CHARACTERIZING THE EFFECT OF USDA REPORT ANNOUNCEMENTS IN THE WINTER WHEAT FUTURES MARKET USING REALIZED VOLATILITY

by

Gabriel David Bunek

A thesis submitted in partial fulfillment
of the requirements for the degree

of

Master of Science

in

Applied Economics

MONTANA STATE UNIVERSITY
Bozeman, Montana

April, 2015

ACKNOWLEDGEMENTS

Foremost, I would like to thank my thesis committee members, Dr. Joseph Atwood, Dr. Anton Bekkerman and most notably my committee chair Dr. Joseph Janzen. Their incite, patience, and enthusiasm was indispensable. I would also like to thank the Montana State Department of Agricultural Economics and Economics faculty and staff as a whole for providing the resources to make this piece of research happen. Furthermore, I would like to thank and apologize to anyone who innocently asked: "What's your thesis about?" and then had to listen to me for the next 20 minutes. Finally, I must thank my wife, Hannah, for her endless support, encouragement, and assistance in editing.

TABLE OF CONTENTS

TABLE OF CONTENTS – CONTINUED

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

The United States Department of Agriculture provides information about fundamental supply and demand conditions for major agricultural commodities. I consider whether USDA's crop reports facilitate price consensus in the winter wheat futures market by testing the hypothesis that uncertainty, as measured by realized price volatility is reduced following the release of USDA reports. This hypothesis was originally developed in studies using implied volatility and found significant decreases. I instead calculate realized daily and intraday volatility using transaction level data from Kansas City Board of Trade futures contracts. Dates on which USDA reports are released are compared to the ten days around the report. Exploiting the full granularity of data, intraminute volatilities are computed to test whether there are distributional differences between report and non-report days. All results suggest that realized volatility does not decrease following USDA wheat report releases but instead increases. Regression analysis shows this result is robust to the inclusion of a limited but relevant set of controls.

INTRODUCTION

The price of a futures contract represents a point estimate of the underlying value of the commodity on which the contract is based. Information about supply and demand affects this estimate and there is a significant literature analyzing how public information influences the futures price. The United States Department of Agriculture's (USDA's) crop reports are a major source of public information to agricultural commodity markets. These reports are closely watched by futures market participants for benchmark estimates of market fundamentals. Most studies have found significant "announcement effects" from these USDA reports; that is, public information of underlying value changes the point estimate represented by the futures price.

The futures price gives no indication of the degree of uncertainty that the market places on the point estimate of the value of the commodity. Uncertainty indicates the degree of dispersion of market participants' price expectations. Greater price uncertainty implies more risk for individuals and firms whose profitability depends on the price of the commodity. The level of uncertainty or price dispersion matters because production and consumption decisions are based on more than just the futures price point estimate, they are also based on the level of risk in the market (Williams, 2001, p. 792). Producers and consumers of the underlying commodity may have heterogeneous preferences for risk (Moschini and Hennessy, 2001).

Public information can effect both the point estimate of a commodity's value by changing the futures price and level of uncertainty about the price. All market participants have access to public information and as a result it could help reduce uncertainty. Because public information provides the same information to the entire market, I hypothesize that it could help to equilibrate knowledge of fundamentals, resulting in a tighter distribution of price estimates. I propose using price variability as an indicator for price uncertainty. As uncertainty is reduced, as indicated by a narrowing of price dispersion, traders may gain a better grasp of the true value of the underlying commodity. The market achieves a consensus on the futures price.

USDA crop reports are valuable to the market for two primary reasons. First, they are a widely-followed public good to which traders have unrestricted access. Secondly, they provide comprehensive supply and demand information over a large range of goods (Vogel and Bange, 1999). For these reasons they should facilitate price consensus in the futures market by collaborating beliefs on commodity fundamentals. In a financial context, variance or changes in the distribution of prices are often called volatility. For this reason, volatility is a important measure of uncertainty and riskiness in the futures market.

A large amount of literature has investigated the announcement effects of USDA crop reports on futures prices, and recent findings conclude that these reports do significantly shift the point estimate upon release. Isengildina-Massa, Irwin, Good, and Gomez (2008a) find using a simple event window analysis that the variance of open to close corn and soybean futures returns is statistically greater on report days than on non-report days. Using regression analysis, Adjemian (2012) confirms these results for corn, cotton, and wheat with robustness to various seasonality and fixed effects. This literature moves to high frequency data analysis with Lehecka, Wang, and Garcia (2014), which uses tick corn data to compare days on which USDA reports are released to those which are not on the minute by minute basis. They find that the variability per minute is larger on report days for only the first 10 to 15 minutes.

While the recent work shows that there is a measurable announcement effect from USDA reports, these studies focus on only the influences of USDA publications on the point estimate futures price. These results confirm that there is a price effect but do not study the impact on uncertainty in the market. Few studies investigate into how these reports influence the dispersion of prices in the futures market based on the same hypothesis that USDA reports should tighten the distribution of prices. Since daily prices fail to provide the granularity of data necessary for such analysis, researchers turn to options to back out daily futures price volatility to evaluate certainty levels.

Two studies are similar to the analysis done in this study. McNew and Espinosa (1994) studies whether USDA production reports reduce implied volatility, inputted from an option pricing model. Using an event window analysis as well, they find that the standard deviation of corn and soybean futures prices significantly reduces the day after a report release. In a very similar project, Isengildina-Massa, Irwin, Good, and Gomez (2008b) show a statistically significant reduction in implied volatility on days that World Agricultural Supply and Demand Estimates (WASDE) are released by the USDA. They use a Black-Scholes option pricing model to get implied volatility estimates for corn and soybeans. Like McNew and Espinosa (1994) they find strong evidence that USDA announcements lead to a significant reduction in uncertainty in the agricultural futures market.

The primary question of this study is whether USDA reports facilitate price consensus and reduce price uncertainty in the wheat futures market. The wheat market is used for several reasons. First, as the third largest crop in the United States with half of its production exported, wheat is a major market in the national and global economy (USDA, 2014). Secondly, it is common knowledge that long run wheat prices are collinear with the other major grains, corn and soybeans. This meaning that broadly speaking these results will be generalizable to other major agricultural markets.

Instead of then using implied volatility this study benefits from using transaction-by-transaction intraday data to make real time measurements of the dispersion of prices, called realized volatility, which I posit could indicate the level of uncertainty in the futures market. I consider a group of USDA reports anchored around the WASDE to study if public information could help market participants reduce uncertainty as measured by the volatility of prices.

Realized volatility (RV) estimates are calculated from intraday returns of Hard Red Winter (HRW) wheat futures prices from the Kansas City Board of Trade. The volatility is "realized" because it is an instantaneous measure of the spread of prices at that time based on actual transactions, thus providing a different estimate of uncertainty than implied volatility which relies on interpolation of volatility from a model based on only daily observations. Over the period of 2008 to 2012, different intervals of time and transactions are used to compare the dynamics of price volatility on 75 days reports are announced to a 5 day window before and after the report's release. Based on Lehecka, Wang, and Garcia (2014)'s results, the effect of the reports on the distribution of prices is likely short lived, thus motivating the analysis over shorter time windows.

It is one of the first to calculate and evaluate realized volatility estimators in agricultural futures markets, especially HRW wheat. This project relies on Pagel, De Jongh, and Venter (2007) for several realized volatility estimators. Furthermore, in using realized volatility this study considers issues of microstucture noise, previously unaddressed in the literature, in order to separate the true price variability from the effects of the trading process (e.g. Bandi and Russell, 2008; McAleer and Medeiros, 2008; Liu, Patton, and Sheppard, 2012).

The analysis begins with several replications, first of Lehecka, Wang, and Garcia (2014) using tick wheat data and then McNew and Espinosa (1994) using realized volatility calculated from the wheat data. Results confirm similar to the former but RV daily estimates diverge greatly from the latter. To investigate the appearance that

RV is larger on report days than non-report days, intraminute tabular analysis is done on RV estimates for each minute of each day. This results show that at no point in the day or the days preceding or following a USDA report is the RV measurement of uncertainty reduced. Finally regression analysis is used to check if this result is robust to a specific set of theoretical controls for price limits, calender fixed effects, contract volume, and price move are included to test the robustness of results.

The replication exercises using the Kansas City Board of Trade data yield interesting results. When using Lehecka, Wang, and Garcia (2014)'s methods, wheat prices react immediately to USDA reports with the price effect lasting fewer than 15 minutes. On the other hand when computing daily volatility estimates similar to Implied Volatility, the results show that volatility instead increases on the day of and following USDA information is made public. Further tabular analysis into this fact reveals that using minute by minute calculation show that there is a short lived increase in volatility following the release of USDA reports. Volatility quickly returns to normal levels, never decreasing. Regression analysis confirms this result is robust to various controls.

The primary conclusions from these results is that USDA information does not reduce realized volatility in the wheat futures market. This is in direct contrast to similar analysis conducted using implied volatility. The first obvious implication is that there is a difference between what implied and realized volatility are measuring. Given the technical difference between the two metrics is seems likely that RV is not exactly a measure of uncertainty but instead something of a point estimate of the dispersion of expectations. Furthermore, the speed at which prices adjusts, less than 10 minutes, attests to the efficiency of the wheat futures market.

INSTITUTIONAL BACKGROUND

The USDA provides reports on current and expected supply and demand conditions for agricultural commodities. Specific reports relevant to the wheat market include the crop production, grain stocks, prospective planting, acreage reports, and the World Agricultural Supply and Demand Estimates (WASDE). The first four reports are all published by the National Agriculture Statistics Service, while the last is provided by the World Agricultural Outlook Board. This group of reports follows the model of Lehecka, Wang, and Garcia (2014), who deemed these reports relevant to the corn futures market. All these reports contain data on wheat, making them by extension relevant to wheat futures.

