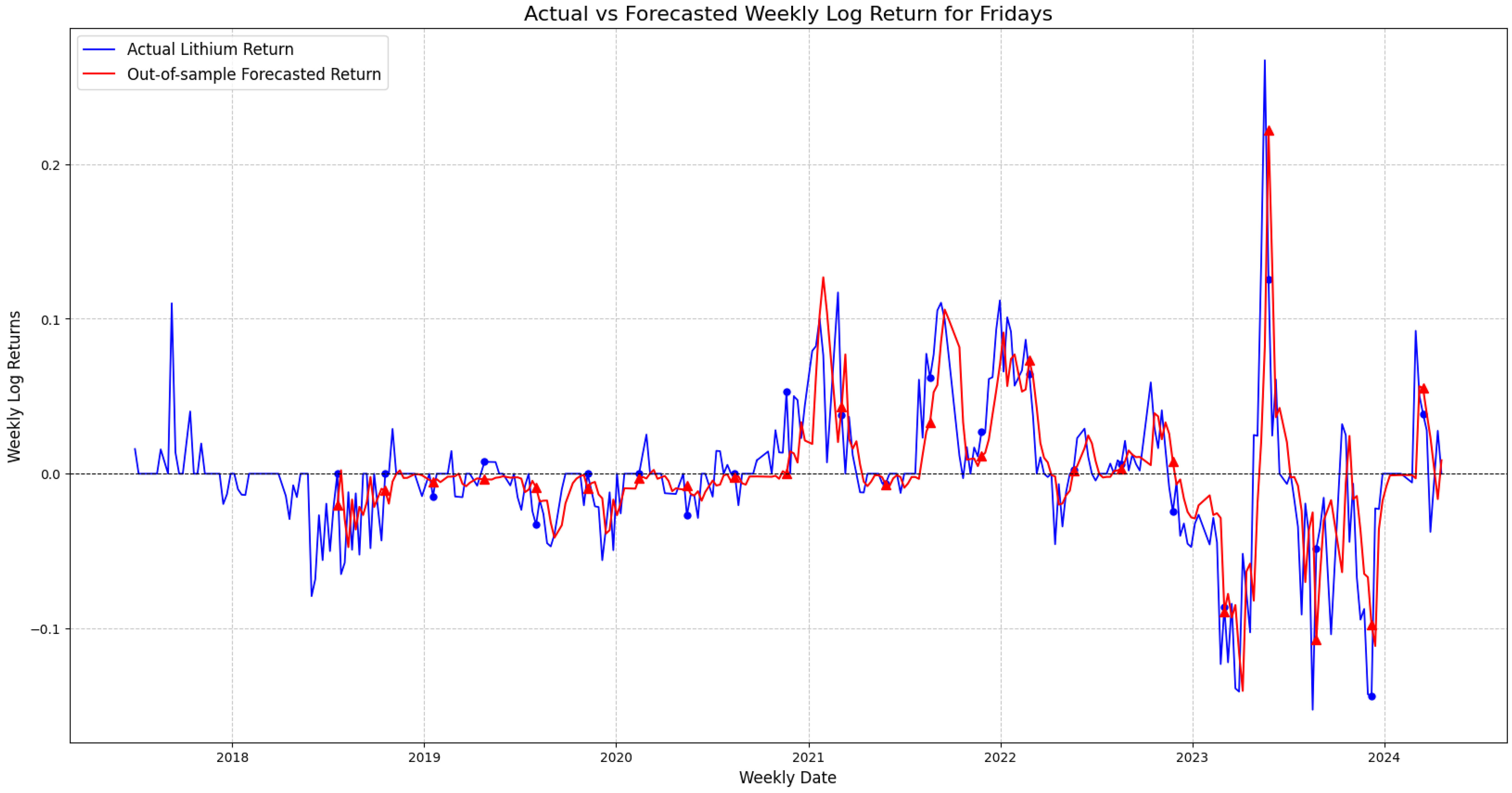
In [121…

```python
def generate_forecast_plot(df, lags, df_name):
    """
    Generates a plot showing the actual observed returns and the out-of-sample
    forecasted returns for the given DataFrame using the AR(2) model with
    separate interactions for zero fractions

    Parameters:
    - df: The input DataFrame containing the data. It must
        include the following columns: 'Log_Return',
        'Zero_Fraction_5', and 'Zero_Fraction_22'.
    - lags: Maximum number of lags to use for HAC standard errors.
    - df_name: A string that specifies the name of the DataFrame,
          used to name the plot titles.

    Returns:
    - forecasted_series: A series of out-of-sample forecast results
    - Plot Actual vs Forecasted Weekly Log Return
    """
    data = df.copy()
    perpared_data = prepare_input_data(data)
    forecasted_returns = []

    for i in range(52, len(perpared_data)):
        train_data = perpared_data[['Log_Return', 'Zero_Fraction_5',     'Zero_Fraction_22']].iloc[:i]
        features = model_features['AR(2) with Separate Weekly and Monthly Zero Interactions']
        model = estimate_model_ar2_z5_z22_separate_interaction(train_data, lags)
        input_data = perpared_data.iloc[i][features]
        forecast = model.predict(input_data)
        forecasted_returns.append(forecast[0])

    # Adjust the index for forecasted returns
    forecasted_index = perpared_data.index[52:]
    forecasted_series = pd.Series(forecasted_returns, index=forecasted_index)

    marker_positions = forecasted_index[::12]

    # Plot the actual vs forecasted returns
```

```python
    plt.figure(figsize=(20, 10))
    plt.plot(perpared_data.index, perpared_data['Log_Return'], label='Actual Lithium Return', color='blue', linewidth=1.3)
    plt.plot(marker_positions, perpared_data.loc[marker_positions, 'Log_Return'], 'bo', markersize=5)
    plt.plot(forecasted_series.index, forecasted_series, label='Out-of-sample Forecasted Return', color='red', linewidth=1.5)
    plt.plot(marker_positions, forecasted_series.loc[marker_positions], 'r^', markersize=7)
    plt.axhline(0, color='k', linestyle='--', linewidth=0.8)
    plt.xlabel('Weekly Date', fontsize=12)
    plt.ylabel('Weekly Log Returns', fontsize=12)
    plt.title(f'Actual vs Forecasted Weekly Log Return for {df_name}', fontsize=16)
    plt.legend(loc='upper left', fontsize=12)
    plt.grid(True, linestyle='--', alpha=0.7)
    plt.show()

    return forecasted_series
```

