Time series - Li2Co3 zeroes calculations and statistics

In [51]:
```
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 [2]:
```
# Data from 2017-05-10 to 2024-04-19
li2co3 = pd.read_csv(r'/content/Lithium Carbonate 99%Min China Spot Historical Data
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 [3]:
```
li2co3.head(10)
```

Out[3]:

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 [4]:

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

Out[4]:

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 [5]:

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

In [6]:

```
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()
```

```python
    weekly_log_return = weekly_log_return.dropna()
    return weekly_log_return
```

In [7]:
```python
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 [8]:
```python
# 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 [9]:
```python
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

Date	Log_Return	Zero_Count_22	Zero_Count_5
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

Date	Log_Return	Zero_Count_22	Zero_Count_5
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

Date	Log_Return	Zero_Count_22	Zero_Count_5
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

Date	Log_Return	Zero_Count_22	Zero_Count_5
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

Date	Log_Return	Zero_Count_22	Zero_Count_5
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

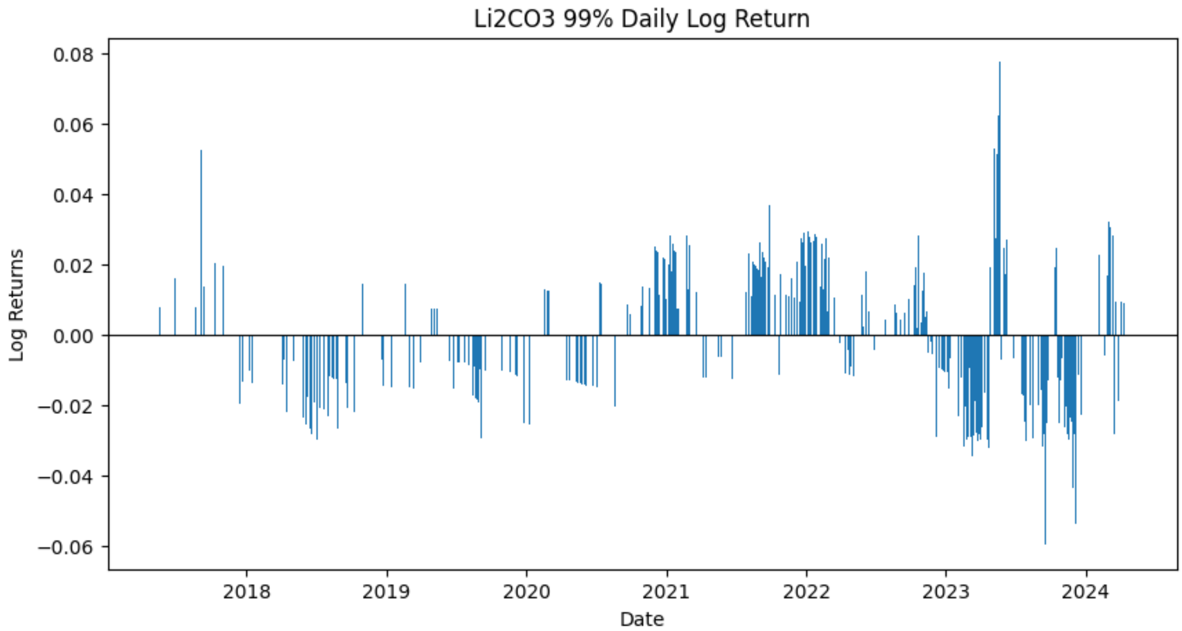
In [10]:
```
summary_statistics_Fridays = Fridays.describe().loc[['mean', 'min', 'max', 'std']]
summary_statistics_Thursdays = Thursdays.describe().loc[['mean', 'min', 'max', 'std'
summary_statistics_Wednesdays = Wednesdays.describe().loc[['mean', 'min', 'max', 'st
summary_statistics_Tuesdays = Tuesdays.describe().loc[['mean', '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)
```

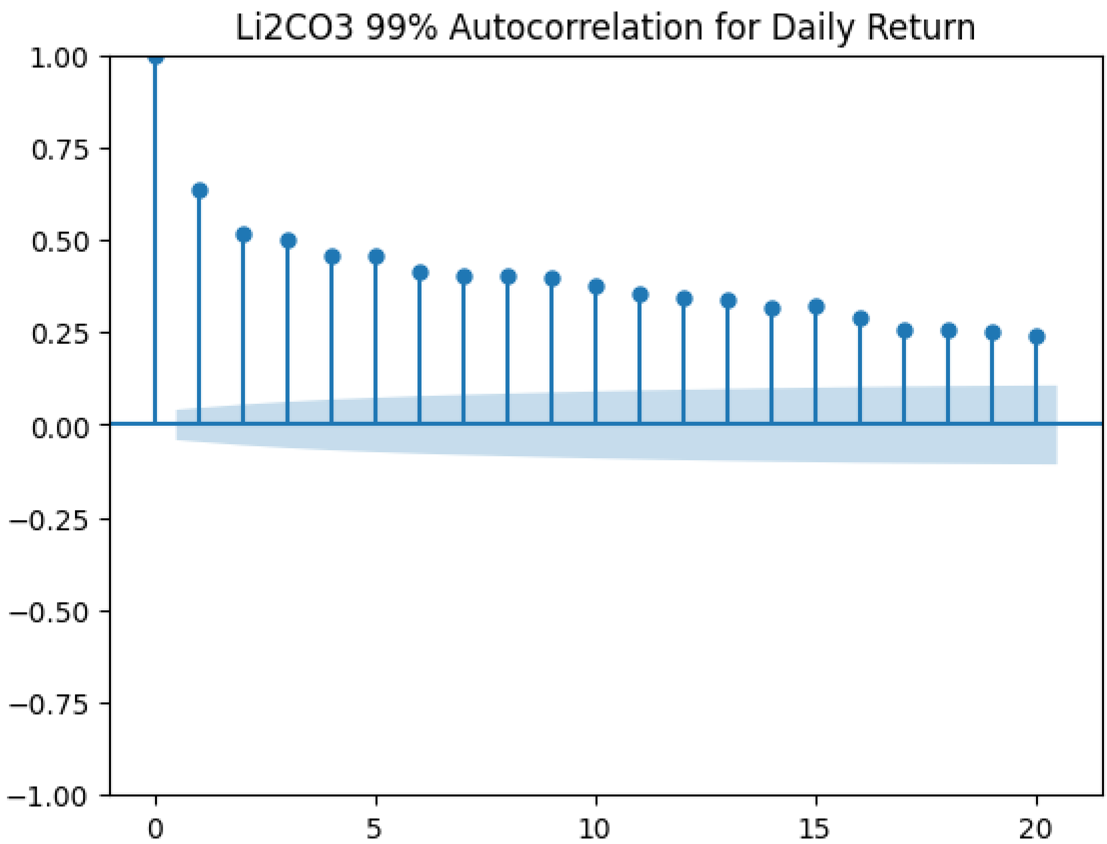
```
Friday to Firday
      Log_Return  Zero_Count_22  Zero_Count_5
mean   -0.000397      15.089783      3.421053
min    -0.152469       0.000000      0.000000
max     0.267022      22.000000      5.000000
std     0.044616       5.608998      1.584804
Thursday to Thursday
      Log_Return  Zero_Count_22  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
```

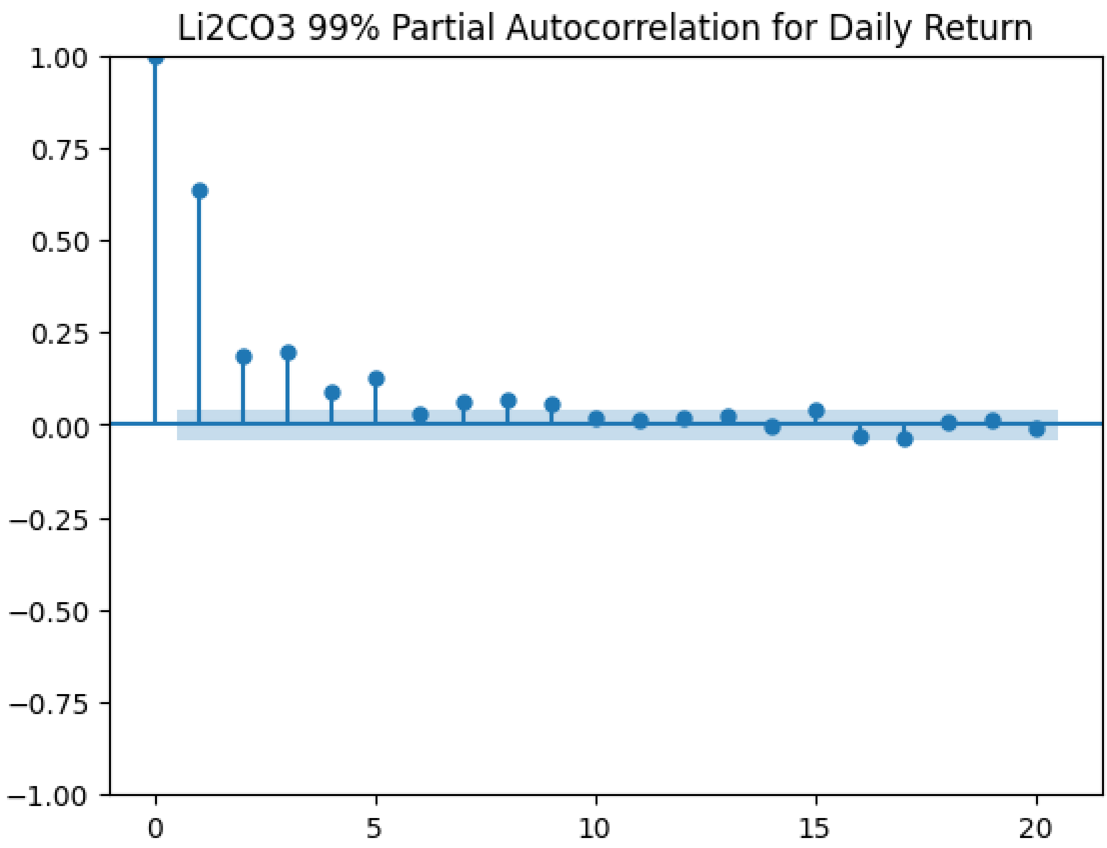
P4. Autocorrelogram and partial autocorrelogram for daily return

In [11]:
```
# 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
acf = plot_acf(li2co3['log_ret'], lags=20, alpha=0.1, title='Li2CO3 99% Autocorrelat
pacf = plot_pacf(li2co3['log_ret'], lags=20, alpha=0.1, method='ywm', title='Li2CO3
```







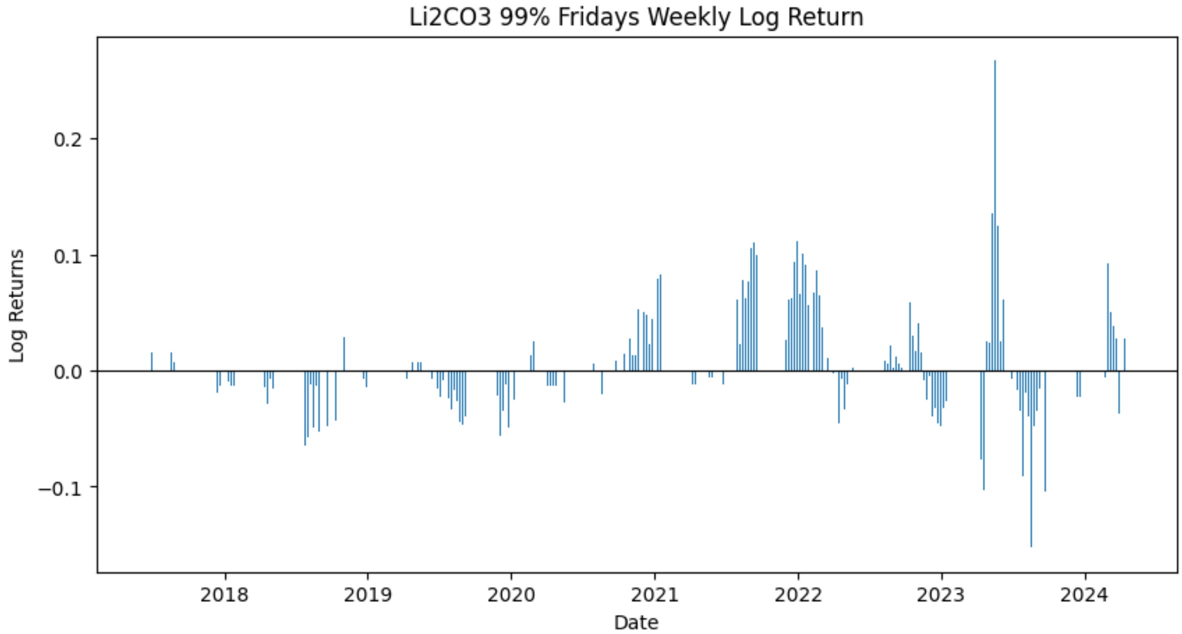
P5. Autocorrelogram and partial autocorrelogram for weekly return

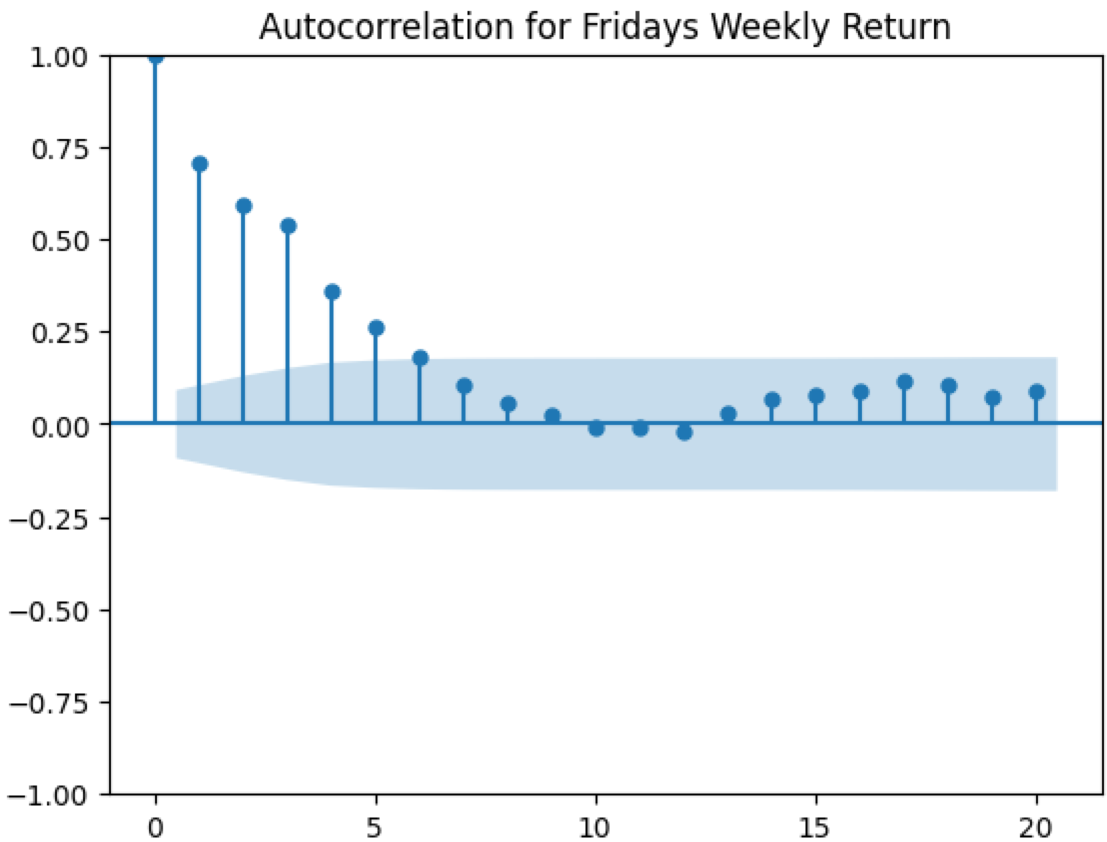
In [12]:

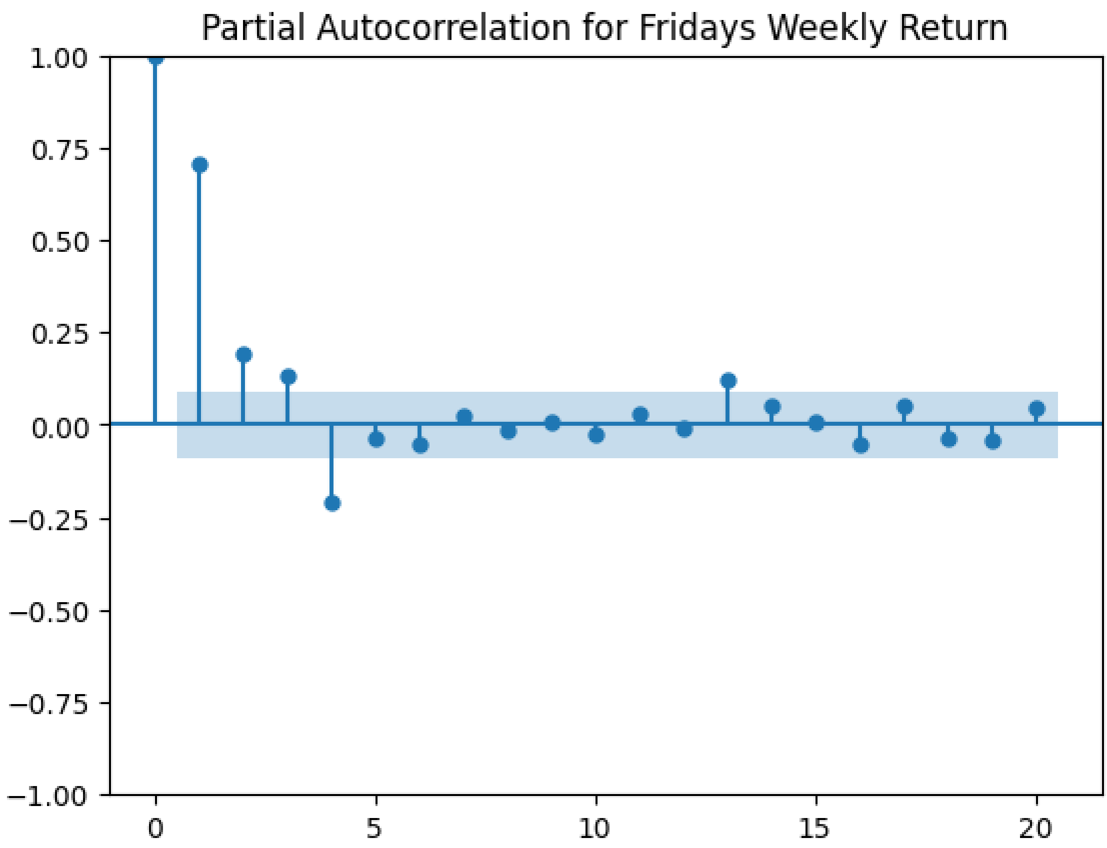
```
def plot_returns_acf_pacf(df, df_name):
    """
    Plots the log returns, autocorrelation, and partial autocorrelation for a given l