USDA History and Relevant Reports

The USDA first published the WASDE, originally the "Agricultural Supply and Demand Estimates," report in September 1973. The report came about because of the previous years embarrassment at the hands of USSR importers who were able to buy up a large portion of the grain crop without having a major impact on the price. This became known as the "Great Grain Robbery," and the USDA realized that separate institutions where aware of the Soviet's activities, but the lack of interagency collaboration prevented the transfer of this knowledge. Had this information been conveyed between USDA agencies and made public, importers and exporters could have adjusted levels and prices to better reflect the change in demand. Hence, to accomplish this, the Outlook and Situation Board began as a means of sharing information between USDA agencies to prevent future mistakes (Allen, 2007, p. 20).

The Agricultural Supply and Demand Estimates, originally only containing domestic numbers, became an important source of information to the market in addition to facilitating information sharing inside the USDA system. The original WAOB (World Agricultural Outlook Board) was not formed until 4 years after the reports began, and it was not until October 1980 that the report officially took on the name "World Agricultural Supply and Demand Estimates" when it began to include international information (Allen, 2007, p. 20). In 1985, the WAOB expanded the report to include estimates for individual countries. During that same year, the USDA began to prepare and release NASS (National Agricultural Statistical Service) crop production and grain stocks at the same time as the WASDE, thus strengthening the continuity between the several reports (Allen, 2007, p. 21).

The World Agricultural Supply and Demand Estimates is one of the most all-encompassing USDA reports. The report focuses on corn, soybeans, and wheat, but also includes other smaller scale grains, sugar, and dairy products. It is a spreadsheet of production, yield, and use for numerous crops grown in the United States and other major crop producing countries. As a result, it is closely watched by grain markets. There are several basic components to the "balance sheet" nature of each WASDE report. Imports, existing stocks, and production forecasts are used to estimate existing supply, while the demand side is measured by exports, domestic use, and year-end stock piles (Vogel and Bange, 1999, p. 9).

WASDE reports are produced by the collaboration of numerous USDA departments in what is called the Interagency Commodity Estimates Committee. The ICEC is composed of analysts from the Economic Research Service, the Foreign Agricultural Service, the Agricultural Marketing Service, and the Farm Service Agency. These analysts employ NASS data, foreign estimates, forecasting models, weather predictions, and even satellite imagery to calculate their numbers (Vogel and Bange, 1999, p. 4).

The members study the relevant information to arrive at estimates individually and then compare with another in order to achieve an unbiased consensus of figures.

The crop production report is the original benchmark indicator for grain supply and demand and contains two basic elements, acres to be harvested and expected yields per acre. These data for each new crop of each grain in the U.S. are collected from National Agricultural Statistics Service's surveys of farmers. The crop production and WASDE reports are released simultaneously between the 9th and 12th day of every month. Originally, the crop report provided the most-followed information since the 1960s, but beginning in the 1980s the WASDE replaced it due to the WASDE having the same information as the crop report and additional world wide estimates (Allen, 2007).

The grain stocks report includes exactly what the title belies: the amount of grain held on site at farms or off site, state by state, for the major grains, i.e., corn, wheat, and soybeans. Their numbers are gathered from quarterly surveys conducted by NASS in December, March, June, and September. NASS then publishes these figures in mid January and at the end of March, June and September.

The annual prospective planting report contains information on how many acres of the major crops are projected to be planted at the beginning of each season in March. The report is produced from an annual survey of farmers and is always published at the end of March, giving markets an initial look at what to expect for production as of the first quarter of the year. Similar to the prospective planting report, the acreage report is also an annual update of actual grains planted in the United States on a state-by-state basis in June. This acreage report gives concrete numbers as to what was actually planted compared to the March report's projections. It results from a large-scale survey conducted by NASS in June via phone and in person based on random sampling (Vogel and Bange, 1999, p. 4). The numbers are then released at the end of June each year.

Report Preparation and Release

The USDA reports under consideration in this study were first released at 3 pm Eastern Standard Time. The WASDE's was released at 3 pm on the second week of every month. Crop production and others were released on other days. The first major change to this came after users felt the WASDE and crop production reports gave conflicting information in early 1984 (Allen, 2007, p. 21). After this, crop production was switched to being released at the same time and compiled together to prevent discrepancies. In May of 1994, the USDA changed the release time of WASDE, crop production, and the other USDA reports to 8:30 am right before commodity markets open, rather than the afternoon, after they close. Likewise, the grain stocks, perspective planting, and acreage releases all changed to this time. This came to conform with other USDA reports and give American markets the advantage of being able to trade on the new information before international markets (Allen, 2007, p. 23).

The USDA commits to the equal access and integrity of its data and report publications. For this reason, each report is prepared in a lock up session where the document of interest is compiled from analyzing and comparing data sources. During these sessions, only a few people, mostly the analysts, are authorized to be in the windowless and arm-guarded room. This prevents information leakage or outside influence from disrupting the report's preparation. As has been alluded to, the USDA streamlined these sessions so that several of the reports are compiled in unison. When the release time changed, nothing was comprised in terms of the integrity for the reports preparation, only that it was done in a lock up session over night rather than during the day.

More recently, the entire group of reports release time again changed to 12 pm ET in January of 2013, essentially as a reaction to the fact that most commodity exchanges were already trading during the report release so it should be released while all such markets are open. During the time changes, no adjustments were made to the range of dates of the release.

The manner that WASDE and other USDA reports are accessed has implications for the reports' value in the market. How users access the reports influences how easily they can reach the information and, more importantly, the speed with which they can access them. Since the late 1990s the primary way of reaching USDA reports is via the USDA website. The amount of web traffic USDA servers can support at one time could prevent some users from reaching the information they desire. This matters because if the market is moving on information that a trader has not yet reached, he could incur significant losses. Initially, WASDE and other USDA reports were physically picked up at the WAOB or NASS office by members of the media who then relayed the information to interested individuals or firms. Reports were also mailed to subscribers. However, with the advent of the internet, the reports are now posted online at the time of publication and are emailed to subscribers.

Since the late 1990s, commodity market participants have had faster and less restrictive access to USDA reports, improving the dissemination of information. Today, the USDA website and its separate archive site hosted by Cornell University receive large amounts of traffic. For the year 2013, the USDA had almost 2 million downloads of their reports and around 150,000 hits per month. The Mann Library archive also received 58,000 hits per month. Currently, there are close to 15,000 subscribers to the WASDE report[1].

While it is clear that people do follow these USDA publications, it is important to understand how such a report could be useful to a futures market, like the Kansas City Board of Trade's HRW wheat. As described in the background and preparation of each report, they each contain projections or actual numbers about how much wheat is being grown, harvested, exported, and stockpiled. The WASDE and crop production reports of each month provide a snapshot of expected per acre yields, total domestic production, and ending stocks for wheat. None of this information explicitly tells the wheat futures market what the price of wheat should be, but it does provide clear information about the current and projected relative scarcity of wheat in the U.S. based on expected use, exports, and carry out or year end stock.

Because wheat futures are merely standardized forward contracts, current and future wheat supply and demand conditions are very important for the worth of that contract now and at its specified delivery date. Wheat contract holders or any others interested in wheat production data must then interpret the information contained in USDA reports as to how it reflects on the value of the wheat, the underlying commodity of their contract. While knowing if USDA reports change traders' perception of the fundamental value of their contracts is ideal, this is naturally unobservable. This analysis instead seeks to capture possible changes in traders' beliefs in wheat market fundamentals by measuring changes in price variability.

[1]Preliminary numbers provided by Office of the Chief Economist via personal correspondence

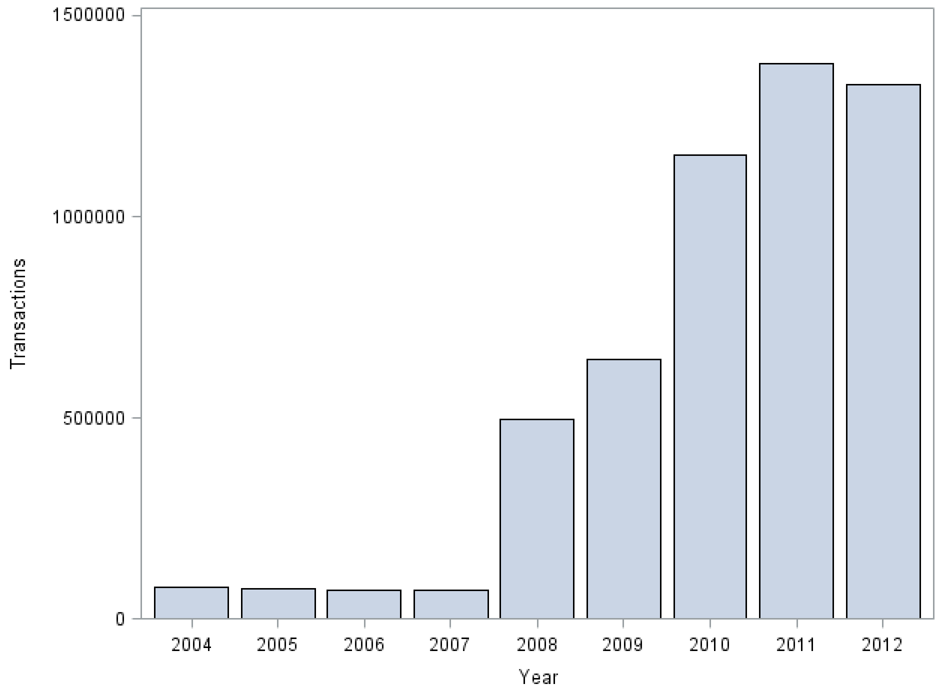
Hard Red Winter Wheat Contract Background

Hard red winter wheat is a staple food of the United States agricultural economy and with a protein content between 10 and 12 percent, it is the primary ingredient in bread flour. Domestic production of HRW wheat has been at or above three quarter billion bushels annually over the last five seasons and usually represents about 40 % of all wheat production in the United States (USDA, 2014). While the U.S.'s share of the global wheat market has eroded due to rising (albeit volatile) overseas production, it still represents roughly 20% of the world wheat export market. As the U.S.'s most-produced single type of wheat, HRW is a major part of the world's supply of wheat (USDA, 2014).