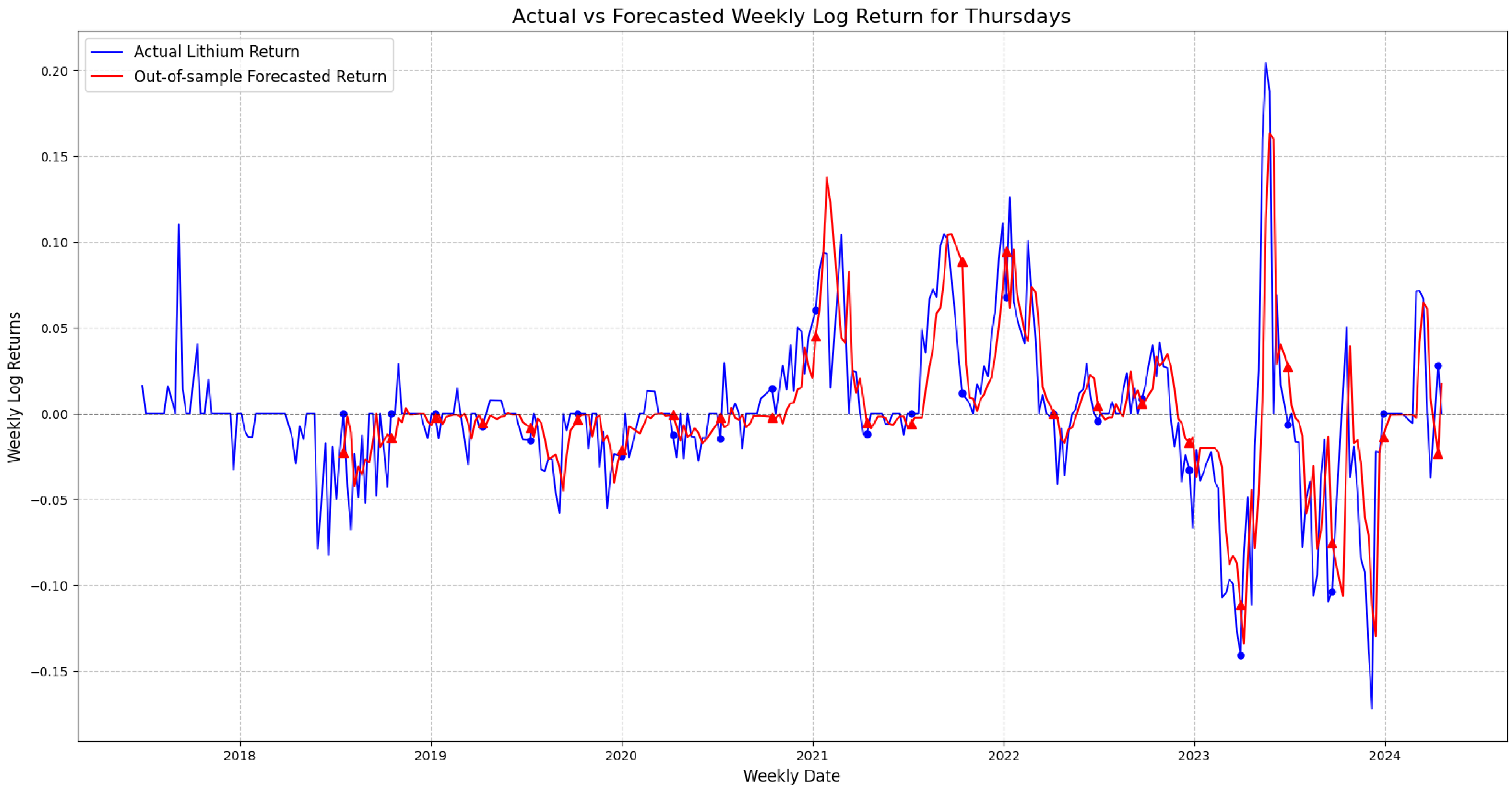
In [122…

```python
fridays_forecast = generate_forecast_plot(Fridays, lags=4, df_name='Fridays')
```



