8-2023

Do FAS Large Export Sales Reports Move Commodity Futures Markets?

William Johnson
University of Arkansas, Fayetteville

Do FAS Large Export Sales Reports Move Commodity Futures Markets?

A thesis submitted in partial fulfillment
of the requirements for the degree of
Master of Science in Agricultural Economics

by

William Johnson
University of Arkansas
Bachelor of Science in Agricultural Food and Life Sciences in Agricultural Business

August 2023
University of Arkansas

This thesis is approved for recommendation to the Graduate Council.

Andrew McKenzie, Ph.D.
Thesis Director

John Anderson, Ph.D.
Committee Member

Eunchun Park, Ph.D.
Committee Member

Abstract

This study sought to determine if the Large Export Reports from the USDA Foreign Agriculture Service have any discernible market-moving impact on particular commodity futures markets. The effect of the reports on futures markets is investigated using four different methods: the average return event study approach, the multiple regression with binary variables model, the autoregressive time series model, and the event spline model. The average returns showed that the FAS's Large Export Sales Reports had no influence on the pricing of wheat futures, but did have a considerable impact on the returns from days before and after the release of the sales report. It became evident that the reports' impact on returns on the futures price depended on the magnitude of sales and the sales destination (China, Other, or various destinations, including China). The size of the sale in the large export sales reports positively impacted both corn and soybean returns, according to the preliminary findings of the multiple linear regressions. Yet, only large corn export reports showed statistically significant destination binary variables. Reports having a country or countries other than China as their destination (Other) and reports with numerous destinations, including China as their destination (Combination), were determined to be statistically significant, where they had a positive and negative connection, respectively. The returns for the whole reporting period were the next thing we wanted to examine. In addition to assessing the entire data period, we also produced new time series datasets to evaluate whether reports that fall in the event window of other reports alter their effects. Compared to the first multiple linear regressions, the original time series data from an ARMA-GARCH model showed the same outcomes. The volume of the sale, and more precisely, the volume of the sale with various destinations, are statistically important in corn reports. The influence of the variables on the returns was not changed by the regression findings using reports

released within the event window. Furthermore, a linear spline model was employed to illustrate how the returns are influenced by the various sales sizes in reports. For corn, we discovered that larger sale sizes have a far more significant influence on returns, but the exact reverse is true for soybeans. Contrary to the primary goal of this study, there is little data to suggest that China alone has a significant market influence. Instead, price shifts occur as soon as the market has a chance to respond to the announcement of the significant sale report, which is often within the first two minutes of trading. Making the reports available in real-time or at another time that does not coincide with trading hours might be advised to evaluate the implications of these reports more thoroughly so as to determine if the sale was known in the market before publication. The findings of the numerous soybean returns analyses indicate that the futures price returns are mostly unaffected by the magnitude of the sale in reports and are unaffected by the destinations. In addition, different representations of the size of reports and particular destinations changed throughout the research, which affected how the results were scaled.

Acknowledgments

I want to start by expressing my gratitude to Dr. Andrew McKenzie for his leadership and unwavering encouragement during my graduate studies. Without Dr. McKenzie's dedication to my success, I would not be as confident in realizing my aspirations. I would also like to thank Dr. Eunchun Park for his assistance in understanding the methods and analysis used in this research. In addition, I would like to thank Dr. John Anderson and the entire faculty for their continued support and amiability while pursuing my studies. Finally, I would like to extend my gratitude to my fellow graduates, friends, and family for their constant encouragement.

Table of Contents

List of Tables

1. Introduction

1.1 Problem Statement and Reasoning

In the agricultural sector of the United States, commodity futures markets are essential because they give farmers, processors, and traders a chance to control price risk and improve pricing transparency. The futures market functions by letting participants purchase and sell standardized contracts for future deliveries of a particular commodity, such as corn, soybeans, or wheat. Futures contracts provide a means for commodity producers and end users to hedge against price fluctuations in the cash market, providing more certainty for farmers and other market participants. In effect, hedging with futures contracts allows commodity producers to lock in a price for their product before it is harvested.

Through its Export Sales Reporting Program, the USDA contributes significantly to the flow of relevant information affecting the commodities futures market. Participants in the market utilize these reports to follow market-relevant supply and demand developments and guide their trading decisions. This program provides weekly updates on U.S. agricultural exports and information on sales, shipments, and destinations. Market participants frequently use the USDA's reports, which are seen as a necessary indication of both market developments and the state of the agricultural industry. The Export Sales Reporting Program is a crucial tool for commodity traders and other market participants in assessing demand for U.S. agricultural products, both domestically and abroad. It also offers valuable information on export trends and market conditions, assisting participants in deciding their futures trading activities.

A portion of the reporting is the Daily Export Sales Reports. As the Branch Chief of the Foreign Agricultural Service (FAS) Export Sales reporting outlined, “a single statistic reveals the significance of the program: in a typical year, the program monitors more than 40 percent of total

U.S. agricultural exports." He continues, "The program also serves as an early alert on the possible impact foreign sales may have on U.S. supplies and prices." Burr, P. (n.d.).

The unexpectedly massive grain purchases by the Soviet Union in 1972, known as "The Great Russian Grain Robbery," originated the need for the Export Sales Reporting Program. The Soviet Union purchased nearly 10 million tons of wheat, corn, and feed grains, severely depleting U.S. grain supplies and causing prices to soar. (The New York Times, 1972). This event had a tremendous impact on the U.S. grain market. According to *The Guardian*, Soviet grain purchases "have sent world grain prices soaring and created severe shortages in some countries" (Smith, 2010). The lack of grain in the United States and other nations had an adverse impact worldwide, driving up prices and upsetting the global grain market.

Before the implementation of the reporting program, it was challenging for the public to learn about export transactions. The program promotes price stability by ensuring that everyone has simultaneous access to the same information. Ultimately, the program's data contributes to improved market stability, efficiency, and price discovery. It is hard to overstate the value of the USDA's Export Sales Reporting Program to the commodity futures market.

1.2 Objective

This research aims to understand how the Daily Large Export Sales Reports affect several commodity futures markets. The first approach to finding a market reaction is to use an event study approach to see if the report's release is a "surprise" to the market. If the report is significant to players in the exchange, prices should unmistakably rise in response to the new knowledge.

2. Literature Review

2.1 Introduction

The influence of public knowledge on various futures markets has been demonstrated in earlier publications that utilize related research. Although the distinctive nature of this research makes it unique, the methods and findings can be compared with other studies from the last several decades. The USDA's *World Agricultural Supply and Demand Estimates* (*WASDE*) reports are often the focus when tracking the volatility of commodity futures markets in relation to the release of public information. While the Foreign Agriculture Service's daily Export Sales Reports are sporadic in their announcement, the *WASDE* reports typically have a scheduled release date of once a month. As a result, the assessment and outcomes of this paper will differ slightly from those of several earlier studies.

2.2 Studies on The Effects of USDA Reports

In 1976 Franklin Edwin Hokana, an economist for the USDA, wrote an article for the *Foreign Agriculture* journal titled "USDA To Continue Monitoring All Export Sales Contracts." In this article, Hokana detailed how the FAS reporting system is meant to be used as an "early-warning" to price shocks. He was researching the relationship between grain prices and basis contract cancellations. The finding that fewer basis contracts are canceled as the market becomes more stable demonstrates the value of the reporting system in helping to comprehend how the price influences the consistency of basis contracts. (Hokana, 1976).

Patterson and Brorsen (1993) examined if the weekly Export Sales reports between 1980-1990 offered the market any new information. The change in futures prices on the days surrounding the report's release is assessed using the event study approach. GARCH models are used to evaluate price fluctuations because it is known that futures pricing variance varies over time. The findings show that traders were anticipating the contents of the report. They find little evidence to support the claim that the report gives the market any new information. (Patterson & Brorsen, 1993).

Colling, Irwin, and Zulauf (1996) analyzed if the weekly Export Sales and Export Inspections reports provide news to the market. Using nearby corn, soybean, and wheat futures prices data between 1988-1991 and regression analysis, they discovered that the reaction to new information is negligible and only accounts for 2% of the overall change in pricing. The authors also find that the level of exports affects the reaction's strength, with larger export levels causing a more vigorous market response. The study emphasizes the value of keeping an eye on USDA reports for traders in the agricultural commodities markets and contends that the data found in these reports may be utilized to make intelligent trading choices. (Colling et al., 1996).

Xie, Isengildina-Massa, Dwyer, and Sharp (2016) look at how publicly available and semi-public information affects the cotton futures market. The study examines how market players make trading decisions in response to USDA reports, weather forecasts, and satellite imaging data. According to the authors, the release of publicly available information significantly influences cotton futures prices, evidenced by the fact that prices immediately fluctuated following the release of the report. Additionally, they discovered that, compared to publicly available information, the market responds less strongly to semi-public information like weather forecasts and satellite images. The study recommends that market players actively monitor public information sources, particularly USDA data, to make knowledgeable trading decisions in the cotton futures market. (Xie et al., 2016).

Adjemian and Irwin (2018) analyze how corn and soybean futures markets responded to the release of USDA reports using high-frequency data to follow price changes in real-time. The study looks at how quickly market players receive and respond to the information and discovers that prices react minutes after the report is released. The authors also suggest market responses are more potent when the USDA announcements include unexpected or surprising information. The analysis shows the difficulties involved in evaluating and responding to USDA reports in real-time, given the complexity of the data and the possibility that different market players would act based on various interpretations of the same data. The study underlines the necessity for traders and analysts to properly evaluate and respond to USDA reports and the significance of timely and accurate information in commodities markets. (Adjemian & Irwin, 2018).

3. Data

The data used in this research is created using a merger of the size and time of the Foreign Agriculture Service's Large Export Sales Reports and time series intraday data of commodity futures prices, volumes, and futures returns collected from Barchart.

3.1 USDA Foreign Agricultural Service Reports

The Foreign Agriculture Service announces a sale of 100,000 metric tons or more of one commodity in a single day to one destination or 200,000 metric tons or more of one commodity in any reporting week to one destination. Exporters must report to FAS the amount traded, the commodity type and classification, the marketing year of the shipment, and the destination by 3:00 p.m. E.T. on the business day after the transaction. The program regularly meets with exporters to check and reinforce suitable reporting methods. Exporters are also required to provide quarterly contract information reports, which confirm stated activities. The information is kept private and is only shared in aggregate form. Anyone who willfully violates the requirements to disclose export sales can be punished with a fine of up to $25,000 or a year imprisonment, or both (though this punishment has never been known to be enforced). The program offers weekly information on the sales activity for 40 American exported agricultural products, such as feed grains, wheat, cotton, and cattle. Burr, P. (n.d.). Corn, soybeans, and wheat were the commodities selected for analysis. The study's first examination concentrated on reports to China. The reports cover a range starting January 3rd, 2011, through August 1st, 2022. In further analysis, all reports for every destination that fell in the date range were included. Tables 3.1, 3.2, and 3.3 and figures 3.1, 3.2, and 3.3 below represent the commodities' export destinations and sizes.

Tables and Figures

Figure 3.1: World Map of Corn Commodity Exports Destination and Total Volume

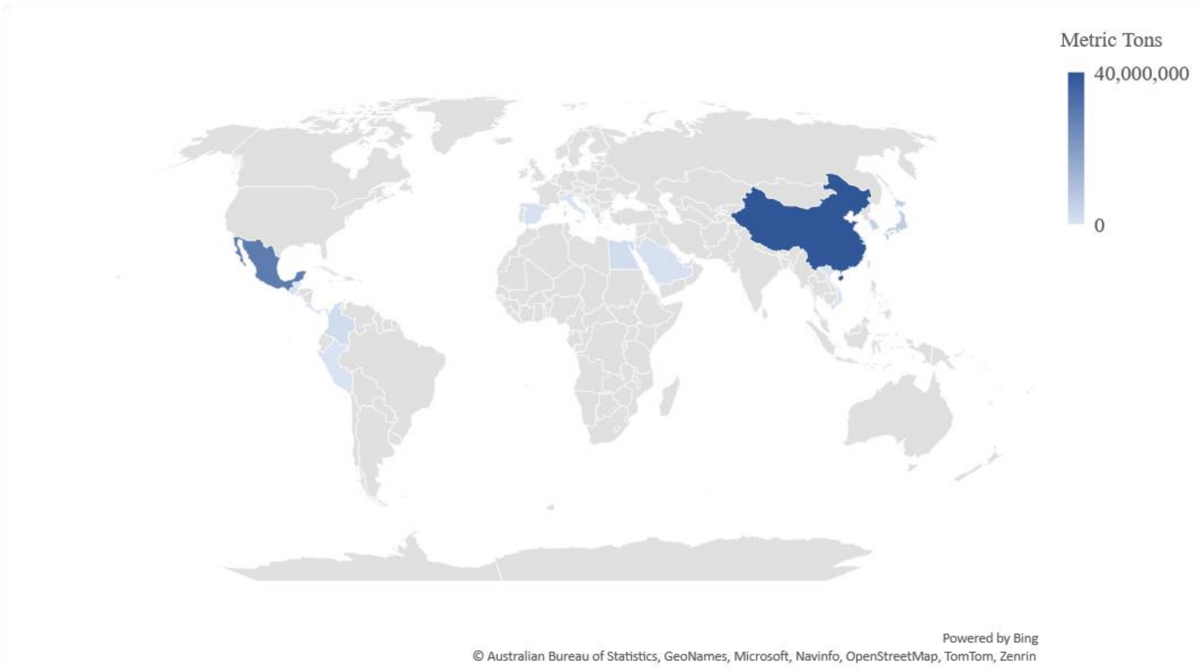


Table 3.1: Corn Commodity Reports Destinations and Sizes (in Metric Tons)

	N	Total	Mean	St. Dev.	Min.	Max.
Country						
China	47	39,530,000	841,064	536,362	120,000	2,108,000
Colombia	15	1,846,000	123,067	28,018	100,000	190,000
Costa Rica	5	723,786	144,757	32,236	106,162	195,338
Egypt	11	1,813,000	164,818	73,414	100,000	340,000
Guatemala	4	504,884	126,221	13,102	114,224	138,403
Israel	2	232,000	116,000	22,627	100,000	132,000
Italy	1	105,000	105,000	N/A	105,000	105,000
Japan	40	5,968,208	149,205	52,823	100,000	278,384
Mexico	111	29,085,254	262,029	317,852	91,440[2]	1,844,040
Panama	2	221,100	110,550	9,122	104,100	117,000
Peru	1	110,000	110,000	N/A	110,000	110,000
Saudi Arabia	4	630,000	157,500	35,000	140,000	210,000
South Korea	29	4,508,000	155,448	49,055	123,000	332,000
Spain	4	534,000	133,500	12,369	120,000	150,000
Taiwan	3	520,000	173,333	75,056	130,000	260,000
Unknown[1]	145	23,699,210	163,443	75,914	100,000	636,524
Vietnam	1	130,000	130,000	N/A	130,000	130,000
Total	375	110,160,442	293,761	358,051	1,893,926	2,108,000

[1] "Unknown" destination is listed as a reportable destination within the Export Sales Report because a destination had not been determined at the time of contract signing. When a destination is specified, exporters can "Change Destinations," reducing the outstanding commitment to unknown destinations and moving them to the designated country before the export is reported.