The Great Plains is some of the most productive crop land in the U.S., and the Kansas City Board of Trade, thanks to its geological location at the junction of two rivers in the Great Plans, rose to prominence as a center for wheat price formation. Futures for HRW wheat are the mainstay of the KCBT and are primarily used by commercial interests, like grain elevators, exporters, and millers to manage price risk from fluctuations in supply and demand. As a result, KBCT HRW wheat futures are an important tool for Unites States agriculture and are heavily traded by commercial and speculative interests with over 6 million contracts traded annually. (KCBT, 2015).

From 2004 to 2008 the Kansas City Board of Trade gradually introduced electronic trading as a method of exchanging futures contracts. This was the first major change to the KCBT in decades. In 2008, KCBT migrated its electronic transactions to a different electronic platform in addition to its traditional pit trading. While they originally used Liffe Connect from the e-Chicago Board of Trade, on January 14th, 2008 HRW wheat began trafficking on the CME Group's Globex trading system. This

Figure 1: Annual Number of Transactions for the Nearby Contract



also changed the hours of trading for the exchange. While the day session remained unchanged, the overnight trading session was expanded to run from 6 pm to 6 am.

When KCBT moved to the CME group's Globex platform, the number of trades in the winter wheat market dramatically increased. The annual number of transactions soared from less than 100,000 to well over 1 million in subsequent years. Figure 1 displays this shift in transaction volume. The increased liquidity is one of the primary motivations for the sample period choice. The change in transaction volume was a result of several things. First, CME Group's wide customer base and greater global recognition opened KCBT HRW futures up to a much broader market. Secondly, the method by which the Globex system automatically matched trades is to split up large orders amongst the oldest bids on the buying side. This by default increases the number of transactions because what used to be one large exchange now became several smaller ones.

As with all agricultural commodities, HRW wheat saw immense price volatility during the 2007 and 2008 period, which drew increased scrutiny from regulators and traders, leading to changes in trading rules. One of these changes was made to the price move limits in the early part of 2008. For years, the maximum amount the price could change from the previous day session's settle price was 30 cents up or down. Following more than 10 days that closed at the price move limit, KCBT announced on February 7th, 2008, the intention to raise the limit move to 40 cents. The next day, in a joint statement between Chicago Mercantile Exchange, the Minneapolis Grain Exchange, and KCBT, they agreed to raise price limits to 60 cents a day with an escalator effect. If any of the 5 nearest contract prices settled at the 60 cent day limit, the next day's limit would increase 50% or to 90 cents total. If again the market closes at the limit move of 90 cents it will increase 50% again to 135 cents. The limit move is capped at 135 cents. On each trading day following the increase, the price limit drops by the 50% it had increased by if the limit is not hit again, until the limit is again only 60 cents.

While the rule change was not approved officially by the Commodity Futures Trading Commission until March 24th, 2008, the KCBT board took immediate action in order to allow the exchange to provide its primary function of price discovery and hedging for the wheat market. Empowered by their by-laws, the changes took effect on February 12th, 2008, in order to deal with the fact that transaction volume had fallen to nearly zero for several days straight since the price limits prevented trading. KCBT wheat futures still trade via Globex; however, toward the end of 2012, the KCBT was purchased by the CME-Group for $126 million. By the middle of 2013, CME-Group moved all KCBT operations to Chicago, eventually selling the KCBT building.

HRW Wheat Futures Market Time Line

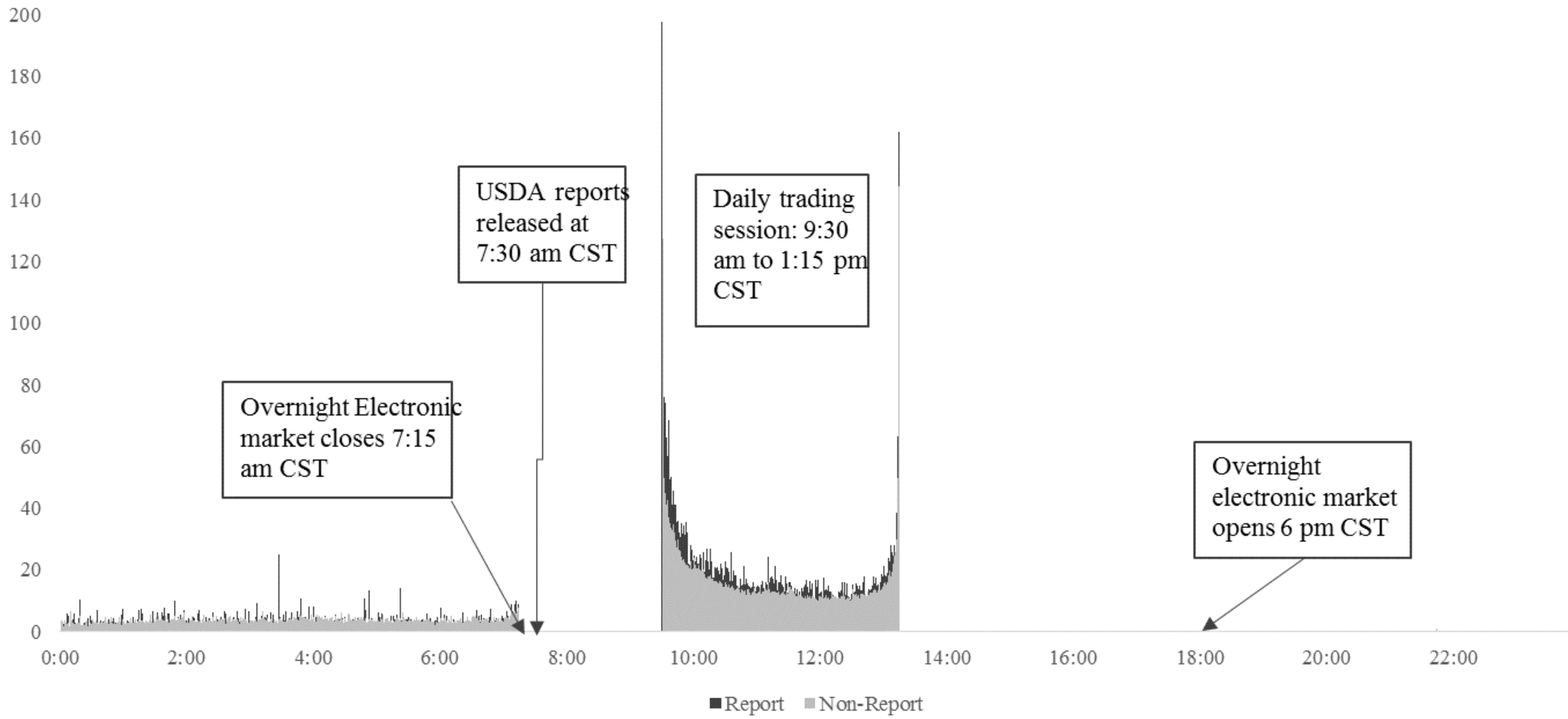
Figure 2 helps visualize and compare average trading report and non-report days at KCBT with reference points to important times, like when USDA reports are released. The daily trading sessions are Monday through Friday, while the night sessions are Sunday through Thursday. All times are in Central Standard Time. Figure 2 not only shows the daily KCBT market time line but compares the average transaction volume per minute for report and non-report days over the 5 year sample period. There is a small number of transactions per minute in the morning session and none during the entire night session from 6 pm to 12 am. This is not entirely surprising given the regional nature of the KCBT. The limited volume of the night session is why it is not considered in this analysis.

There is daily side-by-side trading, meaning both electronic and open outcry, from 9:30 am to 1:15 pm. An overnight electronic session runs from 6 pm to 6 am, until it was expanded on July 1st, 2009 by an hour and 15 minutes. From that point on the night session ran from 6 pm to 7:15 am. While the daily electronic market still closed at 1:15, a ten minute "clean up" period was also added in July of 2009 which takes place in the open pits to finish any transactions that trader had not finalized. Given the sporadic nature of this post market session, transactions in this period are not considered.

As the graph displays, there is on average more transactions on report days than on non-report days for every minute there is activity. In both cases, there is a significantly greater amount of transactions during the day trading session than the overnight session. Compared to the higher number of transactions, still many more transactions happen at the beginning and end of the 9:30 to 1:15 period for both types of days, giving them a distinctive “U” shape. For this reason, my analysis focuses solely on the day trading session of each date the market is open.

As indicated in figure 2, on days that WASDE and other USDA reports are released, it is at 7:30 CST. This gives market participants 2 hours to read and interpret these reports prior to trading on the information. On average much more trading activity takes place on report days than non-report days. This confirms a difference between report and non-report days as indicated by previous research. It also helps to motivate the analysis of this piece of research, as one would anticipate that the dispersion of prices is connected with the number of transactions taking place. The next chapter discusses existing ideas on how public information like the WASDE effects the market and how it applies to this analysis.