    Parameters:
    - df: The DataFrame containing the log returns data.
    - df_name: Name of the DataFrame to use in plot titles to indicate the data sourc
    """
    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'Autocorrelati

    # Plotting the partial autocorrelation
    fig_pacf = plot_pacf(data['Log_Return'], lags=20, alpha=0.1, method='ywm', title
    plt.show()
```

In [13]:

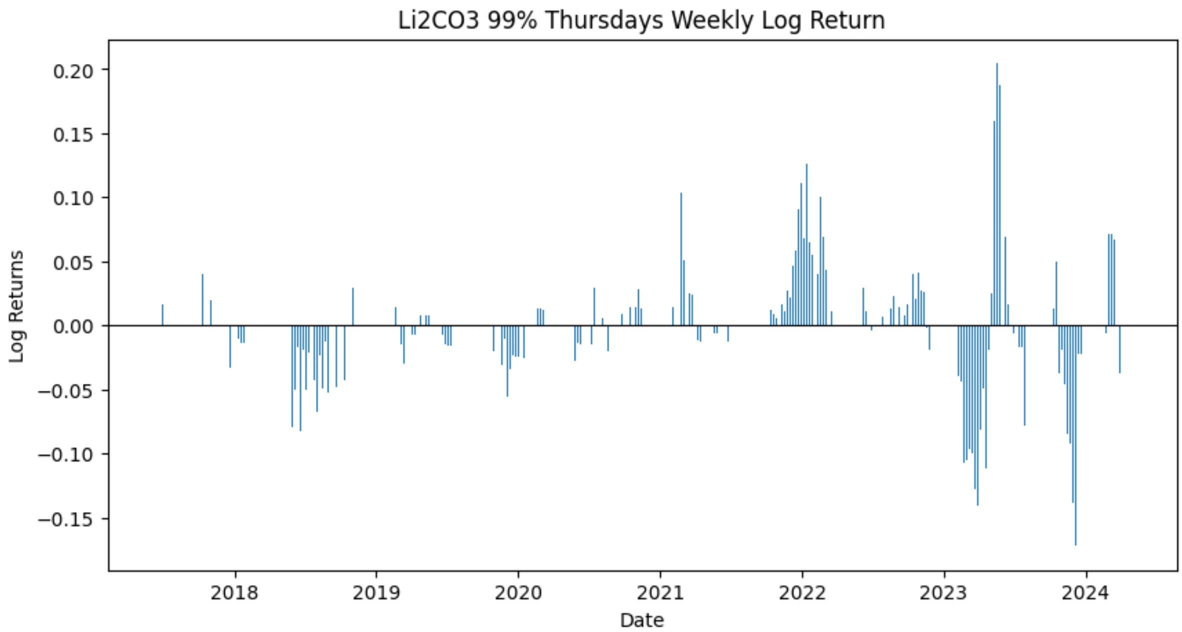
```
# Fridays
plot_returns_acf_pacf(Fridays, 'Fridays')
```

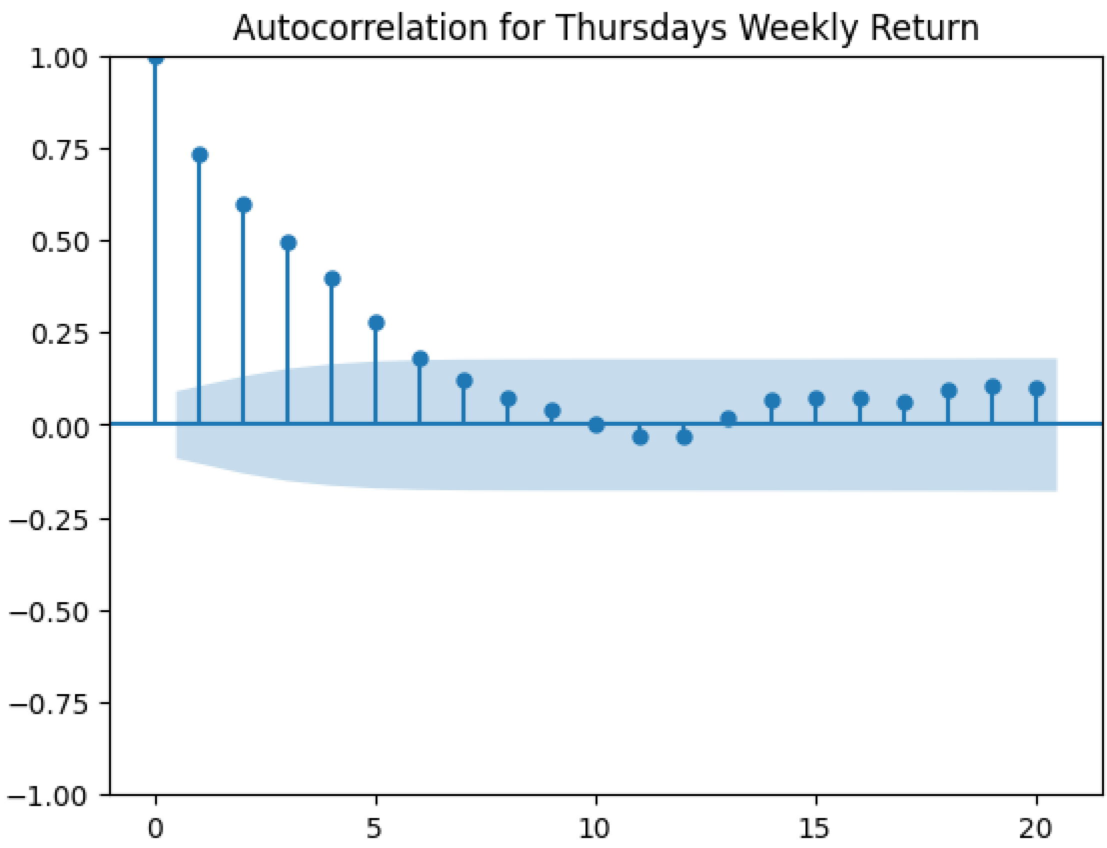

Li2CO3 99% Fridays Weekly Log Return
Log Returns
0.2
0.1
0.0
−0.1
2018
2019
2020
2021
2022
2023
2024
Date

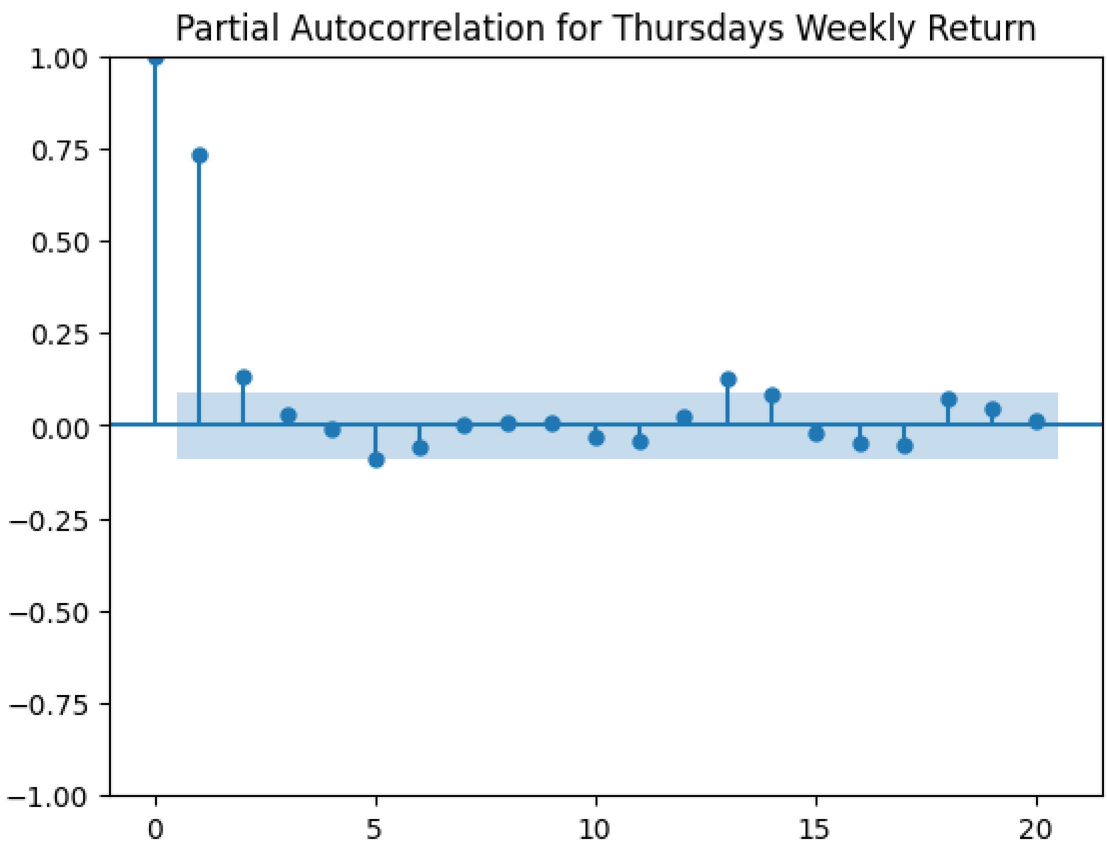

Autocorrelation for Fridays Weekly Return
1.00
0.75
0.50
0.25
0.00
−0.25
−0.50
−0.75
−1.00
0
5
10
15
20



In [14]:
```
# Thursdays
plot_returns_acf_pacf(Thursdays, 'Thursdays')
```

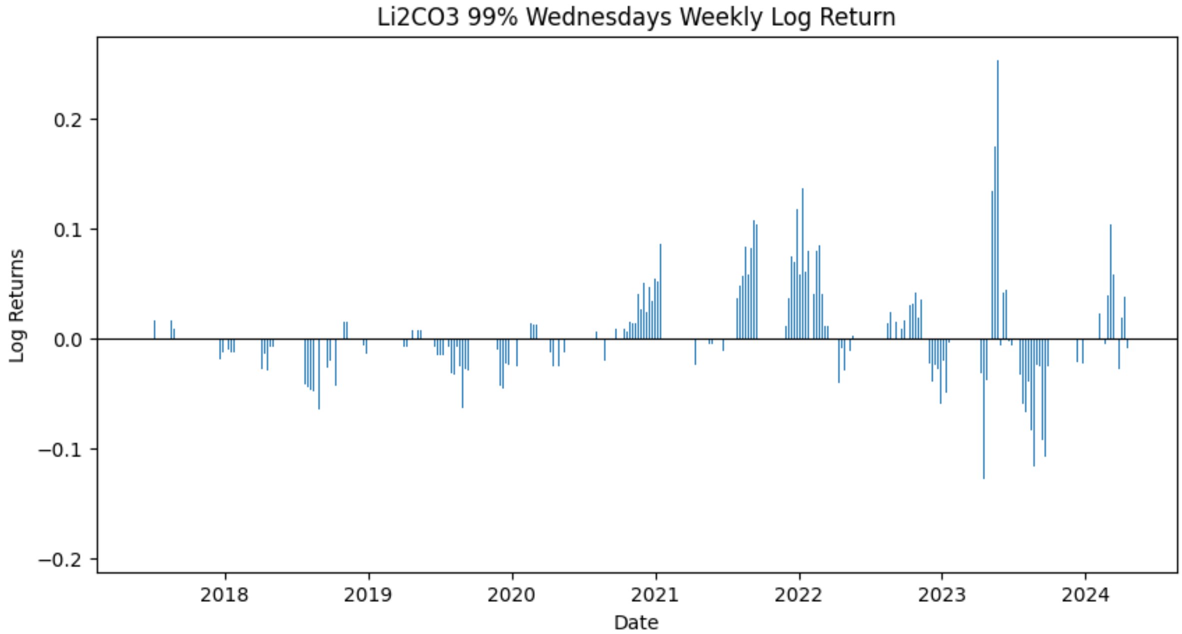


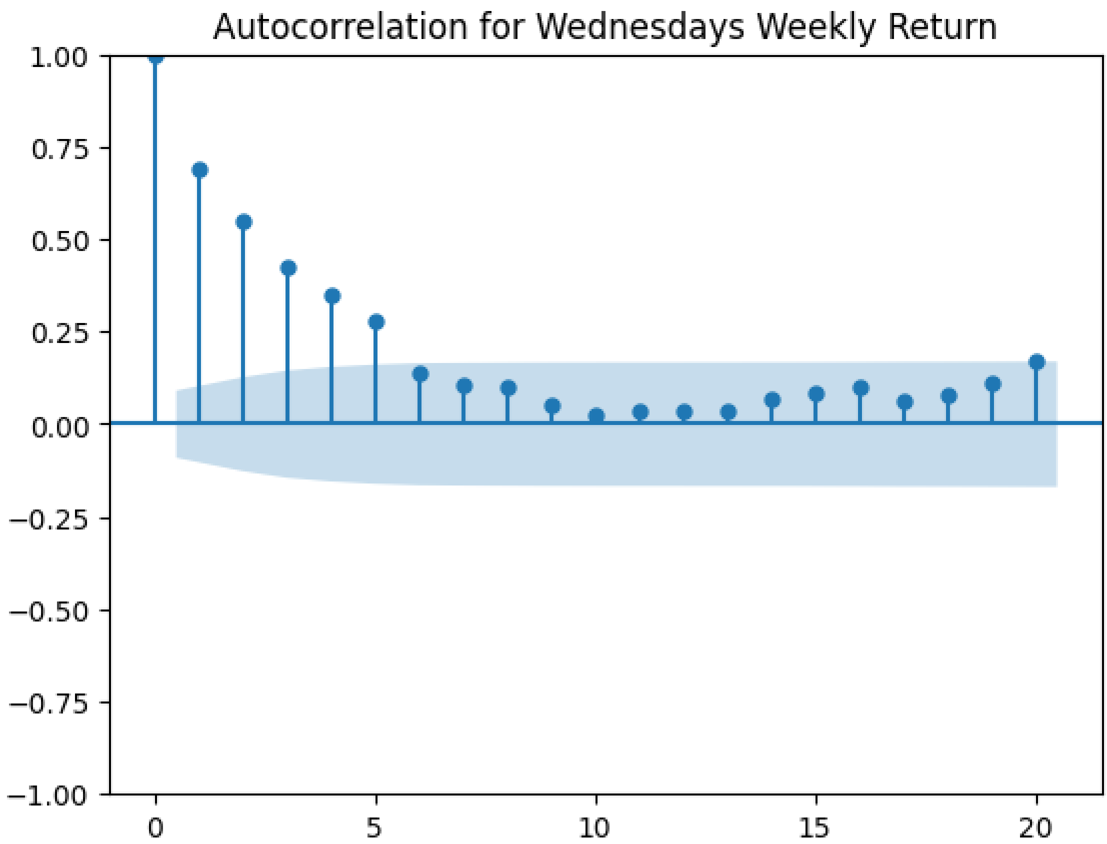


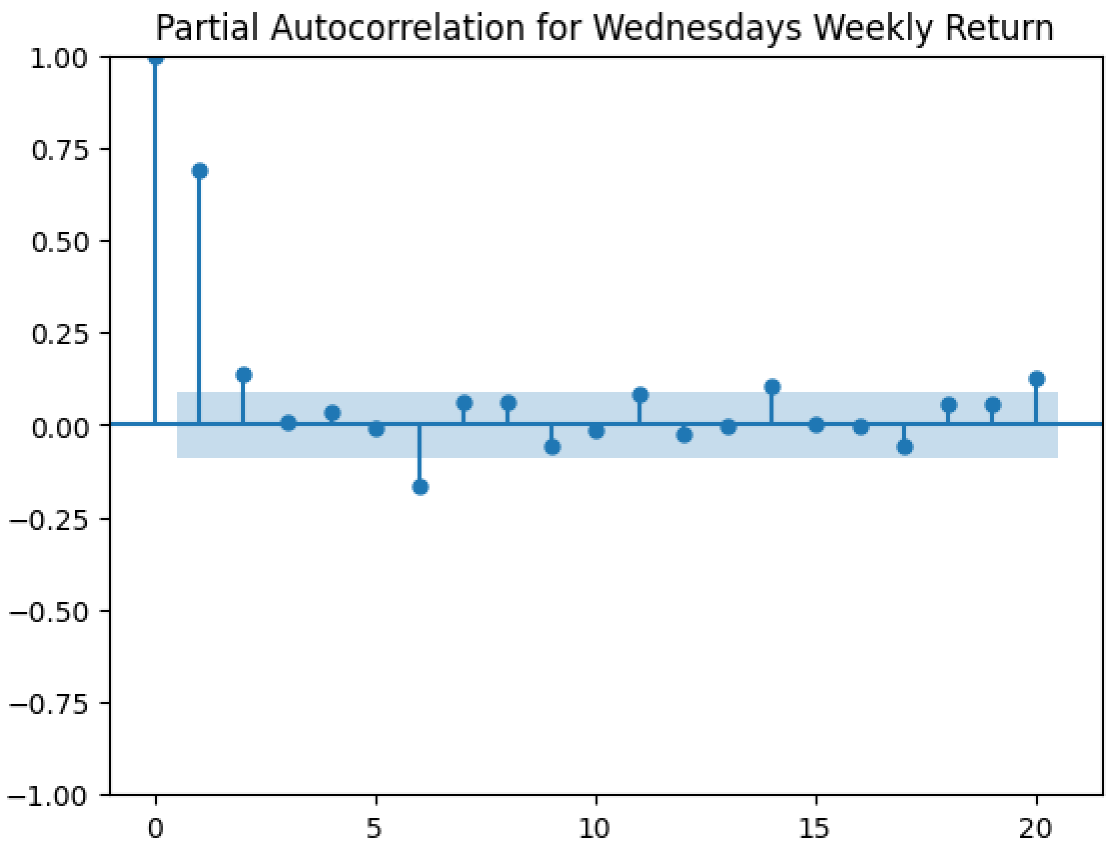


In [15]:

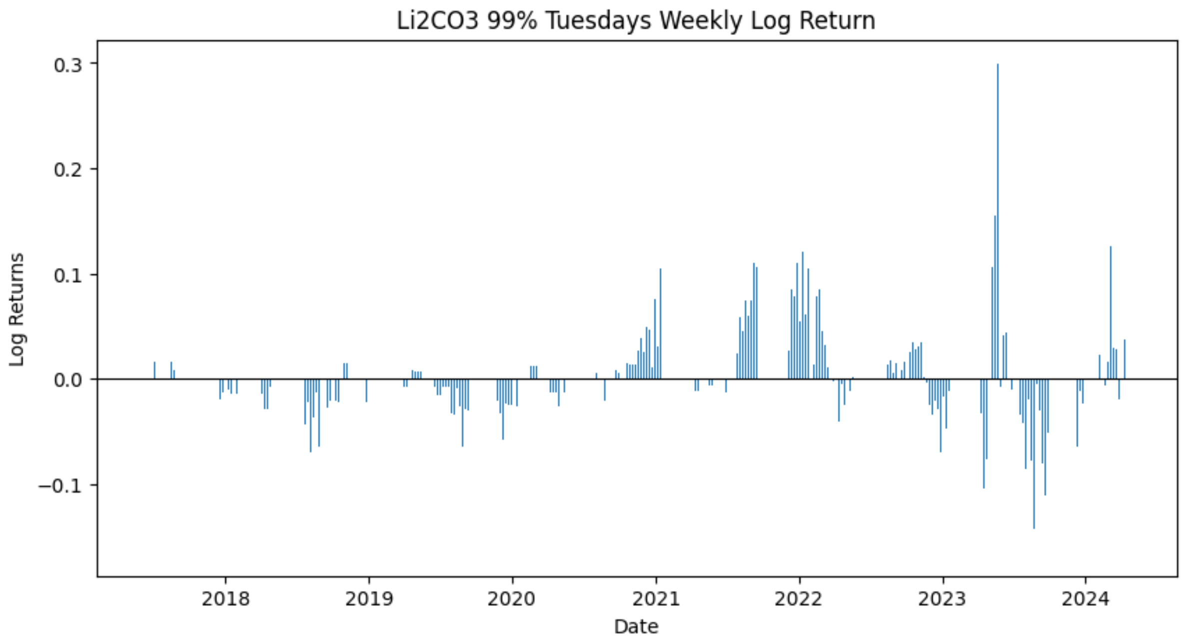
```
# Wednesdays
plot_returns_acf_pacf(Wednesdays, 'Wednesdays')
```

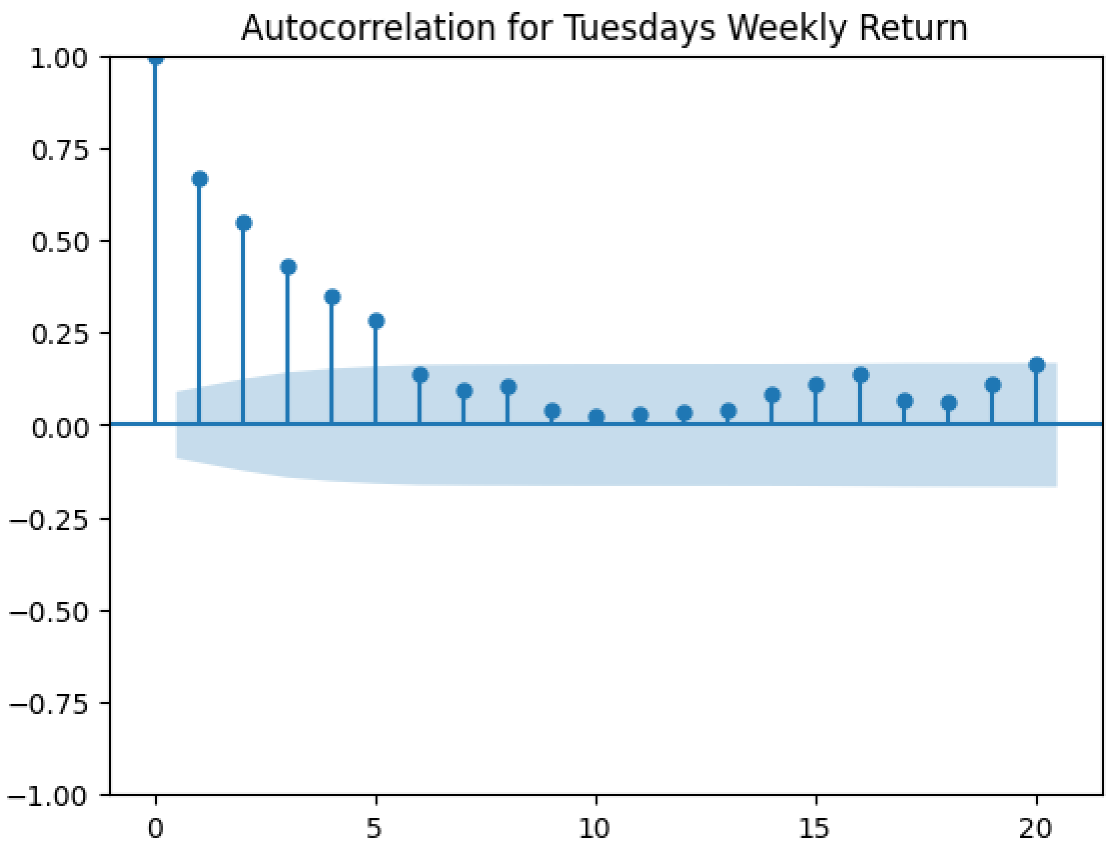

Li2CO3 99% Wednesdays Weekly Log Return
Log Returns
Date
0.2
0.1
0.0
−0.1
−0.2
2018
2019
2020
2021
2022
2023
2024

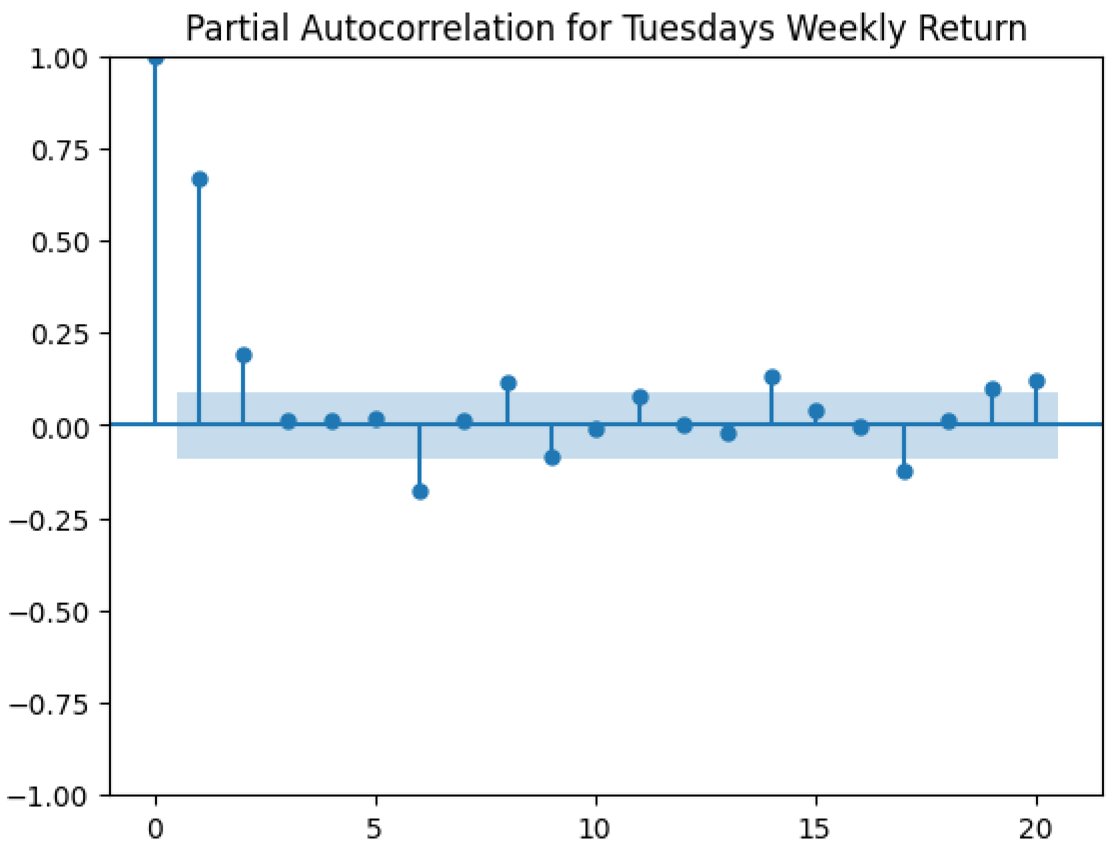

Autocorrelation for Wednesdays Weekly Return
1.00
0.75
0.50
0.25
0.00
−0.25
−0.50
−0.75
−1.00
0
5
10
15
20



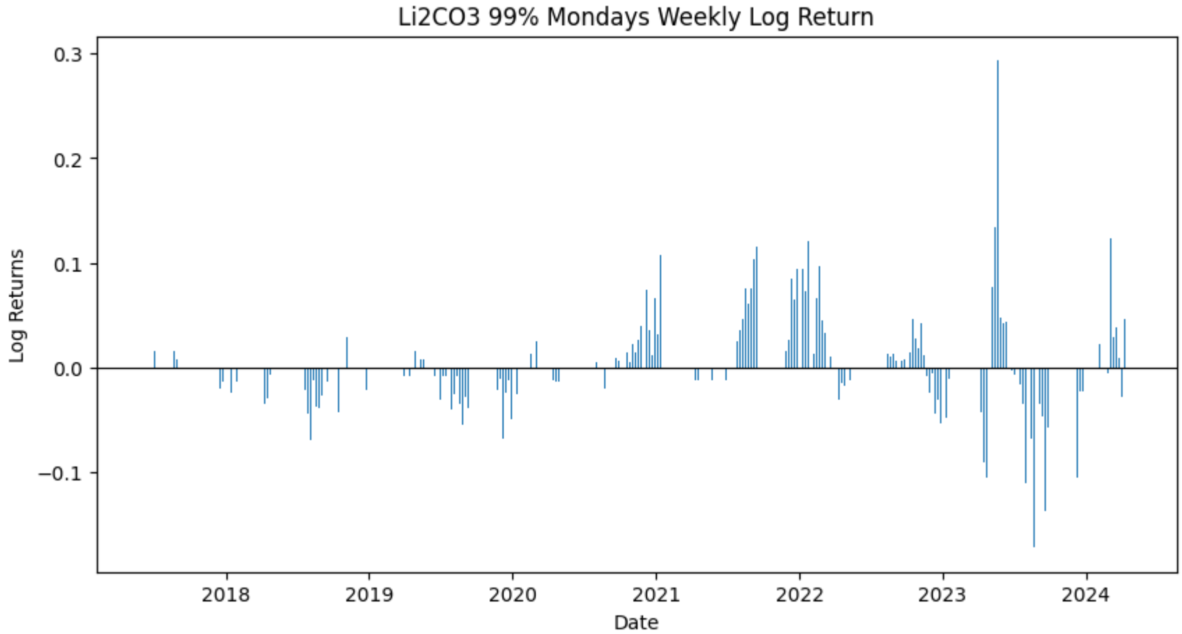
In [16]:
```
# Tuesdays
plot_returns_acf_pacf(Tuesdays, 'Tuesdays')
```

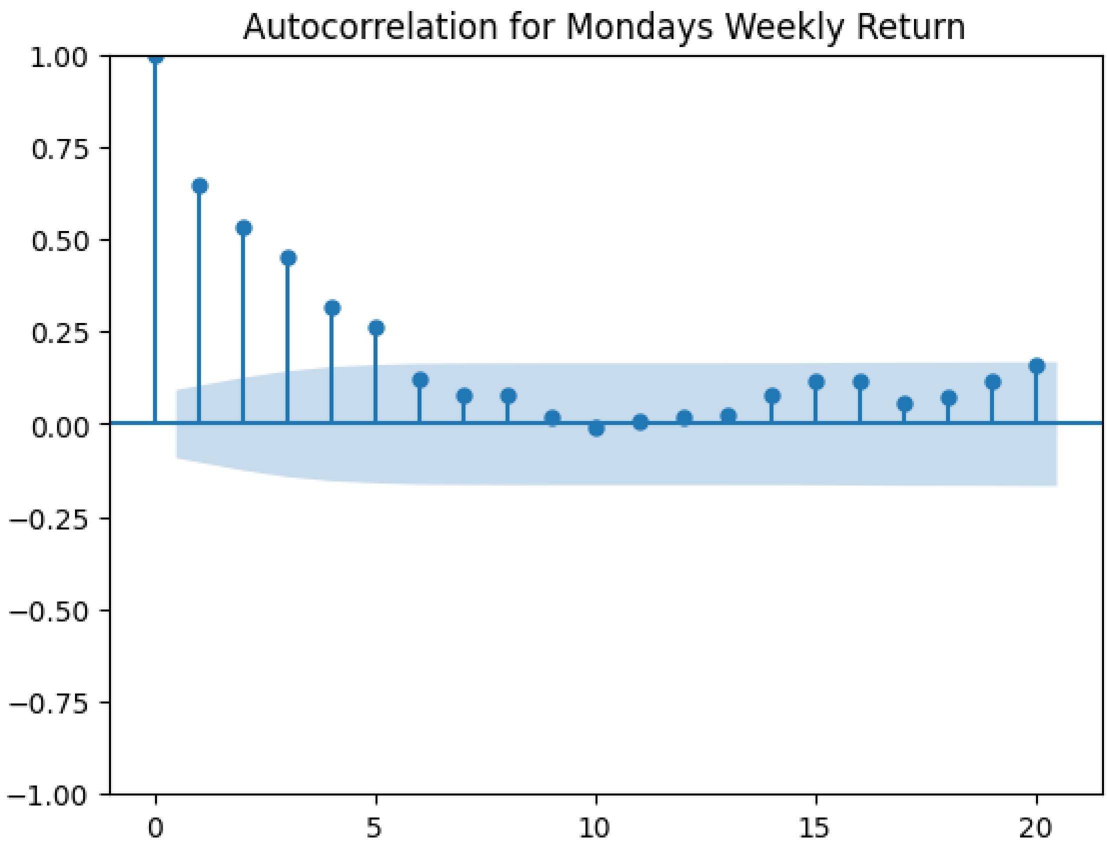


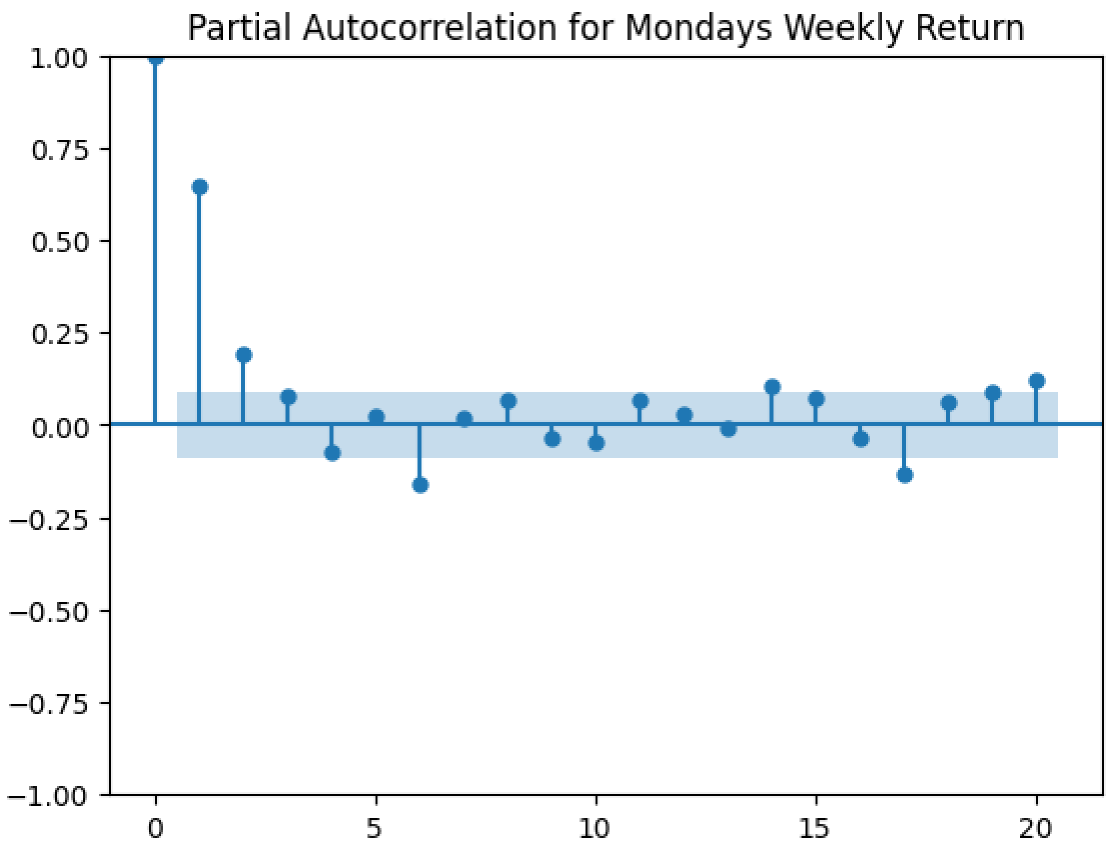




In [17]:
```
# Mondays
plot_returns_acf_pacf(Mondays, 'Mondays')
```


Li2CO3 99% Mondays Weekly Log Return
Log Returns
0.3
0.2
0.1
0.0
−0.1
2018
2019
2020
2021
2022
2023
2024
Date


Autocorrelation for Mondays Weekly Return
1.00
0.75
0.50
0.25
0.00
−0.25
−0.50
−0.75
−1.00
0
5
10
15
20



maxlags = 4 for Fridays

maxlags = 2 for others

P6. Two AR(2) models for daily return

In [18]:
```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.
    """