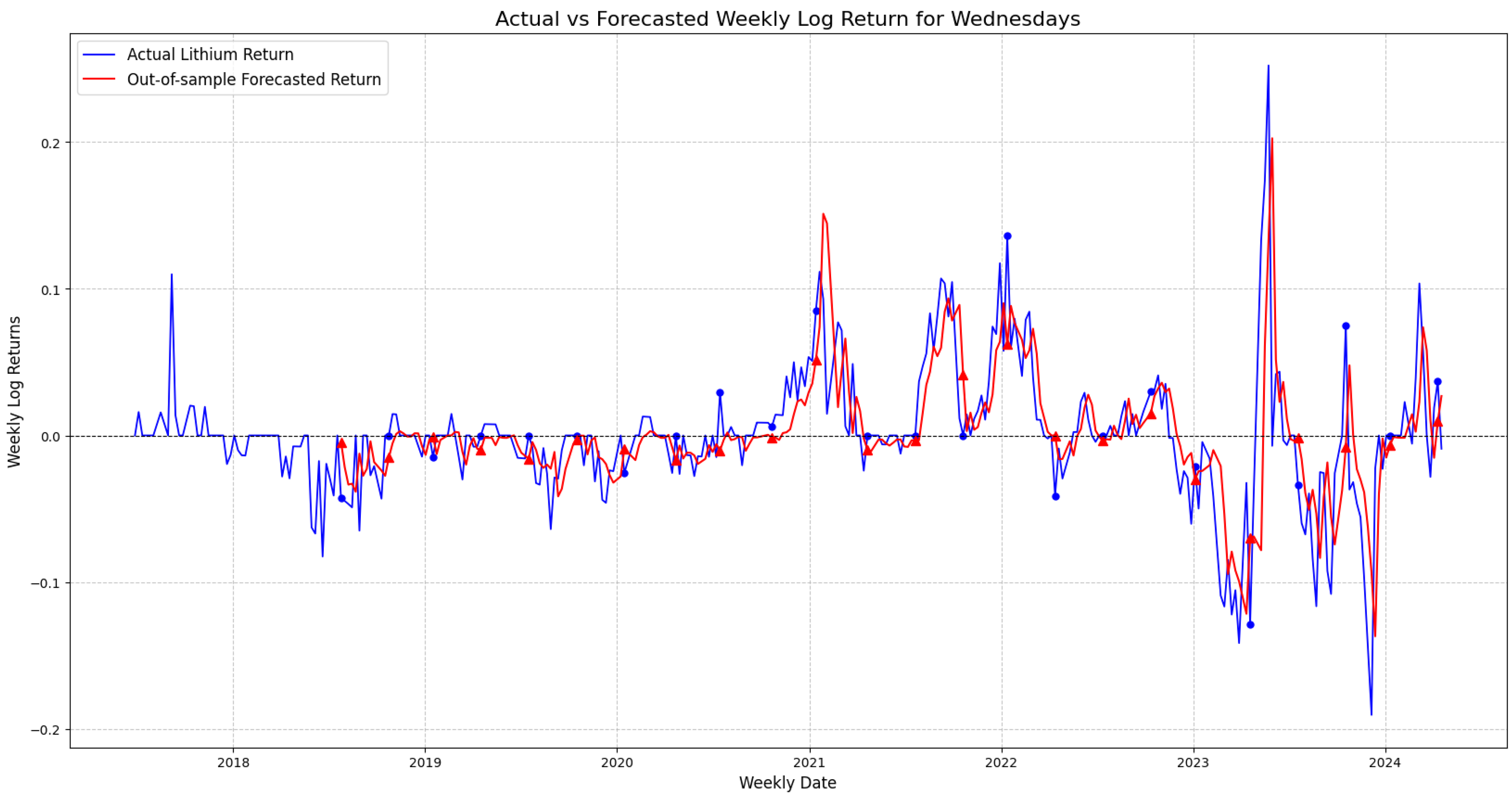

In [123…

```python
thursdays_forecast = generate_forecast_plot(Thursdays, lags=2, df_name='Thursdays')
```