Figure 2: Daily Trading Time Line of Kansas City Board of Trade with Average Number of Transactions per minute on Report and Non-Report Days



LITERATURE REVIEW

Information in General

Information is a basic part of any market because profit and resource allocation motivate the collection of information. In his classic short essay, Friedrich Hayek suggests that the fundamental problem of economics is the allocation of information, or what he prefers to call knowledge. This is because, as Hayek explains "...the knowledge of the circumstances of which we must make use never exists in concentrated or integrated form, but solely as the dispersed bits of incomplete and frequently contradictory knowledge which all the separate individuals possess" (Hayek, 1945, p. 519). Hayek argues that knowledge is dispersed throughout the economy at all times, and markets are essential for individuals to collect information. By aggregating this information, markets provide prices that accurately reflect the relative scarcity of goods and allow individuals to make informed decisions about the allocation of their scarce resources. Kenneth Arrow, in a chapter on inventive activity, begins thinking of information as a commodity. He claims this is the result of uncertainty in the problem of resource allocation. Arrow explains that the potential for greater profits from more accurate information will give information economic value (Arrow, 1962, p. 614).

Public and Private Information

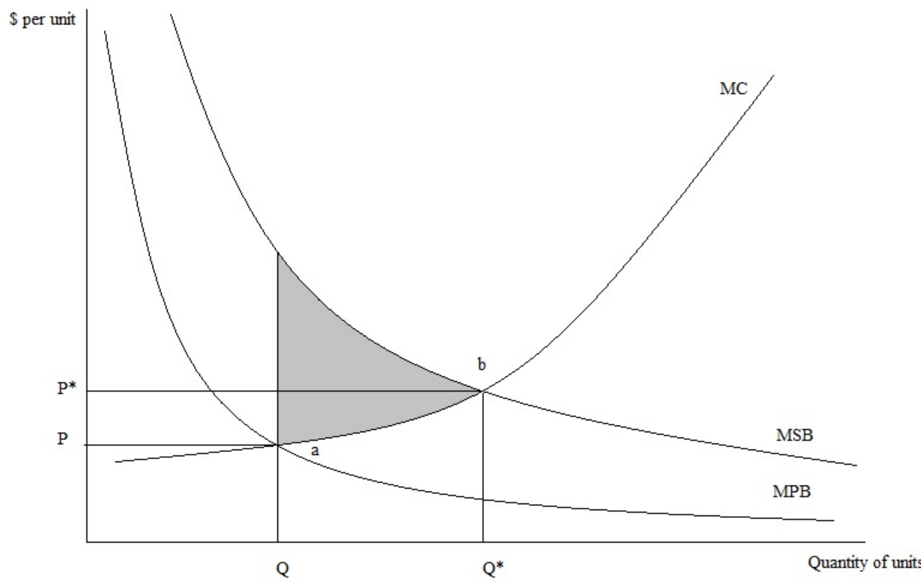
Information is typically categorized as public or private. A useful definition of public information is information that simultaneously becomes known to all individuals and firms and effects the market. It becomes available as soon as it is produced and no individual prior to its release knows it in its entirety by himself. Public information can be stock split announcements, firm mergers, or natural disasters, as well as government publications. On the other hand, private information can be thought of as only available to an individual or a group of individuals and only effects the market through trading. Private information is usually the result of individuals or groups of individuals' analysis of a market or situation which they do not share freely with others (French and Roll, 1986, p. 9).

Profit motivates individuals to collect and use private information. For example, if an individual acquires new information about a blight spreading through the growing wheat crop he expects wheat supply to reduce and as a result the price of wheat to increase. He can then profit by speculating in the futures or physical commodity market.

In the futures market he would buy wheat futures at the current price which does not reflect the coming reduction in supply. He would hold these futures until the prices rises and sell them for a profit. Similarly, in the spot market he can purchase the physical wheat at the current price and store the wheat until it rises. He can then sell the wheat for a profit at the now higher prices. In either case this speculator conveys his private knowledge about the reduction in supply to the market as a whole by his buying behavior. Hence, private information is revealed to the market through the actions of private information holders.

The same holds for if the blight is made public information. The market reacts to the new knowledge and prices move, likely quickly, to reflect the supply and demand shifts. For a given level of information, whether it is public or private, the efficient or equilibrium price is eventually reached through trading allowing for optimal resource allocation. This means information results in private gains, realized by profit and

Figure 3: Market for Information Collection



private allocation, but also provides social gains by conveying details about relative scarcity of goods to the economy, helping to optimize the allocation of resources.

Information has the characteristics that lend itself to public good analysis as depicted in Figure 3. The public benefit refers to the gain in utility for society while the private benefit is the gain in utility by the individual from private acquisition of knowledge. Assume the benefit, either private or social, gained from more information to be decreasing, conforming with diminishing marginal utility. The Marginal Private and Social Benefit curves are thus shown downward slopping. As the literature indicates, there is a cost to information gathering which can be incurred by society as a whole or the individual, but in either case one assume that it is the same per unit and rising. This is shown as the positive sloped Marginal Cost (MC) curve.

Participants in the information market will collect information until their Marginal Private Benefit is equal to their Marginal Cost of the next piece of information, point a shown on the graph. The issue remains that the socially efficient quantity of information, Q^*, may not be reached by private acquisition instead leaving the market with just Q. It could be that the social gains of each unit of information are so much greater than the private benefits that private collection is short of the optimal amount. As shown above, the Marginal Social Benefit lays above the Marginal Private Benefit curve. This leaves the shaded area as lost economic surplus from not acquiring the socially efficient amount of information.

This case motivates the idea of publicly funded information provision such as the supply and demand information provided by the United States Department of Agriculture. Since there are social gains from information, the USDA could seek to increase the overall level of information provision in order to increase the accuracy of prices and optimal resource allocation. Private sources of information may be insufficient to properly inform the entire sector of market fundamentals. As a result the USDA conducts surveys and publishes open access information necessary to convey such valuable news. Graphically, the USDA pushes the quantity of information out to the optimal point, where marginal social benefit equals marginal social cost, shown at point b. As a result, the shaded area now becomes consumer and producer surplus.

Market Efficiency

An important component of information's role in an economy is efficiency of markets. Fama (1970) defines a market as efficient if market prices always fully reflect the available information (p. 383). The notion of information efficiency is crucial to the functioning of any market because an inefficient one would fail to have accurate prices of the underlying commodities traded. This would mean participants would not have a true understanding of the fundamentals in the market leading to resource mis-allocation.

Additionally, Fama explains several criteria for evaluating the efficiency of a market. The *weak form* efficiency test shows a market which only reflects historical prices. Second, he introduces, the *semi-strong form* test where market prices contain information that is obviously public, like announcements by companies or USDA reports. Lastly, he discusses a *strong form* test, in which a market perfectly reflects private information. Fama's review of the literature and empirical work finds general support for the idea of market efficiency. He recognizes the unrealistic assumption of the strong form test, but finds some evidence for the idea, on top of much stronger support for the two less stringent tests (Fama, 1970).

Grossman and Stiglitz (1980) recognize that information acquisition is costly and explore the theoretical implications it brings. The major result is that perfectly efficient or strong form markets cannot exist. This same point is later confirmed by Fama (1991), who updates Market Efficiency to constitute acquiring information until the marginal benefit it equal to the marginal cost. This concept is already presented in Figure 3.

Grossman and Stiglitz argue that "because information is costly, prices cannot perfectly reflect the information which is available, since if it did, those who spent resources to obtain it would receive no compensation" (p. 405). If no profit, i.e. zero expected profit or net benefit, can ever result from new market information but information gathering costs are positive, then the primary motive to acquire information has eroded. While it might be implicit in Hayek's discussion, Grossman and Stiglitz make explicit that the scarcity of information makes it costly to acquire. In order for market participants to seek new information there must be a benefit from having that information.

Methods of Interpreting Public Information

Later theoretical work by the likes of Falk and Orazem (1985) apply much of the financial efficiency theory to agricultural futures markets. They construct a theoretical model to evaluate new information of USDA Crop Forecasts with numerous conclusions. They make an important point about the value of USDA reports evinced by market reaction. Only if the USDA crop forecast actually contains news or unanticipated information does the futures market react. Otherwise, market participants predict the information of the report and have already incorporated it into their understanding and price of the commodity. Hence, there is a *surprise* element to USDA reports, or valid news that makes them valuable to the market which can be measured by changes in the moments of prices.

Their model also suggests that reports during unstable market periods are more useful as well as reports that are released earlier in the season. Falk and Orazem confirm the theoretical story where the social benefits are larger than the private ones, as shown by Figure 3. They find an inverse relationship between information value an market size, meaning that social benefit from more information is greater in major markets which likely already have a large input of private information. Falk and Orazem (1985) provide an important basis for understanding the release of a specific type of public information, USDA reports.

Intending to address a similar question, Morris and Shin (2002) construct a theoretical model in order to measure the social value of public information much like Falk and Orazem (1985). They explore possible negative side effects of public information sources: "Public information has attributes that make it a double-edged instrument. On the one hand, it conveys information on the underlying fundamentals, but it also serves as a focal point of the beliefs of the group as a whole" (Morris and Shin, 2002, p. 1521). If information sources provide a collaborative effect then social welfare in increased, but if public information crowds out private, social utility is reduced.

Morris and Shin imply that the confidence that public information brings can cripple the decision making process of individuals with private information sources. That is public information can undermine the credibility of private sources suggesting that public and private information are substitutes rather than complements. Their model describes public information as being *too effective* and can lead to over reactions in the relevant market (Morris and Shin, 2002, p. 1532). Later studies of USDA information effects test for correlation in returns in order to detect possible under and over reaction to the new information.

The two preceding works represents attempts to grasp the effect of public information release on a market. More recently, difference of opinion models have been taken up as a manner to explain how the market reacts to and interprets public information. In his conclusion, Jack Hirshleifer makes a subtle but important point about the interpretation of information opening the door for further analysis. "With inhomogeneous beliefs, individuals with differing opinions will tend each to believe that revelation of new information will favor his own speculative commitments" (Hirshleifer, 1971, p. 573). Grossman and Stiglitz (1980) hint at similar issues with information interpretation (p. 404). While private information is subject only to opinions of the agent who holds it, public information is subject to interpretation by the market as a whole. Those interpretations can differ greatly despite the open source nature of the information.

