DO AGRICULTURAL COMMODITY PRICE SPIKES ALWAYS STEM FROM NEWS?

BY

ZHOUXIN LI

A thesis submitted in partial fulfillment of the requirements for the

Master of Science

Major in Economics

South Dakota State University

2024

THESIS ACCEPTANCE PAGE

Zhouxin Li

This thesis is approved as a creditable and independent investigation by a candidate for the master's degree and is acceptable for meeting the thesis requirements for this degree. Acceptance of this does not imply that the conclusions reached by the candidate are necessarily the conclusions of the major department.

Zhiguang Wang, PhD
Advisor Date

Joseph Santos, PhD
Director Date

Nicole Lounsbery, PhD
Director, Graduate School Date

ACKNOWLEDGEMENTS

I want to express my deepest gratitude to my supervisor, Dr. Zhiguang Wang, for his insightful guidance, invaluable advice, and constant support throughout my research journey. His expertise and insights have been crucial in shaping this thesis. To my defense committee, Dr. Matthew Diersen and Dr. Jung-Han Kimn, I am extremely grateful for their encouraging words and thoughtful detailed feedback throughout my thesis. I am thankful to them for extending all possible help in the preparation of this manuscript.

Additionally, my heartfelt appreciation goes to my family and friends, whose unwavering support and encouragement have been my anchor throughout this journey. Your faith in my capabilities and your constant love are my source of strength and motivation, I attribute this achievement to each of you.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

ABSTRACT

DO AGRICULTURAL COMMODITY PRICE SPIKES ALWAYS STEM FROM NEWS?

ZHOUXIN LI

2024

This study delves into the occurrence and differentiation of significant price jumps in agricultural commodity markets, challenging the conventional belief that such movements are solely driven by exogenous factors. Existing literature has primarily focused on the impact of news on agricultural commodity prices, neglecting the distinction between endogenous and exogenous price spikes. I aim to identify and categorize both types of price spikes in corn, soybean, and wheat futures markets. I propose a comprehensive methodology involving the collection of agricultural news, non-parametric price jump detection, and differentiation between exogenous (news-driven) and endogenous (non-news related) price spikes. By utilizing intraday price data from the CME Group, I will compare any two consecutive jumps specified by a Bernoulli null hypothesis, and aggregate single jumps into clusters of jumps. I investigate whether endogenous events result from a self-exciting stochastic process. This research lends support to both exogenous and endogenous jumps, providing insights into the efficiency of agricultural commodity markets.

CHAPTER 1 INTRODUCTION

1.1 Background

What are the likely determinants of agricultural commodity price spikes? Sudden and large price spikes in agricultural markets can result from many factors, including the anticipation and release of USDA reports, the announcement or implementation of certain agricultural policies, geopolitics, supply-demand dynamics, weather conditions, etc. Such price spikes may emerge unexpectedly and can be attributed to endogenous factors. But if nothing important is happening in the market, without obvious exogenous features described above, the price may spike simply following some minor fluctuations. I call these jumps self-exciting or endogenous. My research aims to investigate the impact of both exogenous and endogenous factors on the price spikes of agricultural commodities.

The prevailing belief holds that price volatilities are solely caused by exogenous factors. According to the efficient market hypothesis, a cornerstone of modern financial theory, share prices reflect all information in financial markets. Financial markets are subject to exogenous shocks (Subrahmanyam and Titman, 2013) that divert prices from their equilibrium path, leading to greater volatility, bubbles, and even crashes. Agricultural markets, integral to financial markets, also experience substantial unexpected price jumps. A significant amount of price jumps are driven by news events. News-driven jumps have been extensively explored in financial market analysis, particularly in the stock markets, by previous researchers. My study is mainly interested in the influence of news as a representative exogenous element on agricultural futures price spikes. In addition, news occurring at 11 am central time on days of USDA report releases attracts significant attention from the agricultural markets. Thus, it is necessary to summarize the USDA report release dates separately. I examine the extent to which intraday news on these dates correlates with price jumps, complementing the analysis of price jumps associated with daily news.

Numerous empirical studies suggest that movements in an asset's fundamental value fail to account for a significant portion of its price volatility (Hardiman et al., 2013). In the stock market, Joulin et al. (2008) find that neither idiosyncratic news nor market-wide news can explain the frequency and amplitude of price jumps. Most large price jumps seem to be endogenously generated rather than related to the arrival of news in the marketplace. Furthermore, the presence of self-reflexive feedback loops can lead to extreme price displacements from minor or seemingly irrelevant fluctuations, culminating in substantial excess volatilities as discussed by Marcaccioli et al. (2022). Thus, my analysis extends to exploring the role of endogenous factors in driving price fluctuations not attributable to news.

Agricultural commodities sustain billions of people around the world, highlighting the importance of understanding agricultural commodity price spikes. Such knowledge is not only beneficial for humanity but also represents a much-needed area of research. My study aims to contribute to the existing field of work on agricultural commodity price spikes. Past studies have focused on the impact of endogenous or exogenous price spikes in isolation. Drawing on methodologies from stock price analysis, I develop a comprehensive approach that includes the collection of agricultural news, non-parametric price jump detection, and differentiating between these two types of price jumps. My study seeks to fill the existing gap in the literature on agricultural commodity price spikes, offering new perspectives on the efficiency of agricultural commodity markets.

1.2 Research Objective

In this study, I aim to identify and classify the causes of price spikes in corn future markets as either exogenous or endogenous. The specific objectives are to:

(a) collect all related news from major newspapers, magazines, and USDA reports to determine its correlation with futures price spikes,

(b) aggregate price jumps occurring within very close time intervals into clusters, given the prevalence of multiple jumps associated with single news events, and

(c) analyze price spikes unrelated to news to understand the characteristics of endogenous versus exogenous jumps.

In recent decades, commodity markets have seen several developments and changes. Among them, the study of price fluctuations has been the pursuit of economists and analysts, finding various factors lead to the movement of agricultural commodity prices. Research based on the efficient market hypothesis mainly attributes these price movements to exogenous factors. Existing literature explores the power of news on agricultural commodity prices, but there may be potentially important effects of endogenous factors. I propose to combine these two factors and examine the characteristics of their respective effects on price movements. For the exogenous impact component, I expect to find the price spikes significantly influenced by news. For those unrelated to news, I look forward to attributing them to endogenous factors. I follow a non-parametric price jump detection methodology to identify all price spikes (Boudt et al., 2011). Then I compare the observed inter-times between any two consecutive jumps with the one specified by a Bernoulli null hypothesis, transferring single jumps to clusters of jumps. In addition, I investigate whether endogenous events are the result of a self-exciting stochastic process (Marcaccioli et al., 2022).

There are still limitations to this study. I only have access to historical daily news articles for the last ten years. Therefore, I could only utilize the last ten years of trading data to match the timing of the news data. Furthermore, I could not match the daily news precisely to the timestamp of my intraday trading data. Nevertheless, by introducing additional analyses such as those based on USDA reports, I aim to provide comprehensive insights into the dynamics of agricultural commodity price spikes.

The remainder of this thesis is organized as follows: Chapter 2 reviews related literature. Chapter 3 describes my data sources. Chapter 4 presents the empirical methodology. Chapter 5 discusses the findings. Finally, Chapter 6 concludes the study.

CHAPTER 2 LITERATURE REVIEW

In this section, I begin by exploring the literature on the efficient market hypothesis, drawing insights from the stock market to enhance my understanding of agricultural futures markets. I specifically examine both endogenous and exogenous factors, analyzing how these two aspects could impact the prices of agricultural futures.

2.1 Efficient Market Hypothesis

The efficient market hypothesis (EMH) which has been under academic and professional consideration for many years, posits that markets are efficient in reflecting information. This idea dates back to the early 20th century, Louis Bachelier published "Speculation theory" in 1900, discussing the random movements of stock prices and arguing that the expected return of an investment is always equal to zero (Sewell, 2011). However, the formal concept of market efficiency was not clearly defined until Eugene Fama's introduction in his 1965 paper "The Behavior of Stock Market Prices. Roberts (1967) first proposed the concept of the "efficient market hypothesis" and divided market efficiency into the strong and the weak forms. In 1970, Fama expanded this differentiation by adding the semi-strong form of market efficiency. He defined an efficient market as one where market information is "fully reflected" and proposed conducting market efficiency tests alongside asset pricing tests (Fama, 1970). At that time, the idea of market efficiency was popular among academics. EMH asserts that financial markets are efficient. This means that the price remains unaffected if all market participants have access to the same information set (Malkiel, 1992).

Since the 1980s, the EMH has faced various critiques and challenges. Grossman (1976) noted that increased belief in market efficiency makes markets less efficient. He argued that if there is a consensus that the market is efficient, participants begin to act passively and cease to collect information, leading to inefficiency. Later, Shiller (1981) opposed EMH with the concept of excess volatility. He concluded that the actual volatility of stock prices had been higher than that calculated based on fundamental information. De Bondt and Thaler (1985) reconfirmed Shiller’s hypothesis of excess volatility, observing that people tend to overreact to company announcements, a result reflected in stock prices. Subsequently, many studies have concluded that market inefficiency exists, and EMH is now seen as true in relative terms only. EMH fails to explain excess volatility in stock prices, investor overreaction, seasonality in returns, asset bubbles, etc. Furthermore, stock returns are often found to be random, and investors are not capable of consistently earning an excess return.

Despite these critiques, EMH remains a foundational concept in modern finance. It has been adapted and modified in response to empirical challenges and continues to serve as a baseline for understanding and analyzing market behavior. According to EMH, stock prices adjust quickly to new information. This rapid adjustment can lead to short-term volatility as prices react to news, earnings reports, economic data, and other information. Essentially, the market's efficiency in processing information can result in immediate and sometimes sharp price movements.

Drawing from the applications of EMH in the stock market, I can apply its principles to the agricultural futures market. This approach is instructive in analyzing how exogenous factors influence agricultural futures prices, acknowledging that these prices quickly integrate all available information, including environmental, economic, and policy-related data.

2.2 Price Volatility in The Stock Market

The research literature on price volatility associated with the stock market provides a good reference for my study. Why do stock prices change? If the EMH is correct and there are no “noise traders”, theoretically, the price should remain stable between two news items and only fluctuate significantly when news is announced. News releases should be the main determinant of price volatility. Maheu and McCurdy (2004) interpret the individual stock returns as the impact of potential news, which can be measured directly from the price data. They separate the latent news into normal news and unusual news, the latter of which can cause infrequent large volatilities in returns. The effects of these unusual news items are termed price spikes.

However, various pieces of evidence suggest that this scenario is inaccurate. The volatility process is inherently random, with highly non-trivial clustering and long-memory characteristics (Bouchaud and Potters, 2000). In liquid stocks, it seems that most of the volatility stems from trading itself. Hopman (2002) used high-frequency data to separate the volatility into an impact component and a news component, and then found the former is dominant. Similarly, Joulin et al. (2008) found that even if they extend the concept of ‘news’ to a collective market or sector jumps, most large stock price jumps are not related to any publicized news.