    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 [19]:
```python
# lags=5 acording to the daily pacf
daily_return_ar2_model = estimate_ar2_model(li2co3, lags=5)
```

In [20]:
```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:                Wed, 22 May 2024   Prob (F-statistic):           2.62e-74
Time:                        07:43:05   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 [21]:
```python
def estimate_ar2_model_with_zero_dummy(df, lags):
    """
```

```python
    Estimate an AR(2) model for log returns with modifications to account for zero du

    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', 'lo
    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 [22]:
```python
# lags=5 acording to the daily pacf
daily_return_ar2_model_with_zero_dummy = estimate_ar2_model_with_zero_dummy(li2co3,
```

In [23]:
```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:                Wed, 22 May 2024   Prob (F-statistic):          5.35e-211
Time:                        07:43:05   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      0.97
5]
--------------------------------------------------------------------------------
-
const               -0.0010      0.001     -1.594      0.111      -0.002       0.00
0
zero_dummy           0.0010      0.001      1.594      0.111      -0.000       0.00
2
log_ret_lag1         0.6606      0.045     14.796      0.000       0.573       0.74
8
log_ret_lag2         0.3622      0.048      7.575      0.000       0.268       0.45
6
log_ret_lag1_zero   -0.6606      0.045    -14.796      0.000      -0.748      -0.57
3
log_ret_lag2_zero   -0.3622      0.048     -7.575      0.000      -0.456      -0.26
8
==============================================================================
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: ValueWarnin
g: 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 [24]:

```
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 [25]:
```
# 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:                Wed, 22 May 2024   Prob (F-statistic):           3.69e-30
Time:                        07:43:06   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 [26]:
```
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',
```

```python
        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 er
    """