On a topic that has received little attention since, Snehal Banerjee has authored or co-authored the majority of the work on what has become known as difference of opinion models (Banerjee, 2011; Banerjee and Kremer, 2010; Banerjee, Kaniel, and

Kremer, 2009). In particular, Banerjee and Kremer (2010) argue market participants will "agree to disagree" on the interpretation of public information (p. 1270). Studying trade volume, their work shows after a rigorous theoretical model development, difference of opinion can explain clustering and autocorrelation of volume which is consistent with other empirical work (p. 1294). Banerjee and Kremer suggest the difference of opinion framework is superior to asymmetric information models they consider at describing behavior in markets.

Building directly off of Banerjee's various work, the more recent essay of Fishe, Janzen, and Smith (2014) shows theoretically that equilibrium prices differ in a difference of opinion model compared to one with rational expectations. They explain the implications the difference of opinion idea has on prices, stating: "traders do not believe that prices fully incorporate the available information about fundamentals. Rather, they trust their own information signals when choosing positions" (Fishe, Janzen, and Smith, 2014, p. 543). While much of the analysis of information focuses on the differences between public and private, the scant difference of opinion literature shows a new light of interpretation of public information and its influence on agricultural market uncertainty.

Difference of opinion in general can show that conflicting interpretations of information can lead to a larger range of prices or higher volatility, a phenomena unexplained by the simple price reaction to new information. The majority of existing analysis of USDA reports in agricultural commodity markets relies on the market efficiency concepts of Falk and Orazem (1985), which are an application of Fama (1970), to describe the interpretation of information. This theory maintains the market only values information that is unanticipated or a surprise and can be measured by a price reaction. More recent work tests the overreaction hypothesis presented by Morris and Shin (2002); however, little evidence has arisen to support this theory.

Relevant Studies of WASDE and Similar USDA Reports

A long history of literature exists exploring the impact of USDA reports on agricultural markets. Two of the most recent and relevant research on USDA information effects in futures markets lay the foundation for our analysis in addition to providing numerous concepts and techniques. For these reasons a closer attention is payed to Adjemian (2012) and Lehecka, Wang, and Garcia (2014). Beginning with the former, he revisits the the topic of Isengildina-Massa et al. (2008a) using regression analysis in an effort to test the results subject to various seasonal and market factors.

The research goal is to measure the announcement effect of the USDA's World Agricultural Supply and Demand Estimates. Adjemian (2012) argues the WASDE is still important because it provides information to the market which enhances efficiency of resource allocation. Furthermore using along time series and better controls, he hopes to achieve an improved measurement of the WASDE effect. With corn, cotton, and wheat daily futures prices from 1980 to 2010, he forms an similar 11 day event window built around the day WASDE reports are released. Using absolute close to open returns he builds a regression model with indicators of each day of the event window. He also includes various controls for low inventory periods, months and interactions with low and month themselves

Ultimately, Adjemian's results confirm Isengildina-Massa et al. (2008a) but with more detail since using regression analysis he is able to separate out each day of the event window. The additional controls show that the WASDE impact is robust to the highly seasonal and conditional nature of the futures market. The results show a rapid incorporation of WASDE information into corn, wheat, and cotton futures markets with minor spill over effects, implying an efficient market. By using a greater number of controls and a broader set of futures, Adjemian (2012) is able to provide convincing evidence of the WASDE announcement effect as well as new information concerning price discovery in agricultural futures markets.

Adjemian (2012) provides useful concepts to build this current analysis on. It highlights the perils as well as the benefits of using regression analysis with futures time series data. It also confirms the initial part of the hypothesis that the point estimate or futures price of a commodity reacts to new information contained in the WASDE. Lastly, the rigorous regression analysis provides a list of useful concepts like controlling for each day of the even window as well as controls and interaction terms to help capture conditional effects.

Contributing to the analysis of USDA announcement effects, Lehecka, Wang, and Garcia (2014) takes event analysis to the next level of granularity by using high frequency tick data to investigate the impact of a catalog of USDA scheduled publications which include the WASDE reports. They use CBOT nearby corn contracts to compare minute by minute report and non-report days. Lehecka, Wang, and Garcia (2014) seek to assess the quality of price discovery surrounding public information, expanding upon the objectives of Adjemian (2012).

Lehecka, Wang, and Garcia (2014) include quarterly grain stocks, prospective planting, and acreage reports when constructing their 11 day event window. Assuming non-normality of returns and weak form efficiency, they select the last price from every minute of each day and take returns. Their goal is then to construct a more robust measure of return variability unaffected by outliers by using deviations from the median price of each minute on report and non-report days. The result is Equation 1:

$$AAD_m = \frac{1}{D}\sum_{d=1}^{D} |r_{m,d} - median_m| \tag{1}$$

In Equation 1, AAD represents the average absolute deviation from the median for a particular minute in a day, $m = 1...1440$. $D = 1...473$ for the 43 unique report dates and the 430 non-report days before and after them. r is the return for a particular minute, m, for any day in the sample d. $Median$ is the median for that particular minute across all days (Lehecka, Wang, and Garcia, 2014, p. 512). This average absolute deviation from the median is what the authors refer to as their return volatility measure.

Focusing on the opening of the day trading session, they find significant changes in price volatility on report days using a F test for homogeneity of variances and a Kruskal-Wallis test. After a strong reaction in the first 2 minutes there is a sporadic significant difference in return variance for the next 15 minutes when using F test but the Kruskal-Wallis shows significance lasting only 8 minutes."This implies that the market may be slightly less than semi-strong-form-efficient because not all available information is immediately reflected in prices" (Lehecka, Wang, and Garcia, 2014, p. 521). Concerned with possible under or over reaction to USDA reports, Lehecka, Wang, and Garcia (2014) tests for autocorrelation in minute by minute returns but fail to find any evidence to support that concern, similar to results found in previous literature.

Lehecka, Wang, and Garcia (2014)'s results show an immediate increase in return variance following USDA announcements followed by a rapid return to normal volatility. At a glance this appears to answer the question at hand, do WASDE and other USDA reports facilitate price consensus and reduce volatility. But there are fundamental differences between this current study and what Lehecka, Wang, and Garcia (2014) show. Foremost, my proposed measure of volatility shows the spread of prices inside of a day, while Lehecka, Wang, and Garcia (2014)'s, despite using tick data, only measures volatility with prices across the same minute of different days.

Their type of volatility measure is not that different than traditional delayed measures of price variance. This means their measure shows, like previous research, a movement in the point estimate of prices, not the volatility itself in each day. Thus, Lehecka, Wang, and Garcia (2014) has no implications for market uncertainty, the distribution of prices, and by extension, price consensus following information announcements.

Their's is one of the first works with high frequency data on agricultural futures, making their approach and methods a useful guide for future work. My research borrows several concepts from this study. For starters, I begin my analysis with a replication of Lehecka, Wang, and Garcia (2014)'s main result to provide base line analysis. One can assume their results maybe applicable to other markets, hence I confirm the impact of USDA reports on the wheat market to be short lived. Lehecka, Wang, and Garcia (2014) also utilizes a more comprehensive group of USDA publication, prompting this research to adopt their strategy.

Uncertainty Analysis Using Implied Volatility

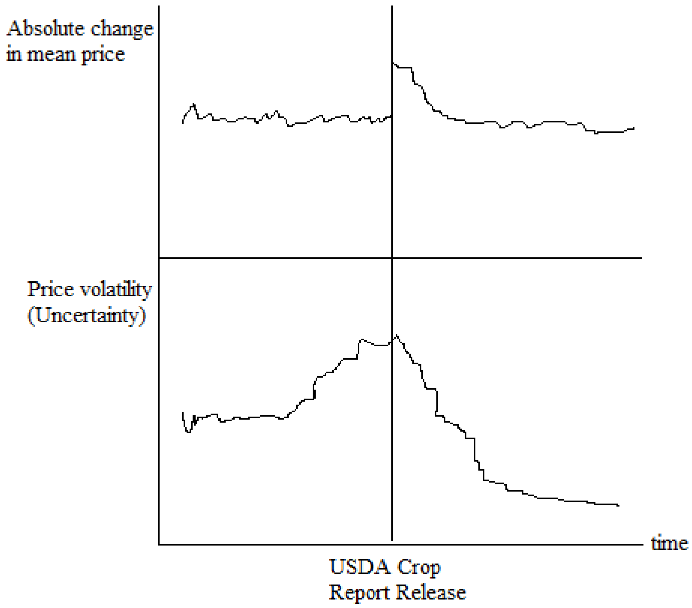
As laid out in the introduction, only a small number of papers address the idea of uncertainty surrounding USDA publications in agricultural futures markets. These works circumvent the lack of up to date volatility measures by combing futures and options data to calculate implied volatility. Two pieces of research, McNew and Espinosa (1994) and Isengildina-Massa et al. (2008b) comprise the most relevant contributions to the analysis in this topic.

The hypothesis for this thesis draws directly from McNew and Espinosa (1994) although the methodology departs significantly from theirs. They seek to bypass the question of whether new information from the USDA effect futures prices but rather ask how it effects market uncertainty about futures prices. They measure uncertainty with the standard deviation of prices backed out of Black's option pricing model. Assuming a log-normal distribution, they expect the absolute change in conditional mean to jump and as well as uncertainty to reduce after the release of unanticipated public information. Consistent with market efficiency, if the information is expected, then one should see no change in the moments of prices (McNew and Espinosa, 1994, p. 480).

They posit that if traders are uncertain whether the report will change their beliefs about the fundamental value of a good, one should see an increase in volatility up until the time of the information release. Following the publication, the news should help to resolve uncertainty about fundamentals resulting in a quick and measurable decrease in the range of prices. Figure 2 intuitively describes the path of volatility and mean price changes over time and is a reproduction of their graph about their theory. The change in volatility should be accompanied by a sudden jump in the mean change of futures prices, but a quick return to normal levels (McNew and Espinosa, 1994, p. 481).