[2] Reports that are less than 100,000 metric tons are from grouped reports with other commodities.

Figure 3.2: World Map of Soybean Commodity Exports Destination and Total Volume

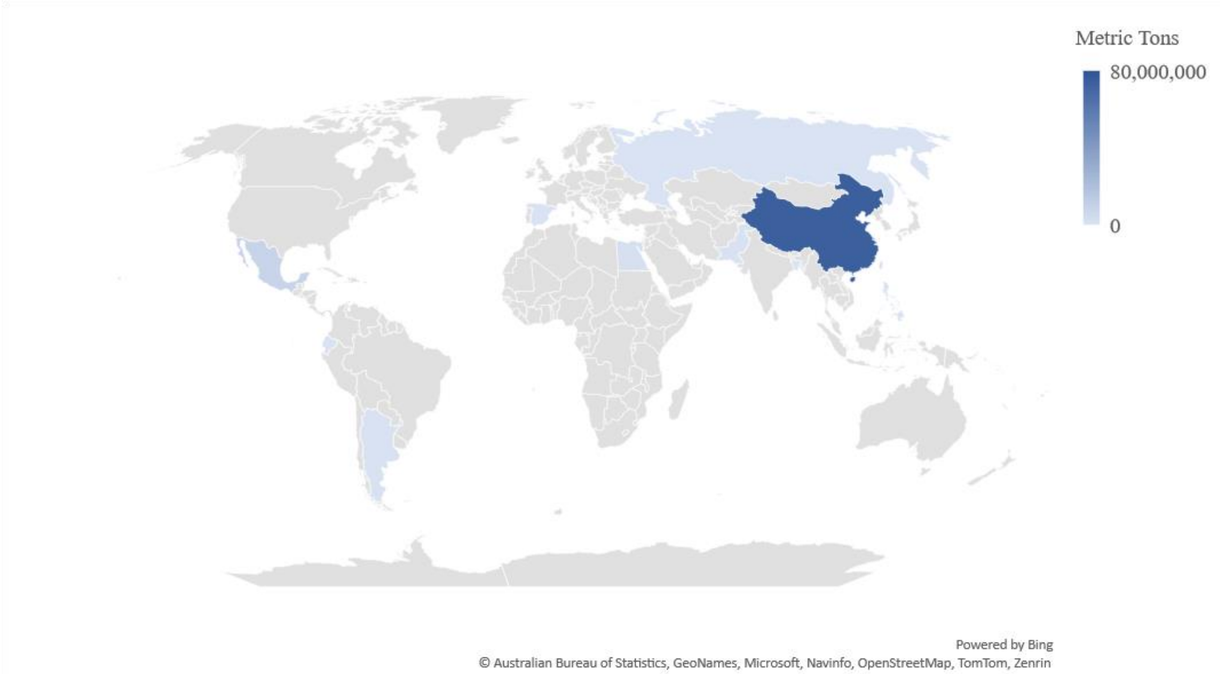


Table 3.2: Soybean Commodity Reports Destinations and Sizes (in Metric Tons)

	N	Total	Mean	St. Dev.	Min.	Max.
Country						
Argentina	4	490,000	122,500	5,000	120,000	130,000
Bangladesh	2	220,000	110,000	0	110,000	110,000
China	292	74,095,300	253,751	289,128	66,000[2]	2,923,000
Ecuador	1	120,000	120,000	N/A	120,000	120,000
Egypt	9	1,508,000	167,556	150,478	100,000	568,000
Mexico	48	8,454,261	176,130	106,677	100,000	671,934
Pakistan	3	683,500	227,833	112,704	132,000	352,000
Philippines	2	266,000	133,000	0	133,000	133,000
Russia	1	120,000	120,000	N/A	120,000	120,000
Spain	1	198,000	198,000	N/A	198,000	198,000
Taiwan	3	342,000	114,000	8,718	104,000	120,000
Unknown[1]	287	55,498,906	193,376	118,570	40,000[2]	1,360,000
Total	535	141,995,967	265,413	256,988	1,343,000	2,923,000

1 "Unknown" destination is listed as a reportable destination within the Export Sales Report because a destination had not been determined at the time of contract signing. When a destination is specified, exporters can "Change Destinations," reducing the outstanding commitment to unknown destinations and moving them to the designated country before the export is reported.

2 Reports that are less than 100,000 metric tons are from grouped reports with other commodities.

Figure 3.3: World Map of Wheat Commodity Exports Destination and Total Volume

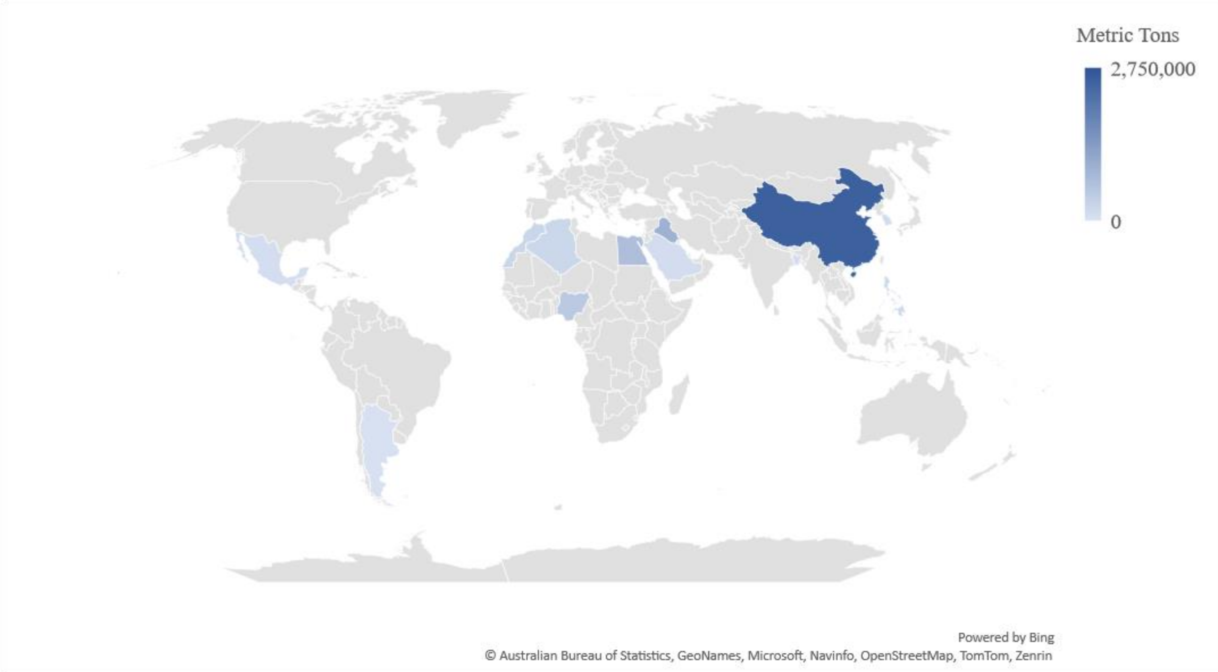


Table 3.3: Wheat Commodity Reports Destinations and Sizes (in Metric Tons)

Country	N	Total	Mean	St. Dev.	Min.	Max.
Algeria	2	240,000	120,000	0	120,000	120,000
Argentina	1	30,000	30,000	N/A	30,000[2]	30,000
Bangladesh	1	120,000	120,000	N/A	120,000	120,000
China	10	2,524,000	252,400	134,246	100,000	480,000
Egypt	6	710,000	118,333	4,082	110,000	120,000
Iraq	6	950,000	158,333	80,104	100,000	300,000
Mexico	1	104,202	104,202	N/A	104,202	104,202
Morocco	2	260,000	130,000	21,213	115,000	145,000
Nigeria	4	506,000	126,500	8,544	120,000	327,300
Philippines	1	193,000	193,000	N/A	193,000	193,000
Saudi Arabia	1	120,000	120,000	N/A	120,000	120,000
South Korea	1	130,000	130,000	N/A	130,000	130,000
Unknown[1]	14	1,835,250	131,089	29,096	100,000	224,000
Total	48	7,722,452	160,884	87,971	1,462,202	480,000

1 "Unknown" destination is listed as a reportable destination within the Export Sales Report because a destination had not been determined at the time of contract signing. When a destination is specified, exporters can "Change Destinations," reducing the outstanding commitment to unknown destinations and moving them to the designated country before the export is reported.

2 Reports that are less than 100,000 metric tons are from grouped reports with other commodities.

3.2 Intraday Futures Prices, Volumes, and Returns

For commodities traded on exchanges, in this example, the Chicago Board of Trade (CBOT) intraday commodity futures prices refer to the current market pricing of futures contracts for certain commodities. Futures contracts are agreements between buyers and sellers to exchange a certain amount of a commodity at a specified date in the future at an agreed . Intraday pricing is dynamic throughout the trading day and represents the market's current supply and demand dynamics. Trading choices are based on these prices, which traders, investors, and analysts use to track market movements, evaluate market risks, and place trades. (McGinnis, n.d.). Using an extension through Barchart, data consisted of intraday minute nearby futures prices and volume for corn, soybean, and wheat commodities that roll on volume[1]. This futures data is collected through The Chicago Board of Trade. The data was used to calculate futures returns and percentage volume changes using nearby contracts with the largest trading volume. The date range was chosen to reflect reliable futures data starting January 3rd, 2011, through August 1st, 2022. When moving earlier than 2011, the futures data was found to be inconsistent, even though the Daily Large Export Sales reports extended before this time. [1]

[1] "Rolling on volume" is when a position in a futures contract that is about to expire is closed out and a new position is opened in a contract that has a later expiration date based on the amount of trading activity in the current contract. This method is frequently used by traders in the futures markets for commodities to avoid taking delivery of the underlying commodity and to keep exposure to the commodity's price movement over time.

4. Methods and Results

4.1 Average Returns Event Study Approach

Using an event study approach, researchers can examine how a particular event affected financial markets, such as changes in stock prices, interest rates, or commodities prices. The event is often something unexpected that is anticipated to significantly impact the markets, such as a large policy announcement or an unforeseen disaster.

The process entails examining the price movement of a financial good during a predetermined time frame, usually right before the event. The time frame before and after the event is referred to as the "event window," which refers to the event itself. The researchers can determine the anomalous returns or the returns that can be attributable to the event by comparing the price behavior of the financial instrument during the event window to a benchmark index, such as the market index. (MacKinlay, 1997).

To determine if the Daily Large Export Sales reports of corn (Figure 4.2), soybeans (Figure 4.3), and wheat (Figure 4.4) to China and other countries are unplanned, unknown occurrences in the market, the first approach (the measurement of intraday tick futures price movements surrounding the release of the reports) was explored.

To do this, the estimate of percentage returns or price changes $R_{it} = ln(P_t + 1/P_t)100$ is calculated on minute intervals ranging from 60 minutes before to 60 minutes after the release for a 3-day event window, as shown in Figure 4.1. The report's release occurs between the close of the night trading and the opening of the day trading. From the beginning of the data to 05/20/2012, the night trading for the Chicago Board of Trade (CBOT) ended at 7:14 a.m. Central Time and the day trading opened at 9:30 a.m. Central Time, so the 60 minutes prior starts at 6:14

to 7:14 a.m. Central time and the 60 minutes after starts at 9:30 to 10:30 a.m. Central Time. From 05/21/12 until 04/06/2013, there was no break in the morning. Still, players in the market treated it similarly to how it was previously, and there were very few observations in that period, so the price returns using the 7:14 to 9:30 a.m. Central Time window are used. From 04/07/2013 to the latest date of the recorded large sales report, where most of the reports in our data occur, there is a window between 7:45 a.m. to 8:30 a.m. Central Time. The 60 minutes prior then are 6:45 a.m. to 7:45 a.m. Central Time and the 60 minutes after are 8:30 a.m. to 9:30 a.m. Central Time. Returns are then averaged across the N large sales events in the sample to obtain the mean return for a minute in event time:

$$R_t = \frac{1}{N}\sum_{i=1}^{N} R_{it} \tag{1}$$

The percentage changes in trading volume on minute intervals ranging from 60 minutes before releasing to 60 mins after the release are calculated. Reports are a "surprise" to the market, and if they include "news," prices should unambiguously rise in response to the new information. After this, controlling the size of the sale is done.

Results

Concerning the initial results presented in panels a–c of figure 4.5, when just looking at the reaction of sales to China, there is an immediate and significant price reaction found in the corn futures market of around 0.16% (panel a). This translates into a 1 cent/bu price increase at current prices. However, there is a small but significant corn futures price reversal (-0.09%) in the second minute after release, equivalent to a 0.5 cent/bu price decrease. For large soybean sales, smaller and only marginally significant soybean futures price reaction (0.03%) at 10% level (panel b) is observed in the second minute after release. This translates into a 0.5 cent/bu price increase at current prices. There are no significant wheat futures price and volume reactions to large wheat sales (panel c7). In all cases, the price reactions are quick, with the markets absorbing the information within the first 1–2 minutes after its release.

Panels a–c of figure 4.6 depict the percentage changes in trading volume for large corn sales on the event day, the day before the report release, and the day after the report release, respectively. On the event day, we observe a price spike that is accompanied by a surge in volume in the opening minute, adding greater weight to our evidence that large corn sales to China are newsworthy and add to price discovery (Figure 4.6, panel a). However, there is an essential caveat to this, typically, volume is always higher on the opening bell irrespective of whether a release occurs, so to better gauge the volume impact of the report, the trading volume 1 trading day prior to (Figure 4.6, panel b) and 1 day after (Figure 4.106, panel c) the report release is shown. Trading volume is significantly higher on the event day than the immediate days before and after the report release. This again adds more evidence to the conclusion that large corn sales to China are newsworthy events to the market.

Now controlling for the sizes of the corn sales to China, exciting results are found. Panel a of figure 4.7 shows average returns over the 120-minute event window for corn sales to China larger than 1,000,000 metric tons. These large sales induce significant increases in first-minute returns following the report releases of around 0.5%. Furthermore, panel b of figure 4.7 shows that the very largest sales to China – those in excess of 1,500,000 metric tons – result in significant returns increases of between (0.6% - 0.75%), and at current prices. This translates into price increases of between 4 to 5 cents/bu.

