Question 3 – Market Impact and Kyle’s Lambda

Part A – Kyle’s Lambda

Kyle (1985) proposed that for short intervals (e.g., 5 seconds), order book thickness can be approximated by:

$$\lambda = \frac{\Delta P_t}{\mathrm{OrderFlow}_t}$$

where:
ΔP_t = price change in interval t
$\mathrm{OrderFlow}_t$ = trade volume × sign of the price change (positive for increases, negative for decreases).

Interpretation:
Smaller λ = thicker order book (price moves little for a given trade volume).
Larger λ = thinner order book (price moves more for the same volume).

Hypothesis

1) λ differs between Brent and WTI

- Eventhough Brent and WTI are linked through arbitrage, differences in market structure and purpose differentiates between the two benchmarks. Brent serves as the global pricing benchmark but trades on a relatively thinner order book, while WTI futures experience more speculation and hedging activity on the CME. These liquidity differences should be reflected in Kyle’s λ, leading me to expect significantly different λ estimates for Brent and WTI.

2) λ varies intraday and is expected to be smaller during core market hours

- Although futures trade nearly 24 hours a day, the majority of volume and liquidity is concentrated during U.S. and European trading hours. Higher trading activity during this window implies deeper order books and tighter spreads, so individual trades exert less price impact. Therefore, I hypothesize that Kyle’s λ will be smaller during market hours.

3) λ decreases near expiry and increases in farther-out contracts

- As futures contracts approach expiry, arbitrage guarantees a convergence to the spot market price.
- In addition, as contracts approach expiry speculators and hedgers will offset their positions, increasing liquidity.

- Consequently, trades exert less influence on prices, leading to smaller λ near expiry.
- By contrast, farther-out contracts trade with lower volumes and thinner order books, so λ is expected to be larger.

Part B – Data Preparation

You are provided with 5-minute interval data (prices and volumes) for November 2025 WTI and Brent crude oil futures (30 days of around-the-clock trading: opens Sunday 5 pm, closes Friday 4 pm)

In [ ]:
```
# install & import libraries
!pip install pandas numpy matplotlib

import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import statsmodels.api as sm
import statsmodels.formula.api as smf
```

In [2]:
```
# load the data files
# CBZ25 = Brent crude oil futures (December 2025)
# CLV25 = WTI crude oil futures (October 2025)

# WTI (West Texas Intermediate) = U.S.
# Brent = Europe, Africa, and the Middle East

brent_data = pd.read_csv('data/cbz25_intraday-5min_historical-data-09-16-2025.cs
wti_data = pd.read_csv('data/clv25_intraday-5min_historical-data-09-16-2025.csv'

# checking data structure
print("Brent:")
print(brent_data.head())
print("\nWTI:")
print(wti_data.head())
```

```
Brent:
               Time   Open   High    Low   Last  Change    %Chg  Volume
0  2025-08-17 17:00  64.87  65.04  64.51  64.54   -0.49  -0.75%     185
1  2025-08-17 17:05  64.55  64.59  64.53  64.58    0.04   0.06%     117
2  2025-08-17 17:10  64.56  64.63  64.54  64.63    0.05   0.08%      61
3  2025-08-17 17:15  64.65  64.71  64.65  64.67    0.04   0.06%      64
4  2025-08-17 17:20  64.68  64.68  64.65  64.65   -0.02  -0.03%      15

WTI:
               Time   Open   High    Low   Last  Change    %Chg  Volume
0  2025-08-17 17:00  62.19  62.19  61.65  61.67   -0.62  -1.00%    1150
1  2025-08-17 17:05  61.68  61.76  61.67  61.73    0.06   0.10%     248
2  2025-08-17 17:10  61.73  61.80  61.71  61.78    0.05   0.08%     171
3  2025-08-17 17:15  61.77  61.87  61.75  61.83    0.05   0.08%     203
4  2025-08-17 17:20  61.83  61.84  61.78  61.83    0.00   0.00%     113
```

In [3]:
```python
# subset data for only what is necessary
brent_data = brent_data[["Time", "Change", "Volume"]]
wti_data = wti_data[["Time", "Change", "Volume"]]

# convert Time column to datetime
brent_data['Time'] = pd.to_datetime(brent_data['Time'])
wti_data['Time'] = pd.to_datetime(wti_data['Time'])

# add benchmark column for Brent and WTI
brent_data['Benchmark'] = 'Brent'
wti_data['Benchmark'] = 'WTI'
```

1.1 Aggregate 5-minute trading volumes into hourly totals, then compute the average hourly volume for each clock hour across all days.

In [4]:
```python
# compute hour column and date column for aggregation at the hour level
brent_data['Hour'] = brent_data['Time'].dt.hour
brent_data['Date'] = brent_data['Time'].dt.date
wti_data['Hour'] = wti_data['Time'].dt.hour
wti_data['Date'] = wti_data['Time'].dt.date
```

In [5]:
```python
print(wti_data.head())
```

```
                 Time  Change  Volume Benchmark  Hour        Date
0 2025-08-17 17:00:00   -0.62    1150       WTI    17  2025-08-17
1 2025-08-17 17:05:00    0.06     248       WTI    17  2025-08-17
2 2025-08-17 17:10:00    0.05     171       WTI    17  2025-08-17
3 2025-08-17 17:15:00    0.05     203       WTI    17  2025-08-17
4 2025-08-17 17:20:00    0.00     113       WTI    17  2025-08-17
```

In [6]:
```python
# check date range; f lets me run formulas inside print statements
print(f"Date range (Brent): {brent_data['Time'].min()} to {brent_data['Time'].ma
print(f"Date range (WTI): {wti_data['Time'].min()} to {wti_data['Time'].max()}")

# check if any values are missing
print(f"\nMissing values (Brent): {brent_data.isna().values.sum()}")
print(f"Missing values (WTI): {wti_data.isna().values.sum()}")

# check how many rows and columns
print(f"\n(rows, columns): {brent_data.shape}")
print(f"(rows, columns): {wti_data.shape}\n")

# how many rows we should have after aggregating
print(f"After aggregation: {brent_data.shape[0]+wti_data.shape[0]}")
```

```
Date range (Brent): 2025-08-17 17:00:00 to 2025-09-12 16:55:00
Date range (WTI): 2025-08-17 17:00:00 to 2025-09-12 15:55:00

Missing values (Brent): 0
Missing values (WTI): 0

(rows, columns): (5106, 6)
(rows, columns): (5487, 6)

After aggregation: 10593
```

