PRICE OVERREACTIONS IN CRUDE OIL FUTURES

NONPARAMETRIC IDENTIFICATION OF NEWS-DRIVEN PRICE DISCONTINUITIES AND THEIR ROLE IN SHORT-TERM OVERREACTIONS

FELIX JOHANSSON

SAGA TRAN

Bachelor Thesis

Stockholm School of Economics

2025

Price overreactions in crude oil futures: Nonparametric identification of news-driven price discontinuities and their role in short-term overreactions

Abstract

This thesis examines whether crude oil futures exhibit short-term overreactions to news events and if it is possible to develop a profitable trading strategy based on this pattern. Using hourly West Texas Intermediary crude oil price data, price jumps are identified through a nonparametric method that isolates discontinuities from local continuous volatility. Logistic regression is used to relate jump probability to news characteristics. We find that news volume and sentiment significantly increase the likelihood of price jumps, while linguistic uncertainty has no explanatory power. We then test a rule-based trading strategy exploiting post-jump mean reversion and find that prices often return towards pre-jump levels, which suggests that many jumps are overreactions. The strategy approximately achieves a 43% return over the period January 2024 until September 2025. This thesis extends news-driven mispricing research to the commodity futures market and expands the empirical setting for testing overreactions. It shows how information flows impact short-term price movements in one of the world's most liquid markets.

Keywords

Crude oil futures, Mean reversion, News flow, Overreaction, Price jump

Authors

Felix Johansson (25818)
Saga Tran (25770)

Tutor

Marieke Bos, Associate Professor, Swedish House of Finance, SSE

Examiner

Ramin Baghai, Professor, Department of Finance, SSE

Acknowledgments

We would like to sincerely thank Marieke Bos for her support and advice throughout the project. We also appreciate the constructive discussions during the peer review seminar, which contributed to meaningful improvements to our thesis.

Bachelor Thesis
Bachelor Program in Business and Economics
Stockholm School of Economics

1 Introduction

Crude oil is the most traded commodity worldwide in terms of both volume and notional value. The benchmark US crude oil futures contract is the West Texas Intermediate (WTI) Light Sweet Crude Oil Futures, which trades under the ticker symbol CL. CL contracts trade on the New York Mercantile Exchange (NYMEX) which is open nearly 23 hours per day, from Sunday evening to Friday afternoon [8]. The oil contracts are highly liquid and among the most volatile commodities. Both demand and supply are variable and sensitive to changes, and intraday fluctuations of several percentage points are not uncommon. Unlike for large-cap equities or crops like wheat and cocoa, whose volatility often peak around earnings or a certain season, oil prices are continuously moved by geopolitical events, OPEC policy, inventory and economic data [31].

A central idea in finance is the efficient market hypothesis (EMH), which states that the prices of publicly traded securities (e.g., equities or crude oil futures) reflect all information available, and prices adjust instantly and correctly to new information. According to this hypothesis, securities trade at their fair market value on the exchanges [12]. Another key idea in finance is the random walk theory, which states that security prices follow a random walk, i.e., the price development occur in a random manner and cannot be systematically predicted [23]. The EMH together with random walk theory suggest that the current price is correct, there should be no mispricing, and there should be no room for predictable reversals following for instance an overreaction in the price movement after a news announcement [12], [23].

Criticism to the EMH and random walk theory argue that price movements are neither entirely random nor unpredictable, instead these movements often exhibit patterns such as momentum or overreactions and mean reversion. Furthermore, behavioral finance research challenges the idea that market participants are fully rational and instead present factors such as biases and emotions (e.g., herd behavior) that can create mispricings, bubbles, and ultimately, price reversals [4].

Recent research show that equity prices tend to systematically overreact to news events in the short-term. The price of an equity might increase or decrease several percentage points following an earnings release, only to then make a partial or full reversal back to the level prior to the news announcement [7], [11], [18], [28]. We find this phenomenon interesting as it challenges fully efficient markets, and suggests that market participants overreact to new information in the short-term. It also raises the question about whether these overreactions can be systematically predicted and then ultimately exploited for profit.

In this thesis we study whether crude oil prices systematically overreact to news events and subsequently mean-revert back to a level close to that before the news event was made public. The news events in question include, but are not limited to, inventory reports (e.g., API), OPEC announcements, as well as geopolitical and other macroeconomic events. We adopt the methodology of Jeon et al. [17], who investigate the relationship between news flow and jumps in equity prices. We find evidence that supports the idea that crude oil futures do tend to overreact to news announcements. We explore whether it is possible to exploit this phenomenon by shorting or longing the future during high volatility in connection with the news announcement.

In the rest of this introductory section, we first present some background literature. Secondly, the objective of the thesis is discussed, followed by the scope and limitations. Lastly, we present the thesis outline.

1.1 Background

This section reviews evidence from literature showing that financial markets often overreact to news and subsequently exhibit price reversals. We summarize findings covering the equity market. Finally, we introduce the key study that forms the foundation of our analysis.

Chen et al. [7] discuss how there seem to be mispricing anomalies in how the stock market reacts to positive and negative news announcements. More specifically, they examine the performance of the Chinese stock market and suggest that there is considerable evidence that the market tends to overreact to positive news and underreact to negative news. The driver of these mispricings is argued to be institutional investor's "limited attention" which is biased towards positive news and away from negative news, creating asymmetric response patterns in the markets.

Studies looking at the US stock market suggest that equities tend to overreact to new information, and that there is a particularly strong overreaction and price reversal when the initial surge is caused by media or is followed by a period of stale news, i.e., news whose content substantially overlaps with earlier reports [11], [28]. For example, Engelberg et al. [11] found that stock recommendations on the popular finance TV show "Mad Money" acts as attention shocks to investors and leads to overnight price surges in the upgraded stocks. However, initial spikes in the stock price often reverses during the coming months. The increased price is not sustained and these temporary spikes following the media coverage suggests mispricing. Tetlock [28] finds a similar systemic pattern that is especially prevalent in stocks following stale news. The staleness of news is defined as "its textual similarity to the previous ten stories about the same firm." Tetlock shows that the stocks return on the day of stale news negatively predicts the return during the following week. That is, if the stock jumps one percentage point following the stale news, the expected return during the next week is negative and vice versa. This again supports the idea that market participants temporarily overreact to new information, and that this overreaction is then followed by a price correction.

Kwon [18] documents strong investor overreactions to more extreme news events, such as CEO changes or M&A activities, while they find that less extreme news events, like earnings or dividend announcements tends to be followed by an underreaction. Here, they find that the size of the reversal as well as the trading volume increases with how extreme the news event is.

The idea of systematic overreactions and price reversals following news announcements in the stock market raises the question of whether other assets exhibit similar patterns on completely different markets, such as commodities like crude oil futures (CL contracts), which is the asset that we will focus on in this study. The futures market is different from the equity market in several ways. Mainly, equities represent ownership in a firm, whereas futures are obligations to buy or sell an asset at a later date. Because futures are traded on margin, meaning only a small portion of the contract's notional value is required upfront, they are more leveraged. The leverage has the potential to magnify the P&L impact of short-term price moves [16].

To explore the idea of systematic overreactions and price reversals in CL futures, we replicate the methodology of Jeon et al. [17], who investigate whether flow of information in financial news influences short-term price jumps. Using a firm-level dataset (SME and LC) they identify a price jump according to the definition given in [19]. Thereafter, they relate the occurrence of a jump to different news characteristics, such as news frequency, sentiment intensity and linguistic uncertainty. Jeon et al. find that "these measures of news flow content are significantly related to nonparametric measures of jump intensity and jump-size distributions and explain an important fraction of variations in the jumps across individual companies". This suggests that information flow triggers price jumps. In this thesis, we identify price jumps in CL contracts using the same nonparametric methodology and examine whether the relationship between news flow and jump intensity observed by Jeon et al. also holds for commodity prices.

1.2 Objective

The main objective of the thesis is to examine whether crude oil futures exhibit systematic overreactions to news events, and if it is possible to develop a profitable trading strategy based on this pattern.

To address this objective, we investigate (i) if price jumps occur around news announcements, (ii) if the magnitude of these price jumps depend on the sentiment of the news event and number of news events in the period prior to the jump, and (iii) given the assumption that price jumps tend to revert back towards a level close to that before the news announcement, is it possible to exploit these mean-reversion patterns through a profitable trading strategy.

Our aim is to extend the literature on price overreactions. There is extensive research looking into price overreactions for equity markets [7], [11], [18], [28]. By replicating the methodology of Jeon et al. [17], but using crude oil price data we examine whether similar anomalies arise on other markets. We thus broaden the context in which the overreaction hypothesis is studied.

In addition, we contribute practically by linking and testing theoretical findings in a controlled market environment. We evaluate whether the assumed mean-reversion after a price jump can be exploited by a simple, rule-based trading strategy. Thus, we not only extend the literature theoretically, but we also examine the extent to which theoretical anomalies might persist practically in the market. The implications of this are interesting not only from an academic viewpoint but also from an industry and market participant perspective.

Lastly, our thesis contributes to the ever-lasting debate of an efficient market. If crude oil futures exhibit predictable reversals after price jumps following news events, it could suggest that the prices adjust inefficiently and incorrectly to new information, which is inconsistent with the strong form of the EMH. However, if no predictable pattern is found, the results would support the idea that markets incorporate new information correctly and nearly immediately, thus reinforcing the argument of the EMH. Either case offers valuable and direct evidence for how well a highly liquid market, such as that for CL contracts, works in practice.

1.3 Scope and delimitations

The scope of the thesis is hourly WTI crude oil futures price data, and a selected subset of news sources with coverage between January 2024 and July 2025. Only publicly available, textual news is considered, and sentiment is measured using the Loughran and McDonald [20] financial dictionary, which was originally developed for US SEC 10-K filings and may not perfectly capture journalistic language.

The thesis focuses on the immediate price movements that occur around a detected jump, rather than attempting to explain the full chain of underlying events or market dynamics leading to a price jump. The analysis relies solely on observable information, which means that private signals, high-frequency order flow, and other unobservable drivers of price are outside the scope of this thesis.

Lastly, we ignore frictions, such as transaction costs, slippage and, liquidity constraints that arise in practice.

1.4 Thesis outline

The thesis outline is as follows. Section 2 presents relevant mathematical background on nonparameteric price jump identification. It also makes an argument for why this approach is chosen over the event study. Section 3 explains our practical applications and implementations of the study. It includes both collection and preprocessing of data, as well as the study design. Section 4 presents our results. Section 5 discusses the results, compares them to Jeon et al. [17], and makes an argument for why this can be a useful trading strategy. Lastly, Section 6 provides a conclusion and a discussion about future work.

2 Theoretical Background

This section outlines the differences between the approach of Jeon et al. [17] and an event study, and makes an argument for using the former. Secondly, it clarifies the difference between a price jump and an overreaction event and states this thesis' main assumptions. Lastly, it also explains the mathematical foundation for the logistic regression (logit) model, the log-likelihood ratio test, the area under the receiver operating characteristic (ROC) curve (AUC), and the permutation robustness test.

2.1 Event study

An event study measures the effect of an economic event on the value of a firm by comparing realized returns around the event to the returns that would be expected absent that event. It relies on the assumption that markets are efficient, which has the consequence that new public information is reflected in prices quickly. This means that the economic or financial impact can be derived from price changes over a short time window, rather than having to wait for financial reports to give a long-run overview. Event studies have been widely applied to both firm-specific and market-wide announcements, such as earnings announcements, debt or equity issues, M&A, regulatory changes, and macro releases [6].

2.1.1 Mathematical foundations

We define the event $e \in \{1, \dots, N\}$ as a scheduled news announcement. We also define three non-overlapping windows: an event window around the event, an estimation window prior to the event, and a post event window after the event, as illustrated in Figure 1.

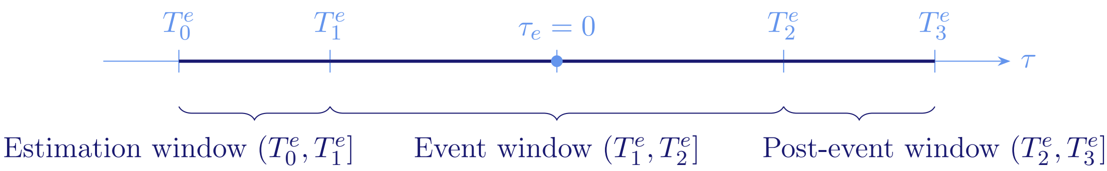

Figure 1: Event study timeline with estimation, event, and post-event windows around the event time $\tau_e = 0$.