Using an 8 day event window for corn and soybean futures and options prices from 1985 to 1991, they compare the daily conditional mean changes and volatilities the four days prior the report to the four days after. McNew and Espinosa do not find a significant increase in volatility leading up to the USDA release of crop production reports, dispelling the theory that the change in conditional mean increases. But they do find a statistically significant decrease in implied volatility on the day immediately after publication, with some residual decline in later days. Hence, they find support for their hypothesis that USDA crop reports reduce uncertainty about market fundamentals. McNew and Espinosa (1994) argues this is a clear sign of the economic value of USDA crop reports.

Figure 4: Announcement Effect of Unanticipated Information in USDA Reports (Source: McNew and Espinosa (1994))



It is not surprising that McNew and Espinosa (1994) finds no results indicating an increase in the absolute mean change of futures prices since work like Fortenberry and Sumner (1993) confirm a reduced impact of USDA reports during the same time period. One expects the absolute mean change in prices to increase as a reaction to newsworthy or unanticipated information in the USDA crop reports, later work confirms this.

The more recent of the two pieces, Isengildina-Massa et al. (2008b) tests an identical hypothesis using an updated corn and soybean futures and options sample running from 1985 until 2002. They compare report and non-report days overall finding that USDA crop reports reduce implied volatility 70% of the time for both goods. They then break the comparison down by each month and find the results are driven by large and statistically significant changes in January, August, September, October and November (Isengildina-Massa et al., 2008b, p. 482).

Isengildina-Massa et al. (2008b) also separates of the impact of WASDE reports that contain only international information and those that contain domestic figures contributed by NASS. The reports containing NASS data have a larger impact across the board. They also consider three sub-periods, 1985-1989, 1990-1995, and 1996 - 2002 finding that the percentages of report releases that reduce uncertainty is smaller in the earlier periods. Furthermore, only the most recent period sees a statistically significant difference when considering all the WASDE reports, not just those with domestic figures. They ignore testing the rise in uncertainty to identify a price reaction, instead referring to their companions paper Isengildina-Massa et al. (2008a) which shows price reaction in the same period and markets to the WASDE.

The results from sub-samples confirm the suspicion about McNew and Espinosa (1994)'s lack of results concerning the mean price change. Additionally, Isengildina-Massa et al. (2008b) argue that implied volatility's strength of results throughout the entire sample suggests it is a more powerful instrument for measuring market reaction to public information. While the hypothesis is borrowed from from their collective research, this work distinguishes itself by employing a higher frequency of data to calculate a different measure of volatility, which allows for the evaluation of changes in uncertainty inside a day. The nuances of these difference are discussed in the next chapter.

HIGH FREQUENCY DATA: THEORY AND METHODOLOGY

In this context, high frequency data refers to intraday data on all transactions or "ticks" in a market. It is most commonly associated with financial equities but is also related to futures and options and any openly traded commodity where ticks are frequent. With the advent of electronic trading in organized exchanges in the late 1980s and early 1990s, the ability to track and record the trading itself has greatly increased. As an immediate consequence, economic analysis of this type of data has turned from non-existent to a massive field of study. High frequency data are needed for the ensuing analysis primarily because they allow for analysis of time-path of prices inside of a day and ultimately provides faster and more accurate measures of prices.

Theory

Benefits of High Frequency Data

Generally speaking, volatility refers to the variability of prices, typically measured using standard deviation or variance. It is of interest in all markets because it indicates the level of risk involved in that investment. Furthermore, price volatility indicates uncertainty as to the fundamental value of the asset traded. The more volatile a market is, the more risky and the less participants actually know about what their commodity is worth. But what do high frequency data gain the practitioner? In brief, the higher granularity of data provides analysts with a faster and more accurate measure of price volatility.

A better estimate of uncertainty is valuable to any market because it provides a better indication of current risk. It is of particular importance in the futures market given the mark-to-market nature of the exchange. Mark-to-market means that the value of the contract is updated daily to the new level determined by the market rather than the book value when the contract was sold or purchased. This means that as the price of the held contract changes in the market the amount that has to be put down on that contract changes. In any formal futures exchange participants must deposit a percentage of the value of the contracts they are trading into an account, called a margin account. This is a sign of ”good faith” meaning the trader possess the monetary value of the contract. The more volatile the market is, the more active a trader has to be in managing their margin account, as it maybe depleted due to frequent price swings. Such price moves can be expensive as participants must constantly refill their margin account (Carter, 2012).

Volatility is the metric of consideration in this study because it is a estimate of investment-loss risk and consensus among beliefs in the market. It can show how USDA reports, like the WASDE, which contain information about wheat supply and demand, impact the uncertainty of the wheat market. With more frequent measure of price dispersion, this study can look at the hypotheses suggested by McNew and Espinosa (1994) and Isengildina-Massa et al. (2008b) in a manner they were unable to. By investigating how USDA reports effect uncertainty over short intraday time horizons, this study can provide concrete real-time evidence to whether information about fundamentals tightens the distribution of prices thereby creating futures price consensus.

In the context of finance and economics, volatility is an absolute measure of the first difference of prices (returns) for a given period. They are called “returns” because the differences in the prices over the given period is the return one earns from holding this instrument. Traditionally, the given time period is restricted by the frequency of the data so volatilities are calculated over months or even years. Alternatively, volatility measures were backed out of less frequent price measurements.

The major drawback to using historical volatility from daily data is the lag it requires to calculate price volatility. Enough days of data must be observed before volatility can be retrospectively estimated. With high frequency data, there are more observations and volatility can be calculated over shorter time windows. Thus, volatilities are computed for separate days or even particular parts of the day.

The second major deficiency of traditional volatility estimates is poor forecasting abilities. Tick-by-tick data provides the practitioner with much more accurate path of prices though time allowing them to make superior predictions of future prices (Andersen and Bollerslev, 1998; Bandi, Russell, and Zhu, 2008). This is because daily price changes tell an analyst less about risk and uncertainty in the market than all the transactions inside that one day do. The added detail of the data assists investors in making market predictions superior to historical methods. This current analysis of WASDE and other USDA report releases benefits from the increased detail of high frequency data in order to observe possible changes in volatility.

High Frequency Data Literature

Early work focuses on the numerous benefits of improved volatility measurement high frequency data afford for research in both finance and economics. Goodhart and O'Hara (1997) offers a valuable overview of the development of high frequency data sets and their initial applications. They asses the value of high frequency data succinctly: "Our ability to analyze the working of (financial) markets is limited by the availability of relevant data" (Goodhart and O'Hara, 1997, p.75). Their paper and even earlier work like Easley and O'Hara (1992) highlight the primary contributions of high frequency data, which is to the growing subfield of market microstructure. The detail of the data allows them to study how information is impounded via the market into prices. Researchers can consider a wider variety of questions, like how prices are formed and how new information influences prices. These two topics are the foundation of a market microstructure and have direct applications to understanding how USDA reports affect the futures market.

"Market microstructure studies the process by which investors' latent demands are ultimately translated into prices and volumes" (Madhavan, 2000, p. 205). High frequency data have greatly advanced the study of market microstructure by providing raw data to study the behavior and structure of markets. One of the core interests of this literature is how new information effects prices, which is called price discovery or formation. Definitionally, the study of USDA information impacting futures prices is such a topic. The growth of microstructure literature has developed numerous tools both theoretical and methodological, diffusing into the study of agricultural commodities markets. This piece of research seeks to exploit more of this finance literature to provide a more accurate picture of how USDA reports, especially the WASDE, influence the relevant futures market upon release.

Initial research like Easley and O'Hara (1992) argue with their model that time plays a key role, previously unavailable in lower frequency data, in conveying information the market. Meaning the time behind trades tells traders as much about price formation and market efficiency as the trades themselves (p. 578). As pointed out by Goodhart and O'Hara (1997), the increased granularity of observations allows for closer observation of volume, changes in prices, and bid and ask quotes (p. 86). This research suggests timing of trades provides qualitative information about speed price discovery and the alacrity of price consensus.

Another contribution from high frequency data is the ability to study price dynamics inside a day. With a large number of observations available, researchers can look at the flow of prices over short intervals of time to study price discovery and market efficiency. Ederington and Lee (1995) is one of the first studies to study price dynamics and find an immediate price adjustment following public information releases. A similar study, Malinowska (2010), extends the previous work's hypotheses to an even finer level with results that show changes in trading behavior as quickly as 10 seconds after information release. Both works confirm the point estimate effect built into the hypothesis at hand. This motivates the consideration of possible information impacts on higher moments of prices at the shortest possible time scale.

Numerous methods of dealing with high frequency price series have emerged; one of the more popular has been Autoregressive Conditional Heteroskedasticity (ARCH) models or some derivation thereof. Engle (2000) develops one of the early adaptations of this model to high frequency data via Maximum Likelihood Estimation in Ultra-High Frequency GARCH model (UHF-GARCH) to see how the timing of trades influences volatility. While this model and others like it are adapted to use a large number of observations, they don't provide the type of estimate amenable to evaluating the certainty level of a market.

Andersen, Bollerslev, Diebold, and Ebens (2001) and Andersen, Bollerslev, Diebold, and Labys (2003) are some of the most advanced research in this field to date and in many ways the benchmark analysis of high frequency data and for high frequency volatility measures. In the former, they introduce knew forms of measuring volatility while considering Dow Jones Industrial Average stocks. They investigate a model free approach to volatility estimation and examine the confounding properties of this approach like long-memory processes and negative correlation. The most important contribution of Andersen et al. (2001) is a method of measuring instantaneous volatility, which has become the standard in the literature. Andersen et al. (2003) on the other hand, is one of the first to build powerful new forecasting models stemming from high frequency data in order to make daily volatility estimates which can be later compared to historical volatility forecast models. This is the first step toward showing the improved predictive ability of high frequency data, that has implications for economic issues like risk management and portfolio allocation.