    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 [27]:

```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:                Wed, 22 May 2024   Prob (F-statistic):           9.64e-45
Time:                        07:43:06   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 [28]:
```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 er
    """

    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)
```

```python
    # 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 [29]:

```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:                Wed, 22 May 2024   Prob (F-statistic):           2.99e-34
Time:                        07:43:06   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 [30]:
```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 er
    """
```

```python
    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 [31]:

```python
# lags=4 acording to the weekly pacf
ar2_model_with_separate_illiquidity = estimate_ar2_model_separate_illiquidity(Friday

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:                Wed, 22 May 2024   Prob (F-statistic):           6.64e-65
Time:                        07:43:06   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 [32]:

```
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 Separa
    'R-squared': [model1_results.rsquared, model2_results.rsquared, model3_results.r
    'Adj. R-squared': [model1_results.rsquared_adj, model2_results.rsquared_adj, mod
    '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.fva
})

# 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 [33]:
```
# 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'
# Double check the range
print(Fridays[['Zero_Count_5', 'Zero_Fraction_5', 'Zero_Count_22', 'Zero_Fraction_22'
```

```
            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 [34]:
```
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_2

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_

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

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'
```

```
       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 [35]:

```
# 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 [36]:
```
# lags have no meaning, just for uniforming the parameter structure with AR models to
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 + z(t,5)

In [37]:
```
# lags have no meaning, just for uniforming the parameter structure with AR models to
def estimate_model_constant_z5(df, lags):
    data = df.copy()
    X = sm.add_constant(data['Zero_Fraction_5'])  # Adding a constant
    Y = data['Log_Return']
    model = sm.OLS(Y, X).fit()
    return model
```

M3: Constant + z(t,22)

In [38]:
```
# lags have no meaning, just for uniforming the parameter structure with AR models to
def estimate_model_constant_z22(df, lags):
    data = df.copy()
    X = sm.add_constant(data['Zero_Fraction_22'])
    Y = data['Log_Return']
    model = sm.OLS(Y, X).fit()
    return model
```

M4: AR(1)

In [39]:
```
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 [40]:
```
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 z(t,5) interaction

In [41]:
```
def estimate_model_ar1_z5_interaction(df, lags):
    data = df.copy()
    data['Log_Return_Lag1'] = data['Log_Return'].shift(1)
    data['Log_Return_Lag1_Zero5'] = data['Log_Return_Lag1'] * data['Zero_Fraction_5'
    data.dropna(inplace=True)
    X = sm.add_constant(data[['Zero_Fraction_5', 'Log_Return_Lag1', 'Log_Return_Lag1
    Y = data['Log_Return']
    model = sm.OLS(Y, X).fit(cov_type='HAC', cov_kwds={'maxlags': lags})
    return model
```

M7: AR(1) with z(t,22) interaction

In [42]:
```
def estimate_model_ar1_z22_interaction(df, lags):
    data = df.copy()
    data['Log_Return_Lag1'] = data['Log_Return'].shift(1)
    data['Log_Return_Lag1_Zero22'] = data['Log_Return_Lag1'] * data['Zero_Fraction_2
    data.dropna(inplace=True)
    X = sm.add_constant(data[['Zero_Fraction_22', 'Log_Return_Lag1', 'Log_Return_Lag
    Y = data['Log_Return']
    model = sm.OLS(Y, X).fit(cov_type='HAC', cov_kwds={'maxlags': lags})
    return model
```

M8: AR(2) with z(t,5) interaction

In [43]:
```
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 er
    """

    data = df.copy()
    data['Log_Return_Lag1'] = data['Log_Return'].shift(1)
    data['Log_Return_Lag2'] = data['Log_Return'].shift(2)
    data['Log_Return_Lag1_Zero5'] = data['Log_Return_Lag1'] * data['Zero_Fraction_5'
    data['Log_Return_Lag2_Zero5'] = data['Log_Return_Lag2'] * data['Zero_Fraction_5'
    data.dropna(inplace=True)
```

```
    X = data[['Zero_Fraction_5', '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 z(t,22) interaction

In [44]:
```
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 er
    """


    data = df.copy()
    data['Log_Return_Lag1'] = data['Log_Return'].shift(1)
    data['Log_Return_Lag2'] = data['Log_Return'].shift(2)
    data['Log_Return_Lag1_Zero22'] = data['Log_Return_Lag1'] * data['Zero_Fraction_2
    data['Log_Return_Lag2_Zero22'] = data['Log_Return_Lag2'] * data['Zero_Fraction_2
    data.dropna(inplace=True)
    X = data[['Zero_Fraction_22', '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 z(t,5) interaction for first lag and z(t,22) for second lag

In [45]:
```
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 er
    """
    data = df.copy()
    data['Log_Return_Lag1'] = data['Log_Return'].shift(1)
    data['Log_Return_Lag2'] = data['Log_Return'].shift(2)
    data['Log_Return_Lag1_Zero5'] = data['Log_Return_Lag1'] * data['Zero_Fraction_5'
    data['Log_Return_Lag2_Zero22'] = data['Log_Return_Lag2'] * data['Zero_Fraction_2
    data.dropna(inplace=True)
```

```
    X = data[['Zero_Fraction_5', 'Zero_Fraction_22', 'Log_Return_Lag1', 'Log_Return_
              '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 [46]:
```
# 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_ar
```

In [47]:
```
model_features = {
    'Model with Weekly Zero': ['const', 'Zero_Fraction_5'],
    'Model with Monthly Zero': ['const', 'Zero_Fraction_22'],
    'AR(1)': ['const', 'Log_Return_Lag1'],
    'AR(2)': ['const', 'Log_Return_Lag1', 'Log_Return_Lag2'],
    'AR(1) with Weekly Zero Interaction': ['const', 'Zero_Fraction_5', 'Log_Return_La
    'AR(1) with Monthly Zero Interaction': ['const', 'Zero_Fraction_22', 'Log_Return_
    'AR(2) with Weekly Zero Interaction': ['const', 'Zero_Fraction_5', 'Log_Return_La
    'AR(2) with Monthly Zero Interaction': ['const', 'Zero_Fraction_22', 'Log_Return_
    'AR(2) with Separate Weekly and Monthly Zero Interactions': ['const', 'Zero_Frac
}
```

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

In [52]:
```
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 [53]:
```
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.
```

```python
    - 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 R²': 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 [54]:
```python
fridays_models = models_comparison(Fridays, lags=4, df_name='Fridays')
```

Model Name	Adj R²	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	0.00022310648329482152	-1090.2855065012473	-1082.7302018548019	0.7024903727390102	0.5932097065481199	371.1396206856793	3.9439332249207275e-80
Model with Monthly Zero	0.0017293477329133333	-1090.7724979312716	-1083.2171932848262	0.7039880876519994	0.5933416062267783	371.93397974655227	2.653985002866106e-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	0.6355013587926379	-1408.8221972849155	-1393.723991102738	0.022901102384244157	0.13515052832489272	41.33018757691817	5.5652952661566914e-09
AR(1) with Monthly Zero Interaction	0.5238905769995136	-1322.8079844918452	-1307.7097783096676	-0.04179607443882941	0.04479371089268994	29.23358707357481	2.0000815386208517e-06
AR(2) with Weekly Zero Interaction	0.6500952020240849	-1414.5814877806774	-1391.9528410418973	0.1300012242351785	0.09485680317697733	37.33948455886217	3.899907165080985e-08
AR(2) with Monthly Zero Interaction	0.5325177617081115	-1321.587709665845	-1298.959062927065	0.015808465406180014	-0.03426826774217348	22.06561473559412	6.321228758613053e-05
AR(2) with Separate Weekly and Monthly Zero Interactions	0.6585345734280297	-1421.4392679155076	-1395.0391800535974	0.02258347894557238	-0.03011749141680086	22.284807738869215	5.690878424002539e-05

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:                Wed, 22 May 2024   Prob (F-statistic):                nan
Time:                        07:44:36   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 spe
cified.
```

```
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:                     1.072
Date:                Wed, 22 May 2024   Prob (F-statistic):              0.301
Time:                        07:44:36   Log-Likelihood:                 547.14
No. Observations:                 323   AIC:                            -1090.
Df Residuals:                     321   BIC:                            -1083.
Df Model:                           1
Covariance Type:            nonrobust
=================================================================================
                    coef    std err          t      P>|t|      [0.025      0.975]
---------------------------------------------------------------------------------
const             0.0052      0.006      0.873      0.384      -0.006       0.017
Zero_Fraction_5  -0.0081      0.008     -1.035      0.301      -0.024       0.007
==============================================================================
Omnibus:                       49.205   Durbin-Watson:                   0.595
Prob(Omnibus):                  0.000   Jarque-Bera (JB):              395.736
Skew:                           0.245   Prob(JB):                     1.17e-86
Kurtosis:                       8.400   Cond. No.                         4.74
==============================================================================

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

```
Parameters for Model: Model with Monthly Zero:
                            OLS Regression Results
==============================================================================
Dep. Variable:             Log_Return   R-squared:                       0.005
Model:                            OLS   Adj. R-squared:                  0.002
Method:                 Least Squares   F-statistic:                     1.558
Date:                Wed, 22 May 2024   Prob (F-statistic):              0.213
Time:                        07:44:36   Log-Likelihood:                 547.39
No. Observations:                 323   AIC:                            -1091.
Df Residuals:                     321   BIC:                            -1083.
Df Model:                           1
Covariance Type:            nonrobust
====================================================================================
                     coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------------
const              0.0079      0.007      1.114      0.266      -0.006       0.022
Zero_Fraction_22  -0.0122      0.010     -1.248      0.213      -0.031       0.007
==============================================================================
Omnibus:                       48.674   Durbin-Watson:                   0.592
Prob(Omnibus):                  0.000   Jarque-Bera (JB):              395.540
Skew:                           0.223   Prob(JB):                     1.29e-86
Kurtosis:                       8.403   Cond. No.                         5.86
==============================================================================

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

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:                Wed, 22 May 2024   Prob (F-statistic):           1.77e-29
Time:                        07:44:36   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:                Wed, 22 May 2024   Prob (F-statistic):           3.69e-30
Time:                        07:44:36   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:                       0.639
Model:                            OLS   Adj. R-squared:                  0.636
Method:                 Least Squares   F-statistic:                     108.7
Date:                Wed, 22 May 2024   Prob (F-statistic):           1.86e-48
Time:                        07:44:36   Log-Likelihood:                 708.41
No. Observations:                 322   AIC:                            -1409.
Df Residuals:                     318   BIC:                            -1394.
Df Model:                           3
Covariance Type:                  HAC
==========================================================================================
=====
                            coef    std err          z      P>|z|      [0.025
0.975]
------------------------------------------------------------------------------------------
-----
const                    -0.0018      0.007     -0.252      0.801      -0.016
0.012
Zero_Fraction_5           0.0007      0.008      0.084      0.933      -0.015
0.016
Log_Return_Lag1           1.2038      0.078     15.431      0.000       1.051
1.357
Log_Return_Lag1_Zero5    -1.2730      0.132     -9.670      0.000      -1.531      -
1.015
==============================================================================
Omnibus:                       81.406   Durbin-Watson:                   1.954
Prob(Omnibus):                  0.000   Jarque-Bera (JB):              474.307
Skew:                           0.892   Prob(JB):                    1.01e-103
Kurtosis:                       8.672   Cond. No.                         100.
==============================================================================
```

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:                       0.528
Model:                            OLS   Adj. R-squared:                  0.524
Method:                 Least Squares   F-statistic:                     74.27
Date:                Wed, 22 May 2024   Prob (F-statistic):           1.98e-36
Time:                        07:44:36   Log-Likelihood:                 665.40
No. Observations:                 322   AIC:                            -1323.
Df Residuals:                     318   BIC:                            -1308.
Df Model:                           3
Covariance Type:                  HAC
==========================================================================================
======
                             coef    std err          z      P>|z|      [0.025
0.975]
------------------------------------------------------------------------------------------
------
const                      0.0011      0.008      0.142      0.887      -0.014
0.016
Zero_Fraction_22          -0.0029      0.009     -0.325      0.745      -0.020
0.015
Log_Return_Lag1            0.9777      0.073     13.445      0.000       0.835
1.120
Log_Return_Lag1_Zero22    -0.8292      0.142     -5.833      0.000      -1.108
```

```
-0.551
==============================================================================
Omnibus:                       59.656   Durbin-Watson:                   2.083
Prob(Omnibus):                  0.000   Jarque-Bera (JB):              358.772
Skew:                           0.574   Prob(JB):                     1.24e-78
Kurtosis:                       8.042   Cond. No.                         137.
==============================================================================

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:                       0.656
Model:                            OLS   Adj. R-squared:                  0.650
Method:                 Least Squares   F-statistic:                     61.77
Date:                Wed, 22 May 2024   Prob (F-statistic):           9.64e-45
Time:                        07:44:36   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_Fraction_5           0.0003      0.008      0.041      0.967      -0.016
0.017
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    -1.0713      0.292     -3.664      0.000      -1.644      -
0.498
Log_Return_Lag2_Zero5    -0.2424      0.237     -1.022      0.307      -0.707
0.223
==============================================================================
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.                         192.
==============================================================================

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:                       0.540
Model:                            OLS   Adj. R-squared:                  0.533
Method:                 Least Squares   F-statistic:                     43.84
Date:                Wed, 22 May 2024   Prob (F-statistic):           2.99e-34
```

```
Time:                        07:44:36   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_Fraction_22         -0.0024      0.009     -0.266      0.790      -0.020
0.015
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   -0.8715      0.286     -3.042      0.002      -1.433
-0.310
Log_Return_Lag2_Zero22    0.2400      0.313      0.766      0.444      -0.374
0.854
==============================================================================
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.                         255.
==============================================================================

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:                       0.665
Model:                            OLS   Adj. R-squared:                  0.659
Method:                 Least Squares   F-statistic:                     89.29
Date:                Wed, 22 May 2024   Prob (F-statistic):           6.64e-65
Time:                        07:44:36   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_Fraction_5           0.0051      0.012      0.410      0.682      -0.019
0.030
Zero_Fraction_22         -0.0056      0.012     -0.447      0.655      -0.030
0.019
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      -1.4432      0.179     -8.052      0.000     -1.795
-1.092
Log_Return_Lag2_Zero22      0.6302      0.241      2.618      0.009      0.158
1.102
==============================================================================
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.                         215.
==============================================================================

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

In [55]:

```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:                       0.528
Model:                            OLS   Adj. R-squared:                  0.524
Method:                 Least Squares   F-statistic:                     74.27
Date:                Wed, 22 May 2024   Prob (F-statistic):           1.98e-36
Time:                        07:44:48   Log-Likelihood:                 665.40
No. Observations:                 322   AIC:                            -1323.
Df Residuals:                     318   BIC:                            -1308.
Df Model:                           3
Covariance Type:                  HAC
==========================================================================================
                             coef    std err          z      P>|z|      [0.025
0.975]
------------------------------------------------------------------------------------------
const                      0.0011      0.008      0.142      0.887     -0.014
0.016
Zero_Fraction_22          -0.0029      0.009     -0.325      0.745     -0.020
0.015
Log_Return_Lag1            0.9777      0.073     13.445      0.000      0.835
1.120
Log_Return_Lag1_Zero22    -0.8292      0.142     -5.833      0.000     -1.108
-0.551
==============================================================================
Omnibus:                       59.656   Durbin-Watson:                   2.083
Prob(Omnibus):                  0.000   Jarque-Bera (JB):              358.772
Skew:                           0.574   Prob(JB):                     1.24e-78
Kurtosis:                       8.042   Cond. No.                         137.
==============================================================================

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 Separate Weekly and Monthly Zero Interactions**, which has the highest value (0.6585), is the winner in this criteria.

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 Separate Weekly and Monthly Zero Interactions** also wins in this criteria, suggesting it provides a strong balance of model fit and complexity.

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 monthly zero interaction** is closest to zero, where the Rho_1 of model **AR(2) with Separate Weekly and Monthly Zero Interactions** is second closest to zero. the Rho_2 of model **AR(1)** is closest to zero, where the Rho_2 of model **AR(2) with Separate Weekly and Monthly Zero Interactions** is still second closest to zero. But both differences to first closest model are very tiny, so **AR(2) with Separate Weekly and Monthly Zero Interactions** can still be treated as the winner 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 second highest p-value (around $5.7 * 10^{-5}$), but is still too low.

Best Model Selection

The model **AR(2) with Separate Weekly and Monthly Zero Interactions** stands out as the best model in three criterias. It has the highest adjusted R-squared, indicating that it explains the variance in the returns most effectively. It also boasts the lowest AIC and BIC scores, suggesting a superior model fit when adjusted for the number of parameters.

Its residuals exhibit relatively low autocorrelation (Rho_1 and Rho_2 are close to zero), and the Ljung-Box test results are quite satisfactory, its p-value close to zero, which might seem counterintuitive but actually shows that the model captures the autocorrelation structure very well.