Figures

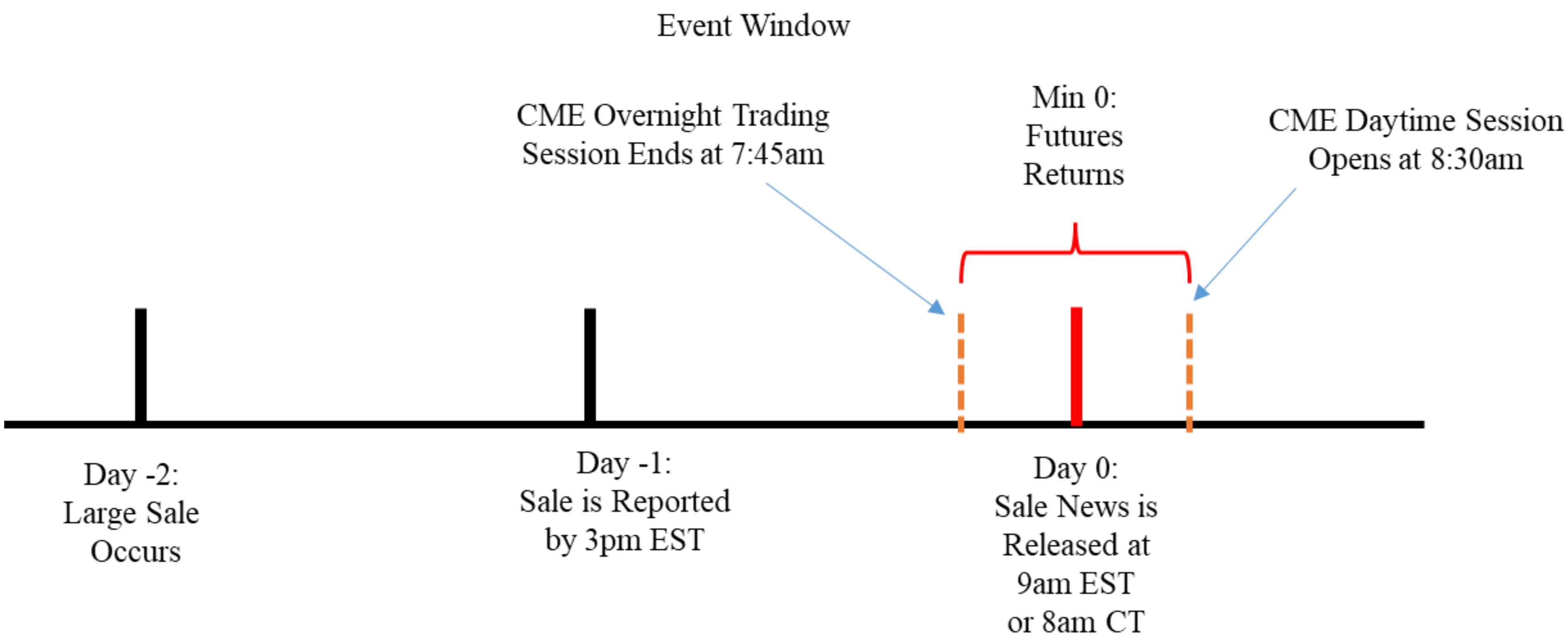


Figure 4.1: Event Window of The Sequence of Intraday Returns and Release of Reports

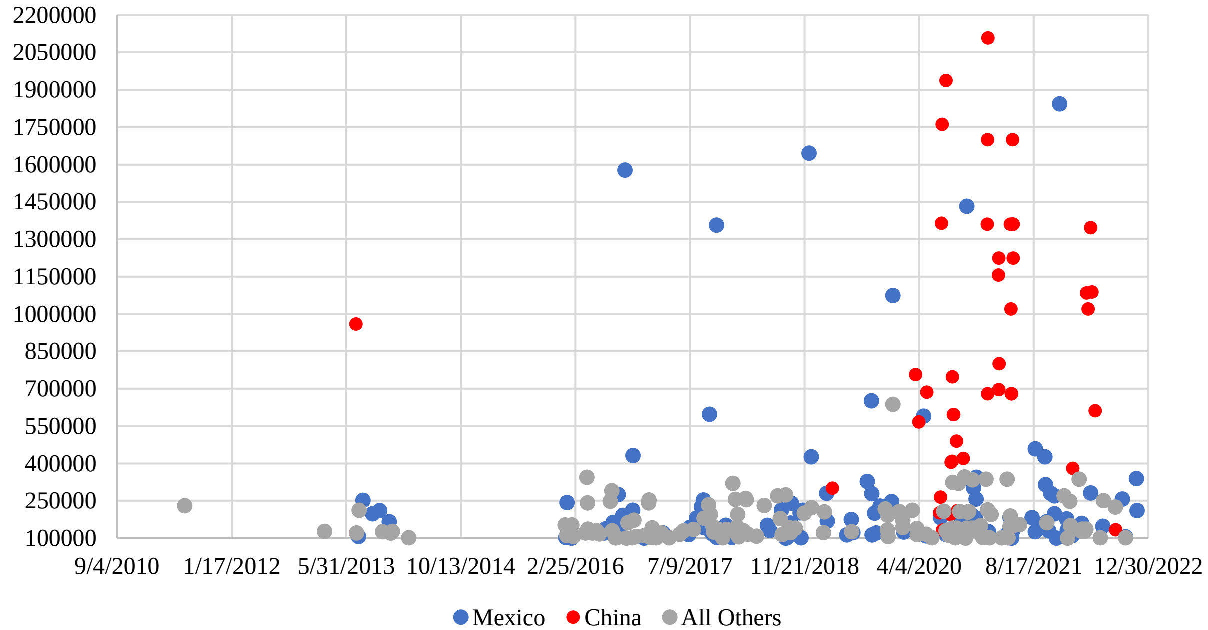


Figure 4.2: Large Daily Corn Sales in Metric Tons Over Time (x-axis: Dates, y-axis: Sale Size in Metric Tons)

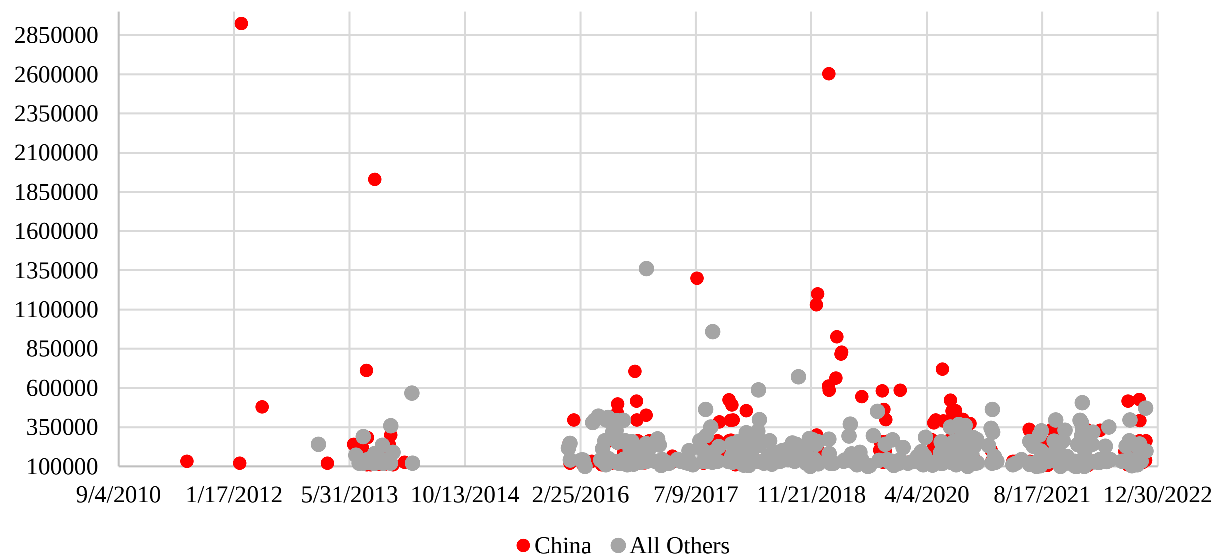


Figure 4.3: Large Daily Soybean Sales in Metric Tons Over Time (x-axis: Dates, y-axis: Sale Size in Metric Tons)

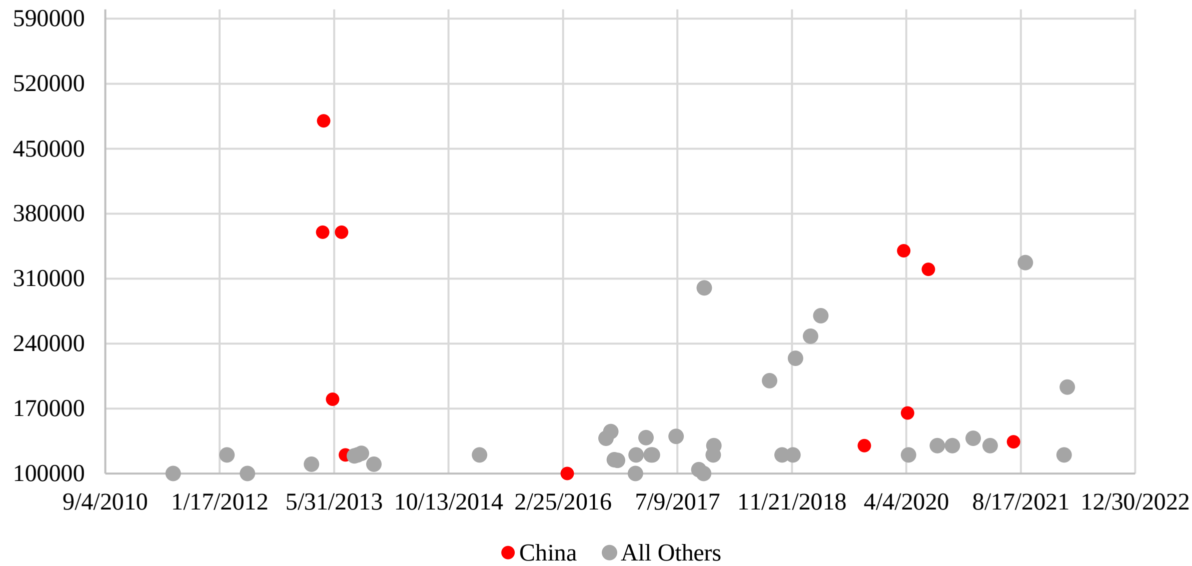


Figure 4.4: Large Daily Wheat Sales in Metric Tons Over Time (x-axis: Dates, y-axis: Sale Size in Metric Tons)

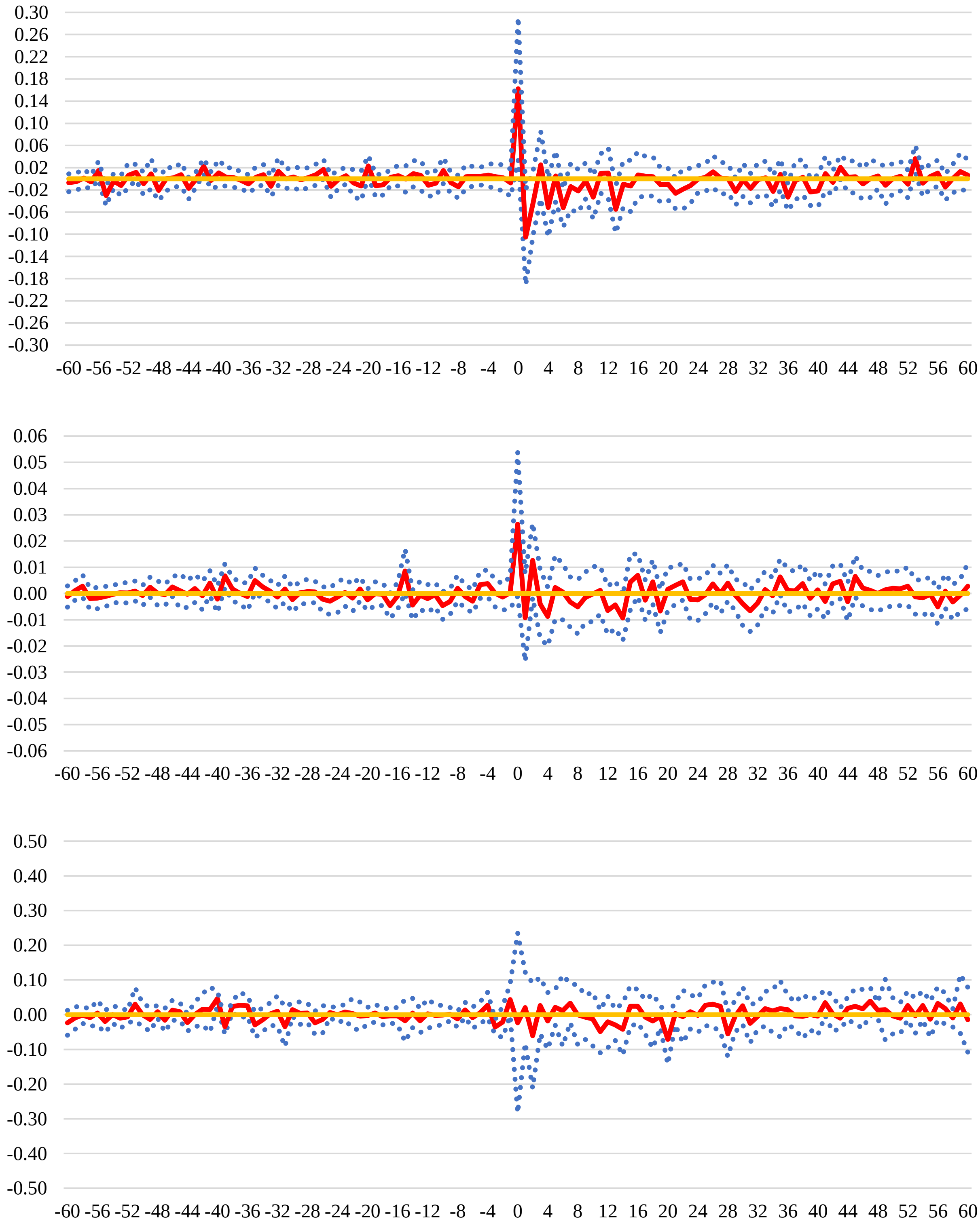


Figure 4.5: Corn (panel a), Soybeans (panel b), and Wheat (panel c) Event Day 0 Returns (x-axis: Minutes Prior and Post to opening minute, y-axis: Returns, with a 95% confidence interval)