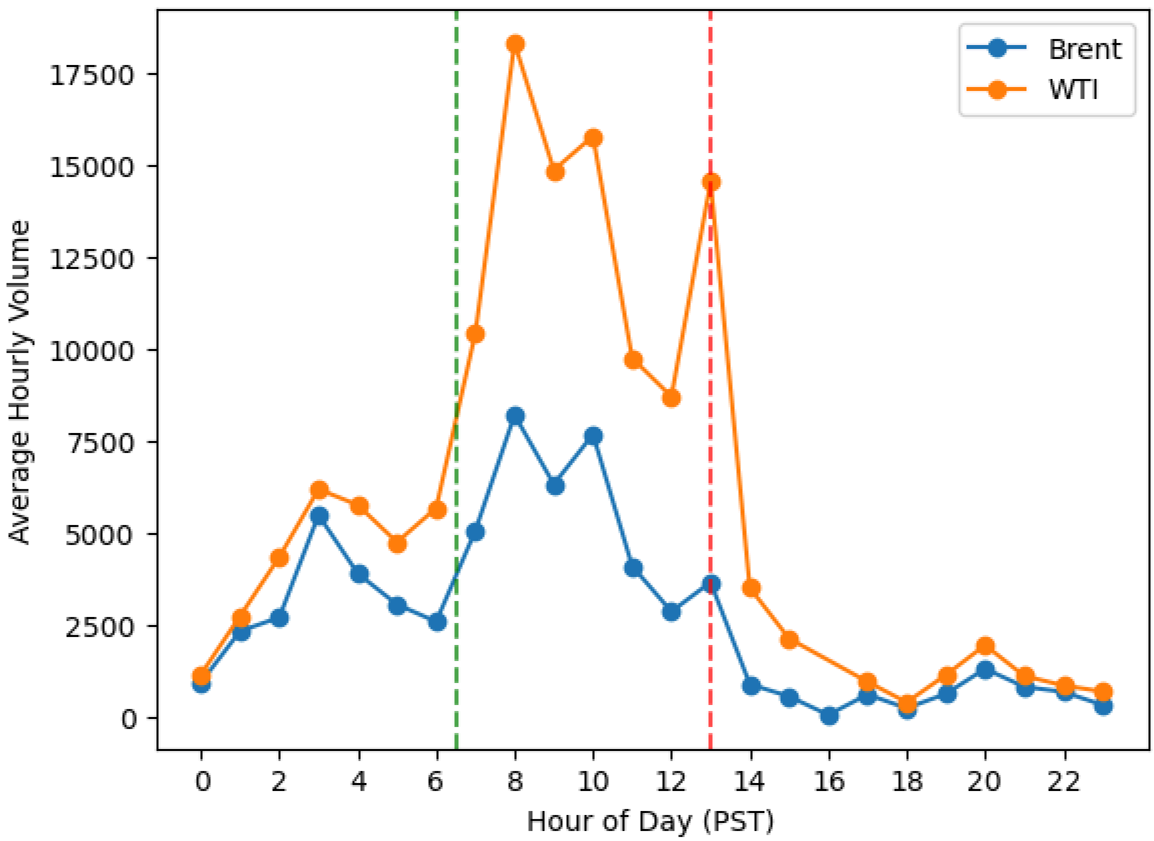
In [7]:
```python
# concatenate datasets before calculation to save time
# ignore_index = TRUE will assign unique indexes in the aggregated dataframe
all_data = pd.concat([brent_data, wti_data], ignore_index=True)

print(f"(rows, columns): {all_data.shape}")
print(all_data.head(10))
```

```
(rows, columns): (10593, 6)
                 Time  Change  Volume Benchmark  Hour        Date
0 2025-08-17 17:00:00   -0.49     185     Brent    17  2025-08-17
1 2025-08-17 17:05:00    0.04     117     Brent    17  2025-08-17
2 2025-08-17 17:10:00    0.05      61     Brent    17  2025-08-17
3 2025-08-17 17:15:00    0.04      64     Brent    17  2025-08-17
4 2025-08-17 17:20:00   -0.02      15     Brent    17  2025-08-17
5 2025-08-17 17:25:00    0.10      66     Brent    17  2025-08-17
6 2025-08-17 17:30:00   -0.09       9     Brent    17  2025-08-17
7 2025-08-17 17:35:00    0.00       1     Brent    17  2025-08-17
8 2025-08-17 17:40:00   -0.02      13     Brent    17  2025-08-17
9 2025-08-17 17:45:00    0.00       3     Brent    17  2025-08-17
```

In [8]:
```python
# aggregate 5-minute volumes into daily hourly sums
# we use .reset_index() to keep volume as its own column and drop Brent/WTI down
volume_hour = all_data.groupby(['Benchmark', 'Date', 'Hour'])['Volume'].sum().re

# calculate average hourly volume for each clock hour across all days
avg_volume_hour = volume_hour.groupby(['Benchmark', 'Hour'])['Volume'].mean().re
```

1.2 Plot average hourly volume over a 24-hour cycle. Compare the pattern with your hypotheses. Adjust if necessary.

In [9]:
```python
# plot
# after inspection, the timezone from the output is likely in PST from Prof. Ver
# we move forward assuming time in PST, where market opens between 6:30am-1:00pm

plt.figure()

# create Brent and WTI dataframes for plotting
brent_hourly = avg_volume_hour[avg_volume_hour['Benchmark'] == 'Brent']
wti_hourly = avg_volume_hour[avg_volume_hour['Benchmark'] == 'WTI']

# plot Brent and WTI
# hour on x, volume on y
plt.plot(brent_hourly['Hour'], brent_hourly['Volume'], label='Brent',  marker='o
plt.plot(wti_hourly['Hour'], wti_hourly['Volume'], label='WTI',  marker='o')
plt.xlabel('Hour of Day (PST)')
plt.ylabel('Average Hourly Volume')
plt.legend()
plt.xticks(range(0, 24, 2))

# reference lines for market open and close in PST
plt.axvline(x=6.5, color='green', linestyle='--', alpha=0.7, label='US Market Op
plt.axvline(x=13, color='red', linestyle='--', alpha=0.7, label='US Market Close

plt.show()
```



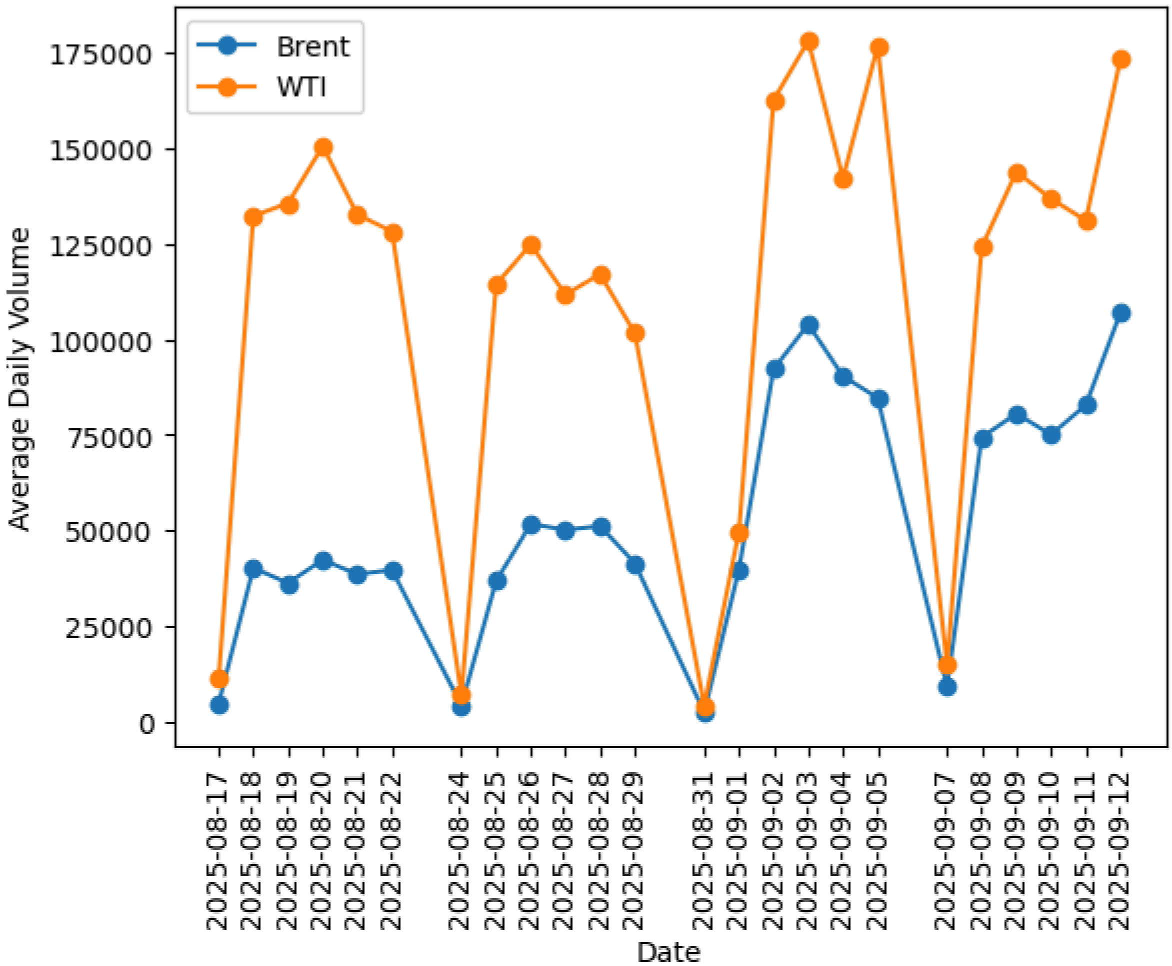
Note: Visually, hypothesis 2 seems promising. Volume is much higher during trading hours. However, we cannot draw formal conclusions as price chage is not included in this graph.