Marcaccioli et al. (2022) analyzed five years of order book data from 300 individual stocks, along with a corresponding news database. They show that price spikes following news releases, referred to as exogenous shocks, exhibit significantly different dynamics compared to those that occur spontaneously, termed endogenous shocks. They used YouTube views and Amazon book sales as examples of exogenous shocks and endogenous shocks, respectively. In their analysis, they employed a non-parametric price jump detection method to identify price spikes and utilized the Hawkes process to categorize extreme events as exogenous or endogenous. They found that endogenous shocks typically exhibit slow power-law growth initially and then have an almost symmetric relaxation. Exogenous shocks are asymmetric around the time of the shock and tend to return more quickly to pre-shock activity levels. This research highlights the importance of considering both endogenous and exogenous influences in stock market price volatility, providing valuable insights for my study of agricultural commodity price spikes. Furthermore, the investigation of news-driven jumps, a major part of unpredictable exogenous shocks in the stock market, can also be applied to the agricultural futures market.

The Hawkes process provides a convenient and practical modeling framework that is consistent with the assumption of a nearly critical system. It stands out as one of the most effective models for exploring the self-exciting properties of earthquake occurrence, allowing for the modeling of aftershock sequences after a mainshock (Kwon et al., 2023). Currently, this method finds broad applications in areas such as social media (Yuan et al., 2019) and crime analysis (Mohler et al., 2011). In the financial market, Ferriani and Zoi (2022) have demonstrated the presence of self-excitation in market jump activities, which leads to the clustering of jumps, though this effect shows little persistence in time. They also apply the non-parametric method to detect price jumps in the high-frequency trading data of the stock market.

Sornette and Helmstetter (2003) observe that systems characterized by long-range persistence and memory tend to display different precursory and recovery patterns in response to shocks, whether of exogenous or endogenous origins. They note that the average volatility profile of endogenous events is significantly more symmetric than that of exogenous events. Although both decay following power-law distributions, they are marked by different relaxation exponents. Marcaccioli et al. (2022) utilize the Hawkes processes to differentiate between exogenous and endogenous extreme events, which can guide my study.

2.3 Agricultural Commodities Price Volatility

A lot of reasons such as geopolitical factors, supply-demand dynamics, weather conditions, and agricultural policies can affect agricultural commodity prices and trade. Nigatu et al. (2020) offer insight into the various factors impacting agricultural markets, detailing how changes in these factors may influence commodity prices and global trade. They enhance the analysis by examining how shifts in demand or supply conditions might impact outcomes in both global and U.S. commodity markets, employing a global agricultural market model for their study. Economic factors like fluctuations in exchange and interest rates can affect prices, given that many commodities are priced in global markets dominated by the US dollar. In addition, agricultural policies and global trade agreements can lead to significant price volatility by restricting or facilitating commodity flow across borders. Understanding these influences is essential for agricultural market participants to make wise decisions.

The effect of news on price jumps is not only present in the stock market, but also in the commodity market. Feuerriegel and Neumann (2013) use abnormal returns (the actual return minus the expected return) to track commodity price trends. Their research provides empirical evidence that news sentiment substantially explains these abnormal returns, there is an asymmetrical impact of sentiment, with commodity markets being usually driven by negative news. Similarly, Lechthaler and Leinert (2012) explore the crude oil market through an analysis of news items from major global sources. They find that these news items significantly impact market dynamics, with forward-looking demand forecasts playing a crucial role in driving price increases.

Yang and Karali (2022) employed an ordinal logistic model to examine if the release of USDA reports increases the likelihood of extreme volatility events in soybean and related markets like soybean oil and meal. Their findings indicate a significant impact of these reports on price volatility, with the largest effects observed when reports are issued in March. The research also differentiates between volatility in individual markets and simultaneous extreme volatility across interconnected markets. It emphasizes the importance of seasonal cycles in agricultural production and market sensitivity to new information. Their study highlights the USDA reports as a critical source of market information, which notably influences commodity price fluctuations and can affect decisions in related markets.

Cao and Robe (2022) investigate how USDA reports influence the price volatility of agricultural commodities like corn and soybeans, particularly focusing on the behavior of option implied volatilities (IVols) in response to these reports. The paper showed a notable drop in commodity price volatility following USDA announcements, suggesting that USDA reports play a key role in resolving market uncertainties about future commodity prices. The observed decrease in IVols persists for several trading days after the announcement, indicating a lasting impact of the USDA reports on market sentiment. This research illustrates USDA reports act as an important informational event for the market, helping to reduce uncertainty and stabilize commodity prices through the dissemination of crucial agricultural data.

Koekebakker and Lien (2004) assume that agricultural futures prices follow a jump-diffusion process. The diffusion term can capture both seasonal and maturity effects. They introduce a model with jumps, seasonality, and maturity effects, demonstrating that the volatility of agricultural futures prices is influenced by calendar time (seasonal effects) and time to maturity (maturity effects). Additionally, they observe that futures prices can experience sudden, sharp jumps. The study concludes that overlooking time-dependent volatility and jump effects in agricultural futures prices could result in significant mispricing of wheat options. More recent studies (Schmitz et al. 2014; Wu et al., 2015; Couleau et al., 2020) provide empirical evidence for jumps in corn futures prices.

In summary, the efficient market hypothesis initially inspired my exploration of exogenous factors affecting prices. Concurrently, I notice the EMH has faced various critiques and challenges, I also pay attention to the impact of endogenous factors on price fluctuations. Much of the existing research focuses on how news events affect stock market prices and the potential influence of endogenous factors. This provides a solid reference for my analysis of agricultural futures prices.

Previous studies on agricultural futures prices have primarily concentrated on exogenous factors such as news and USDA reports, weather, and technical elements, while often overlooking endogenous factors. I plan to represent exogenous factors using news, complemented by USDA reports. Price spikes unrelated to news will be considered endogenous. I will separately discuss these endogenous and exogenous influences on price fluctuations, aiming to bridge the gaps in this research area.

CHAPTER 3 DATA AND SUMMARY STATISTICS

In this chapter, I first outline the data sources, including intraday prices for corn futures and news relevant to the corn market. Then I provide an overview of my datasets through summary statistics, offering insights into price trends, and the occurrence of price jumps. For the news data, I analyze the content character and the frequency of publication, etc.

3.1 Agricultural Futures Data

The analysis of price spikes is based on corn futures trading data. The end-of-5-minute intraday original data of prices for agricultural commodity futures were collected from the Chicago Mercantile Exchange Group (CME Group) from 01/01/2013 to 12/31/2022. Futures markets can be traded virtually 24 hours a day, 6 days per week. To make the results more robust, I exclude trading data from mini agricultural futures contracts, and avoid times when the trading is not active. Consequently, only the normal daytime trading hours, from Monday to Friday between 8:30 a.m. and 1:20 p.m. CT, have been chosen.

3.2 News Data

For consistency with trading data, news data were collected for the same decade from 01/01/2013 to 12/31/2022. The “U.S. Major Dailies” database available through ProQuest offers a very effective news source. It provides access to five prominent U.S. newspapers: The New York Times, The Wall Street Journal, Washington Post, Los Angeles Times, and Chicago Tribune. I confine the source types to newspapers and magazines only. Compared to direct filtering, this selection effectively eliminates a large amount of irrelevant material, such as advertisements and local news that may be restricted to a small area.

Three major factors are involved in each news item including the title, the time posted online, and a list of related agriculture commodities such as wheat, soybeans, and corn. For inclusion in my study, news items must meet at least one of the following criteria:

- The title contains the word "corn" or mentions at least one comparable agricultural commodity (i.e., wheat and soybeans), or
- The title refers to at least one of their relevant agricultural companies' names (i.e., Cargill, Syngenta Group, and Bunge Limited), or
- The title is associated with the USDA or the main producing regions of these crops.

After filtering the keywords noted above, the collected news data was manually screened to remove the residual interfering items irrelevant to the study. Moreover, I additionally collected the release dates of WASDE monthly reports on behalf of USDA reports from Bloomberg for the same decade, to make comparative research with all news data.

3.2 Summary Statistics

Figure 1 plots the corn futures prices from 01/01/2013 to 12/31/2022. Initially, prices plummeted in 2013 and reached a low in 2014. Following this trough, prices fluctuated around $4.00 per bushel until around 2020. Since the second half of 2020, there has been a sharp upward trend with prices climbing steeply. Subsequently, prices continued to be highly volatile and peaked in late 2022. Figure 2 displays the number of corn future price jumps per day. The distribution of jumps is sporadic across the timeframe, and there is no apparent trend in the frequency of jumps over the years. Figure 3 presents the number of corn-related news per day. On most days, there are one or two news pieces, although with occasional spikes of higher activity. Like the price jumps, the frequency of news does not show a clear upward or downward trend over the years; instead, it fluctuates irregularly. I also plot the end-of-5-minute intraday standard deviation of corn futures prices in Figure 4. During regular daytime trading hours, there are 58 such 5-minute intervals. I observe that the standard deviation is relatively large at the opening and closing of the market. To refine my study, I exclude jumps occurring in the first and last 15 minutes of the trading day.

Figure 1: Corn Futures Price Trend (2013-2022)