The estimation windows is used to fit the normal return model, that is, the model that measures the returns that would be expected absent that event. Let τ_e be the event time for event e. Assume $\tau_e = 0$ (without loss of generality) to be the official news release timestamp of event e. Choose integers $T_0^e < T_1^e < \tau_e < T_2^e < T_3^e$ and let:

- $\tau_e \in (T_0^e, T_1^e]$ be the estimation window,
- $\tau_e \in (T_1^e, T_2^e]$ with $\tau_e \in (T_1^e, T_2^e]$ be the event window, and
- $\tau_e \in (T_2^e, T_3^e]$ be the post event window.

Note that T_0^e, T_1^e are negative integers and T_1^e, T_2^e are positive integers [6]. There are two common choices when choosing the normal return model: the constant mean return (CMR) model and the market model [22]. Since this thesis uses intraday crude oil prices, and aims to identify whether they systematically overreact to news events, it is expected that absent the news event prices would remain fairly constant under such a short time window. Therefore, the CMR is the better option for the normal return model in this setting.

In order to define abnormal returns, let p_t be the price at timestamp t, and let $R_t = \log p_t - \log p_{t-1}$ be the hourly log returns. The abnormal returns for an event e is defined as [22]

$$AR_{\tau_e} = R_{\tau_e} - \mu_e \quad \text{for } \tau_e \in (T_1^e, T_2^e] \quad \text{where } \mu_e = \frac{1}{T_2^e - T_1^e} \sum_{t=T_0^e}^{T_1^e} R_t. \tag{1}$$

2.1.2 Event study for unscheduled new releases

Scheduled news events, like the OPEC announcements, have a single release time. This makes them appropriate to use in an event study. However, when the news announcement comes from a news outlet, there are no assurances that the market was not informed prior to that announcement and has already acted on the information [6]. Moreover, announcements can differ by minutes between different outlets, be leaked prior to the official announcement, or even be a consequence of prior price moves. This can lead to the event window capturing both event anticipation and post event actions rather than one discrete shock. MacKinlay [22] stresses that "an important characteristic of a successful event study is the ability to identify precisely the date of the event". Therefore, another approach is to use a non-parametric jump test for identification of overreaction events in crude oil prices.

2.2 Non-parametric jump test

A non-parametric jump test detects sudden, large changes (or discontinuities) in asset prices directly from the returns, without specifying a parametric model for the jump process. This means that the test is not assuming any distribution for how the jumps occur. The idea is to standardize the return over a short interval and compare it to how much the price actually moves during that period. If the observed move is much larger than what is typical, the move is classified as a discontinuity. This lets us identify arrival times and size of price discontinuities from prices alone [19]. In contrast, a classic event study focuses on known events and examines returns around the scheduled time of that event [22].

2.2.1 Intuition behind the Non-parametric jump test

A single observed return can be large, either because a jump occurred or because the market's spot volatility happened to be high at that exact time point. So, making the assumption that a large return implies a price jump would confuse high-volatility periods with true jumps. Therefore, Lee and Mykland [19] standardizes the return by using a measure that explains the local variation excluding any discontinuities. The authors refer to this measure as "instantaneous volatility". The test statistic then becomes a ratio between the log-returns and the instantaneous volatility over a certain time period. The larger the ratio is, the more likely it is that there has been a price jump.

2.2.2 The $\mathcal{L}-$test Statistic

Suppose that there is a fixed time horizon T and let n be the number of observations in $[0, T]$. Let $\Delta t = \frac{T}{n}$ define the distance between two consecutive observations, that is, we have divided the time period into n equal intervals of length Δt. Let K be the window length and let $p(t_i)$ be the asset price at t_i.

Definition 1. *For each time t_i, define*

$$\mathcal{L}(t_i) = \frac{\log p(t_i) - \log p(t_{i-1})}{\hat{\sigma}(t_i)} \tag{2}$$

where

$$\hat{\sigma}(t_i)^2 = \frac{1}{K-2} \sum_{j=i-K+2}^{i-1} \left| \log \frac{p(t_j)}{p(t_{j-1})} \right| \left| \log \frac{p(t_{j-1})}{p(t_{j-2})} \right|. \tag{3}$$

The above definition by Lee and Mykland [19] uses the realized bipower variation, that is, the sum of products of consecutive absolute returns over the window immediately prior to t_i. This estimator stays consistent even if jumps occur, and therefore captures continuous volatility level (the volatility level excluding discontinuities) without being distorted by jumps.

If there is no jump in the specific interval being tested $(t_{i-1}, t_i]$, $\mathcal{L}(t_i)$ follows an asymptotic normal distribution with $\mathcal{L}(t_i) \sim \mathcal{N}(0, \frac{1}{c^2})$, where $c = \sqrt{\frac{2}{\pi}}$. If the distance between t_{i-1} and t_i is small enough (the sampling is happening with a small enough time window), $\mathcal{L}(t_i)$ is asymptotically independent over time [19].

To retain the benefit of using bipower variation, the window length K must grow with the sampling so that the jump effects in window K vanishes. Lee and Mykland [19] recommends $K = 78$ for intraday hourly time series data.

2.3 Clarification on overreaction events

In this thesis, it is specifically important to distinguish between a price jump and an overreaction event. A price jump refers to an unusually large and abrupt change in the asset price over a very short time interval. Lee and Mykland [19] has constructed a non-parametric jump detection test. Note that this test is backward looking, i.e., it only uses information up to the moment of the suspected jump. An overreaction event, however, is an economic or financial interpretation: it occurs when the price moves too far relative to its baseline and subsequently reverses toward its mean level, a pattern documented in the literature on short-term price reactions to news (see, e.g., [28]).

The central assumption of the thesis is that many such reversals are initiated by an initial discontinuity in prices, that is, a price jump. In other words, if a genuine overreaction occurs, it must first manifest as a jump. This assumption allows us to use the statistically identified jumps as proxies for potential overreaction events. By pairing these jump occurrences with news flow and news characteristics, we can examine whether the emergence of overreactions in crude oil prices is systematically related to the flow, tone, or uncertainty of news.

2.4 Logistic regression

Logistic regression is a type of generalized linear model (GLM) designed for binary dependent variables. It models the conditional probability that an outcome $Y_i \in 0, 1$, where $i = 1, \dots, N$ occurs as a nonlinear function of explanatory variables $X_i = \{x_{i1}, \dots, x_{ip}\}$ for $i = 1, \dots, N$. Unlike the linear regression model, which can predict probabilities outside the $[0, 1]$ range, logistic regression uses the logit link to ensure that the predicted probabilities remain bounded between 0 and 1. Let $p_i = \mathbb{P}(Y_i = 1 \mid X_i)$ denote the conditional probability that the event $Y_i = 1$ occurs. The logistic regression model is defined as

$$\log\left(\frac{p_i}{1 - p_i}\right) = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_p x_{ip}, \tag{4}$$

where the left-hand side is the log-odds or logit of the probability of success [27]. The model can be inverted to express p_i as

$$p_i = \frac{1}{1 + \exp\Big(-(\beta_0 + \beta_1 x_{i1} + \cdots + \beta_p x_{ip})\Big)}. \tag{5}$$

Each coefficient β_j represents the expected change in the log-odds of the event for a one-unit increase in x_j, holding other variables constant. The exponential coefficient e^{β_j} can thus be interpreted as the odds ratio, i.e., the change in odds associated with a one-unit increase in x_j. The parameters $\boldsymbol{\beta} = \{\beta_0, \dots, \beta_p\}$ are estimated by maximum likelihood. The likelihood function for a sample of n independent observations is given by

$$\ell(\boldsymbol{\beta}) = \ell(\beta_0, \dots, \beta_p) = \prod_{i=1}^{n} p_i^{y_i}(1 - p_i)^{1 - y_i}. \tag{6}$$

The corresponding log-likelihood is given by

$$\log \ell(\boldsymbol{\beta}) = \sum_{i=1}^{n} \left[y_i \log(p_i) + (1 - y_i)\log(1 - p_i)\right]. \tag{7}$$

The estimates $\hat{\boldsymbol{\beta}} = \{\hat{\beta}_0, \ldots, \hat{\beta}_p\}$ are found by choosing the values that maximizes the log-likelihood. This is typically done through numerical optimization. Inference is based on that, for large samples, the coefficients $\hat{\boldsymbol{\beta}}$ are asymptotically normally distributed, $\hat{\boldsymbol{\beta}} \sim \mathcal{N}(\boldsymbol{\beta}, \widehat{\sigma}(\hat{\boldsymbol{\beta}}))$. The covariance matrix $\widehat{\sigma}(\hat{\boldsymbol{\beta}})$ is estimated using the observed information matrix or robust heteroskedasticity and autocorrelation consistent (HAC) corrections when serial correlation or heteroskedasticity is present [27].

The model's fit is usually assessed using McFadden's pseudo-$\mathcal{R}^2$ measure, the log-likelihood ratio and discrimination metrics such as the area under the ROC curve (AUC)[5], [24], [26]. These complement significance tests for individual coefficients and help evaluate both explanatory and predictive performance. To test robustness, a valuable test is the permutation robustness test [15]. See below for a brief explanation of each test, except for the McFadden's pseudo-$\mathcal{R}^2$ measure which is defined as

$$\mathcal{R}^2 = 1 - \frac{\log L_1}{\log L_0}. \tag{8}$$

Here, L_0 is the null model only containing the intercept and L_1 is the full model including all predictors [24].

In summary, the logistic regression framework is well suited for this thesis's binary outcome, i.e., to distinguish between jump and non-jump outcomes Y_i, since it captures how continuous and discrete predictors (e.g., news intensity, tone, and uncertainty) jointly affect the conditional probability of a price jump.

2.4.1 The log-likelihood ratio test

The log-likelihood ratio test checks whether adding the news variables makes the logistic regression model explain the data better than a model without them. It does so by comparing the likelihood of the null model L_0 (that contains only an intercept) to the likelihood of the full model L_1 that includes the news-based predictors. The test statistic is computed according to

$$-2 \log \frac{L_0}{L_1} = -2(\log L_0 - \log L_1). \tag{9}$$

If the full model has a higher likelihood, the test statistic (9) becomes larger. Under the null hypothesis that the news variables have no effect, this statistic follows a χ^2-distribution. A statistically significant result indicates that including the news variables improves the model's ability to fit the price data [5].

2.4.2 The receiver operating characteristic curve

The ROC curve is a graphical tool for evaluating the discriminatory ability of a binary classifier. It plots the true positive rate against the false positive rate for all possible classification thresholds, thereby tracing out how the sensitivity and specificity of a model change as the decision boundary varies. A classifier with no discriminative power produces a curve that lies along the diagonal line, while a curve bending toward the upper-left corner indicates increasing ability to correctly distinguish positive from negative cases. The area under the ROC curve (AUC) summarizes this performance numerically and can be interpreted as the probability that the model assigns a higher score to a randomly chosen positive instance than to a randomly chosen negative one [26].

2.4.3 The permutation robustness test

The permutation test is another non-parametric procedure used to assess whether an observed association between predictors and a dependent variable could have arisen purely by chance [15].

To implement this test, the model is first fitted to the original data and a test statistic. Throughout this thesis, AUC will be used as the test statistic. The dependent variable is then repeatedly and randomly permuted while the predictors remain fixed, and the model is re-fitted for each permutation. This produces an empirical null-distribution that reflects how the model would perform if no true relationship exists. If the observed AUC substantially exceeds this null distribution, it provides strong evidence that the predictive power of the model is not driven by random variation but by systematic information in the news-based variables [15].

3 Methodology

First, the research process for the thesis is outlined. Second, we present the data collection and preprocessing steps for both the crude oil price data and the news data. Data validity and reliability is then considered. Third, we present the jump detection methodology in which we define three different versions of the jump detector. Lastly, we present the study design and the evaluation framework.