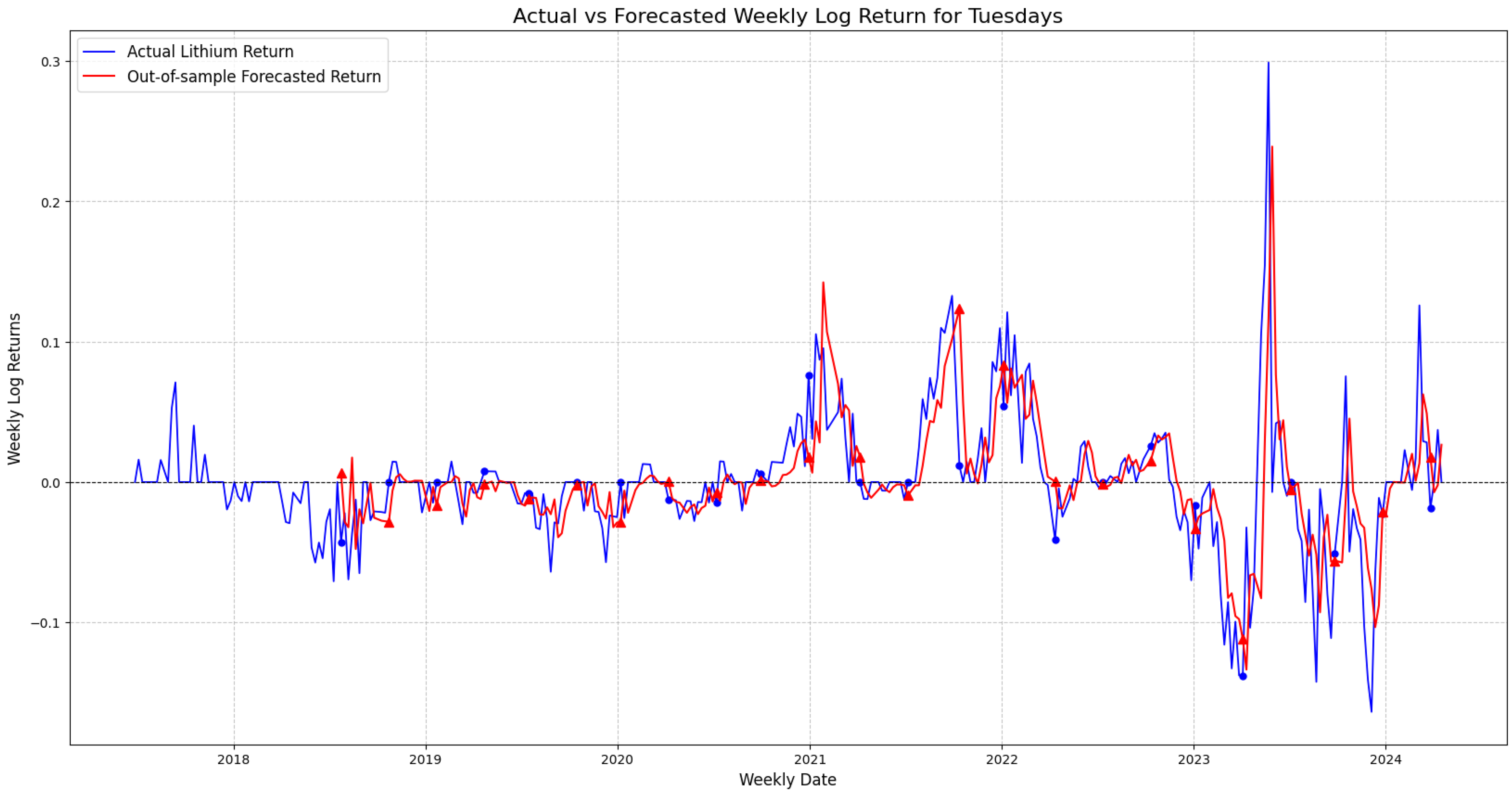

In [124…

```python
wednesdays_forecast = generate_forecast_plot(Wednesdays, lags=2, df_name='Wednesdays')
```



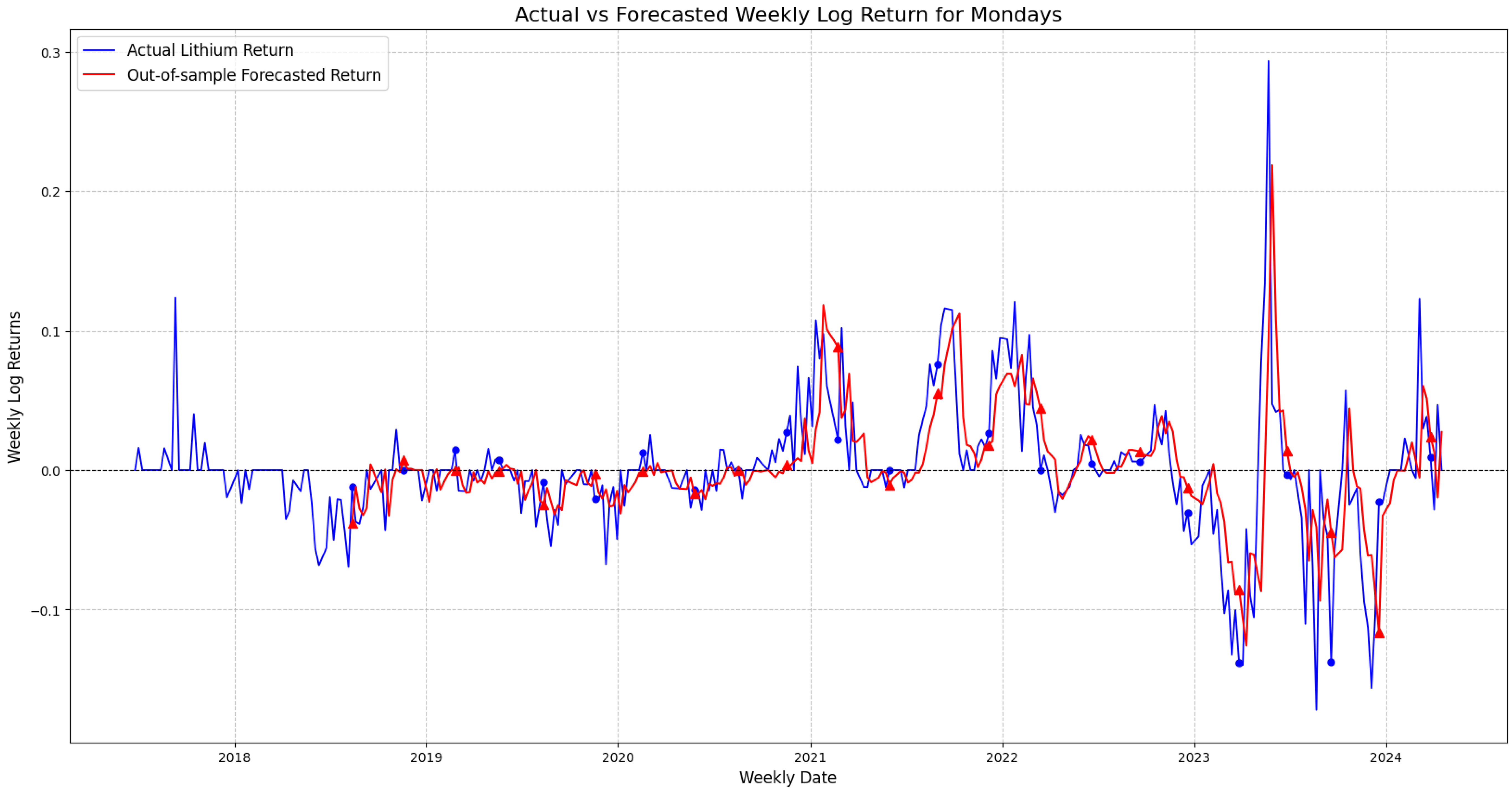

In [125…

```python
tuesdays_forecast = generate_forecast_plot(Tuesdays, lags=2, df_name='Tuesdays')
```