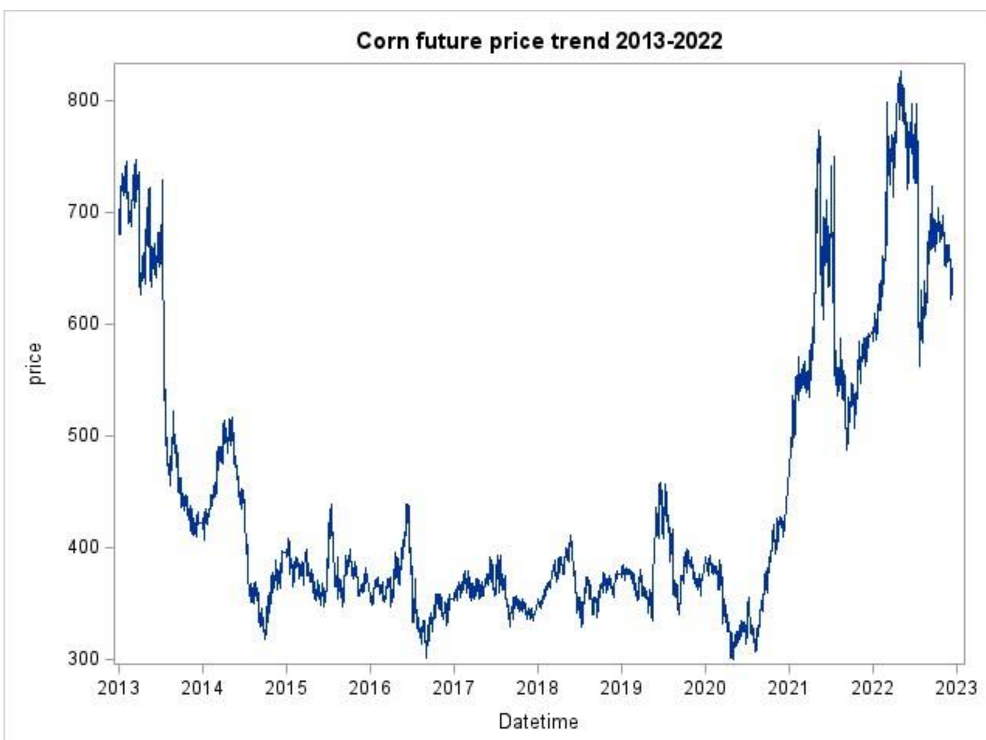


Figure 2: Number of Price Jumps Per Day

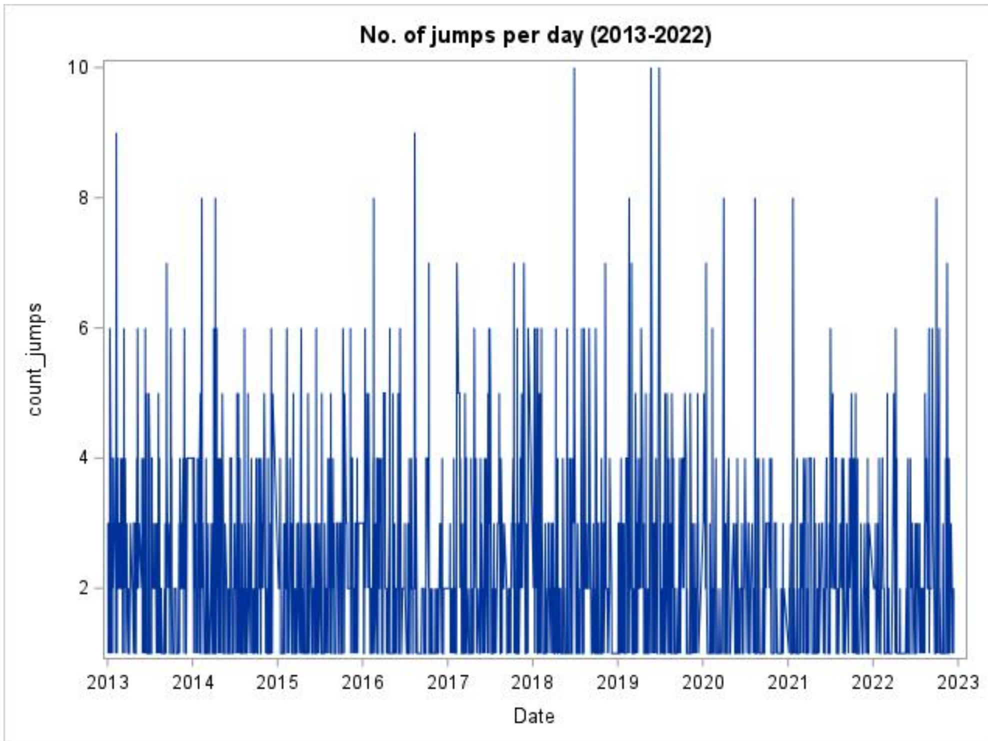


Figure 3: Daily Corn News Volume Statistics (2013-2022)

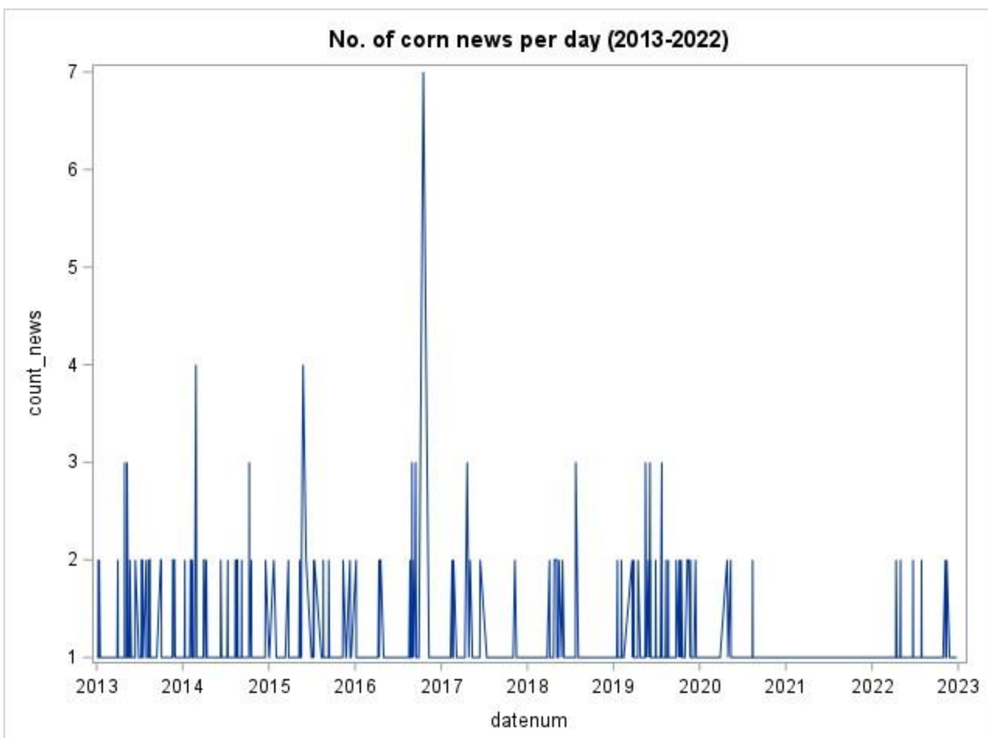


Figure 4: End-of-5-Minute Intraday Standard Deviation of Corn Futures

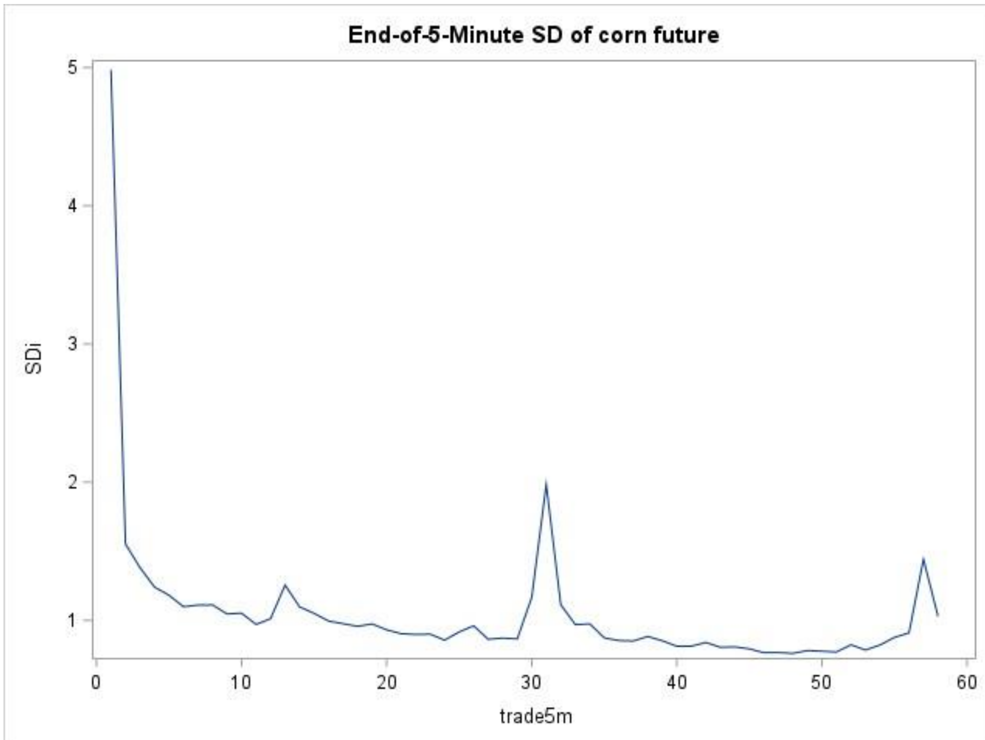


Table 1 provides the time-series summary statistics for corn futures prices over each consecutive end-of-5-minute interval. Each row corresponds to a unique five-minute period and the observed prices of corn futures are expressed in cents per bushel. For each interval, I display the mean, standard deviation, minimum, and maximum price, allowing for a deep analysis of price fluctuations within the specified timeframe. I find the overall average price of corn to be 448.64 cents per bushel, with a standard deviation of 128.11. The skewness of 1.20 and excess kurtosis of 0.12 suggest a deviation from a normal distribution in the price dataset, indicating a possibility of fat tails and associated price jumps.

Table 2 presents the summary statistics for the squared standardized high-frequency returns, denoted as $\bar{r}_t^2$. I set $\bar{r}_t = \frac{r_t}{\sigma_t}$, where r_t is the log-return, $\sigma_t = \sqrt{\frac{\pi}{2K} \sum_{i=1}^{K} |r_{t-i}|\, |r_{t-i+1}|}$ is the square root of the average realized bipower variation (see Appendix A for more details). These values are the result of normalizing the returns to ensure comparability across different periods. The table also lists the mean, standard deviation, minimum, and maximum values of $\bar{r}_t^2$ for each interval. I observed the average end-of-5-minute $\bar{r}_t^2$ is 1.45 and the standard deviation is 8.48. It is noteworthy that the first interval exhibits extremely high values, which justifies the exclusion of jumps from the initial and final 15-minute segments of the trading window.

Table 1: Summary Statistics of End-of-5-Minute Corn Futures Price

Time Interval	Obs.	Mean	Std Dev	Min.	Max.	Skew.	Kurt.
1	2342	447.89	128.31	302.25	821.00	1.22	0.19
2	2262	447.91	127.56	303.00	820.75	1.21	0.16
3	2255	448.33	127.92	303.25	820.25	1.20	0.14
4	2248	448.90	128.66	303.25	820.75	1.20	0.13
5	2244	448.83	128.08	302.00	814.50	1.19	0.11
6	2251	449.23	128.69	302.50	822.00	1.19	0.10
7	2243	448.49	128.12	302.50	821.75	1.20	0.13
8	2238	448.41	128.04	301.50	822.00	1.21	0.16
9	2234	447.78	127.44	302.25	821.00	1.21	0.17
10	2222	448.09	127.11	301.75	820.75	1.20	0.12
11	2217	448.73	127.86	302.50	820.75	1.19	0.12
12	2218	449.92	128.53	304.50	823.00	1.18	0.07
13	2237	449.49	128.70	303.75	826.25	1.19	0.10
14	2205	449.16	128.46	304.00	826.50	1.19	0.10
15	2236	449.11	128.13	303.50	826.75	1.19	0.10
16	2218	448.89	128.43	303.75	826.00	1.19	0.11
17	2219	449.28	128.87	303.25	825.00	1.19	0.09
18	2217	448.28	127.79	303.50	824.75	1.21	0.15
19	2220	448.84	128.48	303.25	825.75	1.20	0.11