1.3 Next, calculate and plot average daily trading volumes. Identify any days with unusually high volumes. If you find spikes, investigate possible causes using news sources and report your findings.

In [10]:

```
# calculate daily trading volumes
daily_volumes = all_data.groupby(['Benchmark', 'Date'])['Volume'].sum().reset_in
print(daily_volumes.head())

# create Brent and WTI dataframes for plotting
brent_daily = daily_volumes[daily_volumes['Benchmark'] == 'Brent']
wti_daily = daily_volumes[daily_volumes['Benchmark'] == 'WTI']

# plot Brent and WTI
# hour on x, volume on y
plt.plot(brent_daily['Date'], brent_daily['Volume'], label='Brent',  marker='o')
plt.plot(wti_daily['Date'], wti_daily['Volume'], label='WTI',  marker='o')
plt.xlabel('Date')
plt.ylabel('Average Daily Volume')
plt.legend()
plt.xticks(daily_volumes["Date"].unique(), rotation=90)

plt.show()
```

```
  Benchmark        Date  Volume
0     Brent  2025-08-17    4755
1     Brent  2025-08-18   40186
2     Brent  2025-08-19   36238
3     Brent  2025-08-20   42300
4     Brent  2025-08-21   38536
```



The volumes separating into chunks make sense as futures market do not trade on Saturdays. The low volume on the first day makes sense as well, since the market opens at 5pm PST on Sundays. Interestingly, even though Fridays end early at 5pm PST, we see similar volumes compared to midweek; a likely explanation is that people are trading primarily during work hours.

There appears to be a spike on September 12th. Online sources show that it's due to a decrease in US crude oil inventories, specifically by 9.3 million barrels during the week ending September 12 (Geiger, 2025). It is possible that this decrease in inventories exceeded analyst projections, causing prices to soar.

https://oilprice.com/Energy/Crude-Oil/US-Crude-Oil-Inventories-Plummet-Pushing-Prices-Higher.html

2.1 Add a new column Q_t = volume × sign of ΔP for each 5-minute interval.

In [11]:
```python
# Order Flow: Q_t = volume × sign(delta P)
# delta P is last trade - this trade
analysis_data = all_data.copy()

# change = delta_P; change also has a sign by default
analysis_data = analysis_data.rename(columns={"Change": "Delta_P"})

# extract sign from delta_P
analysis_data['Sign_Delta_P'] = np.sign(analysis_data['Delta_P'])

# order flow: Q_t = Volume × Sign(ΔP)
analysis_data['Order_Flow'] = analysis_data['Volume'] * analysis_data['Sign_Delt

analysis_data.head()
```

Out[11]:

	Time	Delta_P	Volume	Benchmark	Hour	Date	Sign_Delta_P	Order_Flow
0	2025-08-17 17:00:00	-0.49	185	Brent	17	2025-08-17	-1.0	-185.0
1	2025-08-17 17:05:00	0.04	117	Brent	17	2025-08-17	1.0	117.0
2	2025-08-17 17:10:00	0.05	61	Brent	17	2025-08-17	1.0	61.0
3	2025-08-17 17:15:00	0.04	64	Brent	17	2025-08-17	1.0	64.0
4	2025-08-17 17:20:00	-0.02	15	Brent	17	2025-08-17	-1.0	-15.0

In [12]:
```python
# check for zero price changes (where sign is 0)
zero_changes = analysis_data[analysis_data['Sign_Delta_P'] == 0]
print(f"\nObservations with delta_P = 0: {len(zero_changes)}")

# this is an issue, since Delta_P being 0 doesn't mean no volume ocurred. The vo
# we remove the 0s and take the sign of the last 5 minutes
analysis_data["Sign_Delta_P"] = (
    analysis_data["Sign_Delta_P"].replace(0, np.nan).ffill())

print(analysis_data.head(10))
zero_changes = analysis_data[analysis_data['Sign_Delta_P'] == 0]
print(f"\nObservations with delta_P = 0 now: {len(zero_changes)}")

# recalculate order flow: Q_t = Volume × Sign(ΔP)
analysis_data['Order_Flow'] = analysis_data['Volume'] * analysis_data['Sign_Delt
```

```
Observations with delta_P = 0: 1298
                  Time  Delta_P  Volume Benchmark  Hour        Date  \
0 2025-08-17 17:00:00    -0.49     185     Brent    17  2025-08-17
1 2025-08-17 17:05:00     0.04     117     Brent    17  2025-08-17
2 2025-08-17 17:10:00     0.05      61     Brent    17  2025-08-17
3 2025-08-17 17:15:00     0.04      64     Brent    17  2025-08-17
4 2025-08-17 17:20:00    -0.02      15     Brent    17  2025-08-17
5 2025-08-17 17:25:00     0.10      66     Brent    17  2025-08-17
6 2025-08-17 17:30:00    -0.09       9     Brent    17  2025-08-17
7 2025-08-17 17:35:00     0.00       1     Brent    17  2025-08-17
8 2025-08-17 17:40:00    -0.02      13     Brent    17  2025-08-17
9 2025-08-17 17:45:00     0.00       3     Brent    17  2025-08-17

   Sign_Delta_P  Order_Flow
0          -1.0      -185.0
1           1.0       117.0
2           1.0        61.0
3           1.0        64.0
4          -1.0       -15.0
5           1.0        66.0
6          -1.0        -9.0
7          -1.0         0.0
8          -1.0       -13.0
9          -1.0         0.0

Observations with delta_P = 0 now: 0
```

Part C – Estimation of Kyle's Lambda

3.1 Baseline Model

o Estimate: $\Delta P_t = \lambda Q_t + e_t$
o Report λ and describe what it measures.