This literature highlights the numerous contributions high frequency data has made to the theoretical field of market microstructure primarily through enhanced measures of volatility, price dynamics, market behavior, time, and market efficiency. I have focused only on those contributions relevant to the topic at hand. Thus, the finer guaranty of data provides practitioners and researchers a more accurate estimation of price volatility in addition to a greater flexibility of economic questions to ask and test inside of each day.

Complications of High Frequency Data

Despite the many advantages gained by access of intraday data, there are also complications that come from microstructure effects. Market microstructure effects are institutional factors or artifacts of the trading process which are more prominent in high frequency time series prices. This means that as the granularity of data increases, one sees less of the fundamental price and more effects of the trading process. The fundamental price represents the genuine relative scarcity of goods, the equilibrium interaction of supply and demand that produces prices in a market. While never observable, the fundamental value is further obscured by the microstructure noise component. As a result, volatility measures using tick data contain the indistinguishable fundamental value and microstructure effects.

The high frequency data literature provides a simple structure for understanding the microstructure noise component of the price. Equation 2 shows the basic theory of prices with microstructure noise, a price for a particular time or interval t (e.g. Bandi and Russell, 2008; McAleer and Medeiros, 2008).

$$\tilde{p}_t = p_t^* \vartheta_t \tag{2}$$

In this equation $\tilde{p}_t$ is the price of a transaction at that particular time. But it is composed of the true or fundamental price p_t^* which contains the information important to economic decision makers and is obscured by the microstructure effects, ϑ_t. Neither the microstructure noise or true price are observed - only the transaction price $\tilde{p}_t$. A natural logarithmic transformation is typical of time series price data in order to reduce heteroskedasticity and the impact of outliers. Logging this equation makes the fundamental price and noise term additively separable as shown in Equation 3.

$$\underbrace{ln(\tilde{p}_t)}_{p_t} = \underbrace{ln(p_t^*)}_{m_t} + \underbrace{ln(\vartheta_t)}_{s_t} \tag{3}$$

Combining this equation with some of the basic concepts of market microstructure theory helps to explain the issue of the microstucture noise term on prices (e.g. Hasbrouck, 2007). Relying on the notation and work of Janzen, Smith, and Carter (2014) and the underlying theory of Hasbrouck (1993), I can describe the properties of the equilibrium or fundamental price as part of a semi-martingale random walk process using the underscore notation from Equation 3:

$$m_t = m_{t-1} + w_t \tag{4}$$

In Equation 4, the fundamental price at time t is m_t and is a product of the last time period's fundamental price, m_{t-1}, plus a random incremental shift, w_t. The term represents the change in fundamental value brought on by changes in underlying supply and demand factors. Such changes could be learned by new information, like a USDA report.

Taking the first difference of the observed price p_t to get returns produces Equation 5.

$$\Delta p_t = p_t - p_{t-1} \tag{5}$$

It is merely the current period t's price with the preceding period's price subtracted. Conveniently, the individual terms that compose p_t from Equation 3 can be substituted into the equation as such:

$$\Delta p_t = m_t + s_t - m_{t-1} - s_{t-1} \tag{6}$$

Equation 6 shows the observed return, Δp_t, is comprised of the current period t's true price and noise term with the same parts from the preceding period, $t-1$, subtracted. Using the parts from rearranging and plugging in Equation 4 to Equation 6, a more informative equation is formed. Change in fundamental price is then equal to w_t and the change in the microstructure effects can be collected into the delta notation.

$$\Delta p_t = w_t + \Delta s_t \tag{7}$$

Thus, Equation 7 shows that logged returns, the basic term used in this and most high frequency analysis, are always composed to two parts: the change in the equilibrium value, w_t, and a shift in the microstructure noise term. With this basic equation, the disruptive properties of microstructure noise on price dispersion can now be better illustrated. As shown in Equation 8, taking a measure of the spread of prices like variance produces two terms:

$$Var(\Delta p_t) = Var(w_t) + Var(\Delta s_t) = \sigma_w^2 + \sigma_s^2 \tag{8}$$

As the equation displays, the variance of prices is equal to the spread of both the fundamental value and the microstructure effects. As a result, any volatility measure of high frequency data will necessarily be influenced by microstructure noise. This relationship only holds if changes in the fundamental value and mircrostructure noise are independent. They are by assumption because w_t as described in Equation 4 is random and uncorrelated. Notice the presence of both terms makes any volatility measure larger than if it only contained the equilibrium price.

Microstructure noise can be caused by various institutional influences on the trading process known as irregular pricing, price discreteness, nonsynchronous pricing, and the ask-bid bounce. Zhou (1996) and Andersen and Bollerslev (1997) were some of the first to recognize the issue of microstructure effects with high frequency data while Andersen et al. (2001) brought it to wider attention.

Microstructure noise matters because it makes up more of the observed volatility the smaller the window of observation. This means $\frac{\sigma_s}{\sigma_w + \sigma_s} \to 0$, or the noise to signal ratio goes to zero as the length of time between prices increases. Hence, daily prices contain very little of this microstuture noise; however, as described earlier, observing prices over longer time frames results in lagged calculation of volatility and reduces the accuracy of estimation. But these effects pose a problem only when this analysis focuses on the impact of situation and outlook information on the fundamental value of wheat returns. While this is important, it is not the only concern.

As shown in Equation 8, the problem with microstructure effects, especially the ask-bid bounce, is that when returns are taken any volatility calculation then contains the noise term, σ_s, which obscures the underlying price variability. The statistical consequences of the bid-ask bounce is first-order negative autocorrelation of returns which produces an upward bias in volatility measures (Bai, Russell, and Tiao, 2001; Bandi and Russell, 2005; de Pooter, Martens, and van Dijk, 2008). Also, higher order autocorrelations can result from the return clustering that results from price nonsynchronicity and discreteness (Bandi and Russell, 2005, 2008). Overall this means that volatility estimates will be larger than if only the true equilibrium price was observed.

Price discreteness is the result of transactions lacking the true properties of a continuous variable, hence there is not an flow of infinite prices at every possible time and price throughout the day. Transactions happen in chunks at specific prices or times, resulting in price jumps from one price to another rather than a smooth transition. This is partially because there are limit moves in markets, i.e., the minimum amount a price can change, and because no one bids or asks for a millionth of a penny. Another dimension of this problem is that prices are not spaced continuously throughout the day, but instead occur at discrete times.[1]

Irregular pricing overlaps with discreteness but focuses on the issue of prices clustering with numerous transactions at the same price at the same time or right around one another; Andersen et al. (2001) calls this discrete clustering. In the opposite extreme there ends up gaps throughout the trading day or periods of time where no transactions take place. Bai, Russell, and Tiao (2001), Barndorff-Nielsen and

[1]There seems to be some inconsistency over price discreteness. Typically, it appears synonymous with irregular pricing, but Bandi and Russell (2008) calls it price discreteness, which is "transaction price changes occurring as multiples of ticks" (p.339). Nevertheless, the other works referenced, even Bandi and Russell (2005), imply price discreteness has more to do with the definition employed above, which will be used whenever discussing the topic.

Shephard (2004), and Bandi and Russell (2005) all refer to this as irregular pricing, which is problematic for observation and aggregation methods. de Pooter, Martens, and van Dijk (2008) explains nonsynchronous pricing, which is the result of "not every stock trad[*ing*] in each (intraday) interval or exactly at the end of each interval" (p. 204). This means prices are not observed at the same point in time even if the do appear frequently. This is a problem when considering sampling and aggregation methods, something to be addressed later.

The bid-ask bounce is the most prominent of all of the aforementioned issues and is generally addressed in all the literature. The bid-ask bounce is a phenomenon that appears in high frequency data and is the result of any market being comprised of buyers who want to buy low, and sellers who want to sell high. As a result, there is a gap between these two prices, known as the bid-ask spread. Any time a transaction takes place, either a buyer or seller concedes to a higher or lower price; as a result, transactions bounce in between the high and low of this spread all day. This gives the facade that there are changes in the real value of goods when in fact it is just a natural result of the trading process.

In order to measure and evaluate uncertainty and risk and by extension information's influence on volatility, one expects the behavior of the market, reflected in the microstructure noise, to be equally influenced by USDA reports. In this case, the inability to separate the noise from the fundamental value is less problematic because changes or lack thereof in the noise component is equally indicative of the information effects. Realistically, previous studies which employ implied volatility contain the same microstructure affects of the market and did not correct for it. While these affects are a concern, they do not undermine the nature of the analysis.

Volatility Methodology

I attempt to separate the methods those theories imply to be presented in an orderly and detailed fashion here. These methods consist of measuring the real-time variability of prices in high frequency data in order to provide high frequency measures of the market risk and price uncertainty. They also deal with extracting the fundamental or equilibrium price variability from the noise. A large body of work which deals with this issue is addressed, but as has been indicated, the identification of the true price is not the only objective in this analysis.

Realized Volatility

Volatility is central to the the consideration of information effects, but there are several methods of measuring this variability. The traditional metric of volatility is a standard deviation of returns. Equation 9 shows the basic formula for this standard deviation where r_i represents an observed return over a given interval of time or transactions, and is equivalent to Δp_t from Equation 7. The i will be determined by whatever sampling method (to be discussed later) is used, and $\bar{r}_t$ represents the average return for that same period t. n_t is the number of observations for that period t.

$$s = \sqrt{\sum_{i=1}^{n} \frac{(r_i - \bar{r}_t)^2}{n_t - 1}} \tag{9}$$

Prior to the availability of high frequency data, historic volatility measures, like Equation 9, are calculated *post facto* from daily prices. The Black-Scholes option pricing model as well as Engle's ARCH family of models are the benchmark formula for backing out latent daily volatility estimates. As was mentioned, these models rely on sufficient historic prices in order to calculate volatility in the retrospective time series. As a result, there has always been a delay in calculating the risk in the market. High frequency data allows volatility measurement to become instantaneous, since volatility can be calculated from the most up-to-date transactions. This reflects what is happening in the market right now rather that what happened weeks or months ago. Hence the volatility is "realized" and no longer latent.