20	2204	448.53	128.04	303.25	824.25	1.20	0.11
21	2205	449.21	128.77	302.75	824.00	1.19	0.12
22	2210	448.97	128.10	302.50	822.50	1.20	0.12
23	2211	447.91	127.80	302.50	822.25	1.22	0.17
24	2197	448.17	127.58	302.50	816.50	1.20	0.13
25	2209	448.36	127.61	302.00	822.50	1.20	0.13
26	2218	448.16	127.70	301.75	823.00	1.20	0.13
27	2192	447.93	127.63	301.50	822.50	1.21	0.17
28	2196	450.03	128.88	301.50	822.00	1.18	0.06
29	2189	449.00	128.29	302.00	823.00	1.19	0.10
30	2197	448.84	127.95	301.75	823.50	1.19	0.09
31	2207	448.53	128.22	301.75	823.50	1.20	0.11
32	2201	449.05	128.24	301.75	823.75	1.19	0.11
33	2176	447.99	128.06	301.50	823.25	1.22	0.17
34	2190	449.30	127.94	301.50	823.75	1.18	0.07
35	2182	448.70	128.49	300.75	824.75	1.20	0.13
36	2167	448.25	127.56	300.50	823.75	1.20	0.13
37	2169	447.97	127.67	300.50	824.00	1.21	0.16
38	2172	448.63	127.87	300.50	822.75	1.20	0.12
39	2174	448.12	127.30	301.00	823.50	1.20	0.13
40	2164	449.01	127.89	301.75	824.00	1.18	0.07
41	2175	447.00	126.70	302.50	824.50	1.22	0.18
42	2168	448.76	128.15	302.50	824.00	1.19	0.11
43	2156	448.29	127.76	302.75	824.00	1.20	0.14
44	2135	448.84	127.62	302.50	823.75	1.19	0.10
45	2132	448.45	127.64	303.25	822.50	1.19	0.10
46	2127	448.09	127.06	303.00	822.50	1.20	0.13
47	2139	450.00	128.32	302.50	823.25	1.17	0.04
48	2156	449.67	128.60	303.00	819.25	1.19	0.08
49	2122	449.45	128.12	303.00	823.00	1.18	0.08
50	2144	449.30	128.27	303.00	822.25	1.19	0.11
51	2138	447.93	127.71	302.50	821.00	1.21	0.16
52	2144	448.12	127.57	302.75	821.25	1.20	0.13
53	2149	448.88	128.42	302.75	821.00	1.19	0.11
54	2183	449.22	128.86	302.50	820.25	1.19	0.10
55	2187	448.66	128.39	301.75	819.50	1.21	0.14
56	2196	449.07	128.39	301.75	819.00	1.19	0.11
57	2267	448.08	128.28	302.00	817.75	1.21	0.16
58	1742	446.89	134.47	301.50	818.25	1.20	0.00
Overall	127016	448.64	128.11	826.75	300.50	1.20	0.12

Table 2: Summary Statistics of End-of-5-Minute $\bar{r}_t^2$

Time Interval	Obs.	Mean	Std Dev	Min.	Max.	Skew.	Kurt.
1	2342	24.85	40.13	0	605.55	5.03	42.18
2	2262	2.41	4.23	0	58.60	5.13	42.86
3	2255	1.91	3.37	0	41.03	4.21	26.08
4	2248	1.54	2.51	0	25.18	3.69	19.35
5	2244	1.40	2.76	0	57.08	7.22	96.60
6	2251	1.21	2.26	0	35.04	5.07	42.43
7	2243	1.23	2.15	0	24.75	3.78	20.55
8	2238	1.24	2.43	0	32.67	5.56	47.54
9	2234	1.10	2.14	0	35.61	5.96	58.01
10	2222	1.11	2.16	0	33.87	6.18	63.79
11	2217	0.94	1.80	0	31.34	6.25	72.82
12	2218	1.02	1.92	0	22.97	4.63	31.32
13	2237	1.58	3.01	0	55.41	6.15	69.83
14	2205	1.20	2.31	0	40.16	5.64	56.11
15	2236	1.10	2.22	0	35.71	6.05	58.33
16	2218	0.99	1.72	0	17.25	3.81	20.22
17	2219	0.95	1.99	0	36.40	7.98	105.32
18	2217	0.92	1.82	0	33.03	6.59	74.82
19	2220	0.95	1.75	0	19.33	4.29	26.48
20	2204	0.87	1.61	0	18.80	4.82	34.55
21	2205	0.82	1.78	0	46.73	10.60	217.61
22	2210	0.81	1.52	0	26.68	5.47	54.43
23	2211	0.81	2.00	0	60.33	14.78	374.57
24	2197	0.73	1.36	0	21.96	4.94	42.01
25	2209	0.84	1.61	0	22.71	4.86	36.20
26	2218	0.92	2.18	0	42.28	9.59	140.88
27	2192	0.75	1.47	0	20.26	6.26	59.33
28	2196	0.76	1.41	0	22.78	5.30	48.52
29	2189	0.75	1.47	0	22.35	6.22	60.79
30	2197	1.36	28.81	0	1349.33	46.70	2185.72
31	2207	3.93	27.43	0	778.71	17.60	404.31
32	2201	1.24	5.81	0	195.65	20.82	606.56
33	2176	0.94	2.46	0	49.18	7.98	99.51
34	2190	0.95	2.32	0	52.54	9.46	150.64
35	2182	0.76	1.73	0	31.28	7.96	97.29
36	2167	0.73	1.71	0	32.46	9.81	146.52

37	2169	0.72	1.65	0	39.59	10.46	189.73
38	2172	0.78	1.97	0	49.13	11.66	221.12
39	2174	0.73	1.66	0	41.61	11.28	224.44
40	2164	0.66	1.25	0	16.21	4.77	34.28
41	2175	0.66	1.32	0	26.69	7.08	93.66
42	2168	0.71	1.67	0	25.41	7.13	71.45
43	2156	0.65	1.30	0	23.67	5.88	62.45
44	2135	0.65	1.47	0	31.64	9.18	145.25
45	2132	0.63	1.50	0	35.29	10.34	174.97
46	2127	0.59	1.66	0	53.84	18.65	528.73
47	2139	0.59	1.19	0	21.57	6.62	75.19
48	2156	0.58	1.02	0	8.63	3.67	17.62
49	2122	0.61	1.26	0	21.92	6.33	66.28
50	2144	0.60	1.16	0	18.32	5.22	44.69
51	2138	0.59	1.17	0	18.66	5.66	51.27
52	2144	0.68	1.52	0	42.16	11.83	269.49
53	2149	0.62	1.11	0	12.79	4.21	25.77
54	2183	0.68	1.29	0	15.55	5.02	36.97
55	2187	0.77	1.49	0	28.45	6.31	73.62
56	2196	0.83	1.43	0	20.12	4.93	39.65
57	2267	2.08	4.15	0	70.43	6.75	72.75
58	1742	1.06	1.86	0	25.03	4.50	31.07
Overall	126959	1.45	8.48	0	1349.33	55.87	6343.15

Table 3 shows a small part of my news dataset. This table lists article titles, their publication dates, and a relevance grade manually assigned to each piece. For the grading, a score of 1 denotes high relevance and a score of 3 indicates lesser relevance. Due to the limited amount of news and the publication date without specifying the exact time, a hierarchical analysis of the news relevance levels was not conducted in this study.

Table 3: Examples of News Data

Title	Grade	Date
Egg Producer Hurt by Higher Costs	2	12/29/2021
Nation's Biggest Egg Producer Hurt by Rising Feed, Labor Cos	2	12/28/2021
Biden Raises the Ethanol Gas Tax	1	12/15/2021
Biden Raises the Ethanol Gas Tax; Higher renewable fuel	1	12/14/2021
With food shortage and closed borders, North Korea heads int	2	12/5/2021
Down on the Farm: Carbon Credits: [Special Sections]	1	11/27/2021
We have to rethink farm subsidies before they do more damage	1	10/17/2021
U.S. Farm Belt Revival Prompts Mad Dash for 'Dirt and Iron';	1	9/25/2021
U.S. News: Crops Wither in Summer Scorcher --- Drought, hot	1	8/24/2021
U.S. Crops Wither Under Scorching Heat; Inventories of grain	1	8/23/2021
A Farm-Tech Startup Stumbled, Then Revamped --- Investor fav	2	8/22/2021
How a Billion-Dollar Farm-Tech Startup Stumbled, Then Revamp	1	8/21/2021
Kim Jong Un: N. Korea's food situation is 'tense'	3	6/16/2021
Study: Farms' air pollution causes 17,000 deaths yearly	1	5/11/2021
Corn Is the Latest Commodity to Soar; Prices have risen	1	5/9/2021
Ill. farming, trade in a post-COVID era	1	4/30/2021
If you want a scary story' ...: Agriculture, human health an	2	4/20/2021
Business News: Return of Drivers Lifts Corn Prices	1	3/30/2021
Drivers Are Returning to the Road. That Is Good News for Cor	1	3/29/2021
U.S. News: American Farm Exports to China Surge	1	3/10/2021
For U.S. Farmers, China Is Back and Bigger Than Ever; Record	1	3/9/2021
Argentina Tries to Tax and Spend Its Way Out of an Economic	1	3/8/2021
Business News: U.S. Farmers Poised for Record Planting --- U	1	2/25/2021
U.S. Farmers are Planting More Row Crops Than Ever; Planters	1	2/24/2021

CHAPTER 4 METHODOLOGY

In this chapter, I begin by identifying price spikes within corn futures trading. I adopted the non-parametric jump detection method developed by Lee and Mykland (2008) to categorize significant price changes as jumps. To address spurious effects from closely timed jumps, I grouped them into clusters by utilizing the Bernoulli model. Then I distinguish endogenous and exogenous clusters in corn futures by aligning jump clusters with news occurrence dates.

4.1 Identify Price Spikes

Since my news data is daily, I initially analyzed price spikes using end-of-day trading prices. I collected daily end-of-day data on corn futures from the Chicago Mercantile Exchange Group (CME Group), spanning from 01/01/1990 to 12/31/2022, totaling 8318 entries. However, as corn-related news prior to 2011 was unavailable, my research is limited to 3024 trading entries. This results in a relatively small dataset with a standard deviation of 149.45.

After the preliminary data processing, I identified just 45 price jumps, which is significantly fewer than the 1,045 corn-related news items initially considered. Additionally, using a rolling time window of one month with a daily interval, I was unable to capture the true intraday price fluctuations. The small size of the dataset and high volatility impeded my analysis, I could not observe specific features from the generated figures. Limited data has large errors, yielding results that are contrary to my assumptions and lacking meaningful insights.

For the rest of the thesis, I focus on intra-day price data. Agricultural futures are not traded as frequently as stock markets, so I choose a 5-minute interval rather than a 1-

minute interval as in Marcaccioli et al. (2022) and concentrate on every end-of-5-minute trading price P_t as the first step observation.

Before comparing and analyzing the possible differences between exogenous and endogenous price fluctuations, it is necessary to determine which price changes in $P_t - P_{t-1}$ can be recognized as extremes or jumps. The non-parametric price jump detection methodology proposed by Lee and Mykland (2008) provides a solution to detect jumps and realized jump sizes in asset prices up to the intra-day level. Boudt et al. (2011) make further applications and explanations of this method. They suggest replacing the standard deviation with a robust estimator to enhance the accuracy of the periodic component detection, and subsequently improve the precision of non-parametric intraday jump detection methods. Its steps can be summarized as normalizing the price movement to make the distribution as close as possible to the standard normal distribution even without jumps. After that, by applying extreme value theory, I can obtain a threshold. Once a price movement exceeds this threshold, it can be categorized as a jump within a certain probability level.