3.1 Research process

This study follows an empirical research process consisting of four main stages: (i) data collection, (ii) data preprocessing, (iii) model construction, and (iv) evaluation. We collect two primary datasets: hourly crude oil future prices from NYMEX and a large-scale financial and general news corpus from HuggingFace [13], [29]. The crude oil data provide the basis for constructing the dependent variable (the price jumps), and the news corpus serves as the source for the independent variables news volume, tone, and uncertainty. Second, both datasets are preprocessed to align temporarily. This includes aggregating news to the hourly frequency, handling missing observations, filtering relevant news, and computing sentiment measures using the Loughran and McDonald [21] Master Dictionary. The statistical model is then estimated using logistic regression that links the probability of a jump to news characteristics. Finally, model performance is assessed using HAC standard errors, McFadden's pseudo-$\mathcal{R}^2$, the area under the ROC curve (AUC), as well as the permutation robustness test.

3.2 Data collection and preprocessing

First, we briefly describe the data collection process. Thereafter, the steps applied to process the data are outlined in detail for each of the datasets. Lastly, we shortly discuss the validity and reliability of the data.

3.2.1 Crude oil future price data

This study uses historical data on crude oil futures (WTI, CL1). Note here that CL refers to the generic crude oil futures contract ticker symbol for WTI on NYMEX, while CL1 specifically denotes the front-month (nearest-expiry) WTI futures contract [30]. The data has been gathered from NYMEX via TradingView [29]. The price data sample covers the period 2024-01-01 23.00 to 2025-09-26 16.00 UTC, and includes hourly closing prices in USD, the high and low of each hour and the volume, that is, the number of contracts that exchanged hands per hour. However, to match the news time series, we restrict our sample to the period 2024-01-01 23.00 to 2025-07-03 18:00. Table 1 presents the statistics of the raw data. The price variables (close, high, low) show a spread around a mean of approximately USD 72, with a range between USD 55 and USD 87. With a standard deviation of around USD 6.7, which corresponds to 9% of the mean, we can conclude that the price exhibits quite substantial volatility.

Table 1: Summary statistics of crude oil futures price data

Statistic	close	high	low	Volume
mean	71.89	72.07	71.70	8351
std	6.69	6.68	6.69	9515
min	55.41	56.08	55.12	106
25%	66.89	67.07	66.71	1704
50%	71.40	71.60	71.21	4671
75%	77.55	77.74	77.34	11 856
max	87.46	87.67	87.33	102 784

In order to make sure that the price time series is complete in time (non-trading hours are recorded) and equally spaced (each observation represents the same time interval Δt), we construct an hourly index. From the first to the last timestamp, let $\mathcal{T} = \{t_0, \ldots, t_N\}$ where $\Delta t = t_i - t_{i-1} = 1$ hour (UTC). Re-indexing the time series to $\mathcal{T}$ may lead to possible missing values. Thus, if a consecutive block of missing closing prices is of length smaller than 3 hours, that block is forward-filled with the closing price prior to the missing block. Otherwise, it remains missing. Figure 2 displays the (closing) prices over the full period.

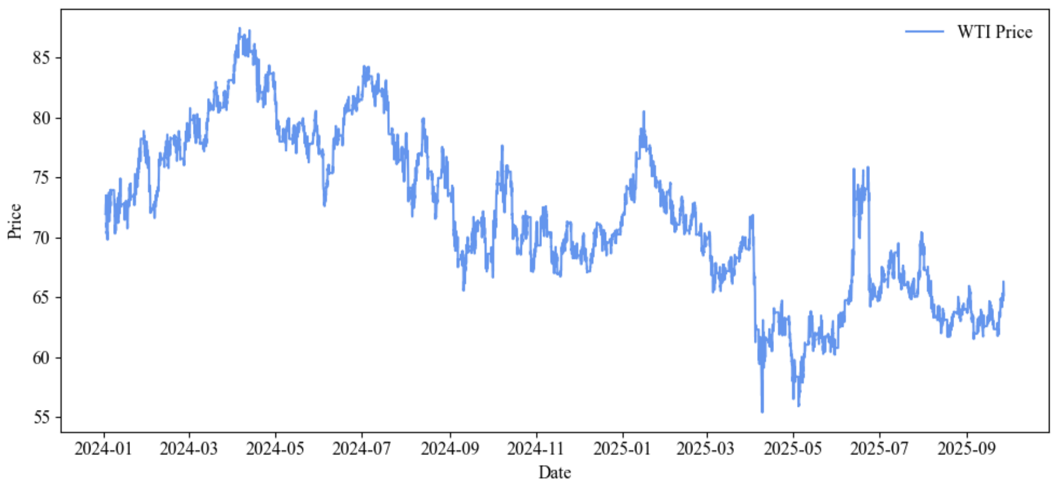

Figure 2: Crude Oil futures price time series

For each observation, the return and log-return is computed from the price p_{t_i} at time t_i according to

$$R(t_i) = \frac{p_{t_i} - p_{t_{i-1}}}{p_{t_{i-1}}}, \quad \log R(t_i) = \log(\frac{p_{t_i} - p_{t_{i-1}}}{p_{t_{i-1}}}). \tag{10}$$

We look for any pricing or return anomalies:

$$\text{bad price}(t_i) = \mathbb{1}\{\, p_{t_i} \leq 0 \ \vee \ p_{t_i} = \infty \,\}\,, \tag{11}$$

$$\text{bad return}(t_i) = \mathbb{1}\{\, |R(t_i)| > 0.5 \ \vee \ |R(t_i)| = \infty \,\}\,. \tag{12}$$

The “bad return” is flagged when the absolute hourly log return is unusually large (above $0.5 \approx 50\%$ price change in one hour). The check identifies no “bad prices” and only one “bad return”. This is expected and not problematic since the single flagged value corresponds to the first row in the re-indexed price series. Because a return requires both p_{t_i} and $p_{t_{i-1}}$, the first row will be NaN.

3.2.2 News data

The Multi-Source Financial & General News corpus is collected from HuggingFace [13], and is a collection of approximately 57.1 million news records over the time period 1990-2025. It joins 24 public news sources into a single-source dataset, where each row represents a news article, containing the following columns:

- UTC ISO-8601 timestamp, yyyy-mm-ddthh:mm:ssz,
- text column typically including a title and an article introduction, and
- extra fields retaining metadata such as URLs, publisher/source names, authors, tickers, categories, origin, and timing details.

To align with the oil price data, the sample is restricted to the time period 2024-01-01 and onward (UTC). Articles outside this interval are discarded before any processing for memory reasons. However, we discovered that the sample became heavily biased to certain days and hours if all news articles between 2024-01-01 23.00 to 2025-09-26 16.00 where included, that is, if we tried to maximize coverage. Therefore, we restricted our sample to the news sources (subsets of the corpus) with the most recent and even coverage. These include The New York Times (NYT), S&P Daily Headlines, and Reddit Finance. The last recorded entry for these subsets are 2025-07-03 18:00. The distribution of news counts over time is displayed in Figure 3.

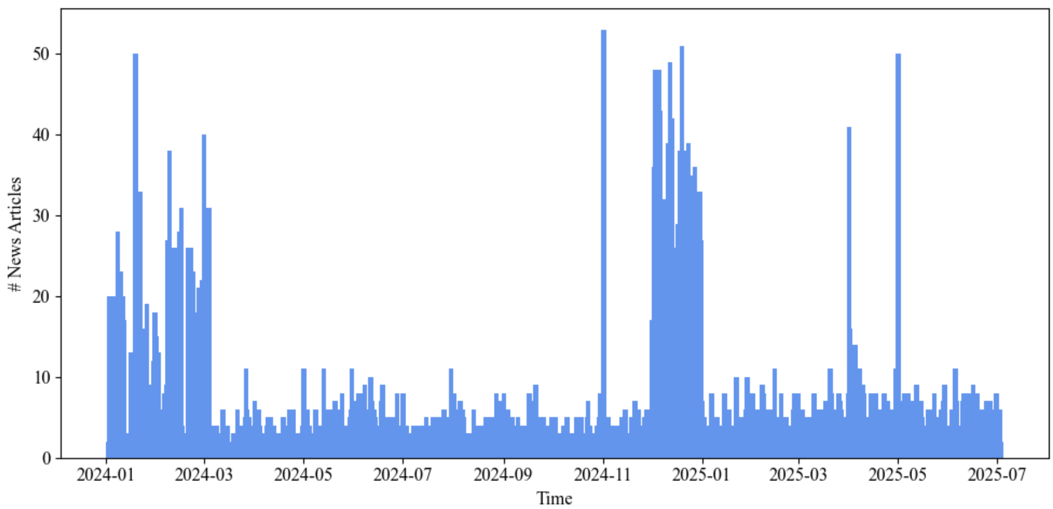

Figure 3: # News articles over time

The articles are read using streaming (row-by-row) and kept if their date string is after 2024-01-01 00:00:00 (UTC). The "extra_fields" column is formatted as a JSON-like string. An AI-based parser is used that

(i) standardizes non-JSON tokens to JSON equivalent ones,

(ii) removes trailing commas,

(iii) attempts JSON parsing, and

(iv) falls back so a (safe) python-literal evaluation when necessary.

The parsed metadata are flattened so that each piece of information has its own column in the dataframe. We then make sure each column in the table has the correct datatype (either integer, date, text, or list). Thereafter, duplicate rows are dropped. Moreover, the "text" column is required to score the article, thus it cannot be empty.

3.2.3 Construction of news flow and sentiment variables

Similar to Jeon et al. [17], we utilize the Loughran and McDonald [21] Master Dictionary, which is a financial-specific word list. It classifies words commonly found in corporate filings into categories among positive, negative, uncertainty. Unlike general sentiment dictionaries, it is tailored to financial language. This is supposed to avoid misclassifications such as treating the word "liability" as negative. In the data, each sentiment column contains a year flag. Positive years (like 2011) marks the year that particular word entered the wordlist, and a negative year (like -2011) means the word was removed in that year. Lastly 0 means a word does not belong to that category. The latest update was made in 2024 [21].

Because our news sample spans two years, we create two sets of words, one as of 2024 and one as of 2025. A word belongs to category c in year Y if and only if it was added on or before year Y and not removed on or before year Y. This ensures that a word that was removed in e.g. 2020 is not counted as positive/ negative/ uncertain.

Let P_Y, N_Y, U_Y define the set of positive, negative, and uncertain words in year Y respectively. For each article, we map the timestamp to a year Y, and the text is tokenized into a set $\mathcal{W} = \{\omega_1, \ldots, \omega_n\}$ of words. Thereafter, the tone and # uncertain words are computed according to

$$\text{Tone} = \text{Pos\%} - \text{Neg\%}, \qquad \text{Unc\%} = \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}\{\omega_i \in U_Y\}, \tag{13}$$

$$\tag{14}$$

where

$$\text{Pos\%} = \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}\{\omega_i \in P_Y\}, \qquad \text{Neg\%} = \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}\{\omega_i \in N_Y\}. \tag{15}$$

Moreover, we remove stopwords (words with little semantic meaning), from the Loughran and McDonald [21] General stop-wordlist, from the text column. Jeon et al. [17] retains all tokens to avoid reducing the apparent tone or uncertainty of longer texts. However, our sample is substantially smaller, which makes the impact of stopwords larger. Removing them ensures that the computed measures are not weakened by frequent but semantically uninformative words.

Each article in the raw dataset is timestamped at the minute or second level, so to link the news data to the price time series, which is sampled at hourly frequency, we aggregate all articles published within each hour. Thus, we round down each timestamp to the nearest full hour. For each hour $t_i \in \mathcal{T} = \{t_1, \ldots, t_N\}$, the score of all articles published during that hour is combined as a weighted average as

$$\text{UncWords}_{t_i} = \begin{cases} 100 \cdot \dfrac{\sum_{a \in \mathcal{A}_{t_i}} n_a \, \text{UncWords}_a}{\sum_{a \in \mathcal{A}_{t_i}} n_a}, & \text{if } N_{t_i} > 0, \\ 0, & \text{if } N_{t_i} = 0, \end{cases} \tag{16}$$

$$\text{AbsTone}_{t_i} = \begin{cases} 100 \cdot \dfrac{\sum_{a \in \mathcal{A}_{t_i}} n_a \, |\text{Tone}_a|}{\sum_{a \in \mathcal{A}_{t_i}} n_a}, & \text{if } N_{t_i} > 0, \\ 0, & \text{if } N_{t_i} = 0. \end{cases} \tag{17}$$

where $\mathcal{A}_{t_i}$ is the set of all articles released in hour t_i and n_a is the number of words of article a. Lastly,

$$\text{NewsCount}_{t_i} = N_{t_i}. \tag{18}$$

Table 2 reports descriptive statistics for the three aggregated hourly news variables NewsCount, UncWords, and AbsTone. They indicate quite sparse and skewed news flow at the hourly level. Note that all variables are zero at the 25th percentile and UncWords is zero even at the 50th percentile.