In [126…

```
mondays_forecast = generate_forecast_plot(Mondays, lags=2, df_name='Mondays')
```




## **New** P14: Unbiasedness test of the forecasts

In [128…

```
def ex_post_regression(df, forecasted_return, lags):
    """
    Perform an ex-post regression to test the unbiasedness of forecasts.

    Parameters:
    - df (Dataframe): A dataframe contains actual returns r(t).
    - forecasted_return (pd.Series): A series of forecasted returns r^(t).
    - lags: Maximum number of lags to use for HAC standard errors.

    Returns:
    - Print a formatted table containing the coefficient, t-statistic, and p-value for the forecasted returns coefficient.
    """
    data = df.copy()
    prepared_data = prepare_input_data(data)
    # Let the index of actual return series match that of forecasted return series
    actual_return = prepared_data['Log_Return'].iloc[52:]
    X = sm.add_constant(forecasted_return)
    model = sm.OLS(actual_return, X).fit(cov_type='HAC', cov_kwds={'maxlags': lags})
    print(model.summary())
    b_coefficient = model.params[0]
    t_statistic = model.tvalues[0]
    p_value = model.pvalues[0]
    results = {
        'b coefficient': b_coefficient,
        't-statistic': t_statistic,
        'p-value': p_value
    }

    table = tabulate(results.items(), headers=['Metric', 'Value'], tablefmt='pretty', showindex=False)
    print(table)
```

In [129…

```
ex_post_regression(df=Fridays, forecasted_return=fridays_forecast, lags=4)
```

```
                            OLS Regression Results
==============================================================================
Dep. Variable:             Log_Return   R-squared:                       0.492
Model:                            OLS   Adj. R-squared:                  0.490
Method:                 Least Squares   F-statistic:                     122.2
Date:                Mon, 03 Jun 2024   Prob (F-statistic):           1.21e-23
Time:                        18:51:50   Log-Likelihood:                 529.42
No. Observations:                 269   AIC:                            -1055.
Df Residuals:                     267   BIC:                            -1048.
Df Model:                           1
Covariance Type:                  HAC
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
const      -2.977e-05      0.002     -0.012      0.990      -0.005       0.005
0              0.8535      0.077     11.056      0.000       0.702       1.005
==============================================================================
Omnibus:                       54.471   Durbin-Watson:                   1.791
Prob(Omnibus):                  0.000   Jarque-Bera (JB):              372.352
Skew:                           0.570   Prob(JB):                     1.40e-81
Kurtosis:                       8.650   Cond. No.                         25.6
==============================================================================

Notes:
[1] Standard Errors are heteroscedasticity and autocorrelation robust (HAC) using 4 lags and without small sample correction
+---------------+-----------------------+
|    Metric     |         Value         |
+---------------+-----------------------+
| b coefficient |   0.8535421296638047  |
|  t-statistic  |   11.055664492292182  |
|    p-value    | 2.0580710313524815e-28 |
+---------------+-----------------------+
```

Based on the regression analysis, the coefficient of the forecasted returns ($b$) is $0.854$, which is significantly different from $1$. The high t-statistic ($11.056$) and the extremely low p-value ($2.06e-28$) provide strong statistical evidence that the coefficient $b$ is not equal to 1. Therefore, we reject the null hypothesis that the forecasted returns are unbiased predictors of the actual returns.

Given that the $b$ coefficient is less than 1, this indicates that the forecasted returns systematically overestimate the actual returns. This overestimation implies that the predictive model tends to predict higher returns than what is observed in reality. As a result, the forecasts are not unbiased and exhibit a systematic bias.

In [130…

```
ex_post_regression(df=Thursdays, forecasted_return=thursdays_forecast, lags=2)
```

```
                            OLS Regression Results
==============================================================================
Dep. Variable:             Log_Return   R-squared:                       0.540
Model:                            OLS   Adj. R-squared:                  0.538
Method:                 Least Squares   F-statistic:                     115.7
Date:                Mon, 03 Jun 2024   Prob (F-statistic):           9.28e-23
Time:                        18:52:59   Log-Likelihood:                 565.19
No. Observations:                 278   AIC:                            -1126.
Df Residuals:                     276   BIC:                            -1119.
Df Model:                           1
Covariance Type:                  HAC
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
const         -0.0005      0.002     -0.219      0.827      -0.005       0.004
0              0.8721      0.081     10.754      0.000       0.713       1.031
==============================================================================
Omnibus:                       32.854   Durbin-Watson:                   1.638
Prob(Omnibus):                  0.000   Jarque-Bera (JB):              194.346
Skew:                           0.064   Prob(JB):                     6.29e-43
Kurtosis:                       7.094   Cond. No.                         25.4
==============================================================================

Notes:
[1] Standard Errors are heteroscedasticity and autocorrelation robust (HAC) using 2 lags and without small sample correction
+---------------+-----------------------+
|    Metric     |         Value         |
+---------------+-----------------------+
| b coefficient |  0.8720799141555767   |
|  t-statistic  |  10.754105458742975   |
|    p-value    | 5.668129349478561e-27 |
+---------------+-----------------------+
```