I examine the return of 5-minute time series $r_t = \log \frac{P_t}{P_{t-1}}$, and assume the mean of the end-of-5-minute price is approximately zero but the variance fluctuates widely (Andersen and Bollerslev, 1998). r_t is also characterized by strong intraday seasonalities and long memory which may affect the result (Blanc et al., 2017). So, it's necessary to determine an appropriate variable instead of r_t to identify the price spikes. I adopt a standardized procedure that would account for both instantaneous fluctuations in variance and potential seasonality. According to Marcaccioli et al. (2022), a jump-robust estimator is essential for this purpose. This estimator can be described as the 'jump-scores' J_t:

$$J_t = \frac{r_t}{\sigma_t f_t}, \quad (1)$$

where $\sigma_t = \sqrt{\frac{\pi}{2K}\sum_{i=1}^{K}|r_{t-i}|\,|r_{t-i+1}|}$ is a normalized robust estimator of the local volatility for a recurring time window of length K=58. The length of 58 represents the total count of 5-minute intervals that occur during the regular daytime trading hours for agricultural futures, from 8:30 a.m. to 1:20 p.m. The factor f_t is supposed to be a deterministic function of periodic variables (see Appendix A for more details).

Under the null hypothesis, which assumes diminishing sample frequency and no jumps, the statistics for the maximum value of $|J_t|$ converges to a Gumbel distribution. Therefore, with a statistical significance of $\alpha = 0.01$, I may reject the null hypothesis that there are no jumps once observe:

$$|J_t| > C_K - S_K \log(\log\frac{1}{1-\alpha}) \approx 4.017, \qquad (K = 58)$$

where the constants $S_K = (2\log K)^{-0.5}$ and $C_K = (2\log K)^{-0.5} - (\log\pi + \log(\log K))/(2(2\log K)^{0.5})$ are designed to take into account the multiple hypothesis tests conducted within each window, which depend on the window size K. By comparing with this critical value, all the price jumps can be recognized and grouped with only 1% spurious jumps.

In short, I classify those price movements as "jumps" whose associated z-scores are about 4 standard deviations away from zero. There are usually huge price swings at the opening and closing of the market, and these jumps are most likely generated independently of the news. To prevent those non-news-related price jumps from impacting the results, I therefore remove the jumps that occur at the beginning and last 15 minutes of the regular daytime trading.

4.2 Clusters of Price Spikes

I applied the price jumps detection method to corn futures. Over 2,403 trading days, I recorded a total of 3,419 jumps. The daily average number of jumps is 1.42. While the initial count of corn-related news items was 1,045, spread across 599 days, the frequency of these jumps significantly exceeds the number of news events that could potentially influence prices.

If I examine the distribution of time intervals between two successive jumps that occurred on the same day, there are obvious departures from the Poisson law that assumes a constant arrival rate of jumps. Kou (2002) has proved that the Poisson law is not typically suitable for describing phenomena like volatility clustering in financial markets. On the contrary, the Power law behavior might describe the distribution more closely. Power law distributions of waiting times have been employed in the studies of human mobility and many social activities (Goh and Barabási, 2008), including widely used for trading in the financial market. For example, Power laws can describe histograms of relevant financial fluctuations, such as fluctuations in stock price, trading volume, and the number of trades (Gabaix et al., 2003). Bai and Zhu (2010) find that a shorter microscopic timescale leads to the power-law tails of the distributions, and a longer timescale leads to an exponential law instead. A common way to model such price jumps is using the self-exciting Hawkes processes (Bacry et al., 2015).

When analyzing the overall price jump results, I notice that some jumps occur too close in time to be considered independent, which could lead to spurious effects possibly caused by long-memory effects. To minimize these issues, I focus on analyzing clusters of jumps instead of individual jumps.

I employ a simple and direct approach to cluster sequential jumps together, comparing the actual time intervals between any two continuous jumps with the period consistent with the Bernoulli null hypothesis. This hypothesis assumes that independent jumps occur with a probability p. Under the null hypothesis, if the observed probability of a given time interval is less than a significance level ϵ, then the two consecutive jumps can be grouped together. I then mark these two jumps as a cluster.

In simple terms, if a jump happens at a time t_1, and a continuous second jump happens at the time t_2, these two jumps will be placed in the same cluster if:

$$t_2 - t_1 < \frac{\log(1-\epsilon)}{\log(1-p)} - 1,$$

Where $\epsilon = 0.05$, I calculate p so that the number of jumps for agricultural future prices in any month can be retained on average by this null model.

To find out all clusters of price jumps, I apply a two-step procedure. Initially, returns are normalized to make comparisons. Returns that deviate from the null hypothesis, which assumes they are independently and identically distributed (i.i.d.) normal random variables, are identified as jumps. Subsequently, employing the Bernoulli null hypothesis, jumps that occur too close together can be pinpointed and grouped into clusters.

After applying the above clustering procedure, a total of 3094 clusters of price jumps were recorded over the same 2403 trading days, including 2769 single jumps. The average number of daily clusters of jumps is 1.29.

Those clusters that occur on the same day and exhibit a normalized inter-time distribution can be described by the exponential distribution. This characterization, to some extent, proves the clusters are independent, so the spurious effects caused by any clustering process are effectively minimized.

4.3 News-Related Price Spikes

Since the news data are stamped to the date, they may also be released consecutively on the same day, with the spurious effect of being too close in time to be independent. Thus, it is still necessary to apply the clustering steps to the news data. The difference is that I directly label news that occurs on the same day as one news cluster. Then if the clusters of jumps happen on the same day as a news cluster, the clusters of jumps would be marked as news-driven or news-related, hence attributing such jumps to be exogenous. For the remaining clusters of jumps, those occurring on days without news releases, I mark them as non-news related or endogenous.

Another effect that I need to consider is the accuracy of news release dates. I note that the dates listed on my news source, the ProQuest-U.S. Major Dailies, are mostly the day that article appeared in that newspaper, likely a day or two after an event. I therefore adjust the news dates to match the trading activities. First, categorize news by weekday and tag the news with the corresponding Sunday through Saturday. Due to the absence of normal daytime trading hours on weekends, I keep the news date posted on Monday unchanged. For news that was released on Tuesday, Wednesday, Thursday, and Saturday, I move its weekday forward one day, which means the actual occurrence on Monday, Tuesday, Wednesday, and Friday, respectively. Since Friday is the release day of the USDA report, I keep it unadjusted. And for the news listed on Sunday, I decided to move it to Monday for matching purposes.

Finally, after merging the adjusted news and clusters of jumps by date, I record 3,094 clusters of jumps that occurred over a total of 2,403 trading days. Of these, there are 994 news-related clusters of jumps on 494 days, which occupy 32.13% of all clusters. The non-news-related clusters account for 67.87%.

CHAPTER 5 RESULTS AND DISCUSSIONS

5.1 The Internal Structure of Clusters

To analyze the differences between endogenous and exogenous clusters of price jumps, first, I compare the internal components of endogenous and exogenous clusters of jumps. After calculating the mean, the average of news-related jumps is 1.12, higher than the 1.09 for non-news-related jumps.

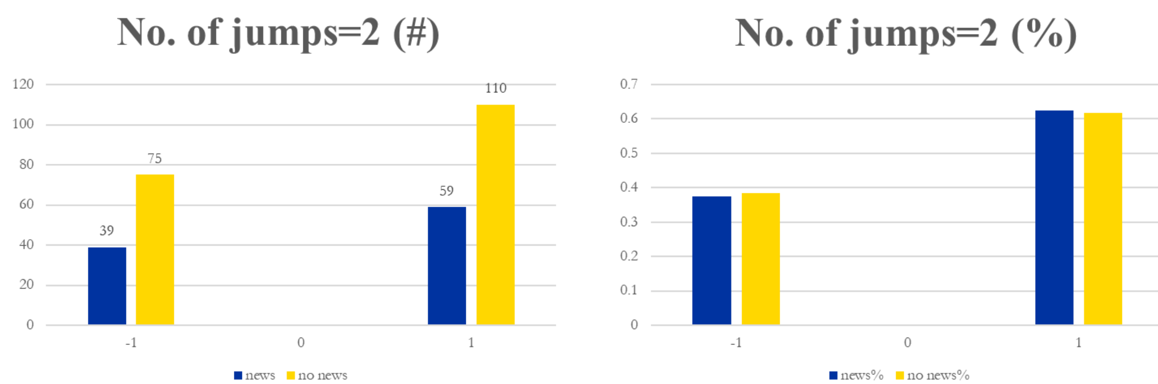
Then I examine the distribution of Kendall's tau correlation between the chronological order and the amplitude-based ranking of jumps in a cluster with N jumps. In any cluster, if the first occurring jump is also the largest, the second jump is the second largest, and so on, the corresponding value $\tau = 1$. Conversely, if a series of jumps take place in the reverse order, i.e., the first jump is the smallest and the last jump is the highest, in such case the value $\tau = -1$.

My dataset is mainly comprised of clusters with a single jump. For those clusters where jumps exceed one, Figure 5 illustrates the distribution of my findings. When the clusters only include two jumps, as shown in the second plot of Figure 5(a), the left and right two columns are $\tau = -1$ and $\tau = 1$, respectively. There are 98 news-related and 185 non-news-related clusters, where the first jump exceeds the second in 60.2% and 59.4% of cases, respectively. When considering clusters with three jumps, I identified 15 news-related and 23 non-news-related clusters, with Kendall's tau being 15.56% and 18.84% respectively.

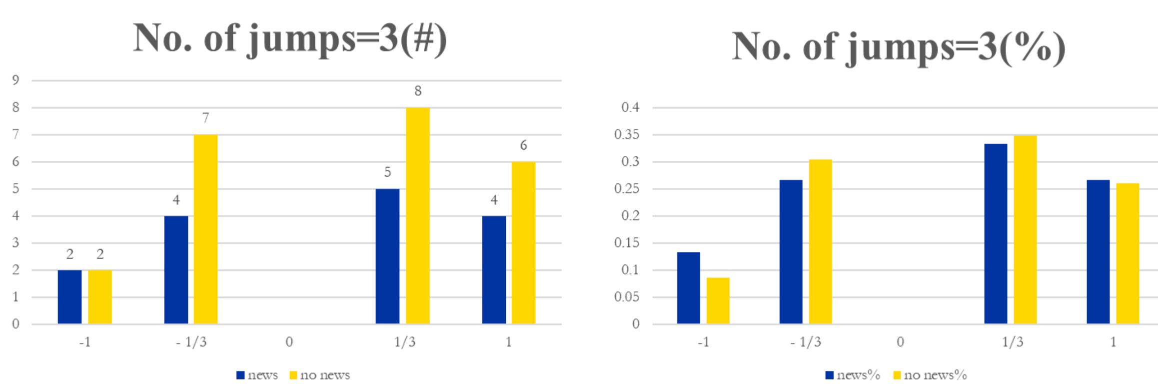
Constrained by the 5-minute interval and the frequency of trading in agricultural futures, my results show that compared to endogenous clusters, news-related clusters are more naturally ranked in time when clusters with two jumps. That reflects the notion that exogenous events are typically abrupt and strong reactions to external shocks, whereas endogenous events arise from a self-exciting stochastic process characterized by gradual accumulation.

Figure 5: Distribution of Kendall's Tau Rank Correlations

(a)



(b)



Based on these observations and taking a reference on the literature about Hawkes processes, I would demonstrate the differences in the instantaneous volatility and price trend profiles of exogenous and endogenous events.