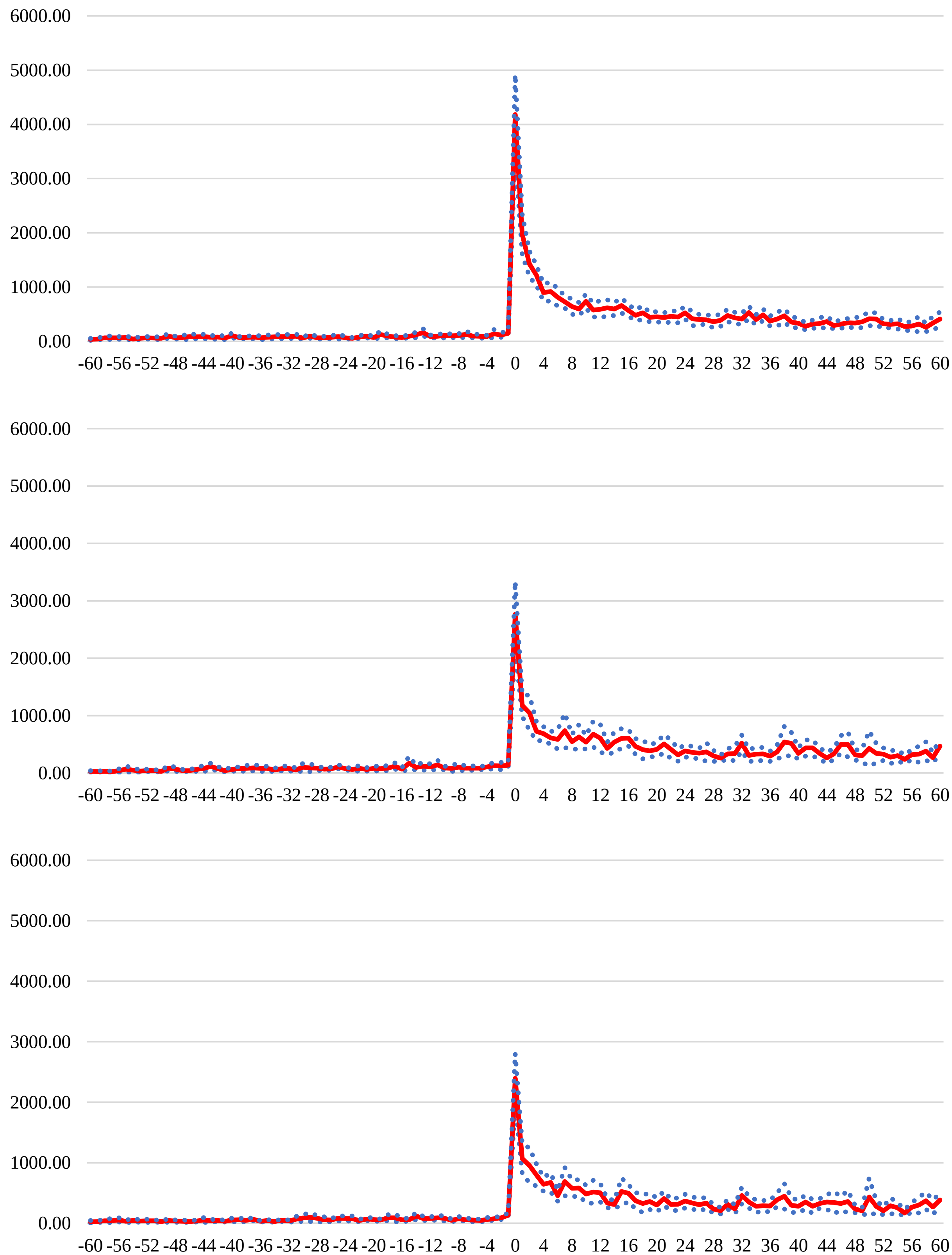


Figure 4.6: Corn Volume Day 0 (panel a), Day -1 (panel b), and Day +1 (panel c) (x-axis: Minutes Prior and Post to opening minute, y-axis: Volume, with a 95% confidence interval)

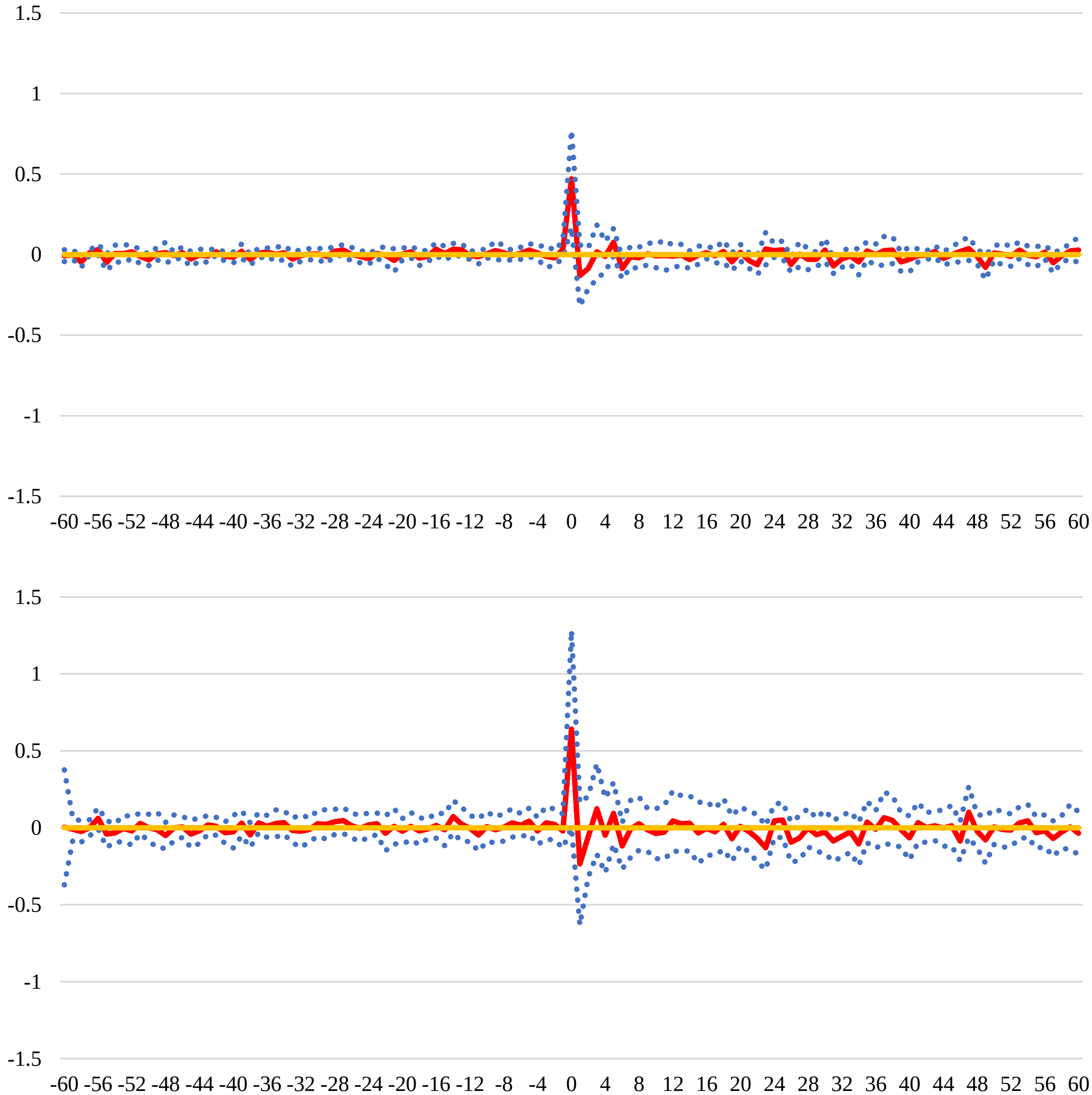


Figure 4.7: Large Corn Sales to China Greater than 1,000,000 MT (panel a) and 1,500,000 MT (panel b) (x-axis: Minutes Prior and Post to opening minute, y-axis: Returns, with a 95% confidence interval)

4.2 Multiple Linear Regression with Binary Variables Model

Multiple linear regression analysis, or multilinear regression with binary variables, looks at the relationship between a dependent variable and two or more independent variables, including one or more binary variables. When categorical variables have only two possible values, often expressed as 0 or 1, binary variables, also known as dummy variables, represent such variables. The dependent variable is often continuous in a multilinear regression with binary variables, and the independent variables may also be continuous or binary. Finding the degree to which each independent variable affects the dependent variable is the objective of the analysis. (Valchanov, 2018).

Since large sales reports of corn and soybean have a significant impact in the first minute of the futures price returns, multiple regression with binary variables model is used on the first-minute returns to understand better the relationship between the country of destination and size of a sale. So again, the percentage returns or price changes $R_0 = ln(P_t + 1/P_t)100$ on the first-minute futures prices are calculated. The large sales report is released between the close of the night trading and the day trading, so the return between the closing price of the night trading and the opening price of the first minute of the day trading (CBOT: 01/03/2011 - 04/06/2013 7:14 to 9:30 a.m. Central Time, 04/07/2013 - 09/01/2022 7:44 to 8:30 a.m. Central Time) is taken. The regression equation below represents a multiple linear regression model where R_0 is the dependent variable, representing futures price returns. Volume size is represented by the independent variable $Sale\ Size$, while D_1 and D_2 are two dummy variables. D_1 is a binary variable that takes the value 1 if the sale in the report comes from China and 0 otherwise. D_2 is a binary variable that accepts the values of 1 if the sales in the report come from China and another country and 0 for observations that originate in any other country. These dummy variables' coefficients, β_1 and β_2, respectively, show how being from China or China and another country affects price returns. The coefficients β_3, β_4, and β_5 represent the effect of volume size and the interaction between volume size and the dummy variables on the price returns. In contrast, the independent variable $Sale\ Size$ represents the natural logarithm of the volume size. The intercept term α is adjusted when D_2 equals 1 and indicates the predicted value of the dependent variable R_0 when all independent variables are equal to zero. With the help of dummy variables that simulate the effects of coming from either China alone or from China and another country, as well as their interaction with volume size, this model aims to explain the relationship between price returns and volume size.

$$R_0 = \alpha + \sum_{i=1}^{2} \beta_i D_i + \beta_3 Sale\ Size + \beta_4 (Sale\ Size * D_1) + \beta_5 (Sale\ Size * D_2) + u_0 \qquad (2)$$

Where

$$D_1 = \begin{cases} 1 \text{ if event day sale is only to China} \\ 0 \text{ otherwise} \end{cases}$$

$$D_2 = \begin{cases} 1 \text{ if event day sale is to China and 1 or more other countries} \\ 0 \text{ otherwise} \end{cases}$$

α: The intercept term represents the expected value of R_0 when all independent variables are zero, i.e., when the volume size is zero and neither D_1 nor D_2 is equal to one.

β_3: This coefficient represents the effect of a one-unit increase in the natural logarithm of the volume size, $Sale\ Size$, on the dependent variable R_0, holding all other independent variables constant. A positive (negative) value of β_3 indicates that an increase in volume size is associated with an increase (decrease) in the expected value of R_0.

β_4: This coefficient represents the difference in the effect of $Sale\ Size$ on R_0 between observations where D_1 = 1 (i.e., observations from China) and observations where D_1 = 0 (i.e., observations from other countries). A positive (negative) value of β_4 indicates that the effect of $Sale\ Size$ on R_0 is stronger (weaker) for China than for other countries.

β_5: This coefficient represents the difference in the effect of $Sale\ Size$ on R_0 between observations where D_2 = 1 (i.e., observations from China and another country) and observations where D_2 = 0 (i.e., observations from other countries). A positive (negative) value of β_5 indicates that the effect of $Sale\ Size$ on R_0 is stronger (weaker) for observations from China and other countries than from other countries.

β_i (for i = 1,2): These coefficients represent the effect of the binary variables (D_1 and D_2) on the dependent variable R_0, holding all other independent variables constant. A positive (negative) value of β_i indicates that the presence of the corresponding binary variable (i.e., China or China and another country) is associated with an increase (decrease) in the expected value of R_0 compared to the reference category (i.e., other countries).

u_0: The error term represents the variability in R_0 that is not explained by the independent variables included in the model.

Results

In the first regression (Table 4.1), the futures price returns of corn are impacted by sales reports with larger volume and multiple sales to China and other countries. The number of observations for corn sales reports is 375, the regression resulted in an r-squared value of 0.096, but due to the complexity of the variables affecting the outcome variable, smaller r-squared values may be more typical and expected. $Sale\ Size$ has a coefficient of 0.06 and is statistically significant at the 5% level with a p-value of 0.03. In short, if the volume size of a product increases by 1 unit on the logarithmic scale, then, assuming that the impacts of the other independent variables are held constant, the price returns are projected to increase by 0.06 units. β_5 which is the size of the sale report multiplied by the dummy variable of the report being a combination of China and other countries, has a coefficient of 0.35, and is statistically significant at the 1% level with a p-value of 0.004. This coefficient shows that the futures price returns increase by 0.35 for every unit increase in the natural logarithm of the volume size when the dummy variable (D_2) is equal to 1. This indicates that the futures price returns are positively correlated with the interaction effect between the natural logarithm of volume size and when the sales report is a combination of China and another country. D_2 is also statistically significant at the 1% level with a p-value of 0.007. The intercept that represents the effect of reports to countries other than China is also statistically significant at the 5% level with a coefficient of -0.72. This means an inverse relationship with corn futures price returns and reports to countries to countries other than China. D_1, China only dummy, and β_4, size of the sale report multiplied by the China-only dummy variable was not statistically significant.

In the second regression (Table 4.1), the futures price returns of soybean are not as impacted by sales reports with larger volume and multiple sales to China and other countries as corn. The number of observations for corn sales reports is 539, and the regression resulted in an r-squared value of 0.003. $Sale\ Size$ has a coefficient of 0.04 and is not statistically significant at the 10% level with a p-value of 0.22. β_5, the size of the sale report multiplied by the dummy variable of the report being a combination of China and other countries, has a coefficient of 0.05 and is not statistically significant at the 10% level with a p-value of 0.39. D_2, D_1, China only dummy, and β_4, size of the sale report multiplied by the China-only dummy variable, were not found to be statistically significant.

In the second regression on corn returns (Table 4.2) the China-only dummy and, as a result, the sale size multiplied by the China-only dummy independent variables are included as they were found to be the only statistically significant variables. The regression resulted in a r-squared value of 0.0997. $Sale\ Size$ has a coefficient of 0.05 and is statistically significant at the 5% level with a p-value of 0.025. If the volume size of a product increases by 1 unit on the logarithmic scale, then, assuming that the impacts of the other independent variables are held constant, the price returns are projected to increase by 0.05 units. β_5, the size of the sale report multiplied by the dummy variable of a combination of China and other countries, has a coefficient of 0.36 and is statistically significant at the 1% level with a p-value of 0.003. This coefficient shows that the futures price returns increase by 0.36 for every unit increase in the natural logarithm of the volume size when the dummy variable (D_2) is equal to 1. This indicates that the futures price returns are positively correlated with the interaction effect between the natural logarithm of volume size and when the sales report is a combination of China and another country. D_2 is also statistically significant at the 1% level with a p-value of 0.005. The predicted corn returns are graphed along the total volume size of each sale in Figure 4.8. We can see in this figure that sales that have a combined destination of China and other countries have a significantly higher impact on returns than reports to just China or other countries.

In the final regression (Table 4.2), the soybean futures price returns and sale sizes are regressed. The futures price returns of soybean are affected by sales reports with larger volumes. Though the simple linear regression resulted in an r-squared value of 0.0076, $Sale\ Size$ has a coefficient of 0.03 and is statistically significant at the 5% level with a p-value of 0.028. So, if the volume size of a product increases by 1 unit on the logarithmic scale, assuming that the impacts of the other independent variables are held constant, the returns are projected to increase by 0.03 units. The predicted soybean returns are graphed along the total volume size of each report in Figure 4.9. In this graph, we see that the larger the sale size of a report, the higher the effect on the return.