Conclusion

The model **AR(2) with Separate Weekly and Monthly Zero Interactions** balances complexity with performance effectively and manages residuals better than simpler models or those considering fewer interaction terms. It should be chosen for further forecasting assuming these results hold consistently across different dataframes (Eg. Wednesdays). This model's 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 [56]:
```
# Thursdays
thursdays_models = models_comparison(Thursdays, lags=2, df_name='Thursdays')
```

```
+----------------------------------------------------------+------------------------
+---------------------+---------------------+-------------------------+-------------
----------+---------------------------+------------------------+
|                         Model Name                       |         Adj R²
|          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    |   0.599727853
4251737   |      383.393733423569      | 8.749828086232154e-83  |
|                  Model with Weekly Zero                  | -0.0016761122890407432
| -1128.0799405331086 | -1120.4696705952756 |   0.7332424684202226    |   0.597721016
2813626   |     382.1233040403662      | 1.6487421833136732e-82 |
|                  Model with Monthly Zero                 | 0.0006466918079647277
| -1128.8507151141202 | -1121.2404451762873 |   0.7333484481909822    |   0.599234481
9360934   |     382.71996220974745     | 1.2244106871190914e-82 |
|                          AR(1)                           |   0.5382307541461196
| -1379.9895675709663 | -1372.3853308202122 |  -0.09593281484946242   |  0.0477102574
2450191   |     4.922570199686606      |  0.17755553347138409   |
|                          AR(2)                           |   0.5447343740533538
| -1378.5015778982604 | -1367.1042999348788 | -0.0038355974340777074 | -0.0197961508
45353307  |     0.5311140933125282     |   0.9120044970665508   |
|              AR(1) with Weekly Zero Interaction          |   0.6702479679711655
| -1489.4636245524944 | -1474.2551510509861 |  0.06042168634871326    |  0.1332308420
7465238   |     14.58159128204214      | 0.0022114762282063453  |
|             AR(1) with Monthly Zero Interaction          |   0.5603364369719341
| -1394.245234857457  | -1379.0367613559488 | -0.004745499387938068  |  0.0681003443
1383816   |     3.302593772187429      |   0.3472818030616484   |
|              AR(2) with Weekly Zero Interaction          |   0.6791693614306187
|  -1491.03240310553  | -1468.2378471787667 |   0.1657314523784008    |  0.0985500691
0051366   |     18.459474942528395     | 0.0003535772329036067  |
|             AR(2) with Monthly Zero Interaction          |   0.5644230234167986
| -1390.1322113036229 | -1367.3376553768596 |  0.02074524262366576    | -0.0017621778
698386106 |     1.2476499669211416     |   0.7416000090119834   |
| AR(2) with Separate Weekly and Monthly Zero Interactions |   0.6901713105605536
| -1501.5674665510562 | -1474.9738179698325 |   0.0579836277267137    |  -0.027446020
14065312  |     3.7344531529929657     |  0.29160293766075845   |
+----------------------------------------------------------+------------------------
+---------------------+---------------------+-------------------------+-------------
----------+---------------------------+------------------------+
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:                Wed, 22 May 2024   Prob (F-statistic):                nan
Time:                        07:44:53   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:                       0.001
Model:                            OLS   Adj. R-squared:                 -0.002
Method:                 Least Squares   F-statistic:                    0.4461
Date:                Wed, 22 May 2024   Prob (F-statistic):              0.505
Time:                        07:44:53   Log-Likelihood:                 566.04
No. Observations:                 332   AIC:                            -1128.
Df Residuals:                     330   BIC:                            -1120.
Df Model:                           1
Covariance Type:            nonrobust
===================================================================================
                      coef    std err          t      P>|t|      [0.025      0.975]
-----------------------------------------------------------------------------------
const               0.0028      0.006      0.487      0.626      -0.009       0.014
Zero_Fraction_5    -0.0051      0.008     -0.668      0.505      -0.020       0.010
==============================================================================
Omnibus:                       39.229   Durbin-Watson:                   0.533
Prob(Omnibus):                  0.000   Jarque-Bera (JB):              227.504
Skew:                           0.186   Prob(JB):                     3.96e-50
Kurtosis:                       7.038   Cond. No.                         4.77
==============================================================================
```

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:                  0.001
Method:                 Least Squares   F-statistic:                     1.214
Date:                Wed, 22 May 2024   Prob (F-statistic):              0.271
Time:                        07:44:53   Log-Likelihood:                 566.43
No. Observations:                 332   AIC:                            -1129.
Df Residuals:                     330   BIC:                            -1121.
Df Model:                           1
Covariance Type:            nonrobust
====================================================================================
                       coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------------
const                0.0066      0.007      0.936      0.350      -0.007       0.020
Zero_Fraction_22    -0.0106      0.010     -1.102      0.271      -0.029       0.008
==============================================================================
Omnibus:                       37.610   Durbin-Watson:                   0.533
Prob(Omnibus):                  0.000   Jarque-Bera (JB):              223.194
Skew:                           0.098   Prob(JB):                     3.42e-49
Kurtosis:                       7.012   Cond. No.                         5.92
==============================================================================
```

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:                Wed, 22 May 2024   Prob (F-statistic):           1.78e-26
Time:                        07:44:53   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:                Wed, 22 May 2024   Prob (F-statistic):           7.71e-26
Time:                        07:44:53   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:                       0.673
Model:                            OLS   Adj. R-squared:                  0.670
Method:                 Least Squares   F-statistic:                     132.8
Date:                Wed, 22 May 2024   Prob (F-statistic):           2.76e-56
Time:                        07:44:53   Log-Likelihood:                 748.73
No. Observations:                 331   AIC:                            -1489.
Df Residuals:                     327   BIC:                            -1474.
Df Model:                           3
Covariance Type:                  HAC
=========================================================================================
=====
                           coef    std err          z      P>|z|      [0.025
0.975]
-----------------------------------------------------------------------------------------
-----
const                   -0.0023      0.006     -0.411      0.681      -0.013
0.009
Zero_Fraction_5          0.0016      0.006      0.257      0.797      -0.011
0.014
Log_Return_Lag1          1.1746      0.062     19.075      0.000       1.054
1.295
Log_Return_Lag1_Zero5   -1.1685      0.093    -12.622      0.000      -1.350      -
0.987
==============================================================================
Omnibus:                       63.863   Durbin-Watson:                   1.879
Prob(Omnibus):                  0.000   Jarque-Bera (JB):              402.043
Skew:                           0.599   Prob(JB):                     4.98e-88
Kurtosis:                       8.265   Cond. No.                         95.4
==============================================================================

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:                       0.564
Model:                            OLS   Adj. R-squared:                  0.560
Method:                 Least Squares   F-statistic:                     54.89
Date:                Wed, 22 May 2024   Prob (F-statistic):           9.21e-29
Time:                        07:44:53   Log-Likelihood:                 701.12
No. Observations:                 331   AIC:                            -1394.
Df Residuals:                     327   BIC:                            -1379.
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      -0.013
0.014
Zero_Fraction_22         -0.0016      0.008     -0.194      0.846      -0.018
0.015
Log_Return_Lag1           0.9773      0.086     11.299      0.000       0.808
1.147
Log_Return_Lag1_Zero22   -0.7416      0.175     -4.229      0.000      -1.085
```

```
-0.398
==============================================================================
Omnibus:                       45.078   Durbin-Watson:                   2.009
Prob(Omnibus):                  0.000   Jarque-Bera (JB):              296.758
Skew:                           0.255   Prob(JB):                     3.63e-65
Kurtosis:                       7.611   Cond. No.                         138.
==============================================================================

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:                       0.684
Model:                            OLS   Adj. R-squared:                  0.679
Method:                 Least Squares   F-statistic:                     74.93
Date:                Wed, 22 May 2024   Prob (F-statistic):           5.41e-52
Time:                        07:44:53   Log-Likelihood:                 751.52
No. Observations:                 330   AIC:                            -1491.
Df Residuals:                     324   BIC:                            -1468.
Df Model:                           5
Covariance Type:                  HAC
==============================================================================
=====
                            coef    std err          z      P>|z|      [0.025
0.975]
------------------------------------------------------------------------------
-----
const                    -0.0022      0.006     -0.378      0.705      -0.014
0.009
Zero_Fraction_5           0.0015      0.006      0.230      0.818      -0.011
0.014
Log_Return_Lag1           0.9734      0.191      5.083      0.000       0.598
1.349
Log_Return_Lag2           0.2702      0.214      1.261      0.207      -0.150
0.690
Log_Return_Lag1_Zero5    -0.9833      0.252     -3.904      0.000      -1.477      -
0.490
Log_Return_Lag2_Zero5    -0.2407      0.266     -0.905      0.365      -0.762
0.281
==============================================================================
Omnibus:                       75.221   Durbin-Watson:                   1.667
Prob(Omnibus):                  0.000   Jarque-Bera (JB):              494.355
Skew:                           0.745   Prob(JB):                    4.49e-108
Kurtosis:                       8.808   Cond. No.                         210.
==============================================================================

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:                       0.571
Model:                            OLS   Adj. R-squared:                  0.564
Method:                 Least Squares   F-statistic:                     35.65
Date:                Wed, 22 May 2024   Prob (F-statistic):           4.90e-29
```

```
Time:                        07:44:53   Log-Likelihood:                 701.07
No. Observations:                 330   AIC:                            -1390.
Df Residuals:                     324   BIC:                            -1367.
Df Model:                           5
Covariance Type:                  HAC
==========================================================================================
                            coef    std err          z      P>|z|      [0.025
0.975]
------------------------------------------------------------------------------------------
const                     0.0001      0.007      0.020      0.984      -0.013
0.014
Zero_Fraction_22         -0.0013      0.008     -0.160      0.872      -0.017
0.015
Log_Return_Lag1           1.0371      0.220      4.723      0.000       0.607
1.467
Log_Return_Lag2          -0.1094      0.224     -0.488      0.626      -0.549
0.330
Log_Return_Lag1_Zero22   -0.9467      0.336     -2.821      0.005      -1.605
-0.289
Log_Return_Lag2_Zero22    0.4184      0.349      1.199      0.231      -0.266
1.102
==============================================================================
Omnibus:                       45.699   Durbin-Watson:                   1.957
Prob(Omnibus):                  0.000   Jarque-Bera (JB):              325.016
Skew:                           0.220   Prob(JB):                     2.65e-71
Kurtosis:                       7.842   Cond. No.                         272.
==============================================================================

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:                       0.696
Model:                            OLS   Adj. R-squared:                  0.690
Method:                 Least Squares   F-statistic:                     91.94
Date:                Wed, 22 May 2024   Prob (F-statistic):           7.16e-67
Time:                        07:44:53   Log-Likelihood:                 757.78
No. Observations:                 330   AIC:                            -1502.
Df Residuals:                     323   BIC:                            -1475.
Df Model:                           6
Covariance Type:                  HAC
==========================================================================================
                            coef    std err          z      P>|z|      [0.025
0.975]
------------------------------------------------------------------------------------------
const                  8.749e-05      0.006      0.015      0.988      -0.011
0.011
Zero_Fraction_5           0.0088      0.009      1.002      0.316      -0.008
0.026
Zero_Fraction_22         -0.0100      0.010     -1.016      0.310      -0.029
0.009
Log_Return_Lag1           1.2065      0.110     10.978      0.000       0.991
1.422
Log_Return_Lag2          -0.1429      0.134     -1.067      0.286      -0.405
0.120
```

```
Log_Return_Lag1_Zero5      -1.3471      0.113    -11.937      0.000     -1.568
-1.126
Log_Return_Lag2_Zero22      0.6594      0.228      2.888      0.004      0.212
1.107
==============================================================================
Omnibus:                       67.071   Durbin-Watson:                   1.882
Prob(Omnibus):                  0.000   Jarque-Bera (JB):              443.159
Skew:                           0.630   Prob(JB):                     5.88e-97
Kurtosis:                       8.536   Cond. No.                         213.
==============================================================================

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

In [57]:

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

Model Name	Adj R²	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	-0.00038518726771097356	-1118.6272949132772	-1111.0170249754442	0.6863735826133497	0.5447237933417762	317.7071084007135	1.4624872089973318e-68
Model with Monthly Zero	0.004448034524227262	-1120.2351940088392	-1112.6249240710063	0.6873530117395088	0.547434706229909	319.7949683853688	5.165740037277218e-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	0.6577493878392358	-1467.2983983490603	-1452.089924847552	0.10739501408397778	0.18625157946673088	18.971940414699226	0.0002770756014247613
AR(1) with Monthly Zero Interaction	0.48422310065644525	-1331.5450196180182	-1316.33654611651	-0.053809630481734115	0.0971912170205857	4.289059496756435	0.2318952254521099
AR(2) with Weekly Zero Interaction	0.6630136335082346	-1464.9980624611542	-1442.203506534391	0.1949892305422475	0.136346997670443	21.360182422532972	8.861414247765491e-05
AR(2) with Monthly Zero Interaction	0.48842531530419786	-1327.2391955511114	-1304.4446396243482	0.025780959023621305	0.022216907739266007	0.40076012111662723	0.9400854268690306
AR(2) with Separate Weekly and Monthly Zero Interactions	0.6742286470353349	-1475.1875703101994	-1448.5939217289756	0.06856758107736555	0.02536327479550779	2.077970566111365	0.5563823081180155

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:                Wed, 22 May 2024   Prob (F-statistic):                nan
Time:                        07:45:03   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 spe
cified.
```

```
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:                    0.8726
Date:                Wed, 22 May 2024   Prob (F-statistic):              0.351
Time:                        07:45:03   Log-Likelihood:                 561.31
No. Observations:                 332   AIC:                            -1119.
Df Residuals:                     330   BIC:                            -1111.
Df Model:                           1
Covariance Type:            nonrobust
===================================================================================
                      coef    std err          t      P>|t|      [0.025      0.975]
-----------------------------------------------------------------------------------
const               0.0050      0.006      0.827      0.409      -0.007       0.017
Zero_Fraction_5    -0.0075      0.008     -0.934      0.351      -0.023       0.008
==============================================================================
Omnibus:                       50.161   Durbin-Watson:                   0.627
Prob(Omnibus):                  0.000   Jarque-Bera (JB):              415.095
Skew:                           0.223   Prob(JB):                     7.30e-91
Kurtosis:                       8.460   Cond. No.                         4.89
==============================================================================

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

```
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:                     2.479
Date:                Wed, 22 May 2024   Prob (F-statistic):              0.116
Time:                        07:45:03   Log-Likelihood:                 562.12
No. Observations:                 332   AIC:                            -1120.
Df Residuals:                     330   BIC:                            -1113.
Df Model:                           1
Covariance Type:            nonrobust
====================================================================================
                       coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------------
const                0.0104      0.007      1.456      0.146      -0.004       0.024
Zero_Fraction_22    -0.0154      0.010     -1.574      0.116      -0.035       0.004
==============================================================================
Omnibus:                       47.761   Durbin-Watson:                   0.625
Prob(Omnibus):                  0.000   Jarque-Bera (JB):              408.682
Skew:                           0.101   Prob(JB):                     1.80e-89
Kurtosis:                       8.432   Cond. No.                         5.93
==============================================================================

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

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:                Wed, 22 May 2024   Prob (F-statistic):           4.63e-19
Time:                        07:45:03   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:                Wed, 22 May 2024   Prob (F-statistic):           1.14e-19
Time:                        07:45:03   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:                       0.661
Model:                            OLS   Adj. R-squared:                  0.658
Method:                 Least Squares   F-statistic:                     99.47
Date:                Wed, 22 May 2024   Prob (F-statistic):           9.08e-46
Time:                        07:45:03   Log-Likelihood:                 737.65
No. Observations:                 331   AIC:                            -1467.
Df Residuals:                     327   BIC:                            -1452.
Df Model:                           3
Covariance Type:                  HAC
==================================================================================
=====
                           coef    std err          z      P>|z|      [0.025
0.975]
----------------------------------------------------------------------------------
-----
const                   -0.0034      0.006     -0.597      0.550      -0.014
0.008
Zero_Fraction_5          0.0024      0.006      0.380      0.704      -0.010
0.015
Log_Return_Lag1          1.2643      0.085     14.822      0.000       1.097
1.431
Log_Return_Lag1_Zero5   -1.4323      0.151     -9.457      0.000      -1.729      -
1.135
==============================================================================
Omnibus:                       92.573   Durbin-Watson:                   1.785
Prob(Omnibus):                  0.000   Jarque-Bera (JB):             1079.314
Skew:                           0.777   Prob(JB):                    4.26e-235
Kurtosis:                      11.709   Cond. No.                         98.8
==============================================================================

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:                       0.489
Model:                            OLS   Adj. R-squared:                  0.484
Method:                 Least Squares   F-statistic:                     30.69
Date:                Wed, 22 May 2024   Prob (F-statistic):           1.67e-17
Time:                        07:45:03   Log-Likelihood:                 669.77
No. Observations:                 331   AIC:                            -1332.
Df Residuals:                     327   BIC:                            -1316.
Df Model:                           3
Covariance Type:                  HAC
==================================================================================
======
                            coef    std err          z      P>|z|      [0.025
0.975]
----------------------------------------------------------------------------------
------
const                     0.0017      0.008      0.206      0.837      -0.014
0.017
Zero_Fraction_22         -0.0034      0.010     -0.350      0.727      -0.022
0.016
Log_Return_Lag1           0.8666      0.114      7.574      0.000       0.642
1.091
Log_Return_Lag1_Zero22   -0.5459      0.221     -2.474      0.013      -0.978
```

```
-0.113
==============================================================================
Omnibus:                       53.068   Durbin-Watson:                   2.105
Prob(Omnibus):                  0.000   Jarque-Bera (JB):              565.120
Skew:                           0.042   Prob(JB):                    1.93e-123
Kurtosis:                       9.401   Cond. No.                         143.
==============================================================================

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:                       0.668
Model:                            OLS   Adj. R-squared:                  0.663
Method:                 Least Squares   F-statistic:                     60.22
Date:                Wed, 22 May 2024   Prob (F-statistic):           3.09e-44
Time:                        07:45:03   Log-Likelihood:                 738.50
No. Observations:                 330   AIC:                            -1465.
Df Residuals:                     324   BIC:                            -1442.
Df Model:                           5
Covariance Type:                  HAC
=====================================================================================
=====
                            coef    std err          z      P>|z|      [0.025
0.975]
-------------------------------------------------------------------------------------
-----
const                    -0.0034      0.006     -0.580      0.562      -0.015
0.008
Zero_Fraction_5           0.0024      0.006      0.374      0.709      -0.010
0.015
Log_Return_Lag1           1.1335      0.217      5.233      0.000       0.709
1.558
Log_Return_Lag2           0.1757      0.237      0.740      0.459      -0.289
0.641
Log_Return_Lag1_Zero5    -1.3349      0.298     -4.477      0.000      -1.919      -
0.750
Log_Return_Lag2_Zero5    -0.1113      0.283     -0.394      0.694      -0.665
0.443
==============================================================================
Omnibus:                      125.791   Durbin-Watson:                   1.609
Prob(Omnibus):                  0.000   Jarque-Bera (JB):             1692.841
Skew:                           1.175   Prob(JB):                         0.00
Kurtosis:                      13.844   Cond. No.                         180.
==============================================================================

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:                       0.496
Model:                            OLS   Adj. R-squared:                  0.488
Method:                 Least Squares   F-statistic:                     21.74
Date:                Wed, 22 May 2024   Prob (F-statistic):           8.93e-19
```

```
Time:                        07:45:03   Log-Likelihood:                 669.62
No. Observations:                 330   AIC:                            -1327.
Df Residuals:                     324   BIC:                            -1304.
Df Model:                           5
Covariance Type:                  HAC
==========================================================================================
======
                          coef    std err          z      P>|z|      [0.025
0.975]
------------------------------------------------------------------------------------------
------
const                   0.0016      0.008      0.193      0.847      -0.015
0.018
Zero_Fraction_22       -0.0032      0.010     -0.322      0.748      -0.023
0.016
Log_Return_Lag1         0.7340      0.260      2.823      0.005       0.224
1.244
Log_Return_Lag2         0.1595      0.273      0.584      0.559      -0.376
0.695
Log_Return_Lag1_Zero22 -0.4169      0.428     -0.974      0.330      -1.256
0.422
Log_Return_Lag2_Zero22 -0.0972      0.449     -0.217      0.829      -0.977
0.783
==============================================================================
Omnibus:                       58.729   Durbin-Watson:                   1.945
Prob(Omnibus):                  0.000   Jarque-Bera (JB):              686.047
Skew:                           0.202   Prob(JB):                    1.06e-149
Kurtosis:                      10.052   Cond. No.                         254.
==============================================================================