5.2 Average Profile of Exogenous and Endogenous Jumps

To demonstrate the distinct characteristics between clusters of jumps generated by news-related and those that are endogenous, I will perform additional analyses to compare these two jumps from the following perspectives:

- Instantaneous jump-score: $|J_t|$;
- Exponential moving average of past excess volatility, defined as:

  $$\sum_t = \kappa|J_t| + (1-\kappa)\sum_{t-1},$$

  where the weighted multiplier κ means the averaging timescale, $\kappa = 2/(N+1)$, and N is the number of 5-minute periods. I chose the decay time of 30 minutes, so the $\kappa = 0.29$. It is important to recognize that the standardized returns J_t, which are labeled as jumps, are not included in my exponential average calculations. Besides, averaging is done by shifting time such that for each cluster, $t = 0$ corresponds to the first jump of each cluster.
- Normalized past price trend:

  $$T_t = \kappa J_t + (1-\kappa)T_{t-1},$$

  where the κ is the same value as above, also the J_t marked as jumps are removed from the evaluation.

After the calculations, I would use these three measurements to analyze how the average profiles of only news-related clusters of jumps differ from those jumps only caused by endogenous factors. Moreover, I am not concerned with the direction of past trends, only focusing on their amplitude, and computing the average over the absolute value of T_t.

Figure 6 depicts the $|J_t|$ before and after jump clusters, the zero point on the x-axis indicates the day when a cluster of price jumps begins. The blue line tracks the trend of prices related to news on days when at least one news-related jump cluster is observed, starting from minute zero and continuing thereafter. To ensure clarity in the trend analysis, days with consecutive clusters, such as those occurring on Tuesday followed by Wednesday and Thursday, are filtered to only include the initial day—Tuesday in this example. The timeline pre-zero reflects the price trend one day before the jump. Meanwhile, the red line plots the average trend of prices on days with non-news-related clusters. It provides a consistent baseline for comparison before and after the zero point on the timeline.

I observe that prior to the zero mark, the news-related price line sits slightly below the non-news-related line. And after zero, it distinctly surpasses the non-news trend line, indicating that prices responded significantly to the news stimulus.

My finding shows some differences from those reported by Marcaccioli et al. (2022). Their research indicates that non-news-related clusters begin with a slow increase in volatility and trends, and then the volatility has an obvious rise up to 75 min before the first jump. They also believe a more frequent final drop in liquidity, driven by intense competition among high-frequency liquidity providers. However, my result shows that non-news-related clusters are relatively lower in the initial 50 minutes. On the days when news-related clusters occur, the non-news-related clusters are only slightly lower, without a significant drop post-shock.

Besides, Marcaccioli et al. (2022) suggest that before the first jump, news-related clusters only have a slight increase in the average profiles, but they will emerge more abruptly after the jumps. My results show before the news-related clusters, the average price movements are only relatively lower than the non-news-related clusters. On the days when news-related clusters occur, they are higher, as demonstrated in Figure 7, and news-related clusters generally register above the non-news-related clusters. This suggests that news-related clusters tend to have higher $|J_t|$ values on the day that the jump happens. This variation can be attributed to two main factors:

- Different financial instruments are traded at different frequencies. The frequency of trading in stocks is generally higher than in agricultural futures, which can contribute to more pronounced clustering effects.
- The accuracy of the data varies. My news dataset is accurate to the date, while the reference study uses news with intraday time stamps.

These differences in market behavior and data granularity could explain the differences in results. A future study can further analyze jump clusters related to the USDA reports using intraday news data.

Figure 6: News-Related and Non-News-Related $|J_t|$

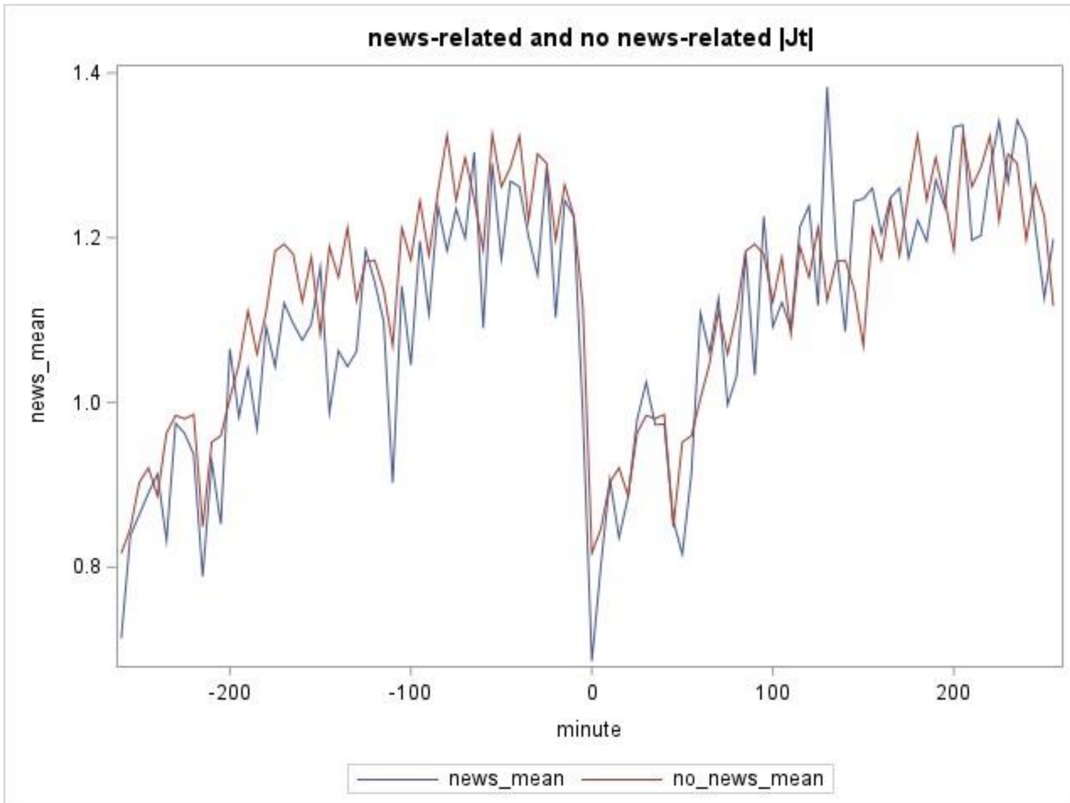
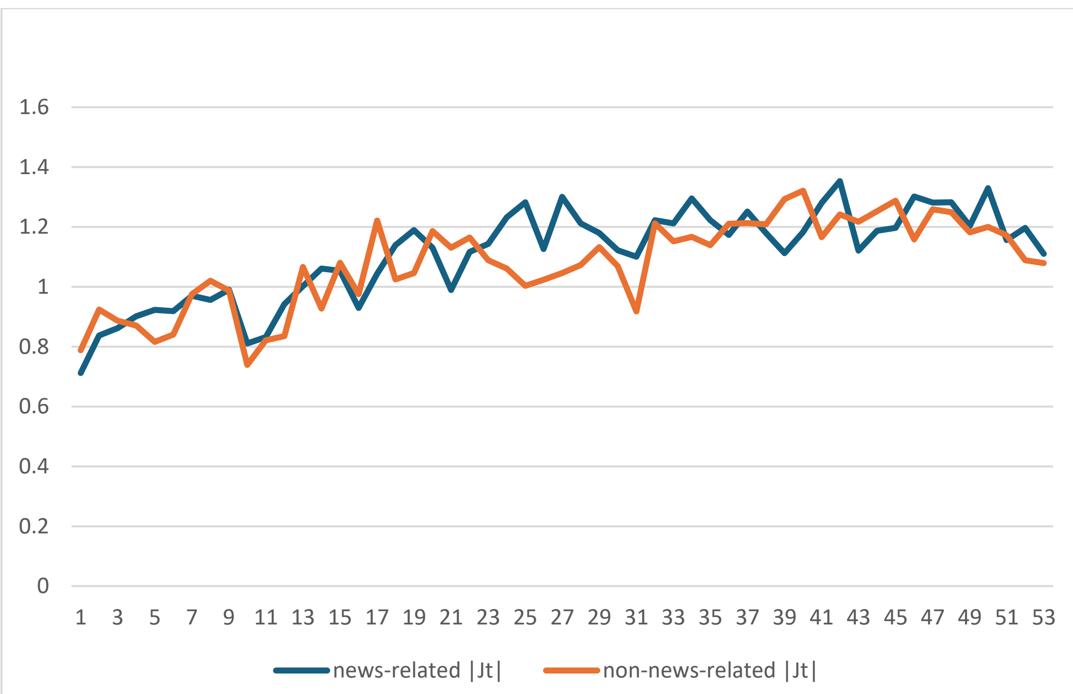


Figure 7: Average $|J_t|$ for Each Time Interval

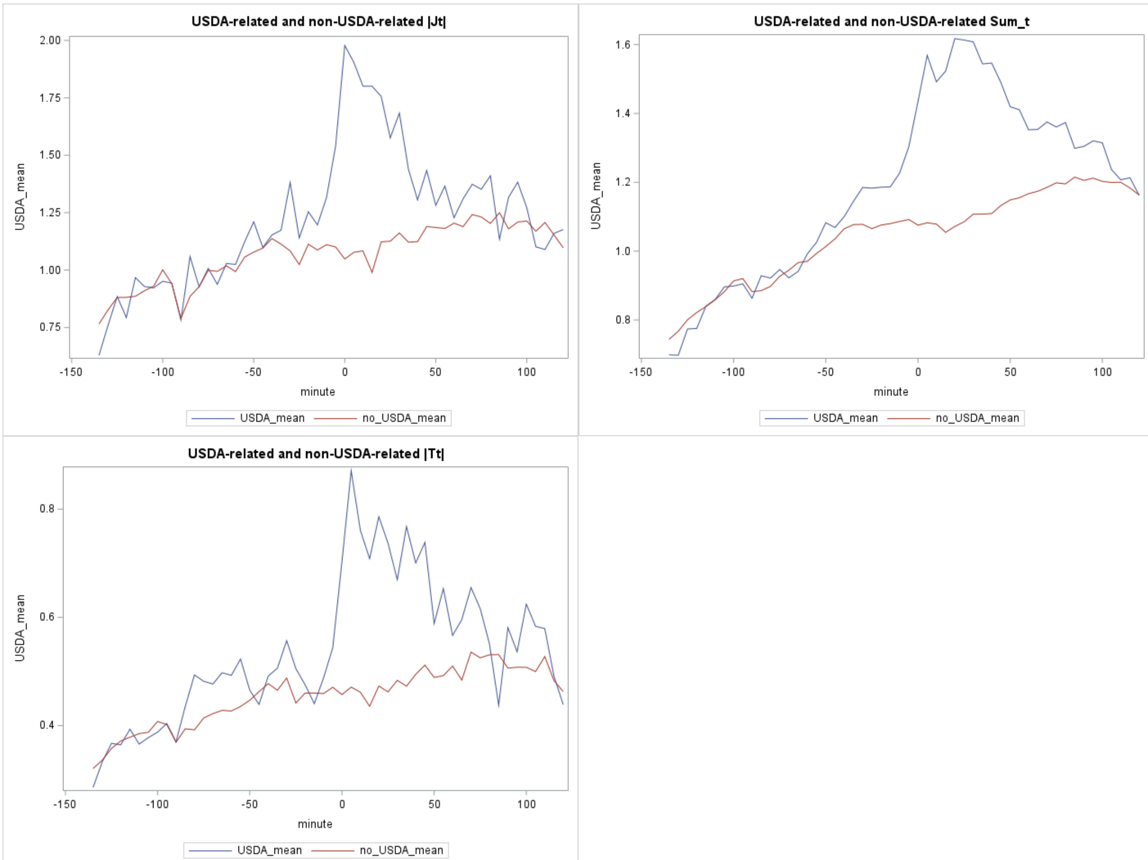


Next, I limit the news to USDA reports and apply the same method on USDA report related (hereafter referred to as USDA-related) jumps and without USDA report related (hereafter referred to as non-USDA-related) jumps. On the release day of USDA WASDE monthly reports, I mark all the clusters of jumps that happen on those dates as USDA-related clusters, and those happen on other trading days without USDA reports are labeled as non-USDA-related clusters. The exact time of the release is 11:00 am CT, which I set to $t = 0$, that clusters occurring before 11:00 on these days are positioned to the left side of $t = 0$, while those after 11:00 are positioned to the right side of $t = 0$. This supplementary analysis implies that both the news data and the price data are intraday, and the results are expected to be more precise than the daily news data.