Tables and Figure

Table 4.1: Multiple Linear Regression Analysis

		CORN				SOYBEANS		
R-Square		0.096				0.003		
Observations		375				535		
Variable	Coefficient	Std Error	T-Stat	P-value	Coefficient	Std Error	T-Stat	P-value
Intercept (Others)	-0.72	0.33	-2.14	0.033**	-0.48	0.41	-1.19	0.24
Size[1]								
Sale Size	0.06	0.03	2.19	0.03**	0.04	0.03	1.23	0.22
Sale Size * China	-0.05	0.07	-0.79	0.433	-0.02	0.04	-0.42	0.67
Sale Size * China and Others	0.35	0.12	2.87	0.004***	0.05	0.06	0.86	0.39
Destination[2]								
China	0.67	0.88	0.77	0.445	0.22	0.53	0.42	0.68
China and Others	-4.47	1.66	-2.70	0.007***	-0.67	0.75	-0.90	0.37

* indicate significance, * at the 10 % level, ** at the 5 % level, and *** at the 1 % level. See footnotes below.
1. Sale Size Variable and Sale Size Variable Multiplied by Destination Binary Variables
2. Destination Binary Variables

Table 4.2: Multiple Linear Regression Analysis

	CORN				SOYBEANS			
R-Square	0.0997				0.0076			
Observations	375				535			
Variable	Coefficient	Std Error	T-Stat	P-value	Coefficient	Std Error	T-Stat	P-value
Intercept	-0.56	0.25	-2.25	0.025**	-0.41	0.20	-2.08	0.038**
Size[1]								
Sale Size	0.05	0.02	2.32	0.021**	0.03	0.02	2.20	0.028**
Sale Size * China and Others	0.36	0.12	3.02	0.003***	-	-	-	-
Destination[2]								
China and Others	-4.63	1.64	-2.83	0.005***	-	-	-	-

* indicate significance, * at the 10 % level, ** at the 5 % level, and *** at the 1 % level. See footnotes below.

1. Sale Size Variable and Sale Size Variable Multiplied by Destination Binary Variable

2. Destination Binary Variable

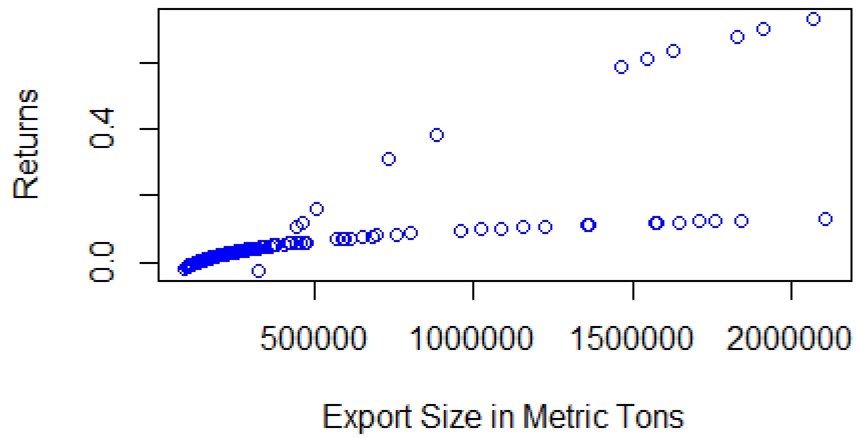


Figure 4.8: Predicted Corn Returns

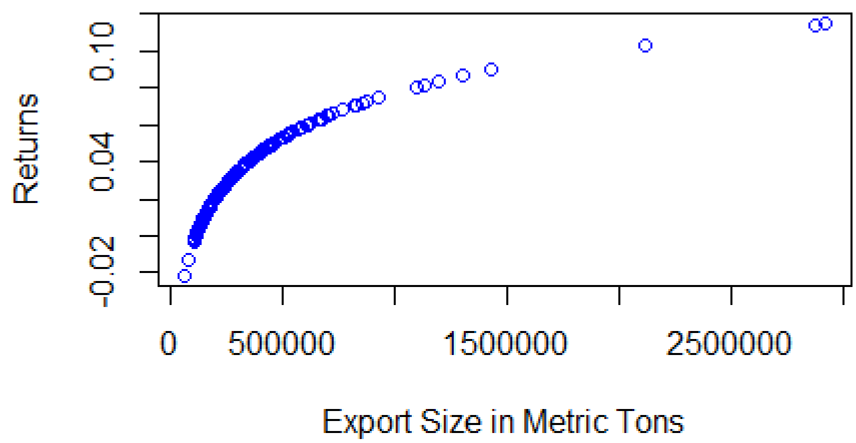


Figure 4.9: Predicted Soybean Returns (x-axis: Sale Size in Metric Tons, y-axis: Returns)

4.3 Autoregressive Moving Average Time series Model with GARCH Errors

A type of time series model known as an autoregressive (A.R.) model uses a variable's historical values to forecast its future values. In an A.R. model, the variable's current value is represented as a linear combination of its previous values and an additional term for unpredictability. The fact that the model is regressing on its previous values is referred to as "autoregressive" behavior. The moving average in a Moving Average (M.A.) model refers to a series of previous error terms or residuals. Using both the variable's own lagged values and the lagged values of the error terms, the M.A. model may characterize a variable's behavior.

An AR model of order p, often known as A.R. (p), is defined as:

$$y_t = \mu + \varphi_1 y_{(t-1)} + \varphi_2 y_{(t-2)} + \cdots + \varphi_p y_{(p-1)} + \varepsilon_t \quad (3)$$

The MA model of order q, often known as M.A. (q), is defined as:

$$y_t = \mu + \varepsilon_t + \theta_1 \varepsilon_{(t-1)} + \theta_2 \varepsilon_{(t-2)} + \cdots + \theta_q \varepsilon_{(q-1)} \quad (4)$$

Then the combination of the two models to create an ARMA (p, q) model, is defined as:

$$y_t = \mu + \varphi_1 y_{(t-1)} + \varphi_2 y_{(t-2)} + \cdots + \varphi_p y_{(p-1)} + \theta_1 u_{(t-1)} + \theta_2 u_{(t-2)} + \cdots + \theta_q u_{(q-1)} + u_t \quad (5)$$

Or

$$y_t = C + \sum_{i=1}^{p} \varphi_i y_{(t-i)} + \sum_{i=1}^{q} \theta_i \varepsilon_{(t-i)} \quad (6)$$

Where y_t is the current value of the appropriate variable at time t. C is the constant term. φ_1, φ_2, ..., and φ_p are the model's parameters for the autoregressive model (or the "lags"). θ_1, θ_2, ..., and θ_q are the model's parameters for the moving average model. ε_t is the error term at time t. The number of historical values included in the linear combination is dependent upon the order of the

ARMA model, p. A higher p and q mean that the model is using more historical data to forecast the current value, which can improve prediction accuracy but also increase the model's complexity. Numerous methods, including the maximum likelihood method, can be used to estimate the parameters of the ARMA model, and methods like the Akaike information criterion (AIC) or Bayesian information criterion (BIC) can be used for determining the model's goodness-of-fit.

The presence of volatility clustering, a typical characteristic of financial time series, is considered by the generalized autoregressive conditional heteroscedasticity (GARCH) model, a type of autoregressive time series model. In a sense, the time series' volatility tends to remain constant and is impacted by the past values' volatility. It is an extension of the Autoregressive Conditional Heteroskedasticity model (ARCH), which models variance as a function of past squared errors to indicate heteroskedasticity or volatility clustering of a time series. The conditional variance of the time series at time t is described as follows in a simple GARCH(p, q) model:

$$y_t = x_t\beta + \varepsilon_t \tag{7}$$

$$\varepsilon_t = \sqrt{h_t}e_t \tag{8}$$

$$h_t \ = \ \omega \ + \sum\nolimits_{i=1}^{q} \alpha_i \varepsilon_{(t-i)}^2 + \sum\nolimits_{i=1}^{p} \beta_j h_{(t-j)} \tag{9}$$

Where

$$\omega > 0, \alpha \geq 0, \beta \geq 0$$

Where y_t is the current value of the appropriate variable at time t. ε_t is the error term at time t. h_t is the conditional variance of the time series at time t. e_t is the error term with a standard normal distribution. Omega (ω) is the constant term representing the long-run variance. The ARCH coefficient, alpha (α), shows how the conditional variance has been impacted by previous squared error terms. The GARCH coefficient, beta (β), shows how previous conditional variances have affected the conditional variance. The GARCH(1,1) model captures the persistence of volatility in a time series in which the squared errors from previous periods and the conditional variances from the previous affect the current conditional variance. To better portray the intricate nature of volatility in a time series, the GARCH model can be expanded to include additional lags of squared errors, conditional variances, and external regressors. (Zivot, 2021).

The equation for a fitted Autoregressive Moving Average - Generalized Autoregressive Conditional Heteroscedasticity (ARMA(2,2)-GARCH(1,1)) is implemented to easily include external regressors and capture both the autocorrelation and volatility clustering contained in the time series data. Let R_t denote the futures price returns at the first minute, this model can be expressed as:

$$R_t = C + \varphi_1 y_{(t-1)} + \varphi_2 y_{(t-2)} + \theta_1 \varepsilon_{(t-1)} + \theta_2 \varepsilon_{(t-2)} + \beta_1(Sale\ Size) + \beta_2(D_1) + \beta_3(D_2)$$

$$+\beta_4(Sale\ Size * D_1) + \beta_5(Sale\ Size * D_2) + \beta_6 D_3 + \sum_{i=7}^{17} \beta_i M_i + \varepsilon_t \tag{10}$$

$$\varepsilon_t = \sqrt{h_t} e_t \tag{11}$$

$$h_t = \omega + \alpha_1 \varepsilon_{(t-1)}^2 + \alpha_2 \varepsilon_{(t-2)}^2 + \gamma_1 h_{(t-1)}^2 + \gamma_2 h_{(t-2)}^2 \tag{12}$$

Where C is the constant term. φ_1 and φ_2 are the model's parameters for the autoregressive model (or the "lags"). θ_1 and θ_2 are the model's parameters for the moving average model. β_1 is the coefficient for the external regressor $Sale\ Size$ that measures the effect that the change in the size of the sale (divided by 1000) of each report has on the futures price returns. β_2 and β_3 are the coefficients for the binary external regressors that measure the effect of the report being from China (D_1) and Combined Countries (D_2) has on the returns. β_2 and β_3 are the coefficients for eternal regressors that measure the effects of the reported size of the sale ($Sale\ Size$) for the destination (D_1 and D_1). These destination dummies were chosen to remain consistent across the models. β_6 represents the coefficient of a dummy variable for days that are not event days (D_3). β_i are the binary external regressors for the months (M_i) that the returns take place to test seasonality. ε_t is the error term at time t. h_t is the conditional variance of the time series at time t. e_t is the error term with a standard normal distribution. ω (omega) is the constant term representing the long-run variance. α (alpha) is the ARCH coefficient that shows how the conditional variance has been impacted by previous squared error terms. γ (beta) is the GARCH coefficient that shows how previous conditional variances have affected the conditional variance.

For this process, both the corn and soybean report and returns data were separated into two different time series sets. The first data sets account for clustering, where daily report observations were removed if they were clustered with other reports where a report was released within the event window (1 day prior) to ensure that returns are not affected by multiple reports that are captured in the market before the release (See Figure 4.1). In addition to clustered reports, observations before 4/08/2013 were removed since the trading hours change. For the corn large exports sales reports, 234 reports are not clustered and released after the change in trading hours; for soybean 298 reports are not clustered and released after the change in trading hours. The subsequent data sets account for all reports and all returns from 1/3/2011 to 12/30/2022, 375 reports for corn and 535 reports for soybeans.

Results

The results presented in Table 4.3 are the first two models that measure the impact of the size and destination of the corn and soybean large export reports (with clustering removed) on the volatility of the returns of first-minute futures prices. The parameters that are significant at the 1% level for the model measuring corn futures price returns are the coefficient for “Sale Size,” both autoregressive coefficients, both moving average coefficients, the constant term in the GARCH model, the ARCH coefficient that shows how the conditional variance in the present is affected by the squared error term from the previous period, and the GARCH coefficient that shows how the conditional variance in the present is affected by the conditional variance from the previous period. The coefficient for “Sale Size” is 0.0002, showing that, assuming all other variables remain constant, a one-unit increase in the “Sale Size” divided by one thousand is related to an anticipated rise in the returns of 0.0002 units. The nearness of the coefficient to zero suggests that the external regressor has little impact on the results. The first autoregressive term's (φ_1) coefficient is 1.432. A positive coefficient denotes a positive correlation between the lagged and current values. The coefficient for the second autoregressive term (φ_2) is -0.607. A negative coefficient indicates that the current value and its second lag have an inverse correlation. The moving average terms (θ_1 and θ_2) have coefficients of -1.387 and 0.600, respectively. While the positive coefficient for θ_2 shows a positive correlation with the second lagged error term, the negative coefficient for θ_1 suggests an inverse correlation between the present value and the first lagged error term. The long-term average volatility is represented by the coefficient for omega (ω), which is 0.0004. The effects of previous squared error terms and previous conditional variances, respectively, on the present conditional variance are shown by the coefficients for alpha (α) and beta (γ). The calculated values for these coefficients are 0.0665 and 0.9326, respectively. No other regressors were found to significantly impact the volatility of the corn futures price returns in the model. Unlike the model for the corn futures price returns, the soybean time series model parameters do not include significant autoregressive or moving average coefficients but do include a significant external regressor. The parameters that are significant at the 1% level for the model measuring soybean futures price returns are the constant term in the GARCH model, the ARCH coefficient, the GARCH coefficient, and the only other significant coefficient (at the 10% level) is that of the dummy variable for reports with combination destinations including China and other countries. The coefficient for omega is zero, signifying that the long-term average volatility is near zero (0.0018). The coefficients for alpha and beta are 0.077 and 0.9, respectively. This implies that the previous squared error terms and conditional variances significantly influence the present conditional variance in the GARCH model. The “Combination” dummy variable's coefficient is estimated to be 0.0702. This indicates that large soybean export sales to the “Combination” destinations positively influence returns.

The next two models measure the impact of the size and destination of the corn and soybean large export reports on the volatility of the returns of first-minute futures prices with the full data represented in table 4.4. The parameters significant at the 1% level for the model measuring corn futures price returns are the constant term in the GARCH model, the ARCH coefficient, the GARCH coefficient, and the “Sale Size” coefficient. The coefficient for “Sale Size” multiplied by the “Combination” destination dummies is significant at the 10% level. The estimated value of the omega coefficient is 0.0002, which indicates a non-zero long-term average volatility in the returns. The coefficients for alpha and beta are 0.04 and 0.959, respectively. This implies that the previous squared error terms and conditional variances significantly influence the present conditional variance in the GARCH model. According to the estimated coefficient for the "Sale Size" variable of 0.0002, "Sale Size" positively affects price returns, which means that larger sale sizes are related to higher price returns. The coefficient for the interaction terms between "Sale Size" and the destination dummy, “Combination,” is 0.003. This indicates that the volume of reports with sales to both China and other countries positively affects returns. The autoregressive and moving average coefficients are not statistically significant, implying that the lagged values of the returns and the error terms from the moving average component have no substantial effect on the present price returns. Unlike the model for the corn futures price returns, the soybean time series model parameters do not include significant external regressors but do include significant autoregressive or moving average coefficients. The parameters significant at the 1% level for the model measuring soybean futures price returns are the constant term in the GARCH model, the ARCH coefficient, and the GARCH coefficient. The second autoregressive term (φ_2) is statistically significant at the 5% level, and the second moving average term (θ_2) is statistically significant at the 10% level. The estimated value of the omega coefficient is 5.8E-05, which indicates a near-zero long-term average volatility in the returns. The coefficients for alpha and beta are 0.029 and 0.97, respectively. This implies that the previous squared error terms and conditional variances significantly influence the present conditional variance in the GARCH model. The coefficient for the second autoregressive term is 0.358, which indicates a positive correlation between the second lagged value and the current value. The coefficient for the second moving average term is -0.274, which indicates a negative correlation with the second lagged error term.