In [13]:
```
# regression using delta_P as the dependent variable and Q_t (order flow) as the
# lambda is the beta1 from the regression output

for benchmark in ["WTI", "Brent"]:
    subset = analysis_data[analysis_data["Benchmark"] == benchmark]

    X = sm.add_constant(subset["Order_Flow"])
    y = subset["Delta_P"]

    model = sm.OLS(y, X).fit()
    print(f"\n{benchmark} Output")
    print(model.summary())
```

WTI Output

```
                            OLS Regression Results                            
==============================================================================
Dep. Variable:                Delta_P   R-squared:                       0.511
Model:                            OLS   Adj. R-squared:                  0.511
Method:                 Least Squares   F-statistic:                     5728.
Date:                Sat, 27 Sep 2025   Prob (F-statistic):               0.00
Time:                        02:20:01   Log-Likelihood:                 10064.
No. Observations:                5487   AIC:                        -2.012e+04
Df Residuals:                    5485   BIC:                        -2.011e+04
Df Model:                           1                                         
Covariance Type:            nonrobust                                         
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const          0.0004      0.001      0.708      0.479      -0.001       0.001
Order_Flow  4.449e-05   5.88e-07     75.683      0.000    4.33e-05    4.56e-05
==============================================================================
Omnibus:                     1289.567   Durbin-Watson:                   2.072
Prob(Omnibus):                  0.000   Jarque-Bera (JB):            59698.007
Skew:                          -0.259   Prob(JB):                         0.00
Kurtosis:                      19.151   Cond. No.                         888.
==============================================================================

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

Brent Output

```
                            OLS Regression Results                            
==============================================================================
Dep. Variable:                Delta_P   R-squared:                       0.478
Model:                            OLS   Adj. R-squared:                  0.478
Method:                 Least Squares   F-statistic:                     4678.
Date:                Sat, 27 Sep 2025   Prob (F-statistic):               0.00
Time:                        02:20:01   Log-Likelihood:                 9472.3
No. Observations:                5106   AIC:                        -1.894e+04
Df Residuals:                    5104   BIC:                        -1.893e+04
Df Model:                           1                                         
Covariance Type:            nonrobust                                         
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const          4.6e-05      0.001      0.087      0.931      -0.001       0.001
Order_Flow  7.893e-05   1.15e-06     68.397      0.000    7.67e-05    8.12e-05
==============================================================================
Omnibus:                     1101.877   Durbin-Watson:                   2.137
Prob(Omnibus):                  0.000   Jarque-Bera (JB):            21429.880
Skew:                          -0.519   Prob(JB):                         0.00
Kurtosis:                      12.982   Cond. No.                         459.
==============================================================================

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

WTI lambda = 4.449e-05
Brent lambda = 7.893e-05

λ represents how much price changes for a one unit increase of order_flow.
Brent is nearly double of WTI; this makes sense as from the figure earlier, descriptively volume is much larger for WTI than Brent.

3.2 Seasonal Control

o Re-estimate the model adding hourly or daily controls (i.e., dummies) to account for systematic variation in liquidity.
o Compare λ estimates with and without these controls - has λ changed materially?

In [14]:

```
# hourly dummies
for benchmark in ["WTI", "Brent"]:
    subset = analysis_data[analysis_data["Benchmark"] == benchmark].copy()
    model = smf.ols("Delta_P ~ Order_Flow + C(Hour)", data=subset).fit()

    print(f"\nResults for {benchmark}:")
    print(model.summary())
```

Results for WTI:

OLS Regression Results

Dep. Variable:	Delta_P	R-squared:	0.514
Model:	OLS	Adj. R-squared:	0.512
Method:	Least Squares	F-statistic:	250.8
Date:	Sat, 27 Sep 2025	Prob (F-statistic):	0.00
Time:	02:20:01	Log-Likelihood:	10080.
No. Observations:	5487	AIC:	-2.011e+04
Df Residuals:	5463	BIC:	-1.995e+04
Df Model:	23		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
Intercept	0.0026	0.002	1.051	0.293	-0.002	0.008
C(Hour)[T.1]	6.955e-05	0.004	0.020	0.984	-0.007	0.007
C(Hour)[T.2]	-0.0034	0.004	-0.969	0.332	-0.010	0.003
C(Hour)[T.3]	0.0004	0.004	0.124	0.901	-0.006	0.007
C(Hour)[T.4]	-0.0049	0.004	-1.378	0.168	-0.012	0.002
C(Hour)[T.5]	-0.0042	0.004	-1.199	0.231	-0.011	0.003
C(Hour)[T.6]	0.0021	0.004	0.587	0.557	-0.005	0.009
C(Hour)[T.7]	-0.0091	0.004	-2.572	0.010	-0.016	-0.002
C(Hour)[T.8]	-0.0033	0.004	-0.926	0.355	-0.010	0.004
C(Hour)[T.9]	-0.0020	0.004	-0.570	0.568	-0.009	0.005
C(Hour)[T.10]	0.0027	0.004	0.771	0.441	-0.004	0.010
C(Hour)[T.11]	-0.0049	0.004	-1.392	0.164	-0.012	0.002
C(Hour)[T.12]	-0.0023	0.004	-0.646	0.518	-0.009	0.005
C(Hour)[T.13]	0.0048	0.004	1.342	0.180	-0.002	0.012
C(Hour)[T.14]	-0.0031	0.004	-0.873	0.383	-0.010	0.004
C(Hour)[T.15]	-0.0031	0.004	-0.879	0.380	-0.010	0.004
C(Hour)[T.17]	-0.0037	0.004	-1.043	0.297	-0.011	0.003
C(Hour)[T.18]	-0.0021	0.004	-0.590	0.555	-0.009	0.005
C(Hour)[T.19]	-0.0028	0.004	-0.795	0.427	-0.010	0.004
C(Hour)[T.20]	-0.0056	0.004	-1.583	0.114	-0.012	0.001
C(Hour)[T.21]	-0.0020	0.004	-0.554	0.579	-0.009	0.005
C(Hour)[T.22]	-0.0026	0.004	-0.749	0.454	-0.010	0.004
C(Hour)[T.23]	-0.0027	0.004	-0.768	0.442	-0.010	0.004
Order_Flow	4.445e-05	5.88e-07	75.572	0.000	4.33e-05	4.56e-05

Omnibus:	1304.870	Durbin-Watson:	2.082
Prob(Omnibus):	0.000	Jarque-Bera (JB):	57017.845
Skew:	-0.317	Prob(JB):	0.00
Kurtosis:	18.780	Cond. No.	2.08e+04

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
[2] The condition number is large, 2.08e+04. This might indicate that there are strong multicollinearity or other numerical problems.

Results for Brent:

OLS Regression Results

Dep. Variable:	Delta_P	R-squared:	0.481

```
Model:                            OLS   Adj. R-squared:                  0.479
Method:                 Least Squares   F-statistic:                     196.3
Date:                Sat, 27 Sep 2025   Prob (F-statistic):               0.00
Time:                        02:20:01   Log-Likelihood:                 9486.4
No. Observations:                5106   AIC:                        -1.892e+04
Df Residuals:                    5081   BIC:                        -1.876e+04
Df Model:                          24
Covariance Type:            nonrobust
=================================================================================
                    coef    std err          t      P>|t|      [0.025      0.975]
---------------------------------------------------------------------------------
Intercept         0.0032      0.002      1.291      0.197      -0.002       0.008
C(Hour)[T.1]     -0.0021      0.003     -0.591      0.554      -0.009       0.005
C(Hour)[T.2]     -0.0057      0.003     -1.648      0.099      -0.013       0.001
C(Hour)[T.3]     -0.0019      0.003     -0.549      0.583      -0.009       0.005
C(Hour)[T.4]     -0.0034      0.003     -0.981      0.326      -0.010       0.003
C(Hour)[T.5]     -0.0053      0.003     -1.522      0.128      -0.012       0.002
C(Hour)[T.6]     -0.0004      0.003     -0.111      0.911      -0.007       0.006
C(Hour)[T.7]     -0.0094      0.003     -2.697      0.007      -0.016      -0.003
C(Hour)[T.8]     -0.0008      0.003     -0.216      0.829      -0.008       0.006
C(Hour)[T.9]     -0.0040      0.003     -1.160      0.246      -0.011       0.003
C(Hour)[T.10]     0.0027      0.003      0.791      0.429      -0.004       0.010
C(Hour)[T.11]    -0.0059      0.003     -1.710      0.087      -0.013       0.001
C(Hour)[T.12]    -0.0025      0.003     -0.725      0.469      -0.009       0.004
C(Hour)[T.13]    -0.0023      0.004     -0.655      0.512      -0.009       0.005
C(Hour)[T.14]    -0.0050      0.004     -1.407      0.159      -0.012       0.002
C(Hour)[T.15]    -0.0029      0.004     -0.830      0.406      -0.010       0.004
C(Hour)[T.16]    -0.0014      0.005     -0.307      0.759      -0.010       0.008
C(Hour)[T.17]    -0.0163      0.007     -2.381      0.017      -0.030      -0.003
C(Hour)[T.18]     0.0043      0.007      0.577      0.564      -0.010       0.019
C(Hour)[T.19]    -0.0019      0.004     -0.535      0.593      -0.009       0.005
C(Hour)[T.20]    -0.0063      0.003     -1.805      0.071      -0.013       0.001
C(Hour)[T.21]    -0.0028      0.003     -0.795      0.427      -0.010       0.004
C(Hour)[T.22]    -0.0025      0.004     -0.699      0.485      -0.009       0.004
C(Hour)[T.23]    -0.0033      0.004     -0.937      0.349      -0.010       0.004
Order_Flow     7.888e-05   1.16e-06     68.256      0.000    7.66e-05    8.11e-05
==============================================================================
Omnibus:                     1062.780   Durbin-Watson:                   2.147
Prob(Omnibus):                  0.000   Jarque-Bera (JB):            18905.404
Skew:                          -0.509   Prob(JB):                         0.00
Kurtosis:                      12.372   Cond. No.                     1.09e+04
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
[2] The condition number is large, 1.09e+04. This might indicate that there are
strong multicollinearity or other numerical problems.
```

In [15]:

```
# days of the week dummies
analysis_data['DayOfWeek'] = analysis_data['Time'].dt.dayofweek

for benchmark in ["WTI", "Brent"]:
    subset = analysis_data[analysis_data["Benchmark"] == benchmark].copy()
    model = smf.ols("Delta_P ~ Order_Flow + C(DayOfWeek)", data=subset).fit()
```

```
    print(f"\nResults for {benchmark}:")
    print(model.summary())
```

Results for WTI:

```
                            OLS Regression Results
==============================================================================
Dep. Variable:                Delta_P   R-squared:                       0.511
Model:                            OLS   Adj. R-squared:                  0.511
Method:                 Least Squares   F-statistic:                     954.8
Date:                Sat, 27 Sep 2025   Prob (F-statistic):               0.00
Time:                        02:20:02   Log-Likelihood:                 10066.
No. Observations:                5487   AIC:                        -2.012e+04
Df Residuals:                    5480   BIC:                        -2.007e+04
Df Model:                           6
Covariance Type:            nonrobust
==================================================================================
====
                        coef    std err          t      P>|t|      [0.025      0.
975]
----------------------------------------------------------------------------------
----
Intercept             0.0018      0.001      1.495      0.135      -0.001
0.004
C(DayOfWeek)[T.1]    -0.0013      0.002     -0.756      0.450      -0.005
0.002
C(DayOfWeek)[T.2]    -0.0016      0.002     -0.935      0.350      -0.005
0.002
C(DayOfWeek)[T.3]    -0.0027      0.002     -1.658      0.097      -0.006
0.001
C(DayOfWeek)[T.4]    -0.0011      0.002     -0.628      0.530      -0.005
0.002
C(DayOfWeek)[T.6]    -0.0019      0.002     -0.797      0.426      -0.007
0.003
Order_Flow         4.449e-05   5.88e-07     75.663      0.000    4.33e-05    4.56
e-05
==============================================================================
Omnibus:                     1287.403   Durbin-Watson:                   2.073
Prob(Omnibus):                  0.000   Jarque-Bera (JB):            59503.701
Skew:                          -0.256   Prob(JB):                         0.00
Kurtosis:                      19.125   Cond. No.                     5.48e+03
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly
specified.
[2] The condition number is large, 5.48e+03. This might indicate that there are
strong multicollinearity or other numerical problems.
```

Results for Brent:

```
                            OLS Regression Results
==============================================================================
Dep. Variable:                Delta_P   R-squared:                       0.479
Model:                            OLS   Adj. R-squared:                  0.478
Method:                 Least Squares   F-statistic:                     780.8
Date:                Sat, 27 Sep 2025   Prob (F-statistic):               0.00
Time:                        02:20:02   Log-Likelihood:                 9475.2
No. Observations:                5106   AIC:                        -1.894e+04
Df Residuals:                    5099   BIC:                        -1.889e+04
Df Model:                           6
```

```
Covariance Type:            nonrobust
=================================================================================
====
                       coef    std err          t      P>|t|      [0.025      0.
975]
---------------------------------------------------------------------------------
----
Intercept            0.0021      0.001      1.752      0.080      -0.000      
0.005
C(DayOfWeek)[T.1]   -0.0023      0.002     -1.374      0.169      -0.006      
0.001
C(DayOfWeek)[T.2]   -0.0013      0.002     -0.765      0.444      -0.005      
0.002
C(DayOfWeek)[T.3]   -0.0033      0.002     -1.913      0.056      -0.007       8
e-05
C(DayOfWeek)[T.4]   -0.0038      0.002     -2.065      0.039      -0.007      -
0.000
C(DayOfWeek)[T.6]   -0.0024      0.003     -0.922      0.357      -0.007      
0.003
Order_Flow        7.896e-05   1.15e-06     68.415      0.000    7.67e-05    8.12
e-05
==============================================================================
Omnibus:                     1092.991   Durbin-Watson:                   2.140
Prob(Omnibus):                  0.000   Jarque-Bera (JB):            21259.581
Skew:                          -0.509   Prob(JB):                         0.00
Kurtosis:                      12.944   Cond. No.                     2.89e+03
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly
specified.
[2] The condition number is large, 2.89e+03. This might indicate that there are
strong multicollinearity or other numerical problems.
```

3.3 Hypothesis Testing

o For each of your three hypotheses, subset the data as needed (e.g., WTI vs. Brent, calm vs. news days).
o Estimate λ for each subset and compare magnitudes.
o State whether differences align with your expectations. Formal statistical testing is optional.

Hypothesis

1) λ differs between Brent and WTI

- Eventhough Brent and WTI are linked through arbitrage, differences in market structure and purpose differentiates between the two benchmarks. Brent serves as the global pricing benchmark but trades on a relatively thinner order book, while WTI futures experience more speculation and hedging activity on the CME. These liquidity differences should be reflected in Kyle’s λ, leading me to expect significantly different λ estimates for Brent and WTI.

2) λ varies intraday and is expected to be smaller during core market hours

- Although futures trade nearly 24 hours a day, the majority of volume and liquidity is concentrated during U.S. and European trading hours. Higher trading activity during this window implies deeper order books and tighter spreads, so individual trades exert less price impact. Therefore, I hypothesize that Kyle's λ will be smaller during market hours.

3) λ decreases near expiry and increases in farther-out contracts

- As futures contracts approach expiry, arbitrage guarantees a convergence to the spot market price.
- In addition, as contracts approach expiry speculators and hedgers will offset their positions, increasing liquidity.
- Consequently, trades exert less influence on prices, leading to smaller λ near expiry.
- By contrast, farther-out contracts trade with lower volumes and thinner order books, so λ is expected to be larger.