Similar results as Friday to Friday

In [131…
```python
ex_post_regression(df=Wednesdays, forecasted_return=wednesdays_forecast, lags=2)
```

```
                            OLS Regression Results
==============================================================================
Dep. Variable:             Log_Return   R-squared:                       0.449
Model:                            OLS   Adj. R-squared:                  0.447
Method:                 Least Squares   F-statistic:                     74.51
Date:                Mon, 03 Jun 2024   Prob (F-statistic):           4.88e-16
Time:                        18:53:02   Log-Likelihood:                 535.45
No. Observations:                 278   AIC:                            -1067.
Df Residuals:                     276   BIC:                            -1060.
Df Model:                           1
Covariance Type:                  HAC
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
const         -0.0002      0.002     -0.094      0.925      -0.005       0.004
0              0.8279      0.096      8.632      0.000       0.640       1.016
==============================================================================
Omnibus:                       51.396   Durbin-Watson:                   1.655
Prob(Omnibus):                  0.000   Jarque-Bera (JB):              529.231
Skew:                           0.282   Prob(JB):                    1.20e-115
Kurtosis:                       9.736   Cond. No.                         26.0
==============================================================================

Notes:
[1] Standard Errors are heteroscedasticity and autocorrelation robust (HAC) using 2 lags and without small sample correction
+---------------+------------------------+
|    Metric     |         Value          |
+---------------+------------------------+
| b coefficient |   0.8279232315944781   |
|  t-statistic  |   8.631861350275683    |
|    p-value    | 6.0361526891087934e-18 |
+---------------+------------------------+
```

Similar results as Friday to Friday

In [132…
```python
ex_post_regression(df=Tuesdays, forecasted_return=tuesdays_forecast, lags=2)
```

```
                            OLS Regression Results
==============================================================================
Dep. Variable:             Log_Return   R-squared:                       0.421
Model:                            OLS   Adj. R-squared:                  0.419
Method:                 Least Squares   F-statistic:                     55.40
Date:                Mon, 03 Jun 2024   Prob (F-statistic):           1.23e-12
Time:                        18:53:26   Log-Likelihood:                 528.36
No. Observations:                 280   AIC:                            -1053.
Df Residuals:                     278   BIC:                            -1045.
Df Model:                           1
Covariance Type:                  HAC
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
const         -0.0002      0.002     -0.092      0.926      -0.005       0.004
0              0.8071      0.108      7.443      0.000       0.595       1.020
==============================================================================
Omnibus:                       63.265   Durbin-Watson:                   1.719
Prob(Omnibus):                  0.000   Jarque-Bera (JB):              802.598
Skew:                           0.436   Prob(JB):                    5.22e-175
Kurtosis:                      11.248   Cond. No.                         25.8
==============================================================================

Notes:
[1] Standard Errors are heteroscedasticity and autocorrelation robust (HAC) using 2 lags and without small sample correction
+---------------+----------------------+
|    Metric     |        Value         |
+---------------+----------------------+
| b coefficient |  0.8070958437789344  |
|  t-statistic  |  7.4432265528796675  |
|    p-value    | 9.82553673773781e-14 |
+---------------+----------------------+
```

Similar results as Friday to Friday

In [133…
```python
ex_post_regression(df=Mondays, forecasted_return=mondays_forecast, lags=2)
```

```
                            OLS Regression Results
==============================================================================
Dep. Variable:             Log_Return   R-squared:                       0.410
Model:                            OLS   Adj. R-squared:                  0.408
Method:                 Least Squares   F-statistic:                     73.91
Date:                Mon, 03 Jun 2024   Prob (F-statistic):           7.13e-16
Time:                        18:53:29   Log-Likelihood:                 502.45
No. Observations:                 268   AIC:                            -1001.
Df Residuals:                     266   BIC:                            -993.7
Df Model:                           1
Covariance Type:                  HAC
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
const         -0.0002      0.002     -0.083      0.934      -0.005       0.005
0              0.8348      0.097      8.597      0.000       0.644       1.025
==============================================================================
Omnibus:                       66.174   Durbin-Watson:                   1.728
Prob(Omnibus):                  0.000   Jarque-Bera (JB):              481.188
Skew:                           0.750   Prob(JB):                    3.25e-105
Kurtosis:                       9.391   Cond. No.                         27.0
==============================================================================

Notes:
[1] Standard Errors are heteroscedasticity and autocorrelation robust (HAC) using 2 lags and without small sample correction
+---------------+-----------------------+
|    Metric     |         Value         |
+---------------+-----------------------+
| b coefficient |  0.8347515603741322   |
|  t-statistic  |   8.596850736326246   |
|    p-value    | 8.193344205106168e-18 |
+---------------+-----------------------+
```

Similar results as Friday to Friday

## New P15: Three MSE Statistics

In [142…
```python
# Define the critical values from McCracken's Table 4 for pi = 2.0
mccracken_critical_values = {
    1: {0.01: 3.951, 0.05: 1.518, 0.10: 0.616},
    2: {0.01: 4.250, 0.05: 1.706, 0.10: 0.506},
    3: {0.01: 4.184, 0.05: 1.612, 0.10: 0.127},
    4: {0.01: 4.096, 0.05: 1.029, 0.10: -0.456},
    5: {0.01: 3.783, 0.05: 0.459, 0.10: -1.072},
    6: {0.01: 3.321, 0.05: -0.109, 0.10: -1.664},
}
```