A Weighted Ljung-Box Test on Standardized Residuals and Standardized Squared Residuals was performed for each regression. The results of these tests show that the model's residuals and squared residuals at various lag levels do not exhibit any discernible autocorrelation. There is no sign of systematic volatility in the residuals at these lags, which suggests that the model appropriately reflects the volatility patterns. The Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC) are both measures used to compare the model's goodness of fit. These measures consider the number of parameters in the model as well as the log-likelihood value of the model. A lower value of AIC or BIC indicates a better model fit. The AIC and BIC for the un-clustered corn returns model are 0.0174 and 0.0757, respectively, and the AIC and BIC for the un-clustered soybean returns model are -0.127 and -0.079, respectively. The AIC and BIC for the total corn returns model are 0.544 and 0.594, respectively, and the AIC and BIC for the entire soybean returns model are 0.251 and 0.291, respectively. Some of the values are negative because they are calculated as -2 times the log-likelihood plus a penalty term based on the number of parameters in the model. Since the AIC is lower than the BIC for all models, it indicates that the models may be overfitting the data, meaning that it is too complex and may not generalize well to new data. (Ri et al., 2023).

Tables and Figures

Table 4.3: ARMA-GARCH Model (Un-clustered Data)

Conditional Variance Dynamics

GARCH Model	GARCH(1,1)							
Mean Model	ARMA(2,2)							
Distribution	norm							
	CORN				**SOYBEANS**			
LogLikelihood	3.273				184.813			
	Coefficient	Std Error	T-Stat	P-value	*Coefficient*	*Std Error*	T-Stat	P-value
Optimal Parameters								
$\boldsymbol{C}$	-0.0296	0.024	-1.25	0.210	-0.0171	0.015	-1.15	0.249
Auto-Regressive and Moving Avg.[1]								
$\boldsymbol{\varphi_1}$	1.4327	0.164	8.74	0***	0.1214	0.266	0.46	0.648
$\boldsymbol{\varphi_2}$	-0.6069	0.195	-3.11	0.002***	0.3026	0.243	1.24	0.213
$\boldsymbol{\theta_1}$	-1.3863	0.170	-8.16	0***	-0.1150	0.271	-0.42	0.671
$\boldsymbol{\theta_2}$	0.6003	0.207	2.91	0.004***	-0.2268	0.246	-0.92	0.357
Size[2]								
Sale Size	0.0002	6.0E-05	2.69	0.007***	0.0001	9.5E-05	1.44	0.149
Sale Size * China	0.0001	2.2E-04	0.25	0.802	-0.0001	1.2E-04	-1.01	0.314
Sale Size * Combination	-0.0004	3.1E-04	-1.42	0.155	-0.0001	1.1E-04	-1.37	0.171
Destination[3]								
China ($\boldsymbol{D_1}$)	-0.0741	0.153	-0.48	0.628	0.0212	0.026	0.80	0.422
Combination ($\boldsymbol{D_2}$)	0.2619	0.249	1.05	0.292	0.0702	0.036	1.94	0.052*
Non-Event Days[4]								
Days ($\boldsymbol{D_3}$)	0.0167	0.019	0.86	0.392	-0.0072	0.018	-0.39	0.696
Month[5]								
February	-0.0242	0.022	-1.08	0.278	0.0055	0.021	0.26	0.795
March	-0.0264	0.023	-1.14	0.253	-0.0059	0.021	-0.28	0.781

April	0.0137	0.024	0.57	0.567	-0.0140	0.021	-0.68	0.499
May	0.0297	0.025	1.19	0.236	0.0058	0.022	0.27	0.787
June	-0.0184	0.030	-0.621	0.535	0.0326	0.030	1.083	0.279
July	0.0234	0.029	0.81	0.417	0.0080	0.028	0.28	0.777
August	0.0008	0.026	0.03	0.977	-0.0017	0.022	-0.08	0.940
September	0.0046	0.024	0.19	0.847	-0.0079	0.021	-0.38	0.707
October	0.0210	0.022	0.96	0.335	0.0027	0.020	0.13	0.895
November	-0.0001	0.021	0.00	0.997	-0.0193	0.021	-0.91	0.364
December	0.0149	0.021	0.71	0.476	0.0285	0.022	1.32	0.188
Variance Terms[6]								
$\boldsymbol{\omega}$ (omega)	0.0004	1.5E-04	2.70	0.007***	0.0018	3.2E-04	5.50	0***
$\boldsymbol{\alpha}$ (alpha)	0.0665	0.007	9.32	0***	0.0771	0.011	6.93	0***
$\boldsymbol{\gamma}$ (beta)	0.9326	0.007	135.43	0***	0.9000	0.012	72.60	0***

* Indicates significance, * at the 10 % level, ** at the 5 % level, and *** at the 1 % level. See footnotes below.

1. Autoregressive (A.R.) Terms, φ, and Moving Average (M.A.) Terms, θ
2. Log of Sale Size Variable and Log of Sale Size Variable Multiplied by Destination Binary Variable
3. Destination Binary Variable
4. Non-Event Days Binary Variable
5. Month Binary Variables
6. Variance Terms: Constant Term (omega), Lagged Squared Error Term (alpha), and Lagged Conditional Variance Term (beta)

Table 4.4: ARMA-GARCH Model (Clustered Data)

Conditional Variance Dynamics

GARCH Model	GARCH(1,1)
Mean Model	ARMA(2,2)
Distribution	norm

	CORN				**SOYBEANS**			
LogLikelihood	-809.576				-367.449			
	Coefficient	Std Error	T-Stat	P-value	*Coefficient*	*Std Error*	T-Stat	P-value
Optimal Parameters								
$\boldsymbol{C}$	-0.0393	0.021	-1.88	0.061*	-0.0103	0.016	-0.66	0.507
Auto-Regressive and Moving Avg.[1]								
$\boldsymbol{\varphi_1}$	0.3450	0.762	0.45	0.65	0.2177	0.182	1.20	0.23
$\boldsymbol{\varphi_2}$	0.3508	0.618	0.57	0.57	0.3582	0.155	2.31	0.02**
$\boldsymbol{\theta_1}$	-0.2915	0.764	-0.38	0.70	-0.2136	0.186	-1.15	0.25
$\boldsymbol{\theta_2}$	-0.3158	0.578	-0.55	0.58	-0.2738	0.157	-1.74	0.08*
Size[2]								
Sale Size	0.0002	5.1E-05	2.98	0.003***	0.0002	1.6E-04	1.46	0.145
Sale Size * China	-0.0001	1.2E-04	-0.47	0.639	-0.0002	1.8E-04	-1.12	0.262
Sale Size * Combination	0.0003	1.5E-04	1.89	0.058*	-0.0002	1.7E-04	-1.40	0.162
Destination[3]								
China ($\boldsymbol{D_1}$)	0.1260	0.094	1.34	0.180	0.0274	0.041	0.67	0.506
Combination ($\boldsymbol{D_2}$)	-0.0875	0.166	-0.53	0.598	0.0652	0.048	1.35	0.177
Non-Event Days[4]								
Days ($\boldsymbol{D_3}$)	0.0218	0.016	1.34	0.181	-0.0176	0.032	-0.54	0.587
Month[5]								
February	-0.0199	0.022	-0.89	0.375	0.0028	0.022	0.13	0.899
March	-0.0237	0.023	-1.01	0.313	-0.0041	0.022	-0.18	0.856
April	0.0183	0.025	0.74	0.460	-0.0283	0.023	-1.23	0.220
May	0.0385	0.026	1.47	0.141	-0.0025	0.024	-0.10	0.917

June	-2.2E-05	0.027	-0.001	0.999	0.0288	0.030	0.973	0.330
July	0.0095	0.028	0.34	0.731	-0.0015	0.025	-0.06	0.952
August	0.0052	0.026	0.20	0.841	0.0020	0.024	0.09	0.932
September	0.0085	0.024	0.35	0.726	-0.0143	0.023	-0.63	0.528
October	0.0291	0.023	1.28	0.201	-0.0004	0.022	-0.02	0.986
November	0.0018	0.022	0.08	0.934	-0.0215	0.022	-0.98	0.328
December	0.0212	0.021	1.00	0.319	0.0169	0.021	0.80	0.427
Variance Terms[6]								
$\boldsymbol{\omega}$ (omega)	0.0002	7.2E-05	2.76	0.006***	0.0003	5.8E-05	4.88	0***
$\boldsymbol{\alpha}$ (alpha)	0.0398	0.002	17.64	0***	0.0288	0.002	13.44	0***
$\boldsymbol{\gamma}$ (beta)	0.9592	0.002	546.59	0***	0.9695	0.002	587.22	0***

* Indicates significance, * at the 10 % level, ** at the 5 % level, and *** at the 1 % level. See footnotes below.

1. Autoregressive (A.R.) Terms, φ, and Moving Average (M.A.) Terms, θ
2. Log of Sale Size Variable and Log of Sale Size Variable Multiplied by Destination Binary Variable
3. Destination Binary Variable
4. Non-Event Days Binary Variable
5. Month Binary Variables
6. Variance Terms: Constant Term (omega), Lagged Squared Error Term (alpha), and Lagged Conditional Variance Term (beta)

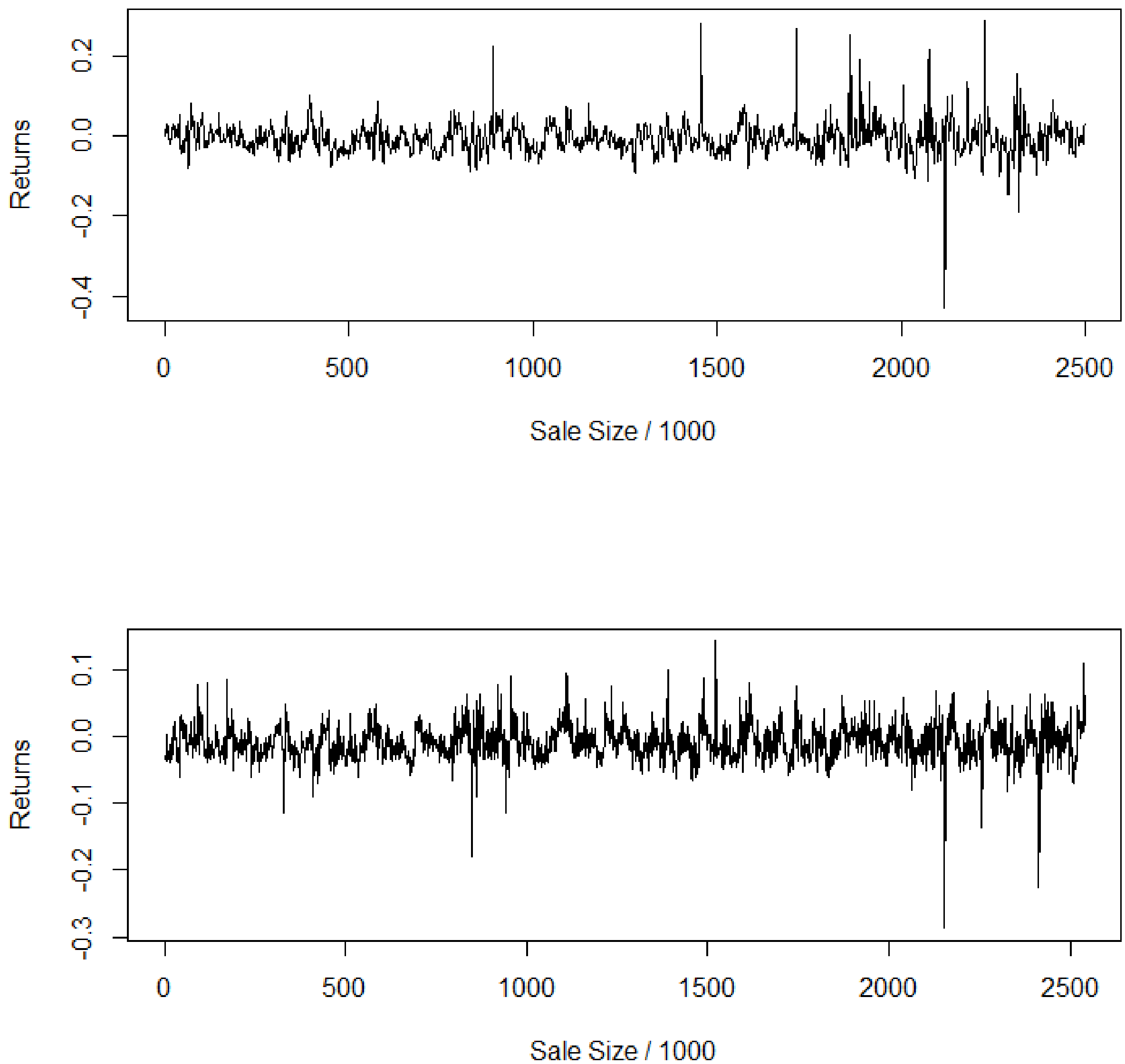


Figure 4.10: Fitted GARCH for Corn Futures Price Returns (panel a) and Soybean Futures Price Returns (panel b) with Clustered Data Removed (x-axis: Sale Size Divided by 1000 in Metric Tons)