Table 2: Summary statistics for independent variables

Statistic	NewsCount	UncWords	AbsTone	$\lvert R(t_{i-1})\rvert$
mean	1.73	0.17	0.88	0.18
std	2.88	0.61	1.75	0.29
min	0	0	0	0
25%	0	0	0	0
50%	1.00	0	0.26	0.08
75%	2.00	0.08	1.21	0.25
max	53.00	33.33	40.00	7.68

3.2.4 Data validity and reliability

As NYMEX is the official marketplace for WTI crude oil futures (and the most liquid for crude oil in general) [8], we assume that the recorded prices accurately represent the true transaction prices at each closing hour. The dataset itself consists of regularly time-stamped market data collected directly from the exchange feeds, which ensures that they have high temporal precision. In turn, this produces a minimal risk of reporting or sampling errors. Therefore, potential sources of measurement noise are limited to missing or misaligned observations, and these are identified and flagged during preprocessing.

The Huggingface corpus provides a broad and diverse set of both financial and general news. However, because the dataset merges 24 different sources collected through unknown scraping methods, the consistency of timestamps, article completeness, and metadata formatting cannot be fully verified. Although duplicate removal and type checks improve internal consistency, uneven temporal coverage reduces the reliability over time. Additionally, the use of UTC timestamps may not align perfectly with publication or market times.

The Loughran–McDonald Master Dictionary was originally developed for formal corporate and financial documents, specifically US SEC 10-K filings between 1994 and 2008, to better capture tone in the language of business. The authors created their word lists by examining the most frequent terms in these filings to identify words that typically have financial, rather than general, connotations of sentiment. While the dictionary was not designed for journalistic text, Loughran and McDonald [20] note that it may be applied to other financial documents such as press releases or news articles, as there is no clear reason to expect greater misclassifications in such contexts. Nonetheless, applying a dictionary derived from corporate filings to news may introduce mismatches, since linguistic features are different between regulatory and journalistic writing.

3.3 Jump detection methodology

We have formulated three different versions of the $\mathcal{L}-$statistic. First, using Theorem 1 from Lee and Mykland [19] (see Appendix 7.1.1), then using the same authors Lemma 1 (see Appendix 7.1.2), which we for the last version have calibrated using the Gilder correction [14].

3.3.1 Pointwise jump detector

Theorem 1 of Lee and Mykland [19] gives us that each trading hour is asymptotically independent of the other. Let $\mathcal{L}(t_i) \sim \mathcal{N}(0, \frac{1}{c^2})$ where $c = \sqrt{\frac{2}{\pi}}$. Define $\lambda_\alpha^{\text{pointwise}}$ as the pointwise threshold for flagging a jump. Specifically,

$$\lambda_\alpha^{\text{pointwise}} = \frac{z_{1-\alpha/2}}{c}, \tag{19}$$

where $z_{1-\alpha/2}$ denotes the critical value such that $\mathbb{P}(Z < z_{1-\alpha/2}) = 1 - \frac{\alpha}{2}$ for $Z \sim \mathcal{N}(0,1)$. Note that this is valid since $\mathcal{N}(0, \frac{1}{c^2}) = \frac{1}{c}\mathcal{N}(0,1)$. A jump is flagged when $|\mathcal{L}(t_i)| > \lambda_\alpha^{\text{pointwise}}$. If we set $\alpha = 5\%$ and $\alpha = 1\%$ replicates the $J_0 95$ and $J_0 99$ indicators of Jeon et al. [17].

3.3.2 Block-wise jump detector

A "block" is defined as one trading day, which, in our dataset corresponds to approximately 24 consecutive observations. Within each day, there are therefore $m \approx 24$ separate tests for jumps. Under the pointwise definition of a jump, a 5% significance level refers to there being a 5% chance of getting a false positive. However, when we test all 24 hours in a day, that error increases to roughly $1 - (1 - 0.05)^{24} \approx 70\%$.

To control this error, Lee and Mykland [19] proposes to calibrate the threshold $\lambda_{\tilde{\alpha}}^{\text{block}}$ to the maximum among all no-jump statistics within each block. Let $(1 - \tilde{\alpha})^m = 1 - \alpha$. This means that

$$\tilde{\alpha} = 1 - (1 - \alpha)^{1/m}. \tag{20}$$

Assume that the condition are as under theorem 1 of Lee and Mykland [19]. Then $c\mathcal{L}(t_i) \sim \mathcal{N}(0,1)$. Define $\lambda_{\tilde{\alpha}}^{\text{block}}$ as the blockwise threshold for flagging a jump. Specifically,

$$\lambda_{\tilde{\alpha}}^{\text{block}} = \frac{z_{1-\tilde{\alpha}/2}}{c}, \tag{21}$$

where $z_{1-\tilde{\alpha}/2}$ denotes the critical value such that $\mathbb{P}(Z < z_{1-\tilde{\alpha}/2}) = 1 - \frac{\tilde{\alpha}}{2}$ for $Z \sim \mathcal{N}(0,1)$. A jump is flagged when $|\mathcal{L}(t_i)| > \lambda_{\tilde{\alpha}}^{\text{block}}$.

3.3.3 Monte Carlo simulated jump detector

Gilder et al. [14] re-examined the $\mathcal{L}-$statistic empirically and found that for finite samples, the theoretical threshold λ is too conservative, i.e., it rejects jumps too often. Jeon et al. [17] adopts this "correction" for their second pair ($J95$ and $J99$) of indicators. However, the paper does not present a concrete method for deriving these corrections, but only states that it has applied the "Gilder correction".

Gilder et al. [14] describes the correction as a simulation-based calibration of the original threshold $\lambda_{\tilde{\alpha}}^{\text{block}}$. We have chosen to apply a Monte Carlo simulation in order to calibrate $\lambda_{\tilde{\alpha}}^{\text{block}}$. Define $\lambda_{\tilde{\alpha}}^{\text{MC}}$ as the Monte Carlo (MC) threshold for flagging a jump. We simulate $200,000$ independent draws of $\max|Z_1, \ldots, Z_m|$, where $Z_i \sim \mathcal{N}(0,1)$ and $m = 24$ for our case. We compute the $(1-\alpha)-$quantile $q_{\alpha}^{m=24}$ (see Appendix 7.2.1). Now, we can estimate

$$\lambda_{\tilde{\alpha}}^{\text{MC}} = \frac{q_{\alpha}^{m=24}}{c}. \tag{22}$$

Finally, a jump is flagged when $|\mathcal{L}(t_i)| > \lambda_{\tilde{\alpha}}^{\text{MC}}$.

3.4 Study design

We will adopt the jump detection statistics defined above at the $\alpha = 5\%$ and $\alpha = 1\%$ thresholds.

We test, for each statistic, whether it is reasonable to follow the recommendation by Lee and Mykland [19] of using a volatility window $K = 78$. We test this by plotting the number of detected jumps while varying the window size K. We choose K where the plot is steady.

For each jump hour t_i, we search for news in a look-back window K ending at t_{i-1}, and construct the NewsCount, AbsTone, and UncWords. Following the methodology of Jeon et al. [17],

$$\text{logit}(p_{t_i}) = \beta_0 + \beta_1 \text{NewsCount}_{t_i} + \beta_2 \big|\text{Tone}_{t_i}\big| + \beta_3 \text{UncWords}_{t_i} + \beta_4 \big|R(t_i)\big| + \epsilon_{t_i} \tag{23}$$

estimates the probability of a jump at time t_i. Hence, we can define the null hypothesis:

$\mathbf{H_0}$: $\beta_1 = \beta_2 = \beta_3 = 0$. That is, news intensity and context do not predict jumps.
$\mathbf{H_1}$: $\beta_j \neq 0$ for any $j \in \{1, 2, 3\}$. That is, at least one of the β's is non-zero.

3.5 Evaluation framework

The empirical model in equation (23) is estimated using logistic regression fitted by maximum likelihood using the `statsmodels.Logit` implementation in Python. Following Jeon et al. [17], all explanatory variables are standardized prior to estimation to have zero mean and unit standard deviation. This normalization allows direct comparison of coefficient magnitudes across variables, and ensures that the estimated coefficients capture the effect of a one–standard-deviation change in each variable on the probability of a jump. Because all variables are standardized, each odds ratio shows how the odds of a jump change when the corresponding predictor increases by one standard deviation. For example, an odds ratio of 1.22 means that a one–standard-deviation increase in that variable raises the odds of a jump by 22%.

In Jeon et al. [17], standard errors are clustered at the firm level to account for cross-sectional dependence and serial correlation within each firm's daily time series. This clustering approach allows the residuals to be correlated within firms over time while assuming independence across firms. In our case, we work with a single time series of hourly crude oil prices, where dependence arises over time rather than across entities. To account for this autocorrelation, we apply HAC standard errors with six hourly lags. The HAC estimator gives asymptotically valid inference in the presence of both heteroskedasticity and short-range serial dependence [2].

Statistical significance is assessed using two-sided z-tests based on the HAC covariance matrix. We report both the raw coefficients and their corresponding odds ratios given by

$$\mathrm{OR}_j = e^{\beta_j}, \tag{24}$$

which under the standardized specification indicate how the odds of a jump change for a one–standard-deviation increase in variable j.

Following Jeon et al. [17], we summarize the logistic regression results by presenting coefficient estimates, HAC-adjusted z-statistics, standardized odds ratios and McFadden pseudo-R^2 values. The pseudo-$\mathcal{R}^2$ measures how much better the fitted model explains the data compared to one with only a constant term, providing an indicator of explanatory power. In addition to these standard metrics, we complement the analysis with the log-likelihood ratio test, the AUC, and a permutation robustness test.

4 Results

In this section, we present the results from this study. Note that we will only be presenting the logistic regression results based on the pointwise $\lambda^{\text{pointwise}}_{\alpha=1\%}$ and Monte Carlo Simulated $\lambda^{\text{MC}}_{\tilde{\alpha}=1\%}$ jump indicators at a 1% threshold, since these are the ones used by Jeon et al. [17]. It is concluded that including $\lambda^{\text{block}}_{\tilde{\alpha}}$ will not add any value since it is closely related to $\lambda^{\text{MC}}_{\tilde{\alpha}=1\%}$. Note further that the results at the 5% confidence level are similar but with lower levels of significance.

First, we present the result from the volatility window $K-$analysis for both indicators. Secondly, we present results from the jump detection itself. Thirdly, we will present detailed results from the logit regression including the log-likelihood ratio, AUC, and permutation robustness test.

4.1 Rolling window size

Figures 4 and 5 show the number of detected jumps as the rolling window size K varies from 15 to 200 with a step-size of 5, holding $\lambda^{\text{pointwise}}_{\alpha=1\%}$ and $\lambda^{\text{MC}}_{\tilde{\alpha}=1\%}$ fixed respectively.

Starting with $\lambda^{\text{pointwise}}_{\alpha=1\%}$, the jump count increases steady and peaks around $K \approx 50-55$, and then declines as K grows larger. For $\lambda^{\text{MC}}_{\tilde{\alpha}=1\%}$, the jump count experiences a peak at around 50, followed by a sharp decline.

For small K, the local volatility might be noisy and inflated by large returns in the nearby hours. This, in turn, will cause the $\mathcal{L}-$statistic to under-estimate the number of jumps, since short-lived volatility spikes will increase the denominator in the statistic and thus tone down otherwise significant price moves. On the other hand, for large values of K, the estimate instead becomes too smooth and slow to adapt to changes. This down-weights extreme price moves and in turn reduces the number of detected jumps.