As the plots show in Figure 8, the USDA-related clusters initially exhibit relatively minor fluctuations and rise slowly, this trend continues, with the average value climbing gradually. Approaching $t = 0$, the USDA-related clusters experience a sudden and sharp increase, appearing as a noticeable spike. After the spike, it shows a quick decline immediately following the peak then fluctuates at a lower level compared to the spike, with the mean generally trending downward but still higher than the initial period's mean.

The non-USDA-related clusters illustrate a different pattern of volatility as compared to the USDA-related clusters. In the beginning, it displays a relatively stable trend with minor variations. With moving toward $t = 0$, it shows a slight increase in volatility, this suggests a steady increase in the mean value, although with some fluctuations. Contrasting with the USDA-related clusters, the endogenous clusters do not exhibit a sharp spike around $t = 0$, it maintains its gradual trend without any dramatic changes. After that time, the non-USDA-related clusters continue to exhibit a pattern of slight movements without significant jumps, and recovery to average baseline volatility is also slower than USDA-related clusters.

Figure 8: USDA-related and non-USDA-Related $|J_t|$, $\sum_t$, and $|T_t|$



5.3 Forecasts of the Hawkes Model

In this part, to diminish the impact of news data limitations, I use the USDA-related clusters to represent the exogenous (EMC) events and the non-USDA-related clusters to stand for endogenous (SEC) events.

Studies from Crane and Sornette (2008) and Sornette et al. (2004) demonstrate that, for endogenous shocks, there exists an asymmetry between the pre-jump growth and the post-jump relaxation. They assume there exists a self-exciting process and Marcaccioli, et al. (2022) simplify the self-exciting Hawkes conditional Poisson process form as

$$\lambda(t) = \lambda_0(t) + \sum_{t_i < t} \phi(t - t_i), \tag{2}$$

where $\lambda(t)$ is the instantaneous rate of price movements, $\lambda_0(t)$ is the price movements that are triggered by exogenous sources, t_i represents the time when previous price movements occurred, and $\phi(\tau)$ is the memory kernel of the system, which represents how previous occurrences increase the likelihood of subsequent events.

Since the agricultural commodity trading price is a publicly available source of information, any shift in the instantaneous fluctuation, whether exogenous or endogenous, could lead some market participants to react. These actions, in turn, will have an impact on the fluctuation itself further, which will then influence other market participants, and the cycle continues. It is worth noting that the effect of traders on other traders, or the effect of previous volatility on future volatility, doesn't happen immediately. This relationship is represented through the memory kernel $\phi(t - t_i)$, which models the time delay in these effects. I conject that equation (2) applies to agricultural commodity prices.

Based on previous work (Sornette et al., 2004; Helmstetter and Sornette, 2002), I assume that a long-memory process is of the form $\phi(\tau) \sim 1/\tau^{1+\theta}$ with $0 < \theta < 1$. The exponent θ is the key parameter of the theory which will be determined empirically from the data. Equation (2) forecasts two distinct patterns for both exogenous and endogenous jumps in cases where the process is relatively stable, specifically when $n :=$

$\int_0^\infty d\tau\phi(\tau) \to 1$. The observed tendencies for the profiles before and after a jump are as follows:

$$|J_t| \propto \begin{cases} \dfrac{1}{\left(t-t_j\right)^{1-\theta}}, & EMC, t > t_j; \\ \dfrac{1}{\left|t-t_j\right|^{1-2\theta}}, & SEC, t \lessgtr t_j, \end{cases} \tag{3}$$

when $t < t_j$, EMC has a relatively flat profile.

The prediction suggests that relaxation from an EMC shock, characterized by a larger exponent $1-\theta$, occurs faster than from an SEC shock, which has an exponent of $1-2\theta$. This reflects the SEC shocks result from a gradual buildup process and affect the network more deeply over a longer period, leading to a more prolonged impact. By taking logarithms on both sides, for an EMC post-jump dynamic with $\theta = 0.6$, I get the value of $p_r^{EMC} \approx 0.4 = 1-\theta$. Similarly, in the case of a SEC jump dynamic where $\theta = 0.4$, $p_r^{SEC} \approx 0.2 = 1-2\theta$.

I observed that the EMC relaxation exponent $p_r^{EMC}$is almost two times as large as p_r^{SEC}. A larger relaxation exponent corresponds to a faster rate of decay, indicating a shorter memory span. Essentially, this means that past events lose relevance and are forgotten more quickly. On the other hand, a smaller exponent suggests a slower decay. This implies that past events have a more prolonged influence, affecting future events for an extended period. Comparing the predicted profiles to the average profiles in Figure 6, it is clear that those jump trends are consistent with my calculation.

In previous studies on other social systems (Crane and Sornette, 2008; Sornette et al. 2004), they concluded that both SEC and EMC jumps have the same value of $\theta =$

0.4 ± 0.1. I observe θ for SEC jumps that is identical to earlier studies, but there is a deviation in the θ related to EMC.

My findings also indicate a slight deviation from the pre/post jump symmetry that the model predicts for SEC jumps. I hypothesize that this asymmetry might be better explained by a modification of equation (2), as recently suggested by Blanc et al. (2017). It means that not only past activity, but also historical price trends should be considered as influencing factors on the current rate of activity. I deduce a possible reason for the observed asymmetry, that some jumps I thought were endogenous might be driven by exogenous factors. This exogenous information could have been missed in my news database. This finding will lead me to further research and deeper investigation.

CHAPTER 6 CONCLUSIONS

Based on the literature related to stock market price movements, this study analyzes the dynamics behind price spikes in the agricultural commodity market, focusing on corn futures. My research challenges the conventional view that such price movements are predominantly driven by exogenous factors, such as news events. By employing a comprehensive methodology, including the collection of corn market-related news, non-parametric price jump detection, clustering the jumps by the Bernoulli null hypothesis, and categorization of price spikes into exogenous and endogenous types, I try to explain the intricate nature of the agricultural commodity market.

My findings reveal that endogenous factors, often overlooked in previous studies, play a significant role in influencing agricultural futures price movements. Considering the significance of agricultural commodities and their impact on economies and livelihoods, this viewpoint is instructive. I combine stock price analysis techniques with factors unique to agricultural markets, fill a gap in the existing research and offer a more comprehensive view of market dynamics.

By analyzing cluster internal structures, I observed that news-related clusters are more naturally ranked in time than endogenous clusters. This suggests that exogenous events are often abrupt responses to external shocks, while endogenous events stem from a gradual, self-exciting stochastic process. Utilizing USDA report-related clusters instead of news-related clusters to enhance temporal precision, I applied the Hawkes model for predictions. This confirmed that exogenous events exhibit a faster decay in their relaxation exponent, indicating a shorter memory span, in contrast to the more prolonged impact of endogenous events.

Acknowledging my study's limitations, particularly the in-depth and temporal accuracy of news data and the number of agricultural futures, my work nonetheless makes significant progress. I separated USDA reports from general news to analyze their effects on intraday price movements, offering a refined view for future agricultural market studies.

In future research, I aim to collect more comprehensive news data with a finer timestamp, and refine the data processing to enhance the precision. I will perform further stratified analysis on the relationship between price spikes and the news relevance grades which are listed in Table 3. In addition, I plan to explore a wide range of agricultural futures, such as soybeans and wheat. To offer a deeper understanding of the complex interplay between news events and market reactions, valuable for both market participants and policymakers.

APPENDIX

Appendix A. Returns Standardization

As log-returns r_t are marked by intraday seasonality and long memory, it cannot be appropriate for identifying price jumps. Then by standardization, I convert r_t to $J_t = \frac{r_t}{\sigma_t f_t}$. $\sigma_t = \sqrt{\frac{\pi}{2K}\sum_{i=1}^{K}|r_{t-i}|\,|r_{t-i+1}|}$ is the square root of the average realized bipower variation over the local window (Nielsen and Shephard, 2004), it can robustly estimate the local volatilities. Based on the standardized high-frequency return, I set $\bar{r}_t = \frac{r_t}{\sigma_t}$ to evaluate the periodic component f_t in intraday volatility (Boudt et al., 2011).

If the standardized scale estimates of returns have an identical periodicity factor, then the non-parametric periodicity estimator can be used to evaluate the periodicity. The set of standardized returns with the same periodicity factor r_i is denoted by $\bar{r}_{1,i}, \dots, \bar{r}_{n_i,i}$. In other word, I can collect the returns of corn future at the given time interval i, $\bar{r}_{1,i}, \dots, \bar{r}_{n_i,i}$ are the returns recorded at the same time of the day and day of the week as r_i. Taylor and Xu (1997) referred the non-parametric periodicity estimator as based on the standard deviation (SD) of all standardized returns belonging to the same local window as $\bar{r}_i$, i.e., $SD_i = \sqrt{\frac{1}{n_i}\sum_{j=1}^{n_i}\bar{r}_{j,i}^2}$. The SD periodicity estimator is defined as

$$\hat{f}_i^{SD} = \frac{SD_i}{\sqrt{\frac{1}{K}\sum_{j\in N_i} SD_j^2}}. \qquad (A.1)$$

Make sure the denominator of Eq. (A.1) meets the standardization condition $\frac{1}{K}\sum_{j\in N_i} f_i^2 = 1$.

When there are no price jumps, since standardized returns $(\overline{r}_{1,i}, \ldots, \overline{r}_{n_i,i})$ follow the normal distribution, SD is available. But if price jumps exist, given that at least one observation in the sample is affected by a big jump, the periodicity estimates then become very large, which leads to a significant bias in the SD estimates. Thus, it's essential to find a robust estimator to replace the SD.

To reduce the effect of bias on the result, I reviewed the relevant literature and experimented with several robust scale estimators. Of these, the Shortest Half scale estimator proposed by Rousseeuw and Leroy (1988) best fulfills my requirements. This method was further developed by Martin and Zamar (1993), they construct asymptotically min-max bias robust estimates of scale and show that a scaled version of the Shorth (the shortest half of the data) is the estimator for which jumps can provide the least maximum bias.