In [16]:
```
# 1) λ differs between Brent and WTI
model_diff = smf.ols("Delta_P ~ Order_Flow * C(Benchmark)", data=analysis_data).
print(model_diff.summary())
```

```
                            OLS Regression Results
==============================================================================
Dep. Variable:                Delta_P   R-squared:                       0.496
Model:                            OLS   Adj. R-squared:                  0.496
Method:                 Least Squares   F-statistic:                     350.8
Date:                Sat, 27 Sep 2025   Prob (F-statistic):          3.31e-217
Time:                        02:20:02   Log-Likelihood:                 19536.
No. Observations:               10593   AIC:                        -3.906e+04
Df Residuals:                   10589   BIC:                        -3.903e+04
Df Model:                           3
Covariance Type:                  HC1
================================================================================
=================
                                     coef    std err          z      P>|z|
[0.025      0.975]
--------------------------------------------------------------------------------
-----------------
Intercept                         4.6e-05      0.001      0.087      0.931      -
0.001       0.001
C(Benchmark)[T.WTI]                0.0003      0.001      0.436      0.663      -
0.001       0.002
Order_Flow                      7.893e-05   3.45e-06     22.905      0.000    7.2
2e-05    8.57e-05
Order_Flow:C(Benchmark)[T.WTI] -3.443e-05   3.97e-06     -8.684      0.000   -4.2
2e-05   -2.67e-05
==============================================================================
Omnibus:                     2391.775   Durbin-Watson:                   2.121
Prob(Omnibus):                  0.000   Jarque-Bera (JB):            78484.424
Skew:                          -0.380   Prob(JB):                         0.00
Kurtosis:                      16.313   Cond. No.                     2.15e+03
==============================================================================

Notes:
[1] Standard Errors are heteroscedasticity robust (HC1)
[2] The condition number is large, 2.15e+03. This might indicate that there are
strong multicollinearity or other numerical problems.
```

Hypothesis 1 Interpretation

To test if Brent and WTI benchmarks have significantly different λs, I estimated Delta_P as a function of order flow, benchmark, and their interaction term. I used HC1 for robust standard errors to adjust for heteroskedasticity. Brent is the baseline benchmark in this regression. The estimated λ for Brent is 7.893e-05 ($p < 0.001$), and the interaction term for WTI is -3.443e-05 ($p < 0.001$). This implies that WTI's λ is significantly smaller by -3.443e-05.

The λ is nearly halved in WTI relative to Brent, which aligns with my expectations. Brent is a global benchmark and trades on a thinner order book, therefore more sensitive to price changes. In contrast, WTI sees more speculation and hedging activities, which provides greater liquidity. The R^2 value of 0.496 gives confidence for the relationship examined in this regression.

In [17]:
```
# 2) λ varies intraday and is expected to be smaller during core market hours
# subset by benchmark
```

```
wti_data = analysis_data[analysis_data["Benchmark"] == "WTI"].copy()
brent_data = analysis_data[analysis_data["Benchmark"] == "Brent"].copy()

# WTI
model_wti_hourly = smf.ols("Delta_P ~ Order_Flow * C(Hour)", data=wti_data).fit(
print("WTI Results:")
print(model_wti_hourly.summary())

# Brent
model_brent_hourly = smf.ols("Delta_P ~ Order_Flow * C(Hour)", data=brent_data).
print("\nBrent Results:")
print(model_brent_hourly.summary())
```

WTI Results:

OLS Regression Results

Dep. Variable:	Delta_P	R-squared:	0.642
Model:	OLS	Adj. R-squared:	0.639
Method:	Least Squares	F-statistic:	72.55
Date:	Sat, 27 Sep 2025	Prob (F-statistic):	0.00
Time:	02:20:02	Log-Likelihood:	10922.
No. Observations:	5487	AIC:	-2.175e+04
Df Residuals:	5441	BIC:	-2.145e+04
Df Model:	45		
Covariance Type:	HC1		