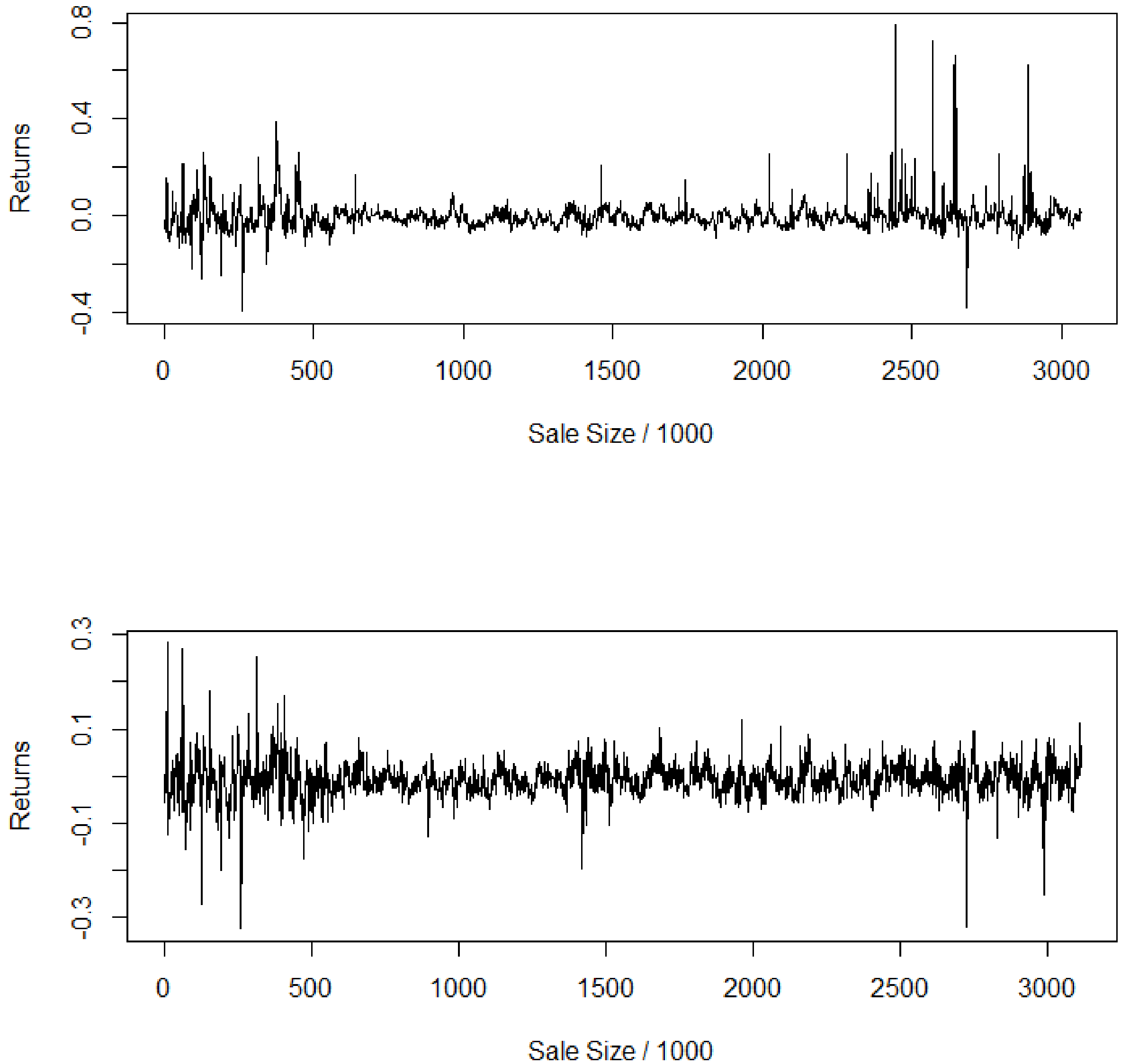


Figure 4.11: Fitted GARCH for Corn Futures Price Returns (panel a) and Soybean Futures Price Returns (panel b) with Full Data (x-axis: Sale Size Divided by 1000 in Metric Tons)

4.4 Event Spline Model

Event spline regression models, known as "B-spline" models, simulate nonlinear interactions between a response variable and one or more predictor variables using piecewise polynomial basis functions. A piecewise polynomial function is a mathematical function created by joining together several polynomial functions defined over various intervals. Different degrees and coefficients can be found in the polynomials. Knots are the locations where the polynomials intersect. (Oh et al., 2007). In a B-spline model, the location of the knots that generate the best fit for the examined data is called the "optimal" knots. Typically, the residual sum of squares or the Akaike information criterion (AIC) determines where to place the knots to minimize some measure of the error or mismatch between the model and the data. Based on the data and the level of flexibility the model wants, the "optimal" knot positions and number of knots are selected. The quantity of knots necessary depends on the degree of the polynomial functions. (Perperoglou et al., 2019)

For the spline models, the corn and soybean report and returns data were separated into two sets. The first data sets account for all reports, 375 reports for corn and 535 reports for soybeans. The following data sets are time series, including all reports and returns from 1/3/2011 to 12/30/2022. Then Multiple Linear Regressions (MLR) are added for each data set to compare. For the un-clustered data, problems occurred with auto-correlation, so the results were excluded.

Let R_0 denote the log of futures price returns at the first minute, and this regression model can be expressed as:

$$R_i = \alpha_1 + \beta_1(Sale\ Size_i) + \beta_2(Sale\ Size_i - Sale\ Size^*)D_i + \beta_3(Sale\ Size_i) * D_c + \beta_4[(Sale\ Size_i - Sale\ Size^*)D_i] * D_c + \beta_5 D_c + u_i \quad (13)$$

where

$Sale\ Size_i$ = volume of sales reports divided by 1000

$Sale\ Size^*$ = threshold value of sales also known as a Knot

D_c = Destination Dummy Variable for Combo (China and 1 or more countries)

$$D_i = \begin{cases} 1 \text{ if } Sale\ Size \text{ is greater than the Knot} \\ 0 \text{ if } Sale\ Size \text{ is less than the Knot} \end{cases}$$

Assuming $E(u_i) = 0$, we see at once that

$$E(R_i|D_i = 0, Sale\ Size_i, Sale\ Size^*) = \alpha_1 + \beta_1(Sale\ Size_i)$$

Which gives the mean sales commission up to target level $Sale\ Size^*$ and

$$E(R_i|D_i = 1, Sale\ Size_i, Sale\ Size^*) = \alpha_1 - \beta_2(Sale\ Size^*) + (\beta_1 + \beta_2)(Sale\ Size_i)$$

(Gujarati, 2008)

Results

In the first models we measure the sale size of the reports on the corn and soybean returns (Table 4.7). To find the "optimal" knot (Knot) for the spline models, we define it as the sale size with the lowest Akaike Information Criterion (AIC) (See panel a in Figures 4.12 and 4.13). The Knot for corn found at the lowest AIC is 1,762,000 metric tons. For the first corn returns spline regression, we find that both "Sale Size" and the sales larger than the Knot are highly statistically significant at the 1% level. The coefficient for the "Sale Size" is 0.0001 with a p-value of 0.001, so for the reports with sale sizes smaller than the Knot, we find that there is a slight but positive reaction to returns. The coefficient for the term representing the sale size beyond the Knot is 0.002 with a p-value of 0.001. This shows that sales larger than the Knot have a much greater impact on returns, as shown in panel b of figure 4.12. Compared to the coefficient for sale size, which is also divided by a thousand, found in the MLR (Table 4.8), the spline better represents how the size of the sales in reports affects the returns. We find similar results from the soybean returns spline regression on sale size. The Knot for soybeans found at the lowest AIC is 706,500 metric tons. Both terms are highly statistically significant at the 1% level. The coefficient for the "Sale Size" is 0.0002 with a p-value of 0.002, so for the reports with sale sizes smaller than the Knot, we find a slight but positive reaction to returns. The coefficient for the term representing the sale size beyond the Knot is -0.0003 with a p-value of 0.005. This shows that sales larger than the Knot have a slightly inverse relationship with returns, as shown in panel b of figure 4.13. Compared to the coefficient for sale size found in the MLR (Table 4.8), the spline is a better representation of how the impact on the returns differs as the sales size increases. For the next model, we include the destination dummy variable that was consistently found to be the most statistically significant among the destinations.

For the following spline models, the combo destination variable (sales to China and one or more other countries) is added to test the significance between the two spline sale size variables (Table 4.9). For the corn spline regression, we find two statistically significant terms. The first statistically significant coefficient is for the term representing the sale size beyond the Knot. The coefficient for this is 0.002 and is statistically significant at the 5% level with a p-value of 0.013. With the "Sale Size" coefficient not being statistically significant in this model, we find that the larger sales have a much greater positive effect on the returns. The other significant coefficient is for the interaction term between "Sale Size" and the destination dummy. This interaction term coefficient is 0.0004 and is statistically significant at the 5% level with a p-value of 0.03. This shows that smaller sales reports have a greater effect if the sales are to China and one or more countries. We can see this effect in panel c of Figure 4.12, where the predicted returns split between the sales to the combo destination and sales to all other destinations and their effects as the sales increase beyond the Knot. Compared to the MLR results, we find that the spline regression results are a better representation of the corn data that shows how, as the size of the sales increases, the effect on returns increases, and the impact of the destination of the reports changes respective to the size. We find three statistically significant terms on soybean returns for the final spline regression. The first coefficient that is statistically significant is for the term representing the "Sale Size." The coefficient for this is 0.0002 and is statistically significant at the 10% level with a p-value of 0.06. With the sale size beyond the knot coefficient not being statistically significant in this model, we find that the smaller sales have a much greater effect on the returns. The larger sales coefficient is not statistically significant but is shown to have a negative coefficient. The next significant coefficient is the interaction term between "Sale Size" and the destination dummy. This interaction term coefficient is 0.0003 and is statistically significant at the 10% level with a p-value of 0.08. This shows that smaller sales reports have a greater effect if the sales are to China and one or more countries. The final coefficient is for the destination dummy variable. This coefficient is negative and statistically significant at the 10% level, meaning that the destination variable negatively impacts returns over all sales reports. We can see this effect in panel c of Figure 4.12, where the predicted returns split between the sales to the combo destination and sales to all other destinations and their differing effects as the sales increase beyond the Knot. Compared to the MLR results, we find that the spline regression results are a better representation of the soybean data that shows how, as the size of the sales increases, the effect on returns decreases. Also, the spline shows how the impact of the destination changes depending on the size before and beyond the Knot.

AIC (Akaike Information Criterion) is a statistical metric used in regression analysis for model selection. Based on the log-likelihood function and the number of parameters employed in the model, it assesses the relative quality of a model for a given set of data. A lower AIC value shows a better model fit. Our first models' AIC values are 63.47 for corn and -92.63 for soybeans. These indicate that the soybean futures price returns model fits the data more closely than the model for corn. Given the substantial difference (156.1) in AIC values between the two models, the model with the lower AIC value offers a far better fit to the data. Even though these models cannot be ideally compared because of the differing observations, the AICs show that the soybean models fit the data better when comparing corn to soybean.

Tables and Figures

Table 4.7: Event Spline Returns

	Corn				Soybeans			
R-Square	0.096				0.015			
	Coefficient	Std Error	T-Stat	P-value	Coefficient	Std Error	T-Stat	P-value
Variable								
(Intercept)	-0.016	0.018	-0.860	0.39	-0.0243	0.0181	-1.35	0.178
Size[1]								
$\boldsymbol{Sale\ Size}$	0.0001	0.0000	3.295	0.001***	0.0002	0.0001	3.13	0.002***
$(\boldsymbol{Sale\ Size} - \boldsymbol{Sale\ Size}^{*})\boldsymbol{D_i}$	0.002	0.0006	3.279	0.001***	-0.0003	0.0001	-2.80	0.005***

* indicate significance, * at the 10 % level, ** at the 5 % level, and *** at the 1 % level. See footnotes below.
1. Splines of Size Variable divided by 1000

Table 4.8: MLR

	Corn				Soybeans			
R-Square	0.072				0.0021			
	Coefficient	Std Error	T-Stat	P-value	Coefficient	Std Error	T-Stat	P-value
Variable								
(Intercept)	-0.0304	0.018	-1.72	0.087	0.0085	0.014	0.61	0.54
Size[1]								
$\boldsymbol{Sale\ Size}$	0.00021	3.8E-05	5.48	8E-08***	5.5E-05	3.7E-05	1.46	0.15

* indicate significance, * at the 10 % level, ** at the 5 % level, and *** at the 1 % level. See footnotes below.
1. Size Variable divided by 1000

Table 4.9: Event Spline Returns (Destination)

	Corn				Soybeans			
R-Square	0.118				0.0168			
	Coefficient	Std Error	T-Stat	P-value	Coefficient	Std Error	T-Stat	P-value
Variable								
(Intercept)	-0.0017	0.0183	-0.09	0.93	-0.0138	0.021	-0.67	0.50
Size[1]								
$\boldsymbol{Sale\ Size}$	0.0001	0.00005	1.41	0.16	0.0002	0.00009	1.85	0.06*
$(\boldsymbol{Sale\ Size} - \boldsymbol{Sale\ Size}^*)\boldsymbol{D_i}$	0.002	0.0008	2.51	0.013**	-0.0002	0.00015	-1.57	0.12
$\boldsymbol{Sale\ Size}$ * Combo	0.0004	0.0002	2.21	0.03**	0.0003	0.00018	1.75	0.08*
$(\boldsymbol{Sale\ Size} - \boldsymbol{Sale\ Size}^*)\boldsymbol{D_i}$ * Combo	-0.0016	0.0013	-1.22	0.22	-0.0004	0.00025	-1.51	0.13
Destination[2]								
Combo	-0.093	0.173	-0.54	0.59	-0.1450	0.072	-2.02	0.04**

* indicate significance, * at the 10 % level, ** at the 5 % level, and *** at the 1 % level. See footnotes below.
1. Spline Size Variables divided by 1000 and Spline Sale Size Variables divided by 1000 Multiplied by Destination Binary Variable
2. Destination Binary Variable

Table 4.10: MLR (Destination)

	Corn				Soybeans			
R-Square	0.108				-0.0015			
	Coefficient	Std Error	T-Stat	P-value	Coefficient	Std Error	T-Stat	P-value
Variable								
(Intercept)	-0.011	0.018	-0.62	0.539	0.0075	0.016	0.48	0.63
Size[1]								
$\boldsymbol{Sale\ Size}$	0.0001	4.4E-05	2.50	0.013**	5.4E-05	5.4E-05	0.99	0.32
$\boldsymbol{Sale\ Size}$ * Combo	0.0004	0.00013	2.74	0.006***	-9E-06	8.3E-05	-0.11	0.91
Destination[2]								
Combo	-0.105	0.158	-0.66	0.508	0.0107	0.041	0.26	0.79

* indicate significance, * at the 10 % level, ** at the 5 % level, and *** at the 1 % level. See footnotes below.
1. Spline Size Variables divided by 1000 and Spline Sale Size Variables divided by 1000 Multiplied by Destination Binary Variable
2. Destination Binary Variable