To establish the Shortest Half (ShortH) scale estimator, I require the appropriate order statistics $\overline{r}_{(1),i}, \ldots, \overline{r}_{(n_i),i}$, and make $\overline{r}_{(1),i} \leq \overline{r}_{(2),i} \leq \cdots \leq \overline{r}_{(n_i),i}$. The shortest half scale, which is composed of $h_i = \lfloor n_i/2 \rfloor + 1$ consecutive order statistics, has the least length of all the "halves". It is expressed as:

$$ShortH_i = 0.741 \cdot \min\{\overline{r}_{(h_i),i} - \overline{r}_{(1),i}, \ldots, \overline{r}_{(n_i),i} - \overline{r}_{(n_i-h_i+1),i}\}.$$

Similar to the SD estimator in Eq. (A.1), the ShortH estimator for the periodicity factor r_i is equal to

$$\hat{f}_i^{ShortH} = \frac{ShortH_i}{\sqrt{\frac{1}{K}\sum_{j\in N_i} ShortH_j^2}}. \tag{A.2}$$

According to the research done by Rousseeuw and Leroy (1988), even though the ShortH estimator is more robust in detecting jumps, its estimated effectiveness is just 37%

under the normal distribution of r_is. The Weighted Standard Deviation (WSD) is more efficient than the ShortH and it can provide a more robust estimator to jump. Weight in the WSD is related to the value of the standardized return divided by the ShortH periodicity estimate, that is $\bar{r}_{l,j}/\hat{f}_i^{ShortH}$. And the WSD estimator can be defined as

$$\hat{f}_i^{WSD} = \frac{WSD_i}{\sqrt{\frac{1}{K}\sum_{j\in N_i} WSD_j^2}} \qquad (A.3)$$

with

$$WSD_j = \sqrt{1.081 \cdot \frac{\sum_{l=1}^{n_j} w_{l,j}\bar{r}_{l,j}^2}{\sum_{l=1}^{n_j} w_{l,j}}}. \qquad (A.4)$$

Where the weight $w_{l,j} = w(\bar{r}_{l,j}/\hat{f}_i^{ShortH})$, and the weight function $w(z) = 1$ if $z^2 \le x$ and 0 otherwise. I follow a two-stage procedure to estimate the periodicity. Firstly, I use $x = 4^2$ to estimate the $\hat{f}_i^{WSD_0}$, that means removing those rescaled returns $\bar{r}_t$ based on 4 standard deviations away from the average. To perform the second-stage estimation $\hat{f}_i^{WSD_1}$, I set the critical value $x = 6.635$ which corresponds to the 99% quantile of the χ2 distribution with one degree of freedom. Without price jumps, $\bar{r}_{l,j}^2 = 0$, generating the $WSD_j = 0$ to average 1% of the returns. If there exist price jumps, more observations will be downweighed. The WSD_j in Eq. (A.4) has a 69% efficiency under normality of r_is, which is more efficient than the ShortH. Then I define the final periodicity factor $f_t = \hat{f}_i^{WSD_0} \cdot \hat{f}_i^{WSD_1}$. I adopt one day as a recurring cycle and mark all the returns of an agricultural commodity future with the same periodicity factor if they happen in the same 5-minute time interval of different days.

REFERENCES

Andersen, Torben G., and Tim Bollerslev. "Deutsche mark–dollar volatility: intraday activity patterns, macroeconomic announcements, and longer run dependencies." the Journal of Finance 53, no. 1 (1998): 219-265.

Bacry, Emmanuel, Iacopo Mastromatteo, and Jean-François Muzy. "Hawkes processes in finance." Market Microstructure and Liquidity 1, no. 01 (2015): 1550005.

Bai, Man-Ying, and Hai-Bo Zhu. "Power law and multiscaling properties of the Chinese stock market." Physica A: Statistical Mechanics and its Applications 389, no. 9 (2010): 1883-1890.

Barndorff-Nielsen, Ole E., and Neil Shephard. "Power and bipower variation with stochastic volatility and jumps." Journal of financial econometrics 2, no. 1 (2004): 1-37.

Blanc, Pierre, Jonathan Donier, and J-P. Bouchaud. "Quadratic Hawkes processes for financial prices." Quantitative Finance 17, no. 2 (2017): 171-188.

Bouchaud, Jean-Philippe, and Marc Potters. Theory of financial risks. Vol. 12. From Statistical Physics to Risk Management: Cambridge University Press, Cambridge, 2000.

Boudt, Kris, Christophe Croux, and Sébastien Laurent. "Robust estimation of intraweek periodicity in volatility and jump detection." Journal of Empirical Finance 18, no. 2 (2011): 353-367.

Cao, An NQ, and Michel A. Robe. "Market uncertainty and sentiment around USDA announcements." Journal of Futures Markets 42, no. 2 (2022): 250-275.

Couleau, A., Serra, T. and Garcia, P. , Are Corn Futures Prices Getting “Jumpy”?. American Journal of Agricultural Economics, 102: 569-588. (2020).

Crane, Riley, and Didier Sornette. "Robust dynamic classes revealed by measuring the response function of a social system." Proceedings of the National Academy of Sciences 105, no. 41 (2008): 15649-15653.

De Bondt, Werner FM, and Richard Thaler. "Does the stock market overreact?" The Journal of finance 40, no. 3 (1985): 793-805.

Fama, Eugene F. "Efficient capital markets." Journal of finance 25, no. 2 (1970): 383-417.

Feng Wu, Robert J. Myers, Zhengfei Guan, Zhiguang Wang. Risk-adjusted implied volatility and its performance in forecasting realized volatility in corn futures prices, Journal of Empirical Finance, Volume 34, 2015, Pages 260-274.

Ferriani, F. and Zoi, P., 2022. The dynamics of price jumps in the stock market: an empirical study on Europe and US. The European Journal of Finance, 28(7), pp.718-742.

Feuerriegel, Stefan, and Dirk Neumann. "News or noise? How news drives commodity prices." (2013).

Gabaix, Xavier, Parameswaran Gopikrishnan, Vasiliki Plerou, and H. Eugene Stanley. "A theory of power-law distributions in financial market fluctuations." Nature 423, no. 6937 (2003): 267-270.

Goh, K-I., and A-L. Barabási. "Burstiness and memory in complex systems." Europhysics Letters 81, no. 4 (2008): 48002.

Grossman, S. (1976), On the efficiency of competitive stock markets where traders have diverse information, The Journal of Finance 31(2), 573–585.

Hardiman, Stephen J., Nicolas Bercot, and Jean-Philippe Bouchaud. "Critical reflexivity in financial markets: a Hawkes process analysis." The European Physical Journal B 86 (2013): 1-9.

Helmstetter, Agnes, and Didier Sornette. "Subcritical and supercritical regimes in epidemic models of earthquake aftershocks." Journal of Geophysical Research: Solid Earth 107, no. B10 (2002): ESE-10.

Hopman, Carl. "Are supply and demand driving stock prices." Quantitative Finance (2002).

Joulin, Armand, Augustin Lefevre, Daniel Grunberg, and Jean-Philippe Bouchaud. "Stock price jumps: news and volume play a minor role." arXiv preprint arXiv:0803.1769 (2008).

Koekebakker, Steen, and Gudbrand Lien. "Volatility and price jumps in agricultural futures prices—evidence from wheat options." American Journal of Agricultural Economics 86, no. 4 (2004): 1018-1031.

Kou, Steven G. "A jump-diffusion model for option pricing." Management science 48, no. 8 (2002): 1086-1101.

Kwon, Junhyeon, Yingcai Zheng, and Mikyoung Jun. "Flexible spatio-temporal Hawkes process models for earthquake occurrences." Spatial Statistics 54 (2023): 100728.

Lechthaler, Filippo, and Lisa Leinert. "Moody oil-What is driving the crude oil price?." (2012).

Lee, Suzanne S., and Per A. Mykland. "Jumps in financial markets: A new nonparametric test and jump dynamics." The Review of Financial Studies 21, no. 6 (2008): 2535-2563.

Maheu, John M., and Thomas H. McCurdy. "News arrival, jump dynamics, and volatility components for individual stock returns." The Journal of Finance 59, no. 2 (2004): 755-793.

Malkiel, Burton G. "Efficient market hypothesis." P. Newman, M. Milgate and J. Eatwell (eds.), New Palgrave Dictionary of Money and Finance (1992).

Marcaccioli, Riccardo, Jean-Philippe Bouchaud, and Michael Benzaquen. "Exogenous and endogenous price jumps belong to different dynamical classes." Journal of Statistical Mechanics: Theory and Experiment 2022, no. 2 (2022): 023403.

Martin, R. Douglas, and Ruben H. Zamar. "Bias robust estimation of scale." The Annals of Statistics 21, no. 2 (1993): 991-1017.

Mohler, George O., Martin B. Short, P. Jeffrey Brantingham, Frederic Paik Schoenberg, and George E. Tita. "Self-exciting point process modeling of crime." Journal of the american statistical association 106, no. 493 (2011): 100-108.

Nigatu, Getachew, Flavius Badau, Ralph Seeley, and James Hansen. "Factors Contributing to Changes in Agricultural Commodity Prices and Trade for the United States and the World". U.S. Department of Agriculture, Economic Research Service, (2020), ERR-272.

Roberts, Harry. "Statistical versus clinical prediction of the stock market." Unpublished manuscript 252 (1967).

Rousseeuw, Peter J., and Annick M. Leroy. "A robust scale estimator based on the shortest half." Statistica Neerlandica 42, no. 2 (1988): 103-116.

Schmitz, A., Wang, Z. and Kimn, J.-H. A Jump Diffusion Model for Agricultural Commodities with Bayesian Analysis. Journal of Futures Markets, 34: 235-260. (2014).

Sewell, Martin. "History of the efficient market hypothesis." Rn 11, no. 04 (2011): 04.

Shiller, R. J. (1981), Do stock prices move too much to be justified by subsequent changes in dividends?, The American Economic Review 71(3), 421–436.

Sornette, Didier, and Agnès Helmstetter. "Endogenous versus exogenous shocks in systems with memory." Physica A: Statistical Mechanics and its Applications 318, no. 3-4 (2003): 577-591.

Sornette, Didier, Fabrice Deschâtres, Thomas Gilbert, and Yann Ageon. "Endogenous versus exogenous shocks in complex networks: An empirical test using book sale rankings." Physical Review Letters 93, no. 22 (2004): 228701.

Subrahmanyam, Avanidhar, and Sheridan Titman. "Financial market shocks and the macroeconomy." The Review of Financial Studies 26, no. 11 (2013): 2687-2717.

Taylor, Stephen J., and Xinzhong Xu. "The incremental volatility information in one million foreign exchange quotations." Journal of Empirical Finance 4, no. 4 (1997): 317-340.

Yang, Yao, and Berna Karali. "Do USDA Reports Contribute to Extreme Price Volatility? The Case of the Soybean Complex." (2022).

Yuan, Baichuan, Hao Li, Andrea L. Bertozzi, P. Jeffrey Brantingham, and Mason A. Porter. "Multivariate spatiotemporal hawkes processes and network reconstruction." SIAM Journal on Mathematics of Data Science 1, no. 2 (2019): 356-382.