In [182…
```python
def calculate_statistics(RMSE_df, forecast_series, df_name):
    """
    Calculate the R², Adjusted R², and McCracken statistics for the given RMSE values of different models.

    Parameters:
    - RMSE_df (pd.DataFrame): A DataFrame containing model names and their corresponding RMSE values.
    - forecast_series (pd.Series): A series containing the forecasted return values.
    - df_name: A string that specifies the name of the DataFrame, used to name the output CSV file.

    Returns:
    - result_df (pd.DataFrame): A DataFrame containing the R², Adjusted R², and McCracken statistics for each model.
    - Print result_df
    - Store the result_df to a CSV file
    """
    df = RMSE_df.copy()

    df['MSE'] = df['RMSE'] ** 2
    mse_mean = df.loc[df['Model'] == 'Model with Only Constant', 'MSE'].values[0]

    # Calculate R² values
    df['R^2'] = 1 - (df['MSE'] / mse_mean)

    # Calculate Adjusted R² values
    T = len(forecast_series)
    df['k'] = df['Model'].apply(lambda x: len(model_features.get(x, ['const'])))
    df['Adj R^2'] = 1 - (1 - df['R^2']) * ((T - df['k']) / (T - 1))

    # Calculate McCracken statistics
    df['McCracken Stat'] = T * (mse_mean - df['MSE']) / df['MSE']
```

```python
    df['k2'] = df['k'] - 1

    # Determine significance using critical values
    df['Significance 1%'] = df.apply(lambda row: row['McCracken Stat'] > mccracken_critical_values[row['k2']][0.01] if row['k2'] in mccracken_critical_values else False, axis=1)
    df['Significance 5%'] = df.apply(lambda row: row['McCracken Stat'] > mccracken_critical_values[row['k2']][0.05] if row['k2'] in mccracken_critical_values else False, axis=1)
    df['Significance 10%'] = df.apply(lambda row: row['McCracken Stat'] > mccracken_critical_values[row['k2']][0.10] if row['k2'] in mccracken_critical_values else False, axis=1)

    df['Significance 1%'] = df['Significance 1%'].apply(lambda x: 'significant' if x else 'not significant')
    df['Significance 5%'] = df['Significance 5%'].apply(lambda x: 'significant' if x else 'not significant')
    df['Significance 10%'] = df['Significance 10%'].apply(lambda x: 'significant' if x else 'not significant')

    # Select only the required columns
    result_df = df[['Model', 'R^2', 'Adj R^2', 'McCracken Stat', 'Significance 1%', 'Significance 5%', 'Significance 10%']]

    print(tabulate(result_df, headers='keys', tablefmt='pretty', showindex=False))

    csv_filename = f'{df_name}_mse_statistics.csv'
    result_df.to_csv(csv_filename, index=True, header=True)
    files.download(csv_filename)

    return result_df
```

In [184…

```python
fridays_mse_statistics = calculate_statistics(fridays_rmse, fridays_forecast, df_name='Fridays')
```

| Model | R^2 | Adj R^2 | McCracken Stat | Significance 1% | Significance 5% | Significance 10% |
|---|---|---|---|---|---|---|
| AR(2) | 0.5137591556318142 | 0.5173878186494872 | 284.22378429466283 | significant | significant | significant |
| AR(1) | 0.5054148290003098 | 0.5072602960562788 | 274.8901442521585 | significant | significant | significant |
| AR(2) with Weekly Zero Interaction | 0.48823321093949523 | 0.4977810987951017 | 256.6300442899527 | significant | significant | significant |
| AR(1) with Weekly Zero Interaction | 0.4878826839913033 | 0.49361534051378875 | 256.27026833717116 | significant | significant | significant |
| AR(1) with Monthly Zero Interaction | 0.4864227103935349 | 0.492171709904055 | 254.77705448409642 | significant | significant | significant |
| AR(2) with Monthly Zero Interaction | 0.4824869373569941 | 0.49214203180928895 | 250.7936427462958 | significant | significant | significant |
| AR(2) with Separate Weekly and Monthly Zero Interactions | 0.4811470958607106 | 0.49276320565487375 | 249.4513719668517 | significant | significant | significant |
| Model with Only Constant | 0.0 | 0.0 | 0.0 | not significant | not significant | not significant |
| Model with Weekly Zero | -0.019201672565752448 | -0.015398681250208712 | -5.06793705232482 | not significant | not significant | not significant |
| Model with Monthly Zero | -0.02228313487554745 | -0.018468645566310338 | -5.8635059867753725 | not significant | not significant | not significant |

In [185…

```python
thursdays_mse_statistics = calculate_statistics(thursdays_rmse, thursdays_forecast, df_name='Thursdays')
```

| Model | R^2 | Adj R^2 | McCracken Stat | Significance 1% | Significance 5% | Significance 10% |
|---|---|---|---|---|---|---|
| AR(1) | 0.555513177047838 | 0.5571178226180624 | 347.4403632341645 | significant | significant | significant |
| AR(2) | 0.5531946880093286 | 0.5564207191428352 | 344.1949304080883 | significant | significant | significant |
| AR(2) with Weekly Zero Interaction | 0.553091895612974 | 0.5611588289051586 | 344.0518206562881 | significant | significant | significant |
| AR(1) with Weekly Zero Interaction | 0.5431201052564556 | 0.5480682629612593 | 330.4750132330659 | significant | significant | significant |
| AR(1) with Monthly Zero Interaction | 0.5346850932516728 | 0.5397246048771059 | 319.4448614653143 | significant | significant | significant |
| AR(2) with Monthly Zero Interaction | 0.5316091783312928 | 0.5400638863036522 | 315.52145076111117 | significant | significant | significant |
| AR(2) with Separate Weekly and Monthly Zero Interactions | 0.5308434996977309 | 0.5410057343613179 | 314.55280449250864 | significant | significant | significant |
| Model with Only Constant | 0.0 | 0.0 | 0.0 | not significant | not significant | not significant |
| Model with Weekly Zero | -0.01905073331557028 | -0.015371849801795623 | -5.197095383560765 | not significant | not significant | not significant |
| Model with Monthly Zero | -0.02197460475834956 | -0.018285165752001564 | -5.977585053853333 | not significant | not significant | not significant |