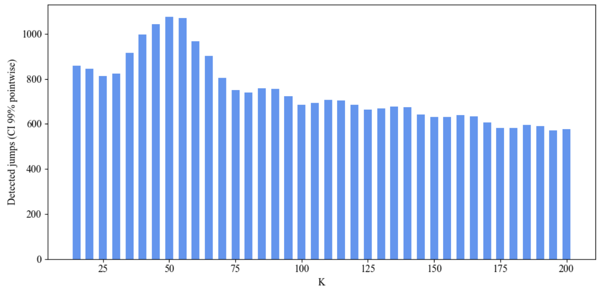

Figure 4: $\lambda^{\text{pointwise}}_{\alpha=1\%}$

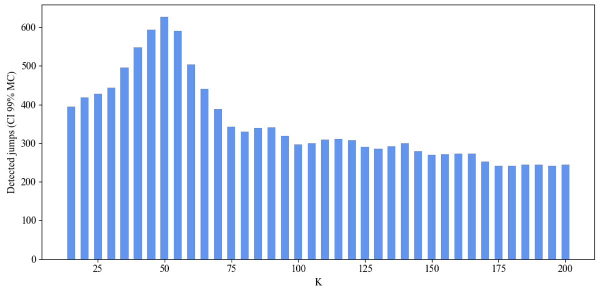

Figure 5: $\lambda^{\text{MC}}_{\tilde{\alpha}=1\%}$

jumps detected based on the value of the volatility window K

Between $K = 70$ and $K = 95$ for both jump indicators, the counts are quite stable. This suggests that the results are not overly sensitive to the exact choice of K within this range. On top of this, the recommended window size by Lee and Mykland [19] for hourly intraday price time series is $K = 78$. Hence, we adopt this value as our baseline.

4.2 Jump detection results

Table 3 summarises the thresholds for the $\mathcal{L}-$statistic and the total number of detected jumps throughout the price time series. The first thing to note is that going from pointwise to block-wise reduces the number of detected jumps quite substantially. This is expected since we move from per-test to per-day control of the error. However, the Monte Carlo simulation reproduces the block-wise thresholds nearly identically. This is a good sign, since they both estimate the same quantile of the daily maximum under the null hypothesis that $c\mathcal{L}(t_i) \sim N(0, 1)$. In other words, this validates our choice of $m = 24$.

Table 3: Jump detection thresholds and counts

Number of observations: 15214						
	$\lambda^{\text{pointwise}}_{\alpha=5\%}$	$\lambda^{\text{pointwise}}_{\alpha=1\%}$	$\lambda^{\text{block}}_{\tilde{\alpha}=5\%}$	$\lambda^{\text{block}}_{\tilde{\alpha}=1\%}$	$\lambda^{\text{MC}}_{\tilde{\alpha}=5\%}$	$\lambda^{\text{MC}}_{\tilde{\alpha}=1\%}$
Threshold	2.4565	3.2283	3.8487	4.4217	3.8501	4.4304
# Jumps Detected	1381	733	496	331	496	328

Notes: λ^J_α where $J = \{\text{pointwise, block, MC}\}$ is the jump threshold for the $\mathcal{L}-$statistic defined in 1. Whenever $\mathcal{L}(t_i) > \lambda^J_\alpha$ one jump is flagged.

Figure 6 shows an arbitrary part of the price time series marked with the detected jumps of $\lambda^{\text{pointwise}}_{\alpha=1\%}$ and $\lambda^{\text{MC}}_{\tilde{\alpha}=1\%}$ respectively. Overall, we can see that both methods successfully identifies the same major price discontinuities. This indicates that the real overreaction events are quite robust under the current choices of hyperparameters $K = 78$ and $m = 24$. However, the pointwise thresholds detects additional jumps that are not confirmed when using the Monte Carlo threshold. Here, the difference between controlling the error of each individual observation versus controlling the error at a daily ($m = 24$) level becomes clear. The $\lambda^{\text{pointwise}}_{\alpha=1\%}$ threshold is less conservative whereas the $\lambda^{\text{MC}}_{\tilde{\alpha}=1\%}$ threshold eliminates jumps that may arise from local volatility noise rather than true discontinuities.

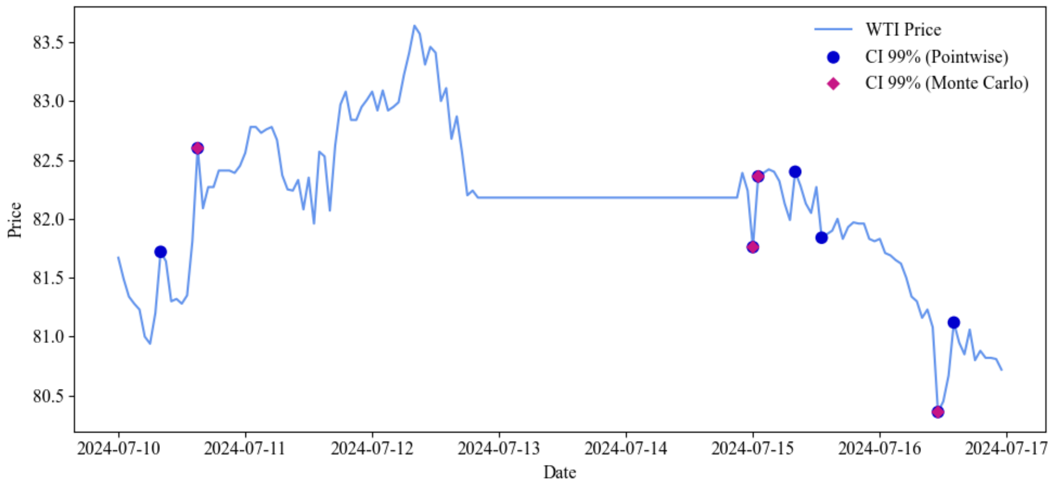

Figure 6: Jumps detected on a subset of the WTI time series using $\lambda^{\text{pointwise}}_{\alpha=1\%}$ and $\lambda^{\text{MC}}_{\tilde{\alpha}=1\%}$

4.3 Logit regression

We are presenting results from running the logit regression using two different jump thresholds, $\lambda^{\text{pointwise}}_{\alpha=1\%}$ and $\lambda^{\text{MC}}_{\tilde{\alpha}=1\%}$. First, we present the results using the Monte Carlo indicator, $\lambda^{\text{MC}}_{\tilde{\alpha}=1\%}$, to determine the whether a jump has occurred or not. Secondly, we will present the same results, but now when using the pointwise indicator $\lambda^{\text{pointwise}}_{\alpha=1\%}$. Finally, we make a comparison of these results and relate these to the output of Jeon et al. [17].

4.3.1 Regression results using $\lambda^{\text{MC}}_{\tilde{\alpha}=1\%}$

Table 4 presents the results of the main regression (see equation (23)). The model is estimated on 13 170 hourly observations, and there was no issues with convergence, i.e., the model was able to find a stable maximum when estimating the log-likelihood. The dependent variable represents whether a price jump has occurred within each hour. Since all explanatory variables were standardized to have mean zero and unit variance, the intercept of -3.87 can be interpreted as the log-odds of a jump when all explanatory variables are at their mean level. This corresponds to a jump probability of approximately

$$p_0 = \frac{e^{-3.87}}{1 + e^{-3.87}} \approx 0.0205 = 2.05\%. \tag{25}$$

The overall fit of the model is quite modest, but statistically significant. The McFadden pseudo$-\mathcal{R}^2$ is 0.0129. The log-likelihood ratio (LLR) test yields a p-value of 4.5×10^{-7}. This confirms that the inclusion of the news significantly improves the explanatory power compared to the intercept-only (null) model. Moreover, the model's ability to discriminate between jump and non-jump hours is assessed using the area under the ROC curve, which equals 0.67. This is above random guessing (AUC=0.5), but still moderate predictive accuracy (see Figure 7).

Table 4: Logistic regression of jump indicator (MC 1% threshold) on news and controls

	HAC (robust) inference					Odds ratios (per 1 SD)	
Variable	Coef.		SE(HAC)	z	p	OR	95% CI
const	-3.8666		0.0677	-57.11	< 0.001	–	–
NewsCount	0.0806	**	0.0337	2.39	0.0166	1.0839	[1.0147, 1.1579]
AbsTone	0.1300	***	0.0461	2.82	0.0048	1.1388	[1.0404, 1.2465]
UncWords	-0.0288		0.0426	-0.68	0.4985	0.9716	[0.8937, 1.0562]
$\lvert R(t_{i-1})\rvert$	0.1771	***	0.0365	4.86	< 0.001	1.1938	[1.1114, 1.2822]
Model summary							
No. observations			13 170				
Pseudo R^2 (McFadden)			0.0129				
Log-likelihood			-1341.5				
LL (null)			-1359.1				
LLR p-value			4.50×10^{-7}				
AUC (in-sample)			0.670				

Notes: Dependent variable is an hourly jump indicator. Predictors are standardized ($\mu = 0$, $\sigma = 1$). HAC standard errors account for heteroskedasticity and serial correlation. Odds ratios (OR) are reported per one standard deviation change in each predictor. The predicted jump probability at the mean values of the independent variables is $\text{logit}^{-1}(-3.8666) \approx 2.05\%$. *Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.*

Turning to the individual coefficients, Table 4 shows that news volume and tone are both positively correlated with the likelihood of a jump. The odds ratio for NewsCount is 1.0839, which means that holding other factors constnat, a one standard deviation increase in the number of articles published within an hour raises the odds of a jump by approximately 8.4%. The corresponding odds ratio for AbsTone is 1.139, implying that a one standard deviation increase in the absolute tone ($= |Pos\% - Neg\%|$) raises the odds of a jump by roughly 13.9%. AbsTone are significant at the 99% level with a p-value of 0.0048. However, NewsCount is just above that threshold (still significant at 95%) with a p-value of 0.0166. On the contrary, the coefficient for UncWords is statistically insignificant since the p-value is almost 0.5.

Lastly, the control variable $|R(t_{i.-1})|$ that represents the absolute lagged return is positive and highly significant. Its odds ratio is 1.194 which implies that higher recent market volatility increases the likelihood of a jump in the next hour by about 19.4%. Since large absolute returns in a short time span is closely related to high short-term volatility, this means that jumps are more likely to occur when the market is already volatile. That is, if recent volatility has been high this increases the likelihood of additional discontinuous price movements.

The histogram of the permutation test, displayed in Figure 8, is concentrated around 0.525. This is as expected under the null hypothesis of no predictive/random relationship. Specifically, none of the permuted models produce an AUC close to the observed value of 0.67. The implied permutation p-value was calculated to 0.001 (that is, significant at the 99% level), which means that the probability of achieving an AUC of 0.67 under random labeling is extremely small.

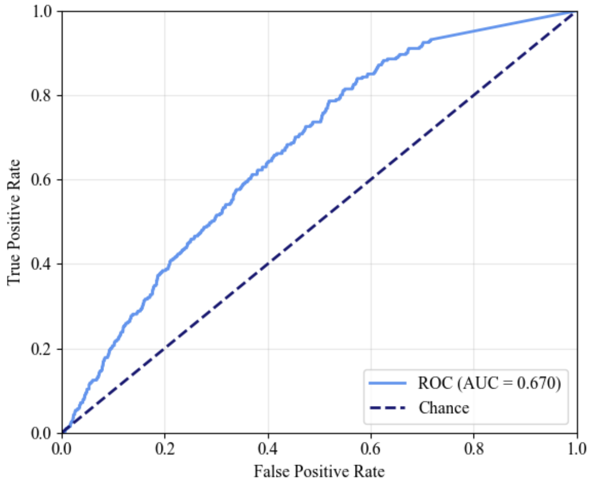

Figure 7: AUC

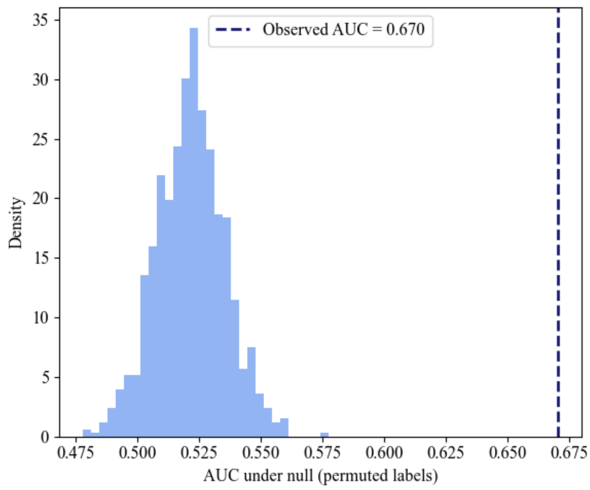

Figure 8: Permutation Robustness test

Overall, the results indicate that both the quantity of news and the intensity of its tone contain information about jump risk that is not captured by past price movements. However, linguistic uncertainty does not appear to play a significant role. The odds ratio of 0.97 implies that the effects are negligible, thus indicating that the frequency of uncertainty-related words does not systematically explain jump occurrences in this set up. Moreover, although the AUC is modest in absolute terms, the permutation test demonstrates that the model captures real and systematic relationship between news count and tone and the likelihood of observing a price jump.