	coef	std err	z	P>|z|	[0.025	0.975]
Intercept	0.0017	0.001	1.544	0.123	-0.000	0.004
C(Hour)[T.1]	-0.0022	0.002	-0.952	0.341	-0.007	0.002
C(Hour)[T.2]	-0.0013	0.002	-0.542	0.588	-0.006	0.003
C(Hour)[T.3]	-0.0008	0.003	-0.315	0.753	-0.006	0.004
C(Hour)[T.4]	-0.0034	0.003	-1.148	0.251	-0.009	0.002
C(Hour)[T.5]	-0.0030	0.003	-1.158	0.247	-0.008	0.002
C(Hour)[T.6]	0.0012	0.002	0.506	0.613	-0.003	0.006
C(Hour)[T.7]	-0.0085	0.003	-2.631	0.009	-0.015	-0.002
C(Hour)[T.8]	-0.0025	0.004	-0.698	0.485	-0.010	0.005
C(Hour)[T.9]	-0.0012	0.003	-0.396	0.692	-0.007	0.005
C(Hour)[T.10]	0.0035	0.004	0.996	0.319	-0.003	0.010
C(Hour)[T.11]	-0.0038	0.002	-1.558	0.119	-0.008	0.001
C(Hour)[T.12]	-0.0013	0.002	-0.523	0.601	-0.006	0.004
C(Hour)[T.13]	0.0022	0.003	0.812	0.417	-0.003	0.007
C(Hour)[T.14]	-0.0018	0.002	-1.006	0.315	-0.005	0.002
C(Hour)[T.15]	-0.0031	0.002	-1.978	0.048	-0.006	-2.79e-05
C(Hour)[T.17]	-0.0032	0.003	-1.163	0.245	-0.009	0.002
C(Hour)[T.18]	-0.0009	0.001	-0.677	0.498	-0.004	0.002
C(Hour)[T.19]	-0.0004	0.002	-0.220	0.826	-0.004	0.003

```
C(Hour)[T.20]                -0.0048      0.002     -2.568      0.010      -0.008
-0.001
C(Hour)[T.21]                -0.0019      0.002     -1.117      0.264      -0.005
0.001
C(Hour)[T.22]                -0.0010      0.002     -0.683      0.495      -0.004
0.002
C(Hour)[T.23]                -0.0018      0.001     -1.340      0.180      -0.005
0.001
Order_Flow                    0.0002   1.33e-05     14.870      0.000       0.000
0.000
Order_Flow:C(Hour)[T.1]   -3.448e-05   1.78e-05     -1.933      0.053   -6.95e-05
4.88e-07
Order_Flow:C(Hour)[T.2]   -9.506e-05   1.55e-05     -6.139      0.000      -0.000
-6.47e-05
Order_Flow:C(Hour)[T.3]      -0.0001   1.71e-05     -7.361      0.000      -0.000
-9.25e-05
Order_Flow:C(Hour)[T.4]      -0.0001    1.6e-05     -7.392      0.000      -0.000
-8.68e-05
Order_Flow:C(Hour)[T.5]      -0.0001    1.5e-05     -7.490      0.000      -0.000
-8.3e-05
Order_Flow:C(Hour)[T.6]      -0.0001   1.43e-05     -8.869      0.000      -0.000
-9.86e-05
Order_Flow:C(Hour)[T.7]      -0.0002   1.47e-05    -10.722      0.000      -0.000
-0.000
Order_Flow:C(Hour)[T.8]      -0.0002   1.37e-05    -11.353      0.000      -0.000
-0.000
Order_Flow:C(Hour)[T.9]      -0.0002   1.36e-05    -11.176      0.000      -0.000
-0.000
Order_Flow:C(Hour)[T.10]     -0.0002   1.38e-05    -10.996      0.000      -0.000
-0.000
Order_Flow:C(Hour)[T.11]     -0.0001   1.39e-05    -10.763      0.000      -0.000
-0.000
Order_Flow:C(Hour)[T.12]     -0.0001   1.41e-05     -9.985      0.000      -0.000
-0.000
Order_Flow:C(Hour)[T.13]     -0.0002   1.35e-05    -13.571      0.000      -0.000
-0.000
Order_Flow:C(Hour)[T.14]     -0.0001   1.45e-05     -8.416      0.000      -0.000
-9.33e-05
Order_Flow:C(Hour)[T.15]     -0.0001   1.59e-05     -6.730      0.000      -0.000
-7.61e-05
Order_Flow:C(Hour)[T.17] -2.901e-05   8.98e-05     -0.323      0.747      -0.000
0.000
Order_Flow:C(Hour)[T.18]  1.122e-05   3.15e-05      0.356      0.722   -5.06e-05
7.3e-05
Order_Flow:C(Hour)[T.19] -5.483e-05   2.13e-05     -2.572      0.010   -9.66e-05
-1.3e-05
Order_Flow:C(Hour)[T.20] -6.006e-05   1.81e-05     -3.326      0.001   -9.54e-05
-2.47e-05
Order_Flow:C(Hour)[T.21] -1.591e-05   3.02e-05     -0.527      0.598   -7.51e-05
4.33e-05
Order_Flow:C(Hour)[T.22]  1.588e-05   2.13e-05      0.745      0.456   -2.59e-05
5.77e-05
Order_Flow:C(Hour)[T.23] -2.921e-06   2.04e-05     -0.143      0.886   -4.29e-05
3.7e-05
==============================================================================
Omnibus:                     1359.776   Durbin-Watson:                   2.111
```

```
Prob(Omnibus):                  0.000   Jarque-Bera (JB):            28551.215
Skew:                          -0.665   Prob(JB):                         0.00
Kurtosis:                      14.096   Cond. No.                     2.22e+04
==============================================================================

Notes:
[1] Standard Errors are heteroscedasticity robust (HC1)
[2] The condition number is large, 2.22e+04. This might indicate that there are
strong multicollinearity or other numerical problems.

Brent Results:
                            OLS Regression Results
==============================================================================
Dep. Variable:                Delta_P   R-squared:                       0.530
Model:                            OLS   Adj. R-squared:                  0.526
Method:                 Least Squares   F-statistic:                     48.78
Date:                Sat, 27 Sep 2025   Prob (F-statistic):               0.00
Time:                        02:20:02   Log-Likelihood:                 9741.2
No. Observations:                5106   AIC:                        -1.939e+04
Df Residuals:                    5058   BIC:                        -1.907e+04
Df Model:                          47
Covariance Type:                  HC1
=================================================================================
===========
                              coef    std err          z      P>|z|      [0.025
0.975]
---------------------------------------------------------------------------------
-----------
Intercept                   0.0035      0.001      2.411      0.016       0.001
0.006
C(Hour)[T.1]               -0.0046      0.003     -1.640      0.101      -0.010
0.001
C(Hour)[T.2]               -0.0064      0.003     -2.354      0.019      -0.012
-0.001
C(Hour)[T.3]               -0.0010      0.003     -0.342      0.732      -0.007
0.005
C(Hour)[T.4]               -0.0030      0.003     -0.987      0.324      -0.009
0.003
C(Hour)[T.5]               -0.0059      0.003     -2.167      0.030      -0.011
-0.001
C(Hour)[T.6]               -0.0016      0.003     -0.610      0.542      -0.007
0.004
C(Hour)[T.7]               -0.0103      0.003     -3.211      0.001      -0.017
-0.004
C(Hour)[T.8]               -0.0017      0.004     -0.432      0.666      -0.009
0.006
C(Hour)[T.9]               -0.0049      0.003     -1.616      0.106      -0.011
0.001
C(Hour)[T.10]               0.0025      0.004      0.653      0.514      -0.005
0.010
C(Hour)[T.11]              -0.0061      0.003     -2.353      0.019      -0.011
-0.001
C(Hour)[T.12]              -0.0032      0.003     -1.180      0.238      -0.008
0.002
C(Hour)[T.13]              -0.0024      0.003     -0.901      0.367      -0.007
0.003
```

C(Hour)[T.14]	-0.0053	0.002	-2.587	0.010	-0.009	-0.001
C(Hour)[T.15]	-0.0033	0.002	-1.676	0.094	-0.007	0.001
C(Hour)[T.16]	-0.0002	0.003	-0.070	0.944	-0.006	0.005
C(Hour)[T.17]	-0.0193	0.016	-1.206	0.228	-0.051	0.012
C(Hour)[T.18]	0.0021	0.006	0.354	0.724	-0.010	0.014
C(Hour)[T.19]	-0.0004	0.003	-0.154	0.878	-0.005	0.005
C(Hour)[T.20]	-0.0066	0.003	-2.617	0.009	-0.011	-0.002
C(Hour)[T.21]	-0.0033	0.002	-1.503	0.133	-0.008	0.001
C(Hour)[T.22]	-0.0019	0.002	-0.990	0.322	-0.006	0.002
C(Hour)[T.23]	-0.0039	0.002	-2.129	0.033	-0.008	-0.000
Order_Flow	0.0002	1.68e-05	8.958	0.000	0.000	0.000
Order_Flow:C(Hour)[T.1]	-8.605e-06	2.08e-05	-0.414	0.679	-4.94e-05	3.22e-05
Order_Flow:C(Hour)[T.2]	-2.347e-05	1.89e-05	-1.244	0.214	-6.05e-05	1.35e-05
Order_Flow:C(Hour)[T.3]	-8.994e-05	1.81e-05	-4.970	0.000	-0.000	-5.45e-05
Order_Flow:C(Hour)[T.4]	-4.655e-05	1.86e-05	-2.507	0.012	-8.3e-05	-1.02e-05
Order_Flow:C(Hour)[T.5]	-4.594e-05	1.9e-05	-2.415	0.016	-8.32e-05	-8.65e-06
Order_Flow:C(Hour)[T.6]	-5.113e-05	1.97e-05	-2.595	0.009	-8.97e-05	-1.25e-05
Order_Flow:C(Hour)[T.7]	-9.317e-05	1.88e-05	-4.943	0.000	-0.000	-5.62e-05
Order_Flow:C(Hour)[T.8]	-7.865e-05	1.84e-05	-4.273	0.000	-0.000	-4.26e-05
Order_Flow:C(Hour)[T.9]	-5.882e-05	1.75e-05	-3.362	0.001	-9.31e-05	-2.45e-05
Order_Flow:C(Hour)[T.10]	-7.742e-05	1.81e-05	-4.289	0.000	-0.000	-4.2e-05
Order_Flow:C(Hour)[T.11]	-6.209e-05	1.87e-05	-3.327	0.001	-9.87e-05	-2.55e-05
Order_Flow:C(Hour)[T.12]	-1.69e-05	2.16e-05	-0.782	0.434	-5.93e-05	2.55e-05
Order_Flow:C(Hour)[T.13]	-9.661e-05	1.88e-05	-5.146	0.000	-0.000	-5.98e-05
Order_Flow:C(Hour)[T.14]	4.137e-05	2.43e-05	1.700	0.089	-6.32e-06	8.91e-05
Order_Flow:C(Hour)[T.15]	2.838e-05	5.03e-05	0.564	0.573	-7.02e-05	0.000
Order_Flow:C(Hour)[T.16]	0.0007	0.000	5.604	0.000	0.000	0.001
Order_Flow:C(Hour)[T.17]	7.632e-05	0.000	0.451	0.652	-0.000	0.000

```
Order_Flow:C(Hour)[T.18]     0.0003   7.27e-05      3.490      0.000       0.000
0.000
Order_Flow:C(Hour)[T.19]  5.777e-05    5.3e-05      1.089      0.276   -4.62e-05
0.000
Order_Flow:C(Hour)[T.20]  1.172e-05   2.57e-05      0.457      0.648   -3.86e-05
6.2e-05
Order_Flow:C(Hour)[T.21] -3.144e-06   2.83e-05     -0.111      0.911   -5.86e-05
5.23e-05
Order_Flow:C(Hour)[T.22]  1.629e-05   2.19e-05      0.746      0.456   -2.65e-05
5.91e-05
Order_Flow:C(Hour)[T.23]  8.378e-06   3.57e-05      0.235      0.814   -6.16e-05
7.83e-05
==============================================================================
Omnibus:                      882.623   Durbin-Watson:                   2.123
Prob(Omnibus):                  0.000   Jarque-Bera (JB):            12859.717
Skew:                          -0.375   Prob(JB):                         0.00
Kurtosis:                      10.738   Cond. No.                     1.15e+04
==============================================================================

Notes:
[1] Standard Errors are heteroscedasticity robust (HC1)
[2] The condition number is large, 1.15e+04. This might indicate that there are
strong multicollinearity or other numerical problems.
```