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:                       0.680
Model:                            OLS   Adj. R-squared:                  0.674
Method:                 Least Squares   F-statistic:                     62.73
Date:                Wed, 22 May 2024   Prob (F-statistic):           2.55e-51
Time:                        07:45:03   Log-Likelihood:                 744.59
No. Observations:                 330   AIC:                            -1475.
Df Residuals:                     323   BIC:                            -1449.
Df Model:                           6
Covariance Type:                  HAC
==========================================================================================
======
                          coef    std err          z      P>|z|      [0.025
0.975]
------------------------------------------------------------------------------------------
------
const                  -0.0021      0.006     -0.336      0.737      -0.014
0.010
Zero_Fraction_5         0.0059      0.008      0.742      0.458      -0.010
0.021
Zero_Fraction_22       -0.0047      0.009     -0.495      0.620      -0.023
0.014
Log_Return_Lag1         1.2954      0.133      9.719      0.000       1.034
1.557
Log_Return_Lag2        -0.1476      0.170     -0.867      0.386      -0.481
0.186
```

```
Log_Return_Lag1_Zero5     -1.5820      0.167     -9.483      0.000     -1.909
-1.255
Log_Return_Lag2_Zero22     0.6466      0.279      2.318      0.020      0.100
1.193
==============================================================================
Omnibus:                      103.307   Durbin-Watson:                   1.862
Prob(Omnibus):                  0.000   Jarque-Bera (JB):             1517.576
Skew:                           0.845   Prob(JB):                         0.00
Kurtosis:                      13.369   Cond. No.                         209.
==============================================================================

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

In [58]:

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

```
+----------------------------------------------------------+------------------------
-+---------------------+---------------------+-------------------------+------------
----------+--------------------------+------------------------+
|                    Model Name                            |        Adj R²
|         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    |  0.5498091033
843614   |     314.5186833119641      | 7.165941752616387e-68  |
|                 Model with Weekly Zero                   | -1.1187337564999567e-05
| -1119.048985976916  | -1111.4267039909626 |   0.6663684172207335    |  0.5470806169
291975   |     313.5534914848556      | 1.1593079726098431e-67 |
|                Model with Monthly Zero                   | 0.00047332969827085236
| -1119.2108520726424 | -1111.588570086689  |   0.6659031915516898    |   0.549412311
126432   |     314.32366766858763     | 7.897443881395843e-68  |
|                         AR(1)                            |   0.44221628864425344
| -1309.095199471638  | -1301.478914491677  |  -0.12692869813086013   |  0.1139968538
880168   |     10.397345606056946     |  0.015473677649263303  |
|                         AR(2)                            |   0.4608973245266801
| -1314.4740652539508 | -1303.0586603472013 | -0.002327121756879495   | -0.0101467210
29427166 |   0.04215333467183039      |   0.9977270952194889   |
|            AR(1) with Weekly Zero Interaction            |   0.6397049321725019
| -1452.650353016808  | -1437.417783056886  |  0.051606429811643866   |  0.2483583351
2121724  |      27.3948706744289      | 4.8653971446982565e-06 |
|           AR(1) with Monthly Zero Interaction            |   0.4506298798857473
| -1312.1746169699295 | -1296.9420470100076 |  -0.08786852316957235   |  0.1316902097
2684173  |      9.17651799578683      |  0.027033752967145667  |
|            AR(2) with Weekly Zero Interaction            |   0.6502198530775722
| -1455.1389195190632 | -1432.3081097055642 |   0.1554979055162424    |  0.1661700084
429255   |     21.682379886804345     | 7.594916831227399e-05  |
|           AR(2) with Monthly Zero Interaction            |   0.46482449585413876
|  -1313.9426618394   | -1291.111852025901  |  0.029356139090624987   | 0.01760915791
9125807  |   0.40900268841602766      |   0.9383765352599154   |
| AR(2) with Separate Weekly and Monthly Zero Interactions |   0.6667421419818788
| -1470.2237941547537 | -1443.5878493723383 | -0.0060668329669526355  |  0.0698299493
6368837  |     2.5793865303526005     |   0.46111482000108095  |
+----------------------------------------------------------+------------------------
-+---------------------+---------------------+-------------------------+------------
----------+--------------------------+------------------------+
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:                Wed, 22 May 2024   Prob (F-statistic):                nan
Time:                        07:45:11   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:                       0.003
Model:                            OLS   Adj. R-squared:                 -0.000
Method:                 Least Squares   F-statistic:                    0.9963
Date:                Wed, 22 May 2024   Prob (F-statistic):              0.319
Time:                        07:45:11   Log-Likelihood:                 561.52
No. Observations:                 334   AIC:                            -1119.
Df Residuals:                     332   BIC:                            -1111.
Df Model:                           1
Covariance Type:            nonrobust
=================================================================================
                    coef    std err          t      P>|t|      [0.025      0.975]
---------------------------------------------------------------------------------
const             0.0046      0.006      0.774      0.439      -0.007       0.016
Zero_Fraction_5  -0.0080      0.008     -0.998      0.319      -0.024       0.008
==============================================================================
Omnibus:                       70.359   Durbin-Watson:                   0.667
Prob(Omnibus):                  0.000   Jarque-Bera (JB):              744.999
Skew:                           0.483   Prob(JB):                    1.68e-162
Kurtosis:                      10.253   Cond. No.                         4.83
==============================================================================
```

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:                     1.158
Date:                Wed, 22 May 2024   Prob (F-statistic):              0.283
Time:                        07:45:11   Log-Likelihood:                 561.61
No. Observations:                 334   AIC:                            -1119.
Df Residuals:                     332   BIC:                            -1112.
Df Model:                           1
Covariance Type:            nonrobust
====================================================================================
                       coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------------
const                0.0064      0.007      0.898      0.370      -0.008       0.021
Zero_Fraction_22    -0.0106      0.010     -1.076      0.283      -0.030       0.009
==============================================================================
Omnibus:                       71.142   Durbin-Watson:                   0.668
Prob(Omnibus):                  0.000   Jarque-Bera (JB):              747.870
Skew:                           0.499   Prob(JB):                    4.00e-163
Kurtosis:                      10.263   Cond. No.                         5.92
==============================================================================
```

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:                Wed, 22 May 2024   Prob (F-statistic):           7.73e-15
Time:                        07:45:11   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:                Wed, 22 May 2024   Prob (F-statistic):           2.89e-17
Time:                        07:45:11   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:                       0.643
Model:                            OLS   Adj. R-squared:                  0.640
Method:                 Least Squares   F-statistic:                     77.71
Date:                Wed, 22 May 2024   Prob (F-statistic):           5.03e-38
Time:                        07:45:11   Log-Likelihood:                 730.33
No. Observations:                 333   AIC:                            -1453.
Df Residuals:                     329   BIC:                            -1437.
Df Model:                           3
Covariance Type:                  HAC
==============================================================================
=====
                           coef    std err          z      P>|z|      [0.025
0.975]
------------------------------------------------------------------------------
-----
const                    0.0015      0.006      0.247      0.805      -0.011
0.014
Zero_Fraction_5         -0.0042      0.007     -0.595      0.552      -0.018
0.010
Log_Return_Lag1          1.2301      0.102     12.009      0.000       1.029
1.431
Log_Return_Lag1_Zero5   -1.3795      0.181     -7.605      0.000      -1.735      -
1.024
==============================================================================
Omnibus:                       85.940   Durbin-Watson:                   1.896
Prob(Omnibus):                  0.000   Jarque-Bera (JB):             1007.035
Skew:                           0.678   Prob(JB):                    2.11e-219
Kurtosis:                      11.411   Cond. No.                         91.3
==============================================================================

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



Parameters for Model: AR(1) with Monthly Zero Interaction:
                            OLS Regression Results
==============================================================================
Dep. Variable:             Log_Return   R-squared:                       0.456
Model:                            OLS   Adj. R-squared:                  0.451
Method:                 Least Squares   F-statistic:                     21.64
Date:                Wed, 22 May 2024   Prob (F-statistic):           8.17e-13
Time:                        07:45:11   Log-Likelihood:                 660.09
No. Observations:                 333   AIC:                            -1312.
Df Residuals:                     329   BIC:                            -1297.
Df Model:                           3
Covariance Type:                  HAC
==============================================================================
======
                            coef    std err          z      P>|z|      [0.025
0.975]
------------------------------------------------------------------------------
------
const                     0.0014      0.008      0.166      0.868      -0.015
0.017
Zero_Fraction_22         -0.0031      0.010     -0.317      0.751      -0.022
0.016
Log_Return_Lag1           0.8338      0.125      6.696      0.000       0.590
1.078
Log_Return_Lag1_Zero22   -0.5250      0.218     -2.411      0.016      -0.952
```

```
-0.098
==============================================================================
Omnibus:                       67.786   Durbin-Watson:                   2.174
Prob(Omnibus):                  0.000   Jarque-Bera (JB):             1204.642
Skew:                          -0.037   Prob(JB):                    2.60e-262
Kurtosis:                      12.317   Cond. No.                         140.
==============================================================================

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:                       0.656
Model:                            OLS   Adj. R-squared:                  0.650
Method:                 Least Squares   F-statistic:                     46.30
Date:                Wed, 22 May 2024   Prob (F-statistic):           4.46e-36
Time:                        07:45:11   Log-Likelihood:                 733.57
No. Observations:                 332   AIC:                            -1455.
Df Residuals:                     326   BIC:                            -1432.
Df Model:                           5
Covariance Type:                  HAC
=====================================================================================
=====
                            coef    std err          z      P>|z|      [0.025
0.975]
-------------------------------------------------------------------------------------
-----
const                     0.0013      0.007      0.188      0.851      -0.012
0.014
Zero_Fraction_5          -0.0038      0.008     -0.486      0.627      -0.019
0.011
Log_Return_Lag1           1.0806      0.203      5.314      0.000       0.682
1.479
Log_Return_Lag2           0.2033      0.189      1.076      0.282      -0.167
0.574
Log_Return_Lag1_Zero5    -1.2771      0.313     -4.077      0.000      -1.891      -
0.663
Log_Return_Lag2_Zero5    -0.1084      0.251     -0.432      0.666      -0.601
0.384
==============================================================================
Omnibus:                      115.235   Durbin-Watson:                   1.688
Prob(Omnibus):                  0.000   Jarque-Bera (JB):             1540.908
Skew:                           1.029   Prob(JB):                         0.00
Kurtosis:                      13.352   Cond. No.                         181.
==============================================================================

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:                       0.473
Model:                            OLS   Adj. R-squared:                  0.465
Method:                 Least Squares   F-statistic:                     18.25
Date:                Wed, 22 May 2024   Prob (F-statistic):           5.64e-16
```

```
Time:                        07:45:11   Log-Likelihood:                 662.97
No. Observations:                 332   AIC:                            -1314.
Df Residuals:                     326   BIC:                            -1291.
Df Model:                           5
Covariance Type:                  HAC
==========================================================================================
                             coef    std err          z      P>|z|      [0.025
0.975]
------------------------------------------------------------------------------------------
const                      0.0013      0.009      0.157      0.876      -0.015
0.018
Zero_Fraction_22          -0.0030      0.010     -0.287      0.774      -0.023
0.017
Log_Return_Lag1            0.6216      0.233      2.670      0.008       0.165
1.078
Log_Return_Lag2            0.2676      0.219      1.223      0.221      -0.161
0.696
Log_Return_Lag1_Zero22    -0.2796      0.390     -0.716      0.474      -1.045
0.485
Log_Return_Lag2_Zero22    -0.2342      0.386     -0.606      0.544      -0.991
0.523
==============================================================================
Omnibus:                       73.378   Durbin-Watson:                   1.940
Prob(Omnibus):                  0.000   Jarque-Bera (JB):             1280.143
Skew:                           0.289   Prob(JB):                    1.05e-278
Kurtosis:                      12.602   Cond. No.                         265.
==============================================================================

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:                       0.673
Model:                            OLS   Adj. R-squared:                  0.667
Method:                 Least Squares   F-statistic:                     54.85
Date:                Wed, 22 May 2024   Prob (F-statistic):           1.49e-46
Time:                        07:45:11   Log-Likelihood:                 742.11
No. Observations:                 332   AIC:                            -1470.
Df Residuals:                     325   BIC:                            -1444.
Df Model:                           6
Covariance Type:                  HAC
==========================================================================================
                             coef    std err          z      P>|z|      [0.025
0.975]
------------------------------------------------------------------------------------------
const                      0.0015      0.006      0.233      0.816      -0.011
0.014
Zero_Fraction_5           -0.0047      0.009     -0.551      0.582      -0.022
0.012
Zero_Fraction_22           0.0015      0.009      0.156      0.876      -0.017
0.020
Log_Return_Lag1            1.2624      0.133      9.484      0.000       1.001
1.523
Log_Return_Lag2           -0.1597      0.134     -1.189      0.234      -0.423
0.104
```

```
Log_Return_Lag1_Zero5     -1.5699      0.176     -8.922      0.000      -1.915
-1.225
Log_Return_Lag2_Zero22     0.7974      0.246      3.242      0.001       0.315
1.279
==============================================================================
Omnibus:                      108.881   Durbin-Watson:                   2.012
Prob(Omnibus):                  0.000   Jarque-Bera (JB):             1523.817
Skew:                           0.930   Prob(JB):                         0.00
Kurtosis:                      13.330   Cond. No.                         213.
==============================================================================

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

In [59]:

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

Model Name	Adj R²	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	-0.0013642233682962246	-1071.5606071650618	-1064.011504073973	0.6480396176913513	0.530393864305114	293.6864814259133	2.3128325060110028e-63
Model with Monthly Zero	-0.0009601028270205259	-1071.6905829280845	-1064.1414798369956	0.6478161168849362	0.5318835116361448	294.54437907639203	1.5082842725604658e-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	0.5965349149049617	-1357.0424231719608	-1341.9566586794408	0.07622468836713626	0.20659773890744296	25.00169788862177	1.542788201211415e-05
AR(1) with Monthly Zero Interaction	0.4490982825589066	-1257.0616868956981	-1241.9759224031782	-0.03388237253119526	0.10406670866775199	11.545775520866492	0.009112736978480765
AR(2) with Weekly Zero Interaction	0.604503175290116	-1356.2161795957159	-1333.6062536209533	0.15480873916245083	0.13287133048486352	20.45510996278377	0.000136592546067836
AR(2) with Monthly Zero Interaction	0.45974285447753893	-1256.4073914369355	-1233.797465462173	0.012894971051207562	0.0007347654678739172	4.315220497958294	0.229376328766992
AR(2) with Separate Weekly and Monthly Zero Interactions	0.6228941722918944	-1370.4743222644406	-1344.0960752938843	0.011265292103581532	0.02770651322372761	4.235255008758069	0.2371558080663643

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:                Wed, 22 May 2024   Prob (F-statistic):                nan
Time:                        07:45:20   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 spe
cified.
```

```
Parameters for Model: Model with Weekly Zero:
                            OLS Regression Results
==============================================================================
Dep. Variable:             Log_Return   R-squared:                       0.002
Model:                            OLS   Adj. R-squared:                 -0.001
Method:                 Least Squares   F-statistic:                    0.5627
Date:                Wed, 22 May 2024   Prob (F-statistic):              0.454
Time:                        07:45:20   Log-Likelihood:                 537.78
No. Observations:                 322   AIC:                            -1072.
Df Residuals:                     320   BIC:                            -1064.
Df Model:                           1
Covariance Type:            nonrobust
===================================================================================
                      coef    std err          t      P>|t|      [0.025      0.975]
-----------------------------------------------------------------------------------
const               0.0033      0.006      0.547      0.585      -0.009       0.015
Zero_Fraction_5    -0.0061      0.008     -0.750      0.454      -0.022       0.010
==============================================================================
Omnibus:                       62.377   Durbin-Watson:                   0.704
Prob(Omnibus):                  0.000   Jarque-Bera (JB):              651.135
Skew:                           0.383   Prob(JB):                    4.05e-142
Kurtosis:                       9.924   Cond. No.                         4.75
==============================================================================

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

```
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:                    0.6921
Date:                Wed, 22 May 2024   Prob (F-statistic):              0.406
Time:                        07:45:20   Log-Likelihood:                 537.85
No. Observations:                 322   AIC:                            -1072.
Df Residuals:                     320   BIC:                            -1064.
Df Model:                           1
Covariance Type:            nonrobust
====================================================================================
                       coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------------
const                0.0049      0.007      0.670      0.504      -0.010       0.019
Zero_Fraction_22    -0.0083      0.010     -0.832      0.406      -0.028       0.011
==============================================================================
Omnibus:                       62.920   Durbin-Watson:                   0.704
Prob(Omnibus):                  0.000   Jarque-Bera (JB):              651.615
Skew:                           0.398   Prob(JB):                    3.19e-142
Kurtosis:                       9.924   Cond. No.                         5.87
==============================================================================

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

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:                Wed, 22 May 2024   Prob (F-statistic):           2.62e-19
Time:                        07:45:20   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:                Wed, 22 May 2024   Prob (F-statistic):           1.07e-20
Time:                        07:45:20   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:                       0.600
Model:                            OLS   Adj. R-squared:                  0.597
Method:                 Least Squares   F-statistic:                     66.88
Date:                Wed, 22 May 2024   Prob (F-statistic):           1.56e-33
Time:                        07:45:20   Log-Likelihood:                 682.52
No. Observations:                 321   AIC:                            -1357.
Df Residuals:                     317   BIC:                            -1342.
Df Model:                           3
Covariance Type:                  HAC
==================================================================================
=====
                           coef    std err          z      P>|z|      [0.025
0.975]
----------------------------------------------------------------------------------
-----
const                    0.0019      0.007      0.259      0.796      -0.013
0.016
Zero_Fraction_5         -0.0046      0.009     -0.540      0.589      -0.021
0.012
Log_Return_Lag1          1.1773      0.103     11.422      0.000       0.975
1.379
Log_Return_Lag1_Zero5   -1.2719      0.182     -6.969      0.000      -1.630      -
0.914
==============================================================================
Omnibus:                      105.213   Durbin-Watson:                   1.846
Prob(Omnibus):                  0.000   Jarque-Bera (JB):              990.319
Skew:                           1.055   Prob(JB):                    9.02e-216
Kurtosis:                      11.342   Cond. No.                         87.8
==============================================================================

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:                       0.454
Model:                            OLS   Adj. R-squared:                  0.449
Method:                 Least Squares   F-statistic:                     35.52
Date:                Wed, 22 May 2024   Prob (F-statistic):           8.09e-20
Time:                        07:45:20   Log-Likelihood:                 632.53
No. Observations:                 321   AIC:                            -1257.
Df Residuals:                     317   BIC:                            -1242.
Df Model:                           3
Covariance Type:                  HAC
==================================================================================
======
                            coef    std err          z      P>|z|      [0.025
0.975]
----------------------------------------------------------------------------------
------
const                    -0.0002      0.008     -0.028      0.978      -0.016
0.016
Zero_Fraction_22         -0.0012      0.010     -0.126      0.900      -0.020
0.018
Log_Return_Lag1           0.9371      0.107      8.727      0.000       0.727
1.148
Log_Return_Lag1_Zero22   -0.8609      0.217     -3.967      0.000      -1.286
```