In [186…

```python
wednesdays_mse_statistics = calculate_statistics(wednesdays_rmse, wednesdays_forecast, df_name='Wednesdays')
```

| Model | R^2 | Adj R^2 | McCracken Stat | Significance 1% | Significance 5% | Significance 10% |
|---|---|---|---|---|---|---|
| AR(1) | 0.47648186367937473 | 0.47837182085020735 | 253.02267278423517 | significant | significant | significant |
| AR(2) | 0.4747391334176516 | 0.47853163064929316 | 251.26082578515528 | significant | significant | significant |
| AR(1) with Weekly Zero Interaction | 0.4616731536053529 | 0.4675034082594466 | 238.41489155121633 | significant | significant | significant |
| AR(1) with Monthly Zero Interaction | 0.45074680084813157 | 0.45669539145266447 | 228.1417947665573 | significant | significant | significant |
| AR(2) with Weekly Zero Interaction | 0.44853384410589636 | 0.45848810684766717 | 226.11071836180543 | significant | significant | significant |
| AR(2) with Monthly Zero Interaction | 0.43418589681330866 | 0.4443991477733572 | 213.32744912912398 | significant | significant | significant |
| AR(2) with Separate Weekly and Monthly Zero Interactions | 0.43306861770772254 | 0.4453487198512376 | 212.35916635265582 | significant | significant | significant |
| Model with Only Constant | 0.0 | 0.0 | 0.0 | not significant | not significant | not significant |
| Model with Monthly Zero | -0.016222192154939696 | -0.012553519981095107 | -4.437778916744636 | not significant | not significant | not significant |
| Model with Weekly Zero | -0.016654587225026773 | -0.01298435405814935 | -4.5541281244744285 | not significant | not significant | not significant |

In [187…

```python
tuesdays_mse_statistics = calculate_statistics(tuesdays_rmse, tuesdays_forecast, df_name='Tuesdays')
```

| Model | R^2 | Adj R^2 | McCracken Stat | Significance 1% | Significance 5% | Significance 10% |
|---|---|---|---|---|---|---|
| AR(2) | 0.435056988582293 | 0.4391067592734593 | 215.62521235079745 | significant | significant | significant |
| AR(1) | 0.42349826770477994 | 0.42556458215745097 | 205.68804621841997 | significant | significant | significant |
| AR(1) with Weekly Zero Interaction | 0.4117222298251434 | 0.4180477972463784 | 195.96563085627776 | significant | significant | significant |
| AR(2) with Weekly Zero Interaction | 0.4069646608288342 | 0.417592534290683 | 192.14724234028245 | significant | significant | significant |
| AR(2) with Separate Weekly and Monthly Zero Interactions | 0.4003169384413602 | 0.41321334836735246 | 186.91330462503035 | significant | significant | significant |
| AR(1) with Monthly Zero Interaction | 0.39749678743025463 | 0.4039753165976713 | 184.72781249707842 | significant | significant | significant |
| AR(2) with Monthly Zero Interaction | 0.3917082359738825 | 0.4026095220675405 | 180.30542670306144 | significant | significant | significant |
| Model with Only Constant | 0.0 | 0.0 | 0.0 | not significant | not significant | not significant |
| Model with Weekly Zero | -0.019546110973747144 | -0.01589182383764043 | -5.367987787646174 | not significant | not significant | not significant |
| Model with Monthly Zero | -0.020825575794964024 | -0.01716670276344079 | -5.712201340614871 | not significant | not significant | not significant |

In [188…

```python
mondays_mse_statistics = calculate_statistics(mondays_rmse, mondays_forecast, df_name='Mondays')
```

| Model | R^2 | Adj R^2 | McCracken Stat | Significance 1% | Significance 5% | Significance 10% |
|---|---|---|---|---|---|---|
| AR(2) | 0.42841774781927766 | 0.4326992628168861 | 200.87390043605484 | significant | significant | significant |
| AR(1) | 0.41686279385073877 | 0.4190468283306985 | 191.58309155015925 | significant | significant | significant |
| AR(1) with Weekly Zero Interaction | 0.40590081223812435 | 0.41257608401072976 | 183.10312473178917 | significant | significant | significant |
| AR(2) with Weekly Zero Interaction | 0.40128570360743265 | 0.41249758181703133 | 179.62585696513366 | significant | significant | significant |
| AR(2) with Separate Weekly and Monthly Zero Interactions | 0.3988810060745892 | 0.412389298072913 | 177.8351885537899 | significant | significant | significant |
| AR(2) with Monthly Zero Interaction | 0.39401606141662493 | 0.4053640752477743 | 174.2559459686519 | significant | significant | significant |
| AR(1) with Monthly Zero Interaction | 0.3881247607201256 | 0.3949997634086635 | 169.99778581564013 | significant | significant | significant |
| Model with Only Constant | 0.0 | 0.0 | 0.0 | not significant | not significant | not significant |
| Model with Monthly Zero | -0.024538196082734665 | -0.020700974374559644 | -6.418732435078348 | not significant | not significant | not significant |
| Model with Weekly Zero | -0.027023489009053492 | -0.023176959087671367 | -7.051732635067792 | not significant | not significant | not significant |