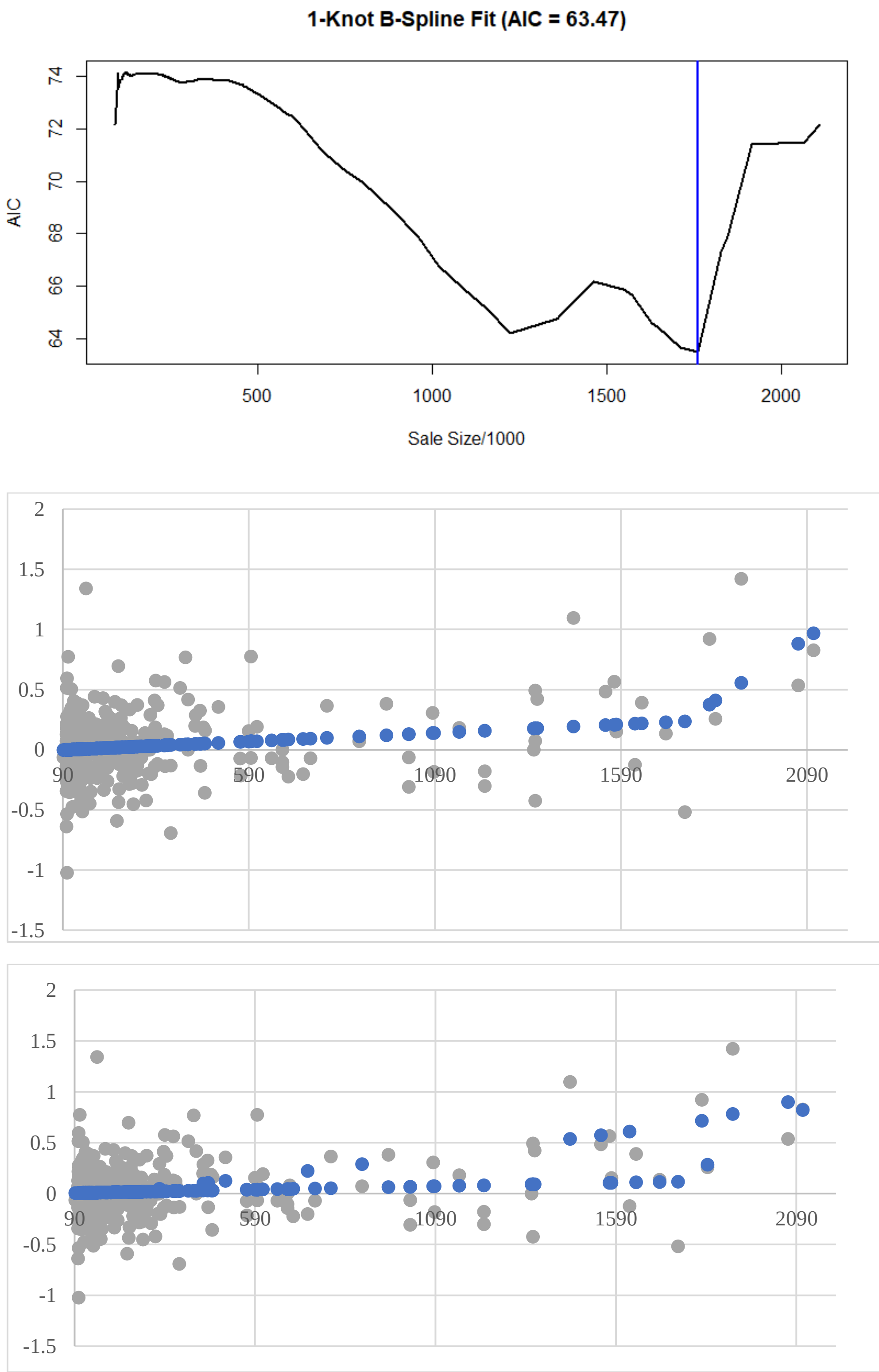
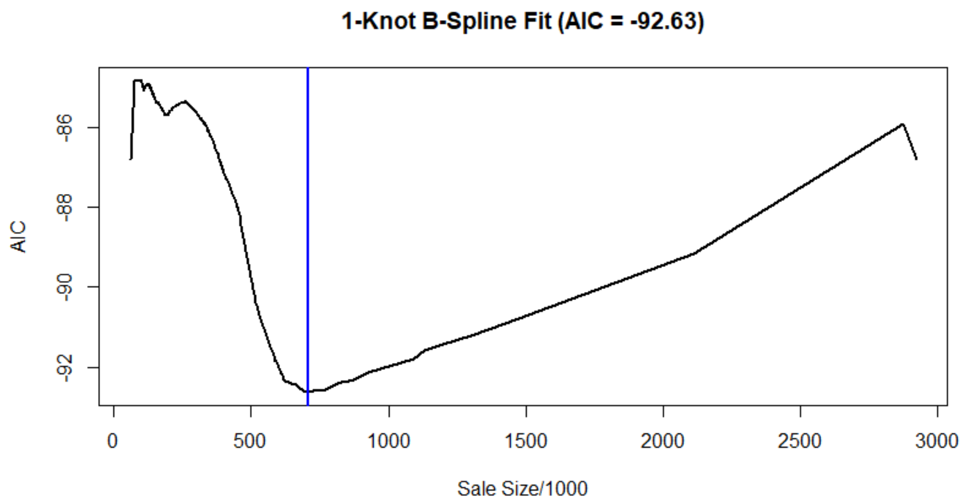
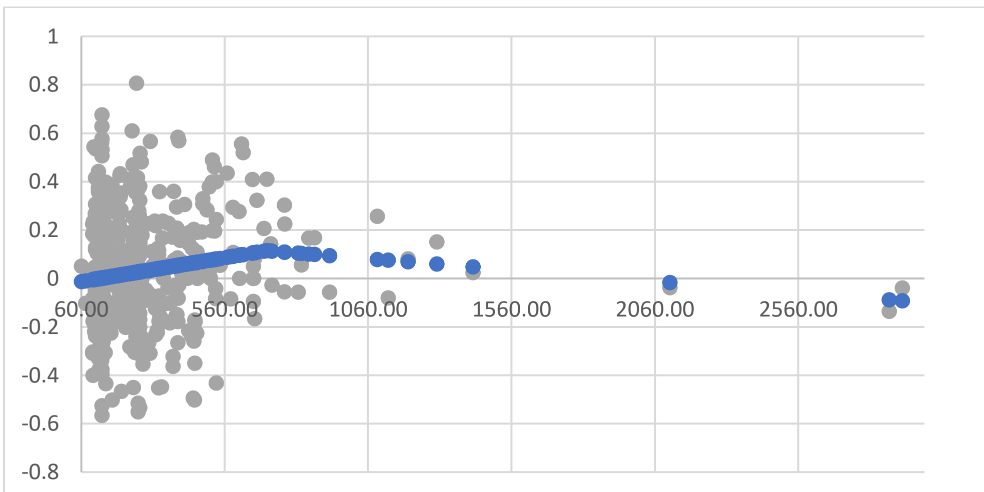


Figure 4.12: Corn Event Spline Optimal Knot (panel a), Actual and Predicted Spline Returns for Sale Size Regression (panel b), and Actual and Predicted Spline Returns for Regression with Destination (panel c) (x-axis: Sale Size in Metric Tons)





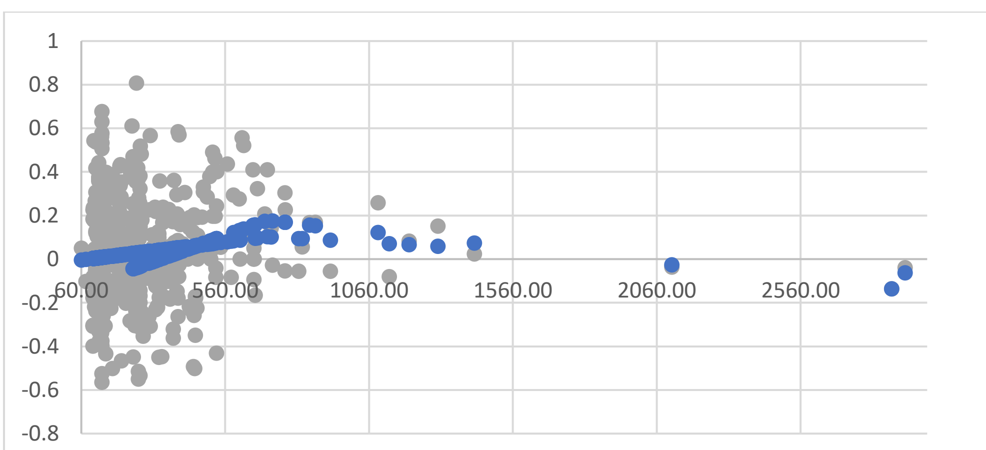


Figure 4.13: Soybean Event Spline Optimal Knot (panel a), Actual and Predicted Spline Returns for Sale Size Regression (panel b), and Actual and Predicted Spline Returns for Regression with Destination (panel c) (x-axis: Sale Size in Metric Tons)

5. Conclusion

5.1 Results and Implications

This research investigated if the USDA Foreign Agriculture Service's Large Export Reports have any substantial influence on specific commodity futures markets. The average return event study approach, the multiple regression with binary variables model, the autoregressive time series model, and the event spline model were utilized to examine whether the reports affect futures markets. When compared to returns from days before and after the release of the sales report, the average returns revealed that the FAS's Large Export Sales Reports have a subtle but significant impact on the prices of corn and soybean futures and no impact at all on the pricing of wheat futures. It became apparent that the volume of sales and the sales destination (China, Other, or several destinations including China) in reports influenced how much the report impacted the returns on the futures price. In the initial results of the multiple linear regressions, we discovered that the size of sales in the large export sales reports positively affected both corn and soybean returns. Still, only large corn export reports had statistically significant destination binary variables. Reports with a destination of a country(ies) other than China (Other) and multiple destinations, including China, (Combination) were found to be statistically significant, where they had an inverse and positive relationship, respectively. Next, we wanted to test all the returns throughout the reports.

Along with measuring the entire period of data, we wanted to know if the reports captured in the event window of another report change the effect of the reports, so additional time series datasets were created to test this. In an ARMA-GARCH model, the original time series data revealed the same results compared to the first multiple linear regressions. In corn reports, the size of the sale and, more specifically, the size of the sale with various destinations are statistically significant. The regression results with reports captured in the event window proved not to affect the impact of the variables on the returns. Finally, a linear spline model was used to understand how the different sizes of the sales in reports influence the returns. We found that larger sale sizes have a much more significant impact on returns for corn, and the complete opposite is true for soybeans.

Contrary to the primary purpose of this study, there is insufficient evidence to imply that China alone impacts the market to any appreciable degree, and any price changes happen as soon after release as the market can react to the large sale report, usually within the first two minutes of trading. Results from the various soybean returns research show that the size of the sale in reports has little to no impact on futures price returns, and particular destinations have no effect. In addition, the volume of reports and specific destinations were represented differently between the analyses resulting in a different scale of results. To better understand the effects of these reports and find whether there is knowledge of the sale in the market before the release, a move to make the reports real-time or at another time that does not fall in the window between trading times could be suggested.

5.2 Study Limitations and Further Research

The timing of the large export reports' release restricts this research's effectiveness and reliability. High volatility at the opening minute of day trading can misattribute the impact of the report's release in the market, which immediately constrains the results of this research to less accurate predictions. Using intraday periodic volatility curves, a method for observing the volatility of futures prices over a trading day could fix this issue. Measuring the volatility during the day and overnight of the two days from the sale's origin to the sale's report could present superior results that show how the sale and report could be captured in the market.

References

Adjemian, M. K., & Irwin, S. H. (2018). USDA announcement effects in real-time. *American Journal of Agricultural Economics*, *100*(4), 1151–1171.

Adonzio, W., & Royales, S. (2012). Impact of the drought on corn exports: Paying the price. BEYOND THE NUMBERS, 1(17).

Andersen, T. G., Su, T., Todorov, V., & Zhang, Z. (2023). Intraday periodic volatility curves. Journal of the American Statistical Association, 1–11. https://doi.org/10.1080/01621459.2023.2177546

Barchart.com. (n.d.). Barchart.Com. Retrieved November 10, 2022, from https://www.barchart.com/

Basic properties of splines and B-splines. (n.d.). https://www.uio.no/studier/emner/matnat/ifi/nedlagte-emner/INF-MAT5340/v05/undervisningsmateriale/kap2-new.pdf

Browse newsroom. (n.d.). USDA Foreign Agricultural Service. Retrieved October 23, 2022, from https://www.fas.usda.gov/newsroom/search?news%5b0%5d=news_type:10102

Burr, P. (n.d.). *USDA's Export Sales Reporting Program: Provides Markets with an Early Alert*.

Colling, P. L., Irwin, S. H., & Zulauf, C. R. (1996). Reaction of wheat, corn, and soybean futures prices to USDA "export inspections" reports. *Applied Economic Perspectives and Policy*,

Corn - 2023 data - 1912-2022 historical - 2024 forecast - Price - Quote - Chart. (n.d.). TRADING ECONOMICS. Retrieved December 16, 2022, from https://tradingeconomics.com/commodity/corn

Forecasting with Exogenous Regressors. (n.d.). Arch 5.5.0. Retrieved December 13, 2022, from https://bashtage.github.io/arch/univariate/univariate_forecasting_with_exogenous_variables.html

Gujarati. (2008). Basic Econometrics (4th ed.). McGraw-Hill Higher Education.

Hokana, F. E. (1976). USDA TO CONTINUE MONITORING ALL EXPORT SALES CONTRACTS. *FOREIGN AGRICULTURE*, *14*(4), 6–11.

Lark, T. J., Hendricks, N. P., Smith, A., Pates, N., Spawn-Lee, S. A., Bougie, M., Booth, E. G., Kucharik, C. J., &; Gibbs, H. K. (2022). Environmental outcomes of the U.S. Renewable Fuel Standard. Proceedings of the National Academy of Sciences, 119(9). https://doi.org/10.1073/pnas.2101084119

Leibold, K., & Hofstrand, D. (2022). Corn and Soybean Price Basis . *Iowa State University: Extension and Outreach*. https://www.extension.iastate.edu/agdm/crops/pdf/a2-40.pdf

MacKinlay, A. C. (1997). Event Studies in Economics and Finance. *Journal of Economic Literature*, *XXXV*, 13–39.

McGinnis, M. (n.d.). Research guides: Commodities: A resource guide: Commodities markets and instruments. Research Guides at Library of Congress. Retrieved November 10, 2022, from https://guides.loc.gov/commodities/markets-instruments

Oh, H.-S., Kim, J. G., &; Hon, W.-T. (2007, August 17). The piecewise polynomial partition of unity functions for the generalized finite element methods (II). UNC Charlotte. https://math.charlotte.edu/preprint-archive/piecewise-polynomial-partition-unity-functions-generalized-finite-element-methodsPatterson, P. M., & Brorsen, B. W. (1993). USDA export sales report: Is it news? *Review of Agricultural Economics*, *15*(2), 367.

Perperoglou, A., Sauerbrei, W., Abrahamowicz, M., & Schmid, M. (2019). A review of spline function procedures in R. BMC Medical Research Methodology, 19(1), 1–16. https://doi.org/10.1186/s12874-019-0666-3

Racine, J. S. (n.d.). A PRIMER ON REGRESSION SPLINES. Retrieved January 10, 2023, from https://cran.r-project.org/web/packages/crs/vignettes/spline_primer.pdf

Ri, X., Chen, Z., & Liang, Y. (2023). Detecting structural change point in ARMA models via neural network regression and LSCUSUM methods. *Entropy*, *25*(1), 133. https://doi.org/10.3390/e25010133

Ruiz, P. (2020, April 16). ML approaches for time series. *Towards Data Science*. https://towardsdatascience.com/ml-approaches-for-time series-4d44722e48fe

SAS Institute. (1996). Forecasting examples for business and economics using the SAS system (pp. 195–204). Sas Inst.

Smith, A. D. (2010, August 7). Global wheat crisis recalls Moscow's "great grain robbery." *The Guardian*. https://www.theguardian.com/world/2010/aug/08/russia-wheat-prices-bread

The New York Times. (1972, September 29). SOVIET GRAIN DEAL IS CALLED A COUP. *The New York Times*. https://www.nytimes.com/1972/09/29/archives/soviet-grain-deal-is-called-a-coup-capitalistic-skill-surprised.html

USDA. (2022). The Importance of the Export Sales Reporting Program. https://www.usda.gov/oce/reports/weekly-export-sales/importance

Valchanov, I. (2018, November 7). *Including a dummy variable into a regression*. 365 Data Science. https://365datascience.com/tutorials/statistics-tutorials/dummy-variable/

Xie, R., Isengildina-Massa, O., Dwyer, G. P., & Sharp, J. L. (2016). The impact of public and semi-public information on cotton futures market. *Applied Economics*, *48*(36), 3416–3431

Zivot, E. (2021). Introduction to computational finance and financial econometrics with R. https://bookdown.org/compfinezbook/introcompfinr/