4.3.2 Regression results using $\lambda^{\text{pointwise}}_{\alpha=1\%}$

Table 5 presents the results of using the pointwise 99% threshold. Model convergence was achieved. When all predictors are standardised and at their average level, the intercept corresponds to a jump level of

$$p_0 = \frac{e^{-3.03}}{1+e^{-3.03}} \approx 0.046 = 4.6\%. \tag{26}$$

Compared to the MC-indicator used in the previous regression, the pointwise indicator identifies a larger number of jumps. This is expected since the pointwise threshold, $\lambda^{\text{pointwise}}_{\alpha=1\%}$, is less conservative and thus classifies more price movements as jumps.

The overall fit of the model improves slightly. The McFadden pseudo-$\mathcal{R}^2$ increases from 0.0129 to 0.0194. This suggests that the explanatory variables account for a very minimal, but slightly larger share of the variation in the (binary) jump outcome. This increase indicates that there might be a better correspondence between the explanatory variables and the jumps that are only defined by $\lambda^{\text{pointwise}}_{\alpha=1\%}$. The LLR is highly significant with a p-value of effectively zeros. This confirms that including the news variables and the lagged return significantly improves the model's fit compared to only using an intercept.

The ROC curve in Figure 9 shows a slightly stronger separation between jump and non-jump hours compared with the MC-indicator model. The AUC is 0.682, which is still moderate but noticeably above random guessing.

Table 5: Logistic regression of jump indicator (pointwise 1% threshold) on news and controls

	HAC (robust) inference					Odds ratios (per 1 SD)	
Variable	Coef.		SE(HAC)	z	p	OR	95% CI
const	−3.0327		0.0481	−63.10	< 0.001	–	–
NewsCount	0.0855	***	0.0244	3.51	< 0.001	1.0839	[1.0384, 1.1426]
AbsTone	0.0935	**	0.0370	2.53	0.0115	1.0980	[1.0212, 1.1806]
UncWords	−0.0064		0.0319	−0.20	0.841	1.0064	[0.9454, 1.0714]
$\lvert R(t_{i-1})\rvert$	0.2461	***	0.0316	7.78	< 0.001	1.2791	[1.2022, 1.3608]
Model summary							
No. observations			13 170				
Pseudo R^2 (McFadden)			0.0194				
Log-likelihood			−2 495.4				
LL (null)			−2 544.8				
LLR p-value			1.75×10^{-20}				
AUC (in-sample)			0.682				

Notes: Dependent variable is an hourly jump indicator. Predictors are standardized ($\mu = 0$, $\sigma = 1$). HAC standard errors account for heteroskedasticity and serial correlation. Odds ratios (OR) are reported per one standard deviation change in each predictor. The predicted jump probability at the mean values of the independent variables is $\text{logit}^{-1}(-3.0327) \approx 4.6\%$. *Significance levels:* * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

In terms of the individual coefficients, NewsCount remains positive and significant. It's odds ratio is 1.089. This reinforces that higher informational intensity (more news) is associated with higher probability of a jump. AbsTone is also still (barely) significant at a 99% threshold with a p-value of 0.0115. It's odds ratio is 1.098, which is lower than in the previous test. This is consistent with there being a larger number of jumps, since more jumps may capture not only highly emotional market respones but also more "routine" volatility that is independent of shifts in tone. UncWords remains insignificant, which may mean that the Loughran and McDonald [21] Master Dictionary may not fully translate from corporate filings language to journalistic contexts. Finally, $\lvert R(t_{i.-1})\rvert$ also remains significant and very positive, with an odds ratio of 1.279.

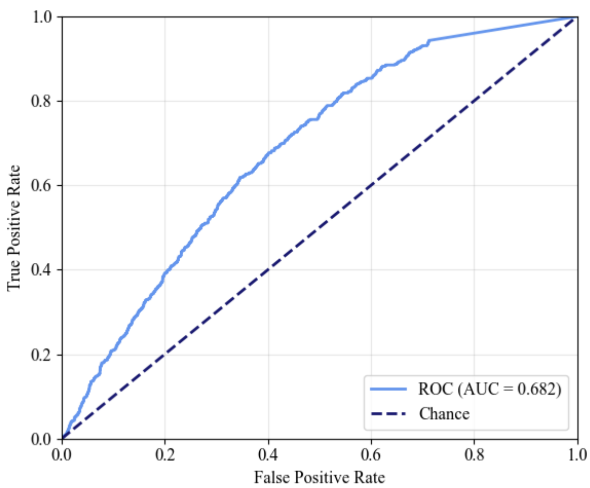

Figure 9: AUC

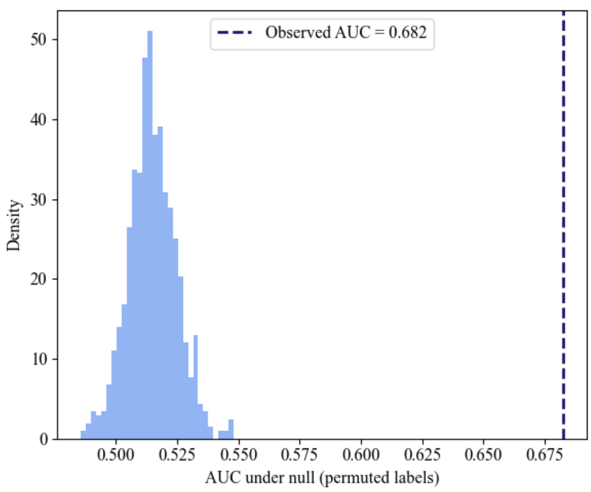

Figure 10: Permutation Robustness test

The permutation test show that the model's predictive performance is far stronger than what would be expected under random chance. The empirical null distribution of AUC values cluster around approximately 0.515 with a p-value of 0.001, while the observed AUC is 0.682. This implies, like for the MC indicator, the robustness of our results.

Overall, the results under the $\lambda^{\text{pointwise}}_{\alpha=1\%}$ threshold are consistent with the results under the $\lambda^{\text{MC}}_{\tilde{\alpha}=1\%}$ threshold. NewsCount and AbsTone are both positively related to the likelihood of a jump, UncWord does not contribute to any additional information, while $|R(t_{i.-1})|$ is the absolute strongest predictor. However, under the Monte Carlo threshold, which is stricter and identifies fewer jump events, the "baseline" jump probability $p_0 = 2.05\%$, while under the pointwise definition, this probability is around 4.6%. However, the model fit metrics (such as $\mathcal{R}^2$ and the LLR test) improves. $\mathcal{R}^2$ increases slightly from 0.0129 to 0.0194 and LLR becomes even more significant. This means that when the jump definition captures more frequent events, the predictors explain a larger share of the observed variation in jump occurrences.

5 Analysis

This section begins by discussing our results and comparing them to those by Jeon et al. [17]. Secondly, it reasons for why it is practically impossible to trade on the news flow in relation to crude oil. Lastly, it presents a potential trading strategy based solely on the crude oil price.

5.1 Systematic overreactions in CL futures

In Section 2.3, we make the central assumption that genuine overreaction events are typically initiated by an abrupt price discontinuity, that is, a jump. The non-parametric $\mathcal{L}-$statistic is backward-looking; it tells us when the realized return over a short interval is unusually large relative to local continuous volatility, but it does not itself tell us whether that move is "too large" in a financial or economic sense. The overreaction label only applies ex post, if the price later reverses toward its pre-jump level. Under this assumption, our jump indicators act as proxies for potential overreaction events: if overreactions in crude oil prices exist and are systematically related to public information, they should manifest first as jumps and show up in the relationship between news flow and jump risk.

The regression results in Section 4 confirm that this mechanism is at least partly at work for crude oil futures. Both for the Monte Carlo and the pointwise 1% thresholds, the probability of detecting a jump increases with the news volume and with the absolute value of the tone, while word uncertainty adds no information. Conditional on the independent variables at their mean, the jump probability is modest around $2-5\%$ per hour (depending on the threshold), but a one standard deviation increase in NewsCount or AbsTone raises the odds of a jump by around $8-14\%$. At the same time, the absolute lagged return $|R_{t_{i-1}}|$ is the strongest single predictor, which indicates that jumps tend to cluster in already volatile states. Both the pseudo$-\mathcal{R}^2$ and the in-sample AUC values are quite low in absolute terms: However, AUC values around $0.67-0.68$ are significantly above what would be expected under random labeling according to the permutation tests. These findings suggest that our statistically identified jump do capture economically meaningful events where prices adjust to new information.

In the framework by Lee and Mykland [19], a jump is a sudden, discrete price movement that is too large to arise from normal continuous volatility and typically reflects an information shock that forces the market to reprice the asset abruptly. Since our jump indicators are based on Lee and Mykland's framework, they respond to news flow in this exact way. This reinforces the idea that potential overreaction events (if they exist) would first appear as price discontinuities before any subsequent reversal occurs.

At the same time, the low but significant explanatory power (pseudo-$\mathcal{R}^2$) and the modest AUC indicate that we are only capturing part of the mechanism: news flow and recent volatility jointly raise the probability of a jump, but they do not determine it fully. Other factors such as order-flow imbalances, inventory constraints, and unobserved private information are likely to contribute as well. This opens up several directions for future research on the mechanism behind overreactions in crude oil and related markets.

5.2 Result comparison

Jeon et al. [17] find that daily jump probability is significantly related to news frequency and to the absolute value of tone, while uncertainty of words have a weaker effect. They also find that news flow helps explain cross-sectional differences in jump intensity, and that these relationships are strongest for equities with high media visibility, measured by the number of distinct news sources covering the firm. Their results show that news flow triggers discrete price adjustments, and that the magnitude of these adjustments grows with the informational intensity of the news.

In their benchmark logistic regression (Table 2, Panel B), a one standard deviation increase in news count increase jump odds by roughly 22% while the absolute value of tone adds about $1-2\%$ and word uncertainty barely 1%. Our values are smaller in magnitude but directionally identical. This might have to do with the fact that Jeon et al. are performing their study on equities at the daily level, while we are looking at data at the hourly level, and naturally, shorter horizons produce smaller absolute effects.

Moreover, Jeon et al. [17] reports a pseudo-$\mathcal{R}^2$ of around $0.6-1.4\%$ while we report a value of $1.3-1.9\%$. Thus, despite different data frequencies and samples, our model has more explanatory power than what Jeon et al. reports. Our findings confirm the central claim of Jeon et al.: higher news frequency and stronger tone significantly increases the likelihood of price jumps, but for the commodity market and in particular the crude oil future market.

5.3 A market perspective

Jeon et al. [17] conduct their analysis at a daily frequency on individual equities. In this environment, information can accumulate in the interval between recorded price observations, especially overnight when corporate announcements, rating changes, or macroeconomic releases can occur. When the market opens, all this accumulated information is incorporated into the opening price, which can generate discrete opening jumps. They state that "explicitly incorporating news processes in models of stock return jumps can potentially help identify jumps due to information arrival as opposed to jumps due to other reasons, such as liquidity or strategic trading based on private information" and that "this may have broad implications for applications such as option pricing and risk management where stock return jump models are frequently used" [17].