```
-0.436
==============================================================================
Omnibus:                       71.199   Durbin-Watson:                   2.066
Prob(Omnibus):                  0.000   Jarque-Bera (JB):              690.368
Skew:                           0.568   Prob(JB):                    1.23e-150
Kurtosis:                      10.094   Cond. No.                         133.
==============================================================================

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:                       0.611
Model:                            OLS   Adj. R-squared:                  0.605
Method:                 Least Squares   F-statistic:                     38.05
Date:                Wed, 22 May 2024   Prob (F-statistic):           1.77e-30
Time:                        07:45:20   Log-Likelihood:                 684.11
No. Observations:                 320   AIC:                            -1356.
Df Residuals:                     314   BIC:                            -1334.
Df Model:                           5
Covariance Type:                  HAC
==============================================================================
=====
                           coef    std err          z      P>|z|      [0.025
0.975]
------------------------------------------------------------------------------
-----
const                    0.0017      0.008      0.218      0.827      -0.013
0.017
Zero_Fraction_5         -0.0042      0.009     -0.465      0.642      -0.022
0.014
Log_Return_Lag1          1.0356      0.187      5.551      0.000       0.670
1.401
Log_Return_Lag2          0.1905      0.187      1.020      0.308      -0.176
0.557
Log_Return_Lag1_Zero5   -1.1651      0.289     -4.030      0.000      -1.732      -
0.599
Log_Return_Lag2_Zero5   -0.1091      0.251     -0.435      0.663      -0.601
0.382
==============================================================================
Omnibus:                      116.273   Durbin-Watson:                   1.690
Prob(Omnibus):                  0.000   Jarque-Bera (JB):             1281.701
Skew:                           1.156   Prob(JB):                    4.81e-279
Kurtosis:                      12.528   Cond. No.                         176.
==============================================================================

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:                       0.468
Model:                            OLS   Adj. R-squared:                  0.460
Method:                 Least Squares   F-statistic:                     23.54
Date:                Wed, 22 May 2024   Prob (F-statistic):           4.29e-20
```

```
Time:                        07:45:20   Log-Likelihood:                 634.20
No. Observations:                 320   AIC:                            -1256.
Df Residuals:                     314   BIC:                            -1234.
Df Model:                           5
Covariance Type:                  HAC
==========================================================================================
                             coef    std err          z      P>|z|      [0.025
0.975]
------------------------------------------------------------------------------------------
const                     -0.0002      0.008     -0.018      0.985      -0.016
0.016
Zero_Fraction_22          -0.0009      0.010     -0.094      0.925      -0.020
0.019
Log_Return_Lag1            0.8894      0.187      4.756      0.000       0.523
1.256
Log_Return_Lag2            0.0208      0.182      0.114      0.909      -0.336
0.378
Log_Return_Lag1_Zero22    -0.9164      0.316     -2.898      0.004      -1.536
-0.297
Log_Return_Lag2_Zero22     0.2921      0.312      0.937      0.349      -0.319
0.903
==============================================================================
Omnibus:                       77.096   Durbin-Watson:                   1.973
Prob(Omnibus):                  0.000   Jarque-Bera (JB):              771.186
Skew:                           0.647   Prob(JB):                    3.46e-168
Kurtosis:                      10.494   Cond. No.                         235.
==============================================================================

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:                       0.630
Model:                            OLS   Adj. R-squared:                  0.623
Method:                 Least Squares   F-statistic:                     44.75
Date:                Wed, 22 May 2024   Prob (F-statistic):           2.16e-39
Time:                        07:45:20   Log-Likelihood:                 692.24
No. Observations:                 320   AIC:                            -1370.
Df Residuals:                     313   BIC:                            -1344.
Df Model:                           6
Covariance Type:                  HAC
==========================================================================================
                             coef    std err          z      P>|z|      [0.025
0.975]
------------------------------------------------------------------------------------------
const                      0.0006      0.007      0.087      0.931      -0.013
0.014
Zero_Fraction_5           -0.0091      0.011     -0.796      0.426      -0.031
0.013
Zero_Fraction_22           0.0072      0.011      0.644      0.519      -0.015
0.029
Log_Return_Lag1            1.2330      0.130      9.493      0.000       0.978
1.488
Log_Return_Lag2           -0.1863      0.138     -1.354      0.176      -0.456
0.083
```

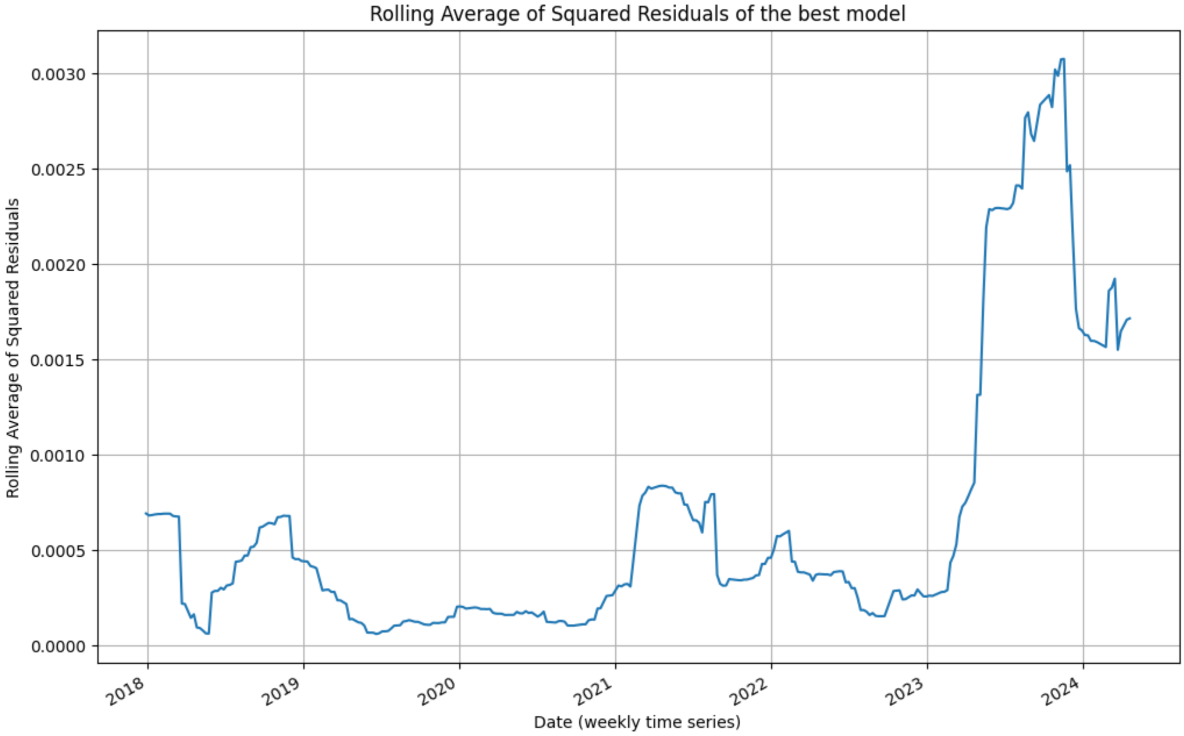
```
Log_Return_Lag1_Zero5      -1.4512      0.201      -7.228      0.000      -1.845
-1.058
Log_Return_Lag2_Zero22      0.7774      0.289       2.689      0.007       0.211
1.344
==============================================================================
Omnibus:                      121.275   Durbin-Watson:                   1.977
Prob(Omnibus):                  0.000   Jarque-Bera (JB):             1199.366
Skew:                           1.262   Prob(JB):                    3.64e-261
Kurtosis:                      12.142   Cond. No.                         201.
==============================================================================

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

The model **AR(2) with Separate Weekly and Monthly Zero Interactions** is still the best model for other 4 dataframes in almost all criterias. Furthurmore, not like Fridays, the mix model shows high enough Ljung-Box p-values ($0.24, 0.29, 0.46, 0.56$) 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 [60]:
```
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 mod
    plt.xlabel('Date (weekly time series)')
    plt.ylabel('Rolling Average of Squared Residuals')
    plt.grid(True)
    plt.show()
```

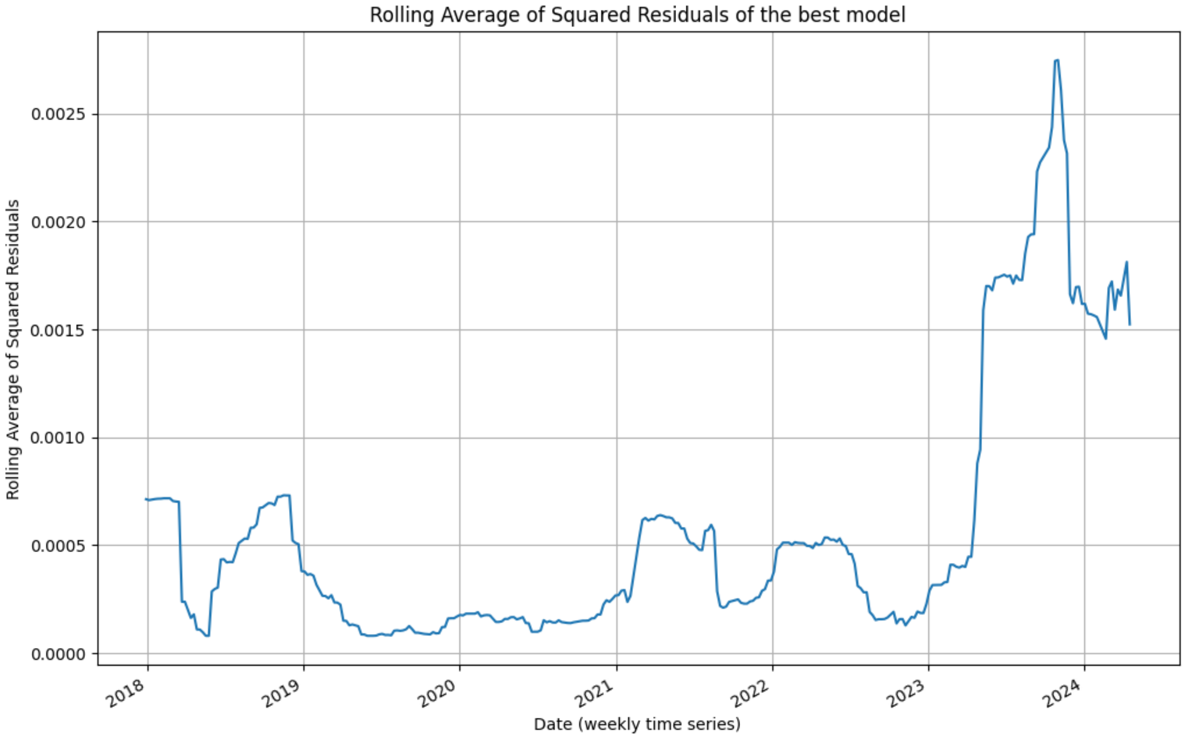
In [61]:
```
plot_rolling_average_of_squared_residuals(model_results=fridays_models['AR(2) with S
```



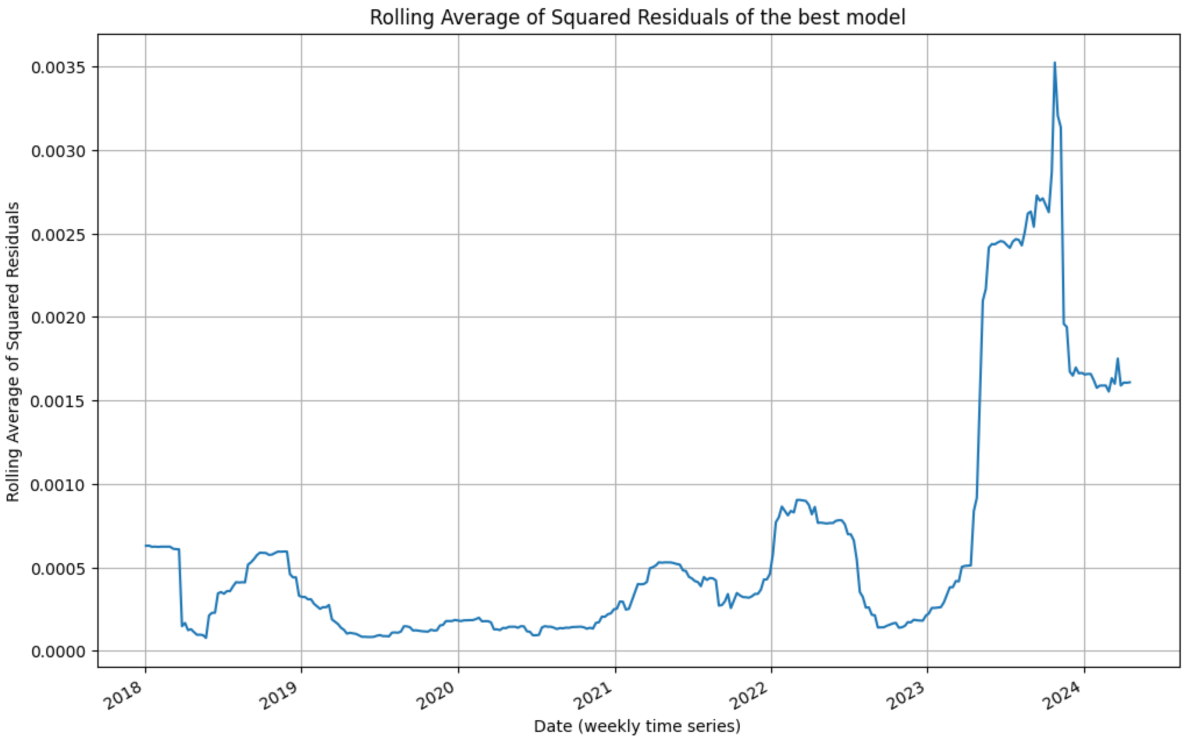
The rolling average follows a totally increasing trend, this indicates a rising volatility, but there also exists a sharp decreasing near the end of year 2023.

Robustness work for P11

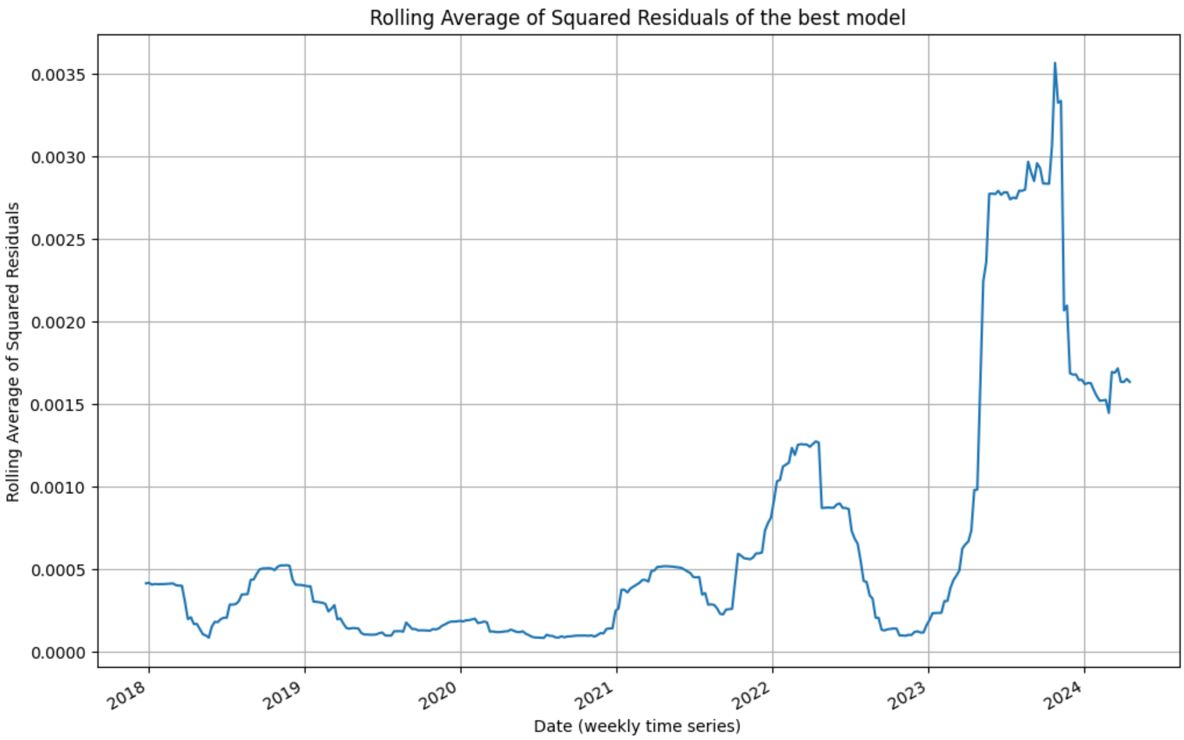
In [62]:
```
# Thursdays
plot_rolling_average_of_squared_residuals(model_results=thursdays_models['AR(2) with
```



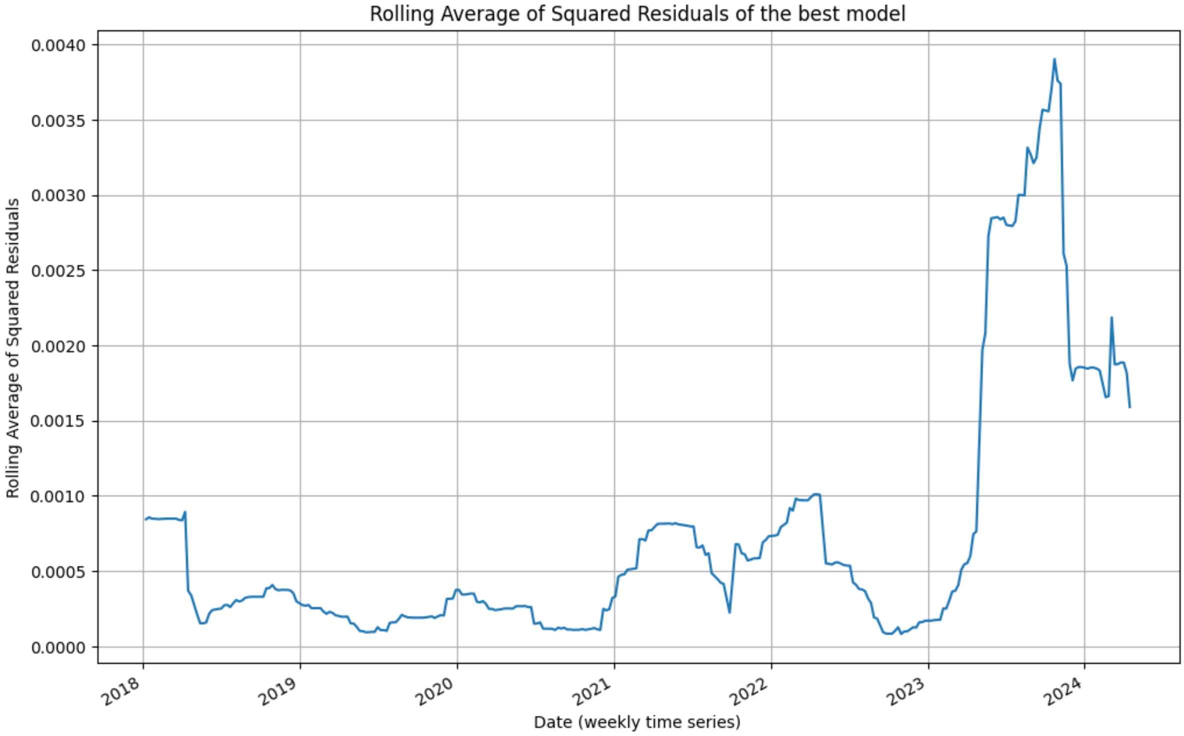
In [63]:
```
# Wednesdays
plot_rolling_average_of_squared_residuals(model_results=wednesdays_models['AR(2) wit
```