Hypothesis 2 Interpretation

To test if there is an intraday difference for λ, I ran separate regressions for Brent and WTI with hourly interactions. For both benchmarks, the baseline λ is 0.0002 ($p < 0.001$).

In WTI, λ is significantly smaller during trading hours, reflecting deeper liquidity and concentrated speculative activity. Notably, the significance of the interaction term clusters during U.S. market open despite futures trading around the clock most days of the week. It's worth mentioning that Hour 17 behaves strangely likely due to it being the opening hour on Sundays. Brent shows a similar but less consistent pattern, consistent with its role as a global benchmark with more dispersed liquidity across time zones.

Overall, the results support the hypothesis that λ is intraday-dependent and smallest during core trading hours, with the effect strongest in WTI where speculative trading dominates.

In [18]:
```
# 3) λ decreases near expiry and increases in farther-out contracts
# set contract expiry dates
expiry_dates = {
    "WTI": pd.Timestamp("2025-09-22"),  # CLV25 - Crude Oil October 2025 Futures
    "Brent": pd.Timestamp("2025-10-31") # CBZ25 - Crude Oil December 2025 Future
}

# calculate daystoexpiry
analysis_data["Expiry"] = analysis_data["Benchmark"].map(expiry_dates)
analysis_data["DaysToExpiry"] = (analysis_data["Expiry"] - analysis_data["Time"]

# subset by benchmark
wti_data = analysis_data[analysis_data["Benchmark"] == "WTI"].copy()
brent_data = analysis_data[analysis_data["Benchmark"] == "Brent"].copy()
```

```python
# WTI
model_wti_expiry = smf.ols("Delta_P ~ Order_Flow * DaysToExpiry", data=wti_data)
print("WTI Results:")
print(model_wti_expiry.summary())

# Brent
model_brent_expiry = smf.ols("Delta_P ~ Order_Flow * DaysToExpiry", data=brent_d
print("\nBrent Results:")
print(model_brent_expiry.summary())
```

```
WTI Results:
                            OLS Regression Results
==============================================================================
Dep. Variable:                Delta_P   R-squared:                       0.512
Model:                            OLS   Adj. R-squared:                  0.512
Method:                 Least Squares   F-statistic:                     170.8
Date:                Sat, 27 Sep 2025   Prob (F-statistic):          6.96e-106
Time:                        02:20:02   Log-Likelihood:                 10073.
No. Observations:                5487   AIC:                        -2.014e+04
Df Residuals:                    5483   BIC:                        -2.011e+04
Df Model:                           3
Covariance Type:                  HC1
================================================================================
==========
                             coef    std err          z      P>|z|      [0.025
0.975]
--------------------------------------------------------------------------------
----------
Intercept                 -0.0007      0.002     -0.425      0.671      -0.004
0.002
Order_Flow              3.844e-05   5.48e-06      7.016      0.000    2.77e-05
4.92e-05
DaysToExpiry            4.581e-05   6.76e-05      0.678      0.498   -8.66e-05
0.000
Order_Flow:DaysToExpiry 2.902e-07   2.49e-07      1.167      0.243   -1.97e-07
7.77e-07
==============================================================================
Omnibus:                     1272.649   Durbin-Watson:                   2.064
Prob(Omnibus):                  0.000   Jarque-Bera (JB):            58772.659
Skew:                          -0.229   Prob(JB):                         0.00
Kurtosis:                      19.027   Cond. No.                     6.02e+04
==============================================================================

Notes:
[1] Standard Errors are heteroscedasticity robust (HC1)
[2] The condition number is large, 6.02e+04. This might indicate that there are
strong multicollinearity or other numerical problems.

Brent Results:
                            OLS Regression Results
==============================================================================
Dep. Variable:                Delta_P   R-squared:                       0.516
Model:                            OLS   Adj. R-squared:                  0.516
Method:                 Least Squares   F-statistic:                     295.4
Date:                Sat, 27 Sep 2025   Prob (F-statistic):          8.40e-177
Time:                        02:20:02   Log-Likelihood:                 9664.0
No. Observations:                5106   AIC:                        -1.932e+04
Df Residuals:                    5102   BIC:                        -1.929e+04
Df Model:                           3
Covariance Type:                  HC1
================================================================================
==========
                             coef    std err          z      P>|z|      [0.025
0.975]
--------------------------------------------------------------------------------
----------
```

```
Intercept                  -0.0026      0.004     -0.632      0.527     -0.011
0.005
Order_Flow                 -0.0001   2.09e-05     -5.524      0.000     -0.000
-7.44e-05
DaysToExpiry             4.214e-05   6.63e-05      0.635      0.525   -8.79e-05
0.000
Order_Flow:DaysToExpiry  3.435e-06    3.6e-07      9.551      0.000    2.73e-06
4.14e-06
==============================================================================
Omnibus:                     1030.150   Durbin-Watson:                   2.108
Prob(Omnibus):                  0.000   Jarque-Bera (JB):            20016.987
Skew:                          -0.437   Prob(JB):                         0.00
Kurtosis:                      12.660   Cond. No.                     2.02e+05
==============================================================================

Notes:
[1] Standard Errors are heteroscedasticity robust (HC1)
[2] The condition number is large, 2.02e+05. This might indicate that there are
strong multicollinearity or other numerical problems.
```

Hypothesis 3 Interpretation

Lastly, to test if λ decreases near contract expiry, I found the contract expiration date online for the respective contracts and calculated days to expiry for each. I then subsetted my data for each benchmarks separately.

I regressed Delta_P on an interaction term between order_flow and days to expiry and found that interestingly, the effect is significant for Brent ($p < 0.001$) but not WTI ($p < 0.243$). Both coefficients are positive, which is the relationship I hypothesized: as days to expiry increases, lambda increases. There is a positive correlation between days to expiry and λ, which also means that as days to expiry decrease, λ decreases.

The differences between the two benchmarks is interesting. One potential explanation for this difference is: WTI experiences more hedging and speculation activity, which increases liquidity relatively to Brent. If this activity is high enough, it can mask the effects of contract expiry.