The futures market in general, and specifically crude oil futures, is fundamentally different. The front-month WTI contract trades 23 hours per day on a centralized electronic order book with deep liquidity and a higher share of participation from algorithmic market makers and high-frequency traders. Information about inventories, OPEC policy decisions, geopolitical tensions, and macroeconomic developments is absorbed by the market practically instantaneously [9]. Because trading is continuous and the market rarely closes, it is reasonable to assume that there is no meaningful interval during which information can accumulate unpriced. Consequently, while our hourly results show that news volume and tone increase the probability that the next hour contains a jump, this does not provide a practical opportunity to take a position ahead of the market. By the time a trader observes and processes a headline, the information is already incorporated in the price. Thus, whereas the equity setting in Jeon et al. [17] operates on a clock where "news today" may translate into a predictable jump at the next day's open, the crude oil futures market absorbs information so quickly that trading directly on the news itself is not feasible for anyone without millisecond execution capabilities.

Note here that there is a distinction between prices adjusting immediately to new information and prices adjusting correctly to new information.

These structural differences explain why the natural trading focus in our context changes from anticipation of the jump to the behavior of prices after the jump. If our central assumption is correct and many overreactions begin with a jump, then any mispricing that survives beyond the first milliseconds should be most visible in the post-jump dynamics rather than in the initial spike. Modern financial markets limit the possibility of trading directly on these events and make it practically impossible to trade on news. High-frequency and algorithmic trading systems respond to news headlines within milliseconds [1]. To gain even marginal speed advantages, many firms invest in proximity to exchange servers and even in R&D to develop new cables aimed at reducing communication latency [25]. This effectively eliminates the opportunity for human or even standard automated execution to benefit from the initial market reaction. Instead of trying to compete with algorithms on speed around the news release, it is more plausible to examine whether prices exhibit systematic mean reversion after discontinuities and whether such patterns can be translated into rules for entering and exiting positions.

This motivates the trading strategy presented in the next subsection, which does not attempt to trade on the news directly. Instead, it assumes that a detected jump reflects a mispricing and takes the opposite position of the initial price movement, shorting after positive jumps and longing after negative jumps, with a pre-set take profit and stop-loss level for both directions. This tests whether the post-jump price development contains exploitable mean-reverting behavior. As stated earlier, prior research suggests that sudden price jumps often reflect short-term overreactions, after which prices partially or fully mean-revert toward their previous levels (see e.g., [7]). If the jumps identified using the nonparametric test indeed do represent such temporary overreactions, then the subsequent mean reversion may be a source of predictable and systematic return.

5.4 A potential trading strategy

The trading strategy is built around the idea that price jumps may be followed by short-term reversals. The backtest therefore constructs a systematic rule that enters a position immediately after a detected jump and exits once the price has moved sufficiently in either direction.

Let p_{t_i} denote the hourly price and R_{t_i} the hourly return. A trade is initiated whenever a jump occurs at time t_i, that is, when the jump indicator $J_{t_i} = 1$. The direction of the trade depends on the sign of the jump return

$$d_{t_i} = \begin{cases} +1, & \text{if } R_{t_i} > 0, \\ -1, & \text{if } R_{t_i} < 0. \end{cases} \tag{27}$$

If the return is zero or missing, no trade is initiated. The entry price p_{t_i} and the trade begins immediately at time t_i.

After entering the position, the strategy evaluates the subsequent price path $p_{t_{i+1}}, p_{t_{i+2}}, p_{t_{i+3}}, \ldots$. At each future hour t_{i+j} the price change relative to entry is computed as

$$\text{move}_{t_{i+j}} = \frac{p_{t_{i+j}} - p_{t_i}}{p_{t_i}}. \tag{28}$$

The position is closed as soon as the move meet either of the prespecified thresholds

$$\text{move}_{t_{i+j}} \geq \text{TARGET} \qquad \text{or} \qquad \text{move}_{t_{i+j}} \leq -\text{ STOP LOSS}. \tag{29}$$

If the move is in the same direction as the trade, the exit is classified as "take profit". If the move is far enough in the opposite direction, the exit is a "stop loss". If neither threshold is reached before the end of the sample, the position is closed at the last available price.

For each trade, the raw return between entry and exit is

$$R^{raw} = \frac{p_{exit} - p_{entry}}{p_{entry}}. \tag{30}$$

The strategy return, which also accounts for long or short direction is

$$R^{strat} = d_{t_i} R^{raw}. \tag{31}$$

5.4.1 Backtesting

The backtest records all trades initiated by the jump signal and summarizes their outcomes in terms of the number of trades, total return, mean return per trade, standard deviation of returns, and win rate. Each trade is treated as an independent unit of notional (that is, all trades are assumed to be executes with equal capital and allowed to overlap). Using a profit target of 0.7% and a stop loss of 0.3%, the backtesting procedure generates a total of 733 trades over the full sample (2024-01-01 until 2025-09-26) using the pointwise jump indicator $\lambda^{pointwise}_{\alpha=1}$. The total cumulative return (P&L), which is equal to the sum of all R^{strat}, is 42.88%. On a per-trade basis, the average return, which is the mean of R^{strat}, is small at 0.06%, but the distribution of the return is relatively wide with a standard deviation of 0.89%. These numbers are shown in table 6.

Table 6: Backtesting Results

Statistic	Value
Number of trades	733
Total P&L (per 1 notional)	42.88%
Average P&L per trade	0.06%
Std. of P&L per trade	0.89%
Win rate	54.2%

Notes: The trading strategy is applied on the period 2024-01-01 until 2025-09-26. It uses a target of 0.7%, a stop loss of 0.3%. Jumps are defined by $|\mathcal{L}(t_i)| > \lambda^{pointwise}_{\alpha=1}$.

The win rate is 54.2%, which is slightly above the 50% benchmark implied by a symmetric random walk [23]. This suggests that the jump-based entry rule combined with asymmetric exit thresholds (target > stop loss) captures a weak but systematic tendency for short-term price continuation or reversal following detected jumps.

The buy/sell pattern, displayed in Figure 11, shows that the trading strategy consistently enters positions immediately after detected price jumps, with long trades following downward jumps and short trades following upward jumps. Many of these trades close profitably as prices partially revert toward pre-jump levels. This increases confidence in that the jump-based entry rule can capture short-term overreactions. However, the pattern also reveals instances where the market continues to move in the same direction as the jump, which causes the strategy to take the wrong position and hit the stop loss.

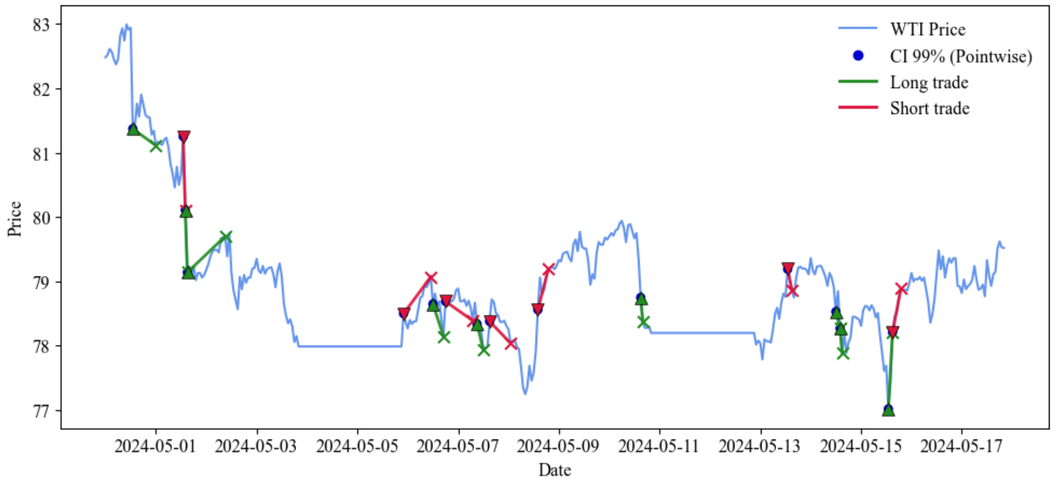

Figure 11: Backtesting trading strategy on a subset of the WTI time series

The clustering of trades in certain intervals in Figure 11 reflects periods of elevated volatility, where jumps occur in succession, while quieter periods produce few or no trades. The asymmetric exit thresholds are also visible: losing trades typically close quickly, while winning trades exit slightly further from the entry point when the larger profit target is reached. Overall, the strategy does exploit some degree of mean reversion after jumps, but the signal is noisy and highly sensitive to local market conditions.

5.4.2 Comparison to HFT and market indices

Certain institutional actors, particularly high-frequency trading (HFT) firms and systematic hedge funds operate with effectively negligible marginal trading costs. These firms execute millions of trades per day, and for such participants the economic relevance of a strategy is determined not by the P&L per individual trade, but by the aggregate performance across extremely high turnover and tight execution spreads [3]. In this segment of the market, even return increments on the order of a few basis points can be meaningful when scaled across large volumes and low-latency execution infrastructure [1].

Since our proposed strategy produces a relatively small average P&L per trade it would likely be eroded by transaction costs (e.g. commissions or bid-ask spreads) for the standard investor or retail trader facing higher frictions. However, the underlying return pattern may still be exploitable for actors in the market who operate in a near-zero–cost, HFT environment.

Over the full sample period, the backtested trading strategy generates a cumulative return of approximately 43%, which exceeds the performance of the Dow Jones Industrial Average, whose cumulative return over the same period is roughly 23% [10]. While this comparison highlights that the strategy outperforms a major equity benchmark in absolute terms, the interpretation should be made cautiously. The strategy's returns come from intraday trades with short holding periods and no overnight exposure, whereas the Dow Jones return reflects a long-only buy-and-hold portfolio [10]. In addition, as previously stated our cumulative return does not account for transaction costs, slippage, or funding requirements, all of which would reduce realized performance.

Another potential advantage of the strategy is that its return profile may be largely independent of broad market movements. Since the trading rules are based on intraday price discontinuities rather than longer-term market direction, the resulting returns could, in principle, exhibit low correlation with major equity indices. If this were the case, the strategy might offer diversification benefits when combined with more traditional long-only exposures, as uncorrelated return streams can improve overall portfolio efficiency.

6 Conclusion and Future Work

The objective of this thesis is to examine whether crude oil futures exhibit systematic overreactions to news events and whether it is possible to develop a profitable trading strategy based on this pattern. First, we conclude our results and address our main objective. Then, we present possible future directions of research. We conclude with a final word.

6.1 Conclusion

The empirical results demonstrate that crude oil prices do react discontinuously to news and that these discontinuities share several properties with those documented for equities in Jeon et al. [17]. However, there are also important structural differences between the two markets that primarily determine the feasibility of trading on them.

Starting with the relationship between news flow and the probability of a price jump, our findings indicate that for crude oil futures, a clear connection exist. Both news volume and the absolute tone significantly increase the likelihood of a jump, whereas linguistic uncertainty adds no predictive power. Hence, we have managed to produce the same outcome on the commodity market as Jeon et al. [17] have done for the equity market. The similarity in the direction/sign of the coefficients and pseudo-$\mathcal{R}^2$ values suggests that the mechanism through which news triggers discrete price moves appears to operate in a comparable way across markets, despite differences in asset class and sampling frequency. However, the effect sizes/magnitude of the coefficients are naturally smaller at the hourly frequency than at the daily frequency, which is consistent with the idea that shorter periods allow less time for information to accumulate. The overall pattern supports the interpretation that news-driven price jumps are not unique to equities and can also arise in highly liquid, continuously traded futures market.

The trading strategy based on detected jumps yields a positive performance. With a total of 733 trades made, the cumulative return of 42.88% reflects that mean-reverting price behavior does exist following jumps, at least on average. The win rate of 54.2% is only slightly above the random benchmark of 50%. However, it indicates that the strategy of triggering an entry with a price jump and having asymmetric exit thresholds (stop-loss vs take profit) captures a weak but systematic tendency for prices to correct after the initial discontinuity. Although the average return per trade is small which limits the economic value for investors facing "normal" transaction costs, the underlying pattern may still be relevant in HFT environments where execution costs are negligible. Overall, the results support the broader idea motivating this thesis: that price jumps in crude oil sometimes reflect temporary overreactions which subsequently mean-revert, and that this post-jump behavior can be exploited through a simple rule-based strategy.