In [64]:
```
# Tuesdays
plot_rolling_average_of_squared_residuals(model_results=tuesdays_models['AR(2) with 
```



In [65]:
```
# Mondays
plot_rolling_average_of_squared_residuals(model_results=mondays_models['AR(2) with S
```



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

In [66]:
```
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_Count_22', 'Zero_Count_5', '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'] = data['Log_Return'].shift(1)
    input_data['Log_Return_Lag2'] = data['Log_Return'].shift(2)

    # Create interactions for lagged Log_Return with Zero_Fractions
    input_data['Log_Return_Lag1_Zero5'] = input_data['Log_Return_Lag1'] * data['Zero
    input_data['Log_Return_Lag1_Zero22'] = input_data['Log_Return_Lag1'] * data['Zer
    input_data['Log_Return_Lag2_Zero5'] = input_data['Log_Return_Lag2']  * data['Zer
    input_data['Log_Return_Lag2_Zero22'] = input_data['Log_Return_Lag2']  * data['Ze

    input_data.dropna(inplace=True)
```

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

    return input_data
```

In [67]:
```
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 [161…
```
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 i
    for i in range(52, len(perpared_data)):
        current_sample = perpared_data[['Log_Return', 'Zero_Fraction_5',        'Zer

        for name, model_func in models:

            if name == 'Model with Only Constant':  # Special case for constant mode
                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 [162…

```
fridays_forecast = forecast_and_compute_RMSE(df=Fridays, lags=4, df_name='Fridays')
```

```
+----------------------------------------------------------+----------------------+
|                          Model                           |         RMSE         |
+----------------------------------------------------------+----------------------+
| AR(2) with Separate Weekly and Monthly Zero Interactions | 0.028574111727738244 |
|           AR(1) with Weekly Zero Interaction             | 0.028699248262294284 |
|           AR(2) with Weekly Zero Interaction             | 0.028750511204274645 |
|          AR(1) with Monthly Zero Interaction             | 0.033110236042147606 |
|                          AR(2)                           | 0.03319323376216553  |
|                          AR(1)                           | 0.03347683497575105  |
|          AR(2) with Monthly Zero Interaction             |  0.0337186149256691  |
|                Model with Only Constant                  | 0.047601851591042715 |
|                 Model with Weekly Zero                   | 0.048232338814404634 |
|                Model with Monthly Zero                   | 0.048241569860663525 |
+----------------------------------------------------------+----------------------+
```

The model **AR(2) with Separate Weekly and Monthly Zero Interactions** still wins in $RMSE$ criteria. This mix model has the lowest RMSE, which signifies highest accuracy in predictive performance. It means that the mix model's predictions are, on average, closer to the actual observed weekly log return.

AR(1) with Weekly Zero Interaction and **AR(2) with Weekly Zero Interaction** also have very lose $RMSE$ to the mixed model.

Robustness work for P12

In [70]:

```
# Thursdays
thursdays_forecast = forecast_and_compute_RMSE(df=Thursdays, lags=2, df_name='Thursd
```

Model	RMSE
AR(1) with Weekly Zero Interaction	0.02644831863828377
AR(2) with Separate Weekly and Monthly Zero Interactions	0.02645577475084485
AR(2) with Weekly Zero Interaction	0.026687454429006337
AR(1) with Monthly Zero Interaction	0.03100346819508346
AR(1)	0.031228836428829586
AR(2)	0.03131017690997134
AR(2) with Monthly Zero Interaction	0.03152634740024679
Model with Only Constant	0.04684102151231067
Model with Weekly Zero	0.047454280581655434
Model with Monthly Zero	0.04748348409918562

In [71]:
```
# Wednesdays
wednesdays_forecast = forecast_and_compute_RMSE(df=Wednesdays, lags=2, df_name='Wedn
```

Model	RMSE
AR(1) with Weekly Zero Interaction	0.027619063383236837
AR(2) with Separate Weekly and Monthly Zero Interactions	0.028173505929623775
AR(2) with Weekly Zero Interaction	0.028430185568081297
AR(1)	0.0344754974330957
AR(2)	0.034532832194297196
AR(1) with Monthly Zero Interaction	0.03482136156950505
AR(2) with Monthly Zero Interaction	0.03567962652680273
Model with Only Constant	0.047647999920785465
Model with Monthly Zero	0.048149164570744966
Model with Weekly Zero	0.04824941542810454

In [72]:
```
# Tuesdays
tuesdays_forecast = forecast_and_compute_RMSE(df=Tuesdays, lags=2, df_name='Tuesdays
```

Model	RMSE
AR(2) with Separate Weekly and Monthly Zero Interactions	0.029554669343551537
AR(2) with Weekly Zero Interaction	0.029972662026528443
AR(1) with Weekly Zero Interaction	0.030109608823150988
AR(2)	0.0363193179205231
AR(1)	0.03668898286049331
AR(1) with Monthly Zero Interaction	0.03711598115118176
AR(2) with Monthly Zero Interaction	0.0373343424740437
Model with Only Constant	0.048320938820529416
Model with Weekly Zero	0.04892653471571601
Model with Monthly Zero	0.04893450220541475

In [73]:
```
# Mondays
mondays_forecast = forecast_and_compute_RMSE(df=Mondays, lags=2, df_name='Mondays')
```

Model	RMSE
AR(2) with Separate Weekly and Monthly Zero Interactions	0.031247302347774532
AR(1) with Weekly Zero Interaction	0.03165887030234597
AR(2) with Weekly Zero Interaction	0.03215754992225925
AR(2)	0.03668184362733399
AR(1) with Monthly Zero Interaction	0.0368788773003294
AR(1)	0.037050763630292645
AR(2) with Monthly Zero Interaction	0.037297806583013175
Model with Only Constant	0.04851899399219849
Model with Weekly Zero	0.0491806017065333
Model with Monthly Zero	0.04926107862652646

The mix model **AR(2) with Separate Weekly and Monthly Zero Interactions** has second lowest $RMSE$ for Thursdays dataframe and Wednesdays dataframe.

The mix model has first lowest $RMSE$ for Fridays, Tuesdays, Mondays dataframe.

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

In [74]:

```
model_test = fridays_models['AR(2) with Separate Weekly and Monthly Zero Interaction
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  22.284808   0.000057
```

large 'lb_stat' and low 'lb_pvalue'

Compute the autocorrelations of the residuals at the first three lags

In [75]:

```
# Fridays
model_test = fridays_models['AR(2) with Separate Weekly and Monthly Zero Interaction
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: [ 0.02258348 -0.03011749  0.25872257]
Fridays' third autocorrelation: 0.25872257471076554
```

In [76]:

```
# Mondays
model_test = mondays_models['AR(2) with Separate Weekly and Monthly Zero Interaction
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: [0.01126529 0.02770651 0.11017399]
Mondays' third autocorrelation: 0.11017399321047239
```

In [77]:
```
# Tuesdays
model_test = tuesdays_models['AR(2) with Separate Weekly and Monthly Zero Interactio
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: [-0.00606683  0.06982995  0.05248715]
Tuesdays' third autocorrelation: 0.05248715210374084
```

In [78]:
```
# Wednesdays
model_test = wednesdays_models['AR(2) with Separate Weekly and Monthly Zero Interact
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: [0.06856758 0.02536327 0.02979565]
Wednesdays' third autocorrelation: 0.029795653661547118
```

In [79]:
```
# Thursdays
model_test = thursdays_models['AR(2) with Separate Weekly and Monthly Zero Interacti
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: [ 0.05798363 -0.02744602  0.08398479]
Thursdays' third autocorrelation: 0.08398478977334237
```

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 (0.258) than other four data ($0.11, 0.052, 0.03, 0.084$), where first two autocorrelations have no significant difference between these five data.

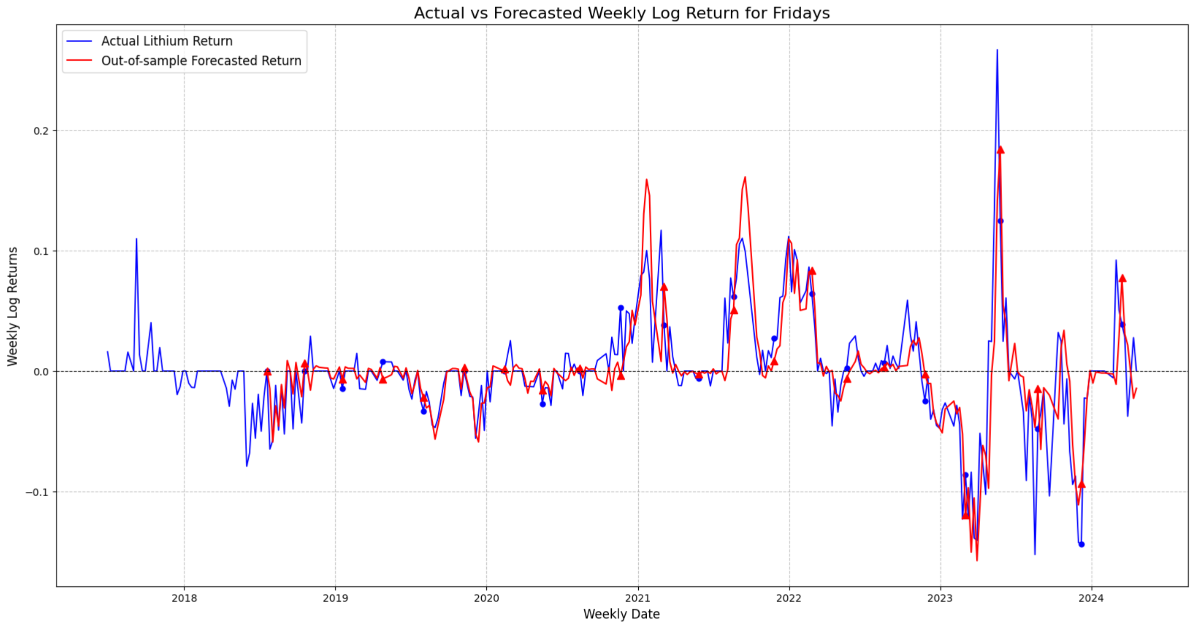
New P13: Forecast Graph

In [179…
```
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:
    - None: This function generates and displays a plot, but does not return any valu
    """
    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_Fracti
        features = model_features['AR(2) with Separate Weekly and Monthly Zero Intera
        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
    plt.figure(figsize=(20, 10))
    plt.plot(perpared_data.index, perpared_data['Log_Return'], label='Actual Lithium
    plt.plot(marker_positions, perpared_data.loc[marker_positions, 'Log_Return'], 'bo
    plt.plot(forecasted_series.index, forecasted_series, label='Out-of-sample Forecas
    plt.plot(marker_positions, forecasted_series.loc[marker_positions], 'r^', markers
    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()
```

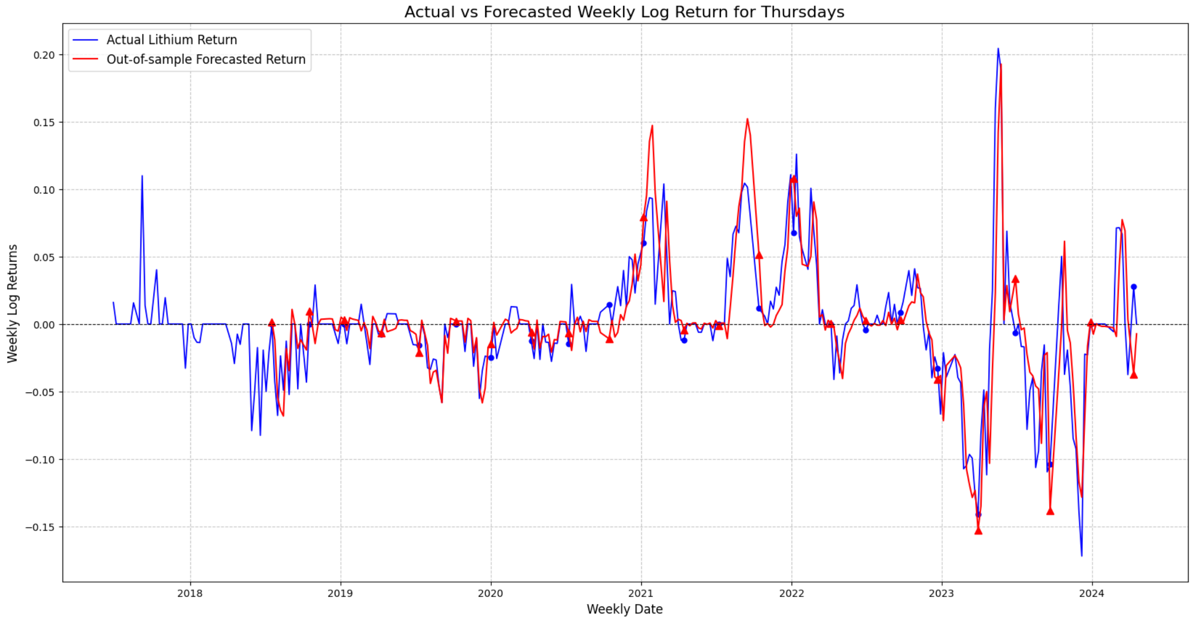
In [180…

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



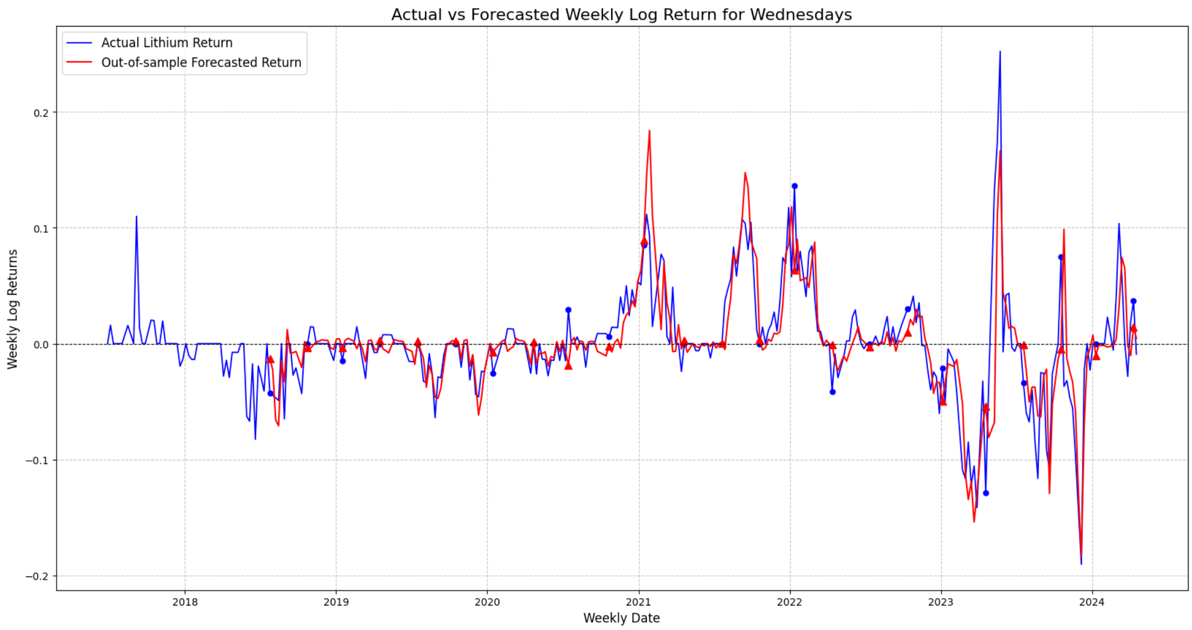
In [181…

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



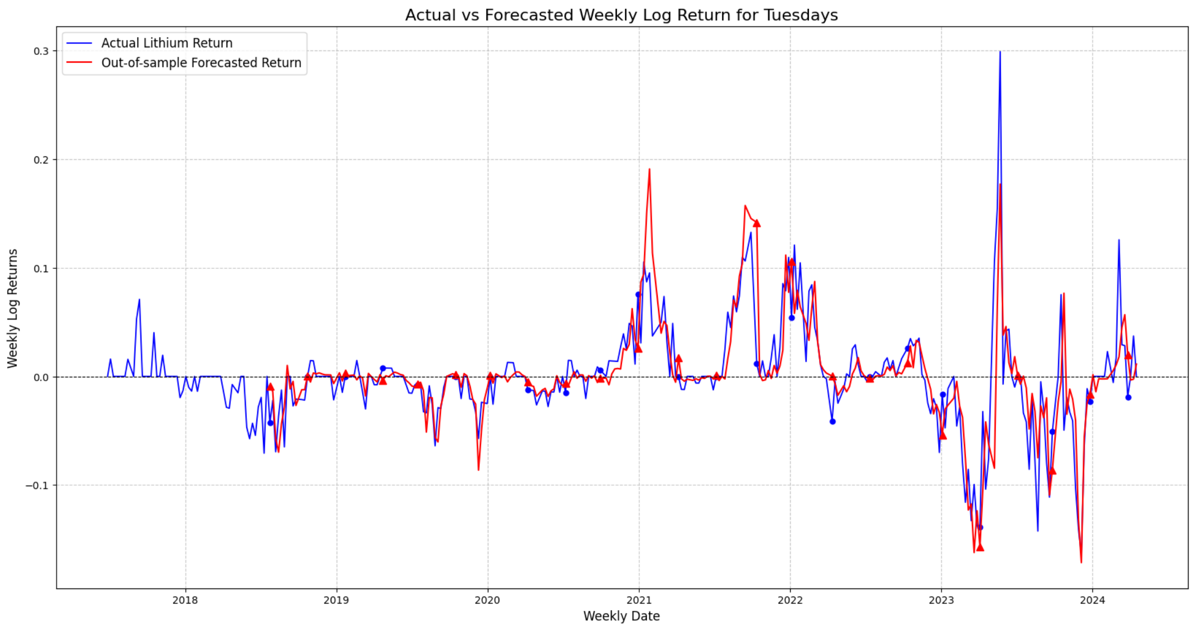
In [182…

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



In [183…

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



In [184…

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