To conclude, we first show that crude oil futures do exhibit price jump around period of increased news activity, which confirms that news volume and tone carry information about the likelihood of a price discontinuity. Second, we find that the magnitude of the jump is related to the news characteristics: higher news volume or absolute tone increases the probability of a price jump. This aligns with the findings of Jeon et al. [17]. Third, by backtesting a trading strategy that takes the opposite position of the initial jump, we demonstrate that post-jump mean reversion do generate positive returns. This suggests that, although crude oil prices react instantly to new information, they do not always adjust fully or correctly. This challenges the idea of the strong EMH. The results therefore contribute to the literature on price overreactions by demonstrating that these anomalies extend from equities to commodities. Additionally, we connect the literature on price discontinuities with short-term overreaction events.

6.2 Future work

While this study provides evidence of systematic relationship between news flow, price jumps, and price reversals, several limitations remain which open paths for future research.

Most importantly, this thesis does not attempt to explain why overreactions arise in crude oil prices. Although the statistical link between news characteristics and jumps is clear, the underlying mechanism is not taken into account. Behavioral factors such as investor sentiment, limited attention, or herd behavior may play a role, especially in markets where macroeconomic or geopolitical narratives can change quickly. Structural frictions, inventory constraints, liquidity imbalances, and the behavior of high-frequency trading algorithms, may also be a potential contributor to price jumps. Future research could therefore explore the behavioral and structural causes to these overreaction events, perhaps by combining order-flow data with sentiment measures or by examining how different categories of news events affect price discontinuities.

Another interesting extension of this methodology would be to replace the dictionary-based sentiment measures with NLP techniques. The Loughran–McDonald dictionary was developed for corporate filings rather than journalistic text [20]. By instead applying word-embedding and machine learning (ML) classifiers to process the news text into categories, it might be possible to capture contextual nuance, shifts in tone, or other semantics that the dictionary does not. Incorporating these approaches has the potential to strengthen the observed relationship between news flow and price jumps.

Finally, more sophisticated trading strategies, such as dynamic exit thresholds, or ML applications could be tested to determine whether the economic significance of the mean-reversal pattern can be strengthened under more advanced decision rules.

6.3 Final word

This thesis shows that price discontinuities in crude oil futures are significantly related to news volume and tone, and that these jumps are often followed by short-term mean reversion that can be exploited in a systematic trading strategy. At the same time, the mechanisms driving these patterns remain open to interpretation, and understanding the behavioral and structural forces behind crude oil overreactions is a potential direction for future research.

References

[1] Aldridge, I., *High-Frequency Trading: A Practical Guide to Algorithmic Strategies and Trading Systems*, 2nd ed. Hoboken, NJ: Wiley, 2013, Copyright © 2013 by Irene Aldridge, ISBN: 9781118343500. DOI: 10.1002/9781119203803.

[2] Andrews, D. W. K., "Heteroskedasticity and autocorrelation consistent covariance matrix estimation," *Econometrica*, vol. 59, no. 3, pp. 817–858, 1991. DOI: 10.2307/2938229.

[3] Bank for International Settlements, Markets Committee, "High-frequency trading in the foreign exchange market," Bank for International Settlements, Basel, Switzerland, Markets Committee Report, Sep. 2011, ISBN 92-9131-885-X (print); 92-9197-885-X (online).

[4] Barberis, N. and Thaler, R., "A survey of behavioral finance," in *Handbook of the Economics of Finance*, Constantinides, George M., Harris, Milton, and Stulz, René M., Eds., vol. 1B, Elsevier, 2003, pp. 1053–1128. DOI: 10.1016/S1574-0102(03)01027-6.

[5] Bewick, V., Cheek, L., and Ball, J., "Statistics review 14: Logistic regression," *Critical Care*, vol. 9, no. 1, pp. 112–118, 2005. DOI: 10.1186/cc3045.

[6] Campbell, J. Y., Lo, A. W., and MacKinlay, A. C., "The econometrics of financial markets," in Princeton, NJ: Princeton University Press, 1997, ch. 4. Event-Study Analysis, pp. 149–180, Chapter PDF hosted by the U.S. NRC, ISBN: 978-0-691-04301-2.

[7] Chen, S., Wu, J., Liang, W., and Yang, H., "News shock, limited institutional attention and stock market response: Evidence from china," *Journal of Asian Economics*, vol. 100, no. C, p. 101 993, 2025. DOI: 10.1016/j.asieco.2025.101993.

[8] CME Group, *Wti crude oil futures - fact card*, https://www.cmegroup.com/trading/files/WTI-Fact-Card.pdf, Accessed: 2025-01-15, 2024.

[9] CME Group, *Light sweet crude oil (wti) — markets: Energy, crude oil, crude oil futures*, https://www.cmegroup.com/markets/energy/crude-oil/light-sweet-crude.html, Accessed: 2025-12-07.

[10] *Dow jones industrial average (dji) — tradingview chart*, https://www.tradingview.com/symbols/TVC-DJI/?timeframe=12M, Accessed: 2025-12-07.

[11] Engelberg, J., Sasseville, C., and Williams, J., "Market madness? the case of *mad money*," *Management Science*, vol. 58, no. 2, pp. 351–364, 2012. DOI: 10.1287/mnsc.1100.1290.

[12] Fama, E. F., "Efficient capital markets: A review of theory and empirical work," *The Journal of Finance*, vol. 25, no. 2, pp. 383–417, 1970. DOI: 10.2307/2325486.

[13] Ferrell, B., *Financial-news-multisource (revision b509ef6)*, 2025. DOI: 10.57967/hf/6432. [Online]. Available: https://huggingface.co/datasets/Brianferrell787/financial-news-multisource.

[14] Gilder, D., Shackleton, M. B., and Taylor, S. J., "Cojumps in stock prices: Empirical evidence," *Journal of Banking & Finance*, vol. 40, pp. 443–459, 2014, ISSN: 0378-4266. DOI: 10.1016/j.jbankfin.2013.04.025.

[15] Good, P. I., *Permutation Tests: A Practical Guide to Resampling Methods for Testing Hypotheses* (Springer Series in Statistics), 2nd ed. New York: Springer, 2013. DOI: 10.1007/978-1-4757-3235-1.

[16] Hull, J. C., *Options, Futures, and Other Derivatives*, 11th ed. Pearson, 2022.

[17] Jeon, Y., McCurdy, T. H., and Zhao, X., "News as sources of jumps in stock returns: Evidence from 21 million news articles for 9000 companies," *Journal of Financial Economics*, vol. 145, no. 2, pp. 1–17, 2022, ISSN: 0304-405X. DOI: 10.1016/j.jfineco.2021.08.002.

[18] Kwon, S. Y., "Extreme events and overreaction to news," Working paper, 2023. [Online]. Available: https://scholar.harvard.edu/files/spencerykwon/files/extreme_news_overreaction.pdf.

[19] Lee, S. S. and Mykland, P. A., "Jumps in financial markets: A new nonparametric test and jump dynamics," *The Review of Financial Studies*, vol. 21, no. 6, pp. 2535–2563, Dec. 2007, ISSN: 0893-9454. DOI: 10.1093/rfs/hhm056.

[20] Loughran, T. and McDonald, B., "When is a liability not a liability? textual analysis, dictionaries, and 10-ks," *The Journal of Finance*, vol. 66, no. 1, pp. 35–65, 2011. DOI: 10.1111/j.1540-6261.2010.01625.x.

[21] Loughran, T. and McDonald, B., *Loughran–mcdonald master dictionary*, https://sraf.nd.edu/loughranmcdonald-master-dictionary/, Accessed: 2025-10-26, 2024.

[22] MacKinlay, A. C., "Event studies in economics and finance," *Journal of Economic Literature*, vol. 35, no. 1, pp. 13–39, Mar. 1997, BU-hosted PDF (working-paper date 1996). Published version: JEL 35(1), 13–39 (1997).

[23] Malkiel, B. G., *A Random Walk Down Wall Street*. W. W. Norton & Company, 1973.

[24] McFadden, D., "Conditional logit analysis of qualitative choice behavior," in *Frontiers in Econometrics*, Zarembka, Paul, Ed., Academic Press, 1973, pp. 105–142. [Online]. Available: file://eml.berkeley.edu/reprints/mcfadden/zarembka.pdf.

[25] Meyer, G., Bullock, N., and Rennison, J., "How high-frequency traders set up shop — and why it matters," *Financial Times*, 2018, Accessed: 2025-12-07 via FT.com.

[26] Nakas, C. T., Bantis, L. E., and Gatsonis, C., *ROC Analysis for Classification and Prediction in Practice*. CRC Press, 2023, ISBN: 978-1-48223-370-0.

[27] Shalizi, C. R., *Lecture 12: Logistic regression*, Lecture notes, Carnegie Mellon University, 2017. [Online]. Available: https://www.stat.cmu.edu/~cshalizi/uADA/12/lectures/ch12.pdf.

[28] Tetlock, P. C., "All the news that's fit to reprint: Do investors react to stale news?" *The Review of Financial Studies*, vol. 24, no. 5, pp. 1481–1512, 2011. DOI: 10.1093/rfs/hhr045.

[29] TradingView, *Nymex: Cl1! – wti crude oil futures continuous contract*, https://www.tradingview.com/symbols/NYMEX-CL1!/, Accessed: September 26, 2025, 2025.

[30] TradingView, *What are 1! and 2! continuous futures contracts?* https://www.tradingview.com/support/solutions/43000483493-what-are-1-and-2-continuous-futures-contracts/, accessed: 2025-12-06, 2025.

[31] US Energy Information Administration, *What drives crude oil prices? – spot prices*, https://www.eia.gov/finance/markets/crudeoil/spot_prices.php, Accessed: 2025-01-15, 2024.

7 Appendix

7.1 Appendix A

7.1.1 A1. Thm 1

Theorem 1. *Let $\mathcal{L}(i)$ be as in Definition 1 and the window size $K = O_p(\Delta t^{\alpha})$, where $-1 < \alpha < -0.5$. Suppose the process follows (1) or (2) and Assumption 1 is satisfied. Let $\bar{A}_n$ be the set of $i \in \{1, 2, \ldots, n\}$ so that there is no jump in $(t_{i-1}, t_i]$. Then, as $\Delta t \to 0$,*

$$\sup_{i \in \bar{A}_n} |\mathcal{L}(i) - \hat{\mathcal{L}}(i)| = O_p\left(\Delta t^{\frac{3}{2}-\delta+\alpha-\epsilon}\right), \tag{32}$$

where δ satisfies $0 < \delta < \frac{3}{2} + \alpha$ and

$$\hat{\mathcal{L}}(i) = \frac{U_i}{c}. \tag{33}$$

Here $U_i = \frac{1}{\sqrt{\Delta t}}(W_{t_i} - W_{t_{i-1}})$, a standard normal variable, and a constant $c = E|U_i| = \frac{\sqrt{2}}{\sqrt{\pi}} \approx 0.7979$ [19].

7.1.2 A2. Lemma 1

Lemma 1. *If the conditions for $\mathcal{L}(i)$, K, c, and $\bar{A}_n$ are as in Theorem 1, then as $\Delta t \to 0$,*

$$\frac{\max_{i \in \bar{A}_n} |\mathcal{L}(i)| - C_n}{S_n} \longrightarrow \xi, \tag{34}$$

where ξ has cumulative distribution function $P(\xi \leq x) = \exp(-e^{-x})$, and

$$C_n = \frac{(2\log n)^{1/2}}{c} - \frac{\log \pi + \log(\log n)}{2c(2\log n)^{1/2}} \quad \text{and} \quad S_n = \frac{1}{c(2\log n)^{1/2}}, \tag{35}$$

where n is the number of observations [19].

7.2 Appendix B

7.2.1 B1. Monte Carlo Simulation Algorithm

Built with inspiration from Gilder et al. [14]

Algorithm 1 MC simulation of jump indicator threshold

1: $B \leftarrow 200\,000$
2: $c \leftarrow \sqrt{2/\pi}$
3: Simulate $Z \sim \mathcal{N}(0,1)^{B\times m}$
4: **for** $i = 1$ to B **do**
5: $\quad M_i \leftarrow \max_{j=1,\ldots,m} \left\{|Z_{ij}|\right\}$
6: **end for**
7: $q_{1-\alpha} \leftarrow$ `numpy.quantile`$(\{M_1, \ldots, M_B\}, 1-\alpha)$
8: $\lambda_\alpha^{\mathrm{MC}} \leftarrow \frac{q_{1-\alpha}}{c}$
9: Return $\lambda_\alpha^{\mathrm{MC}}$