Realized volatility represents a whole new set of estimators that has developed in recent years. Andersen, Bollerslev, Diebold, and Ebens (2001) is one the first to introduce and work with realized volatility. In finance, the standard measure of realized volatility is the sum of squared returns. While there is some diversity in how to employ this method, I rely on the basic estimator described by Pagel, De Jongh, and Venter (2007). In this case, the square root of the realize volatility estimator is taken, making it more similar to the traditional standard deviation.

$$RV_t = \sqrt{SSR} = \sqrt{\sum_{i=1}^{t} r_{i,t}^2} \tag{10}$$

In Equation 10, r is the return for each interval i of a chosen sampling method for a given time period t. The returns are squared and summed. The assumption is that the expected value of each return $\bar{r} = 0$. This gives a very similar measure to a variance, though it does not show deviations from the mean nor is it averaged per observation. Two issues exist with using the sum of squared returns for analysis. The first is that it is not comparable over different amounts of time or number of transactions because it is merely a summation of observations. Secondly, in practice it is identical to a standard deviation, which does not have the comparability issue. This concept will be revisited in the next chapter. For a more complete treatment of types of realized volatility estimators see Pagel, De Jongh, and Venter (2007) or Liu, Patton, and Sheppard (2012).

The age-old maxim of "more data are always better" does not hold for high frequency data because of microstructure effects. Using the raw tick-by-tick data produces upward bias due to the presence of this noise component as shown by Equation 8. In an effort to balance the loss of accuracy and microstruture effects, nearly all empirical work with high frequency data employs a sampling scheme. Because the primary interest of observing volatility in other research is to look at only the fundamental price variability of price, the sampling seeks to reduce the effects of market microstructure. As a result, certain observations are picked out of the raw data based on a rule. This sampling mitigates the disruptive effects of the microstructure noise, reducing the upward bias of volatility measurements. All future references to "sampling" refer to the process of drawing prices in a regular manner from the total data.

Sampling Methods

Over a short time, the microstructure and high frequency data literature has introduced numerous sampling methods. These methods are separated into two basic categories: calender time and transaction time. Alternatively they can be titled clock time and tick time.[2] The practical goal of either sampling method is to produce a time series with consistently spaced observations, making it able to produce consistent volatility estimates (e.g. Wasserfallen and Zimmermann, 1985; Hansen and Lunde, 2006; Russell and Engle, 2010). Calender time sampling is the more commonly used

[2] McAleer and Medeiros (2008) makes a distinction between transaction time and tick time sampling, the former being strictly transaction to transaction, the latter being price move to price move. Only they and Oomen (2005) make this distinction. The literature uses them interchangeably. For clarity, when using transaction time I am referring to the strictly transaction to transaction sampling method. Their suggestion of tick time sampling is ignored on the bases that it systematically leaves out all zero returns which I posit contain significant information about volatility in the market as well as indications of price consensus.

and means to sample the raw data at regular time intervals. Common choices are every minute, 5 minutes, or 15 minutes. Transaction time sampling refers to drawing observations from the raw data at fixed intervals of the number of transactions, such as every 5 transactions.

When considering calendar time sampling, irregular pricing and price discreteness make it rare that transactions fall at identical points in each minute. Interpolation methods must be employed to manufacture the desired equidistantly spaced time series. Two types of issues must be dealt with. The first is having more than one observation in the interval of interest, and the second is not having any observations for one or more intervals. In the work that does describe their method, linear interpolation and the previous tick methods are most common.

Linear interpolation, as described by Andersen and Bollerslev (1997), constructs artificial prices by weighting the nearest observations to the desired point and taking a linear interpolation between those observations to calculate the needed new observation (p. 151). All of Andersen and Bollerslev's subsequent collaborations appear to use this same method. This can also work to fabricate missing observations in other intervals. Other methods include simply taking the two nearest observations, interpolating a line between them and taking observations at the desired time scale or frequency from that line.

The previous tick method takes the first observed price of each interval as the sampled transaction for that interval. For sampling intervals missing observations, the previous observation, or "previous tick," is used to fill in. The initial version of this method was introduced by Wasserfallen and Zimmermann (1985),[3] and Hansen and Lunde (2006) modified it to its current form, comparing it to linear interpolation.

[3]However their model is a mixture of interpolation and previous ticks. They propose averaging quotes appearing in the same minute but filling in missing ones with the previous observation.

Their results suggest it to be a superior method to that of linear interpolation since in the probability limit, realized volatility of the linear interpolation goes to zero as sampling frequency increases. They explain this as a natural result of their being no variance in a straight line (Hansen and Lunde, 2006, p. 129).

The growing research using high frequency data appears to favor the previous tick method as described by Hansen and Lunde (2006). Their method is standard practice in many applications using high frequency data (e.g. Lehecka, Wang, and Garcia, 2014; Zhang, Mykland, and Ait-Sahalia, 2011; McAleer and Medeiros, 2008; Pagel, De Jongh, and Venter, 2007). The above treatment of interpolation methods and sampling frequency is only relevant to calendar time sampling. While it is the most commonly used, serious inquiry has been done comparing it to transaction time sampling. The latter has the benefit of containing no interpolated data. It also spares the researcher from dealing with choosing interpolation methods since sampling comes form the number of transactions rather than the amount of time.

Early work by Zhou (1996) argues that transaction time sampling reduced heteroskedasticity in the high frequency returns (p. 48). Oomen (2005) also finds that transaction time sampling is generally superior based on smaller mean square errors and improved bias correction from microstructure effects. He proposes a transaction-by-transaction sampling frequency to match the calendar time sampling methods frequency in order to make them comparable. Likewise, Liu, Patton, and Sheppard (2012) find that transaction to transaction sampling can improve forecasting accuracy for moderate sampling frequency. Overall, the most common transaction time sampling frequency is every 5th transaction. Russell and Engle (2010), a technical yet theoretical treatment of high frequency methods, states "By the very nature of market microstructure field, these theories often need to be examined at the transaction by transaction frequency" (Russell and Engle, 2010, p.28).

Optimal Sampling

In order to grapple with the complexities of sampling high frequency data, a nuanced literature has developed to find optimal number of observations and manner to sample them. The influence of microstructure distortions can be mitigated by reducing the frequency of the data. The researcher must balance the trade-off between bias caused by these microstructure effects and sampling error, reduced accuracy, or power of volatility estimation due to using less data. (Bandi and Russell, 2008, 2005; de Pooter, Martens, and van Dijk, 2008; Andersen et al., 2000b). In this small body of work, research typically develops a theoretical model of RV from which they can run an optimization problem to calculate the best sampling method seeking to minimize the mean squared errors (MSE) or integrated variance (McAleer and Medeiros, 2008). It then enables them to back out an optimal sampling frequency.

Zhou (1996) is one of the first to attempt this using a theoretical model of exchange rates which is a product of a Brownian motion term, deterministic functions, and what he calls a microactivities functions (p. 47). Using his variance equation he is able to back out the optimal number of observations which seek to minimize the covariance between returns. Looking at the data in tick time, Zhou decides on a frequency between every 3 and 5 ticks.

Andersen et al. (2000b) approaches the optimization problem by employing what they call a "volatility signature plot" which compares the mean volatility of various sampling frequencies (p. 107). Their signature plots confirm that the microstructure component of the variance gets larger the more frequent the data is sampled. They suggest that the trade off between sampling error and microstructure effects levels off between 15 and 20 minute returns (Andersen et al., 2000a, p. 106). Bandi and Russell in a series of published and unpublished papers develop a theoretical model for price volatility from which they minimize the mean squared errors in order to optimize the sampling frequency (Bandi and Russell, 2005, 2008; Bandi, Russell, and Zhu, 2008).

The most common sampling frequency is 5 minutes. The vast majority of research utilizing intraday data takes returns every 5 minutes without introduction or explanation. Andersen et al. (2001) justify their use: "The five-minute horizon is short enough that the accuracy of the continuous record asymptotics underlying our realized volatility measures work well, and long enough that the confounding influences from market microstructure frictions are not overwhelming" (p. 50). Initially, the availability of higher frequency data restricted the frequency of sampling like Epps (1979), which deal directly with sampling frequency use 10 minute intervals since it was the highest granularity available. But Wasserfallen and Zimmermann (1985) investigate the different effects 1 to 10 minute sampling intervals have only a few years later.

During the growth of high frequency data analysis, ad hoc methods of accumulation are common, but 5 minute returns quickly became the standard method. Later reviews like Pagel, De Jongh, and Venter (2007) explain it similarly to Andersen et al. (2001) as a balance between the noisy tick-to-tick and unbiased, albeit less accurate, 15 to 20 minute intervals. More recently, Liu, Patton, and Sheppard (2012) scrutinize the 5 minute convention compared to numerous other estimates and found that only under very specific circumstance did a few other calendar time sampling methods outperform 5 minute returns in forecasting accuracy.

This analysis seeks to benefit from more than enhance the existing high frequency data literature. I rely on the methods which have achieved consensus thanks to a rigorous vetting processes from different sources. There are many other methods of high frequency data accumulation, sampling, interpolation, and volatility. For a more comprehensive review, see Andersen et al. (2000a); Pagel, De Jongh, and Venter (2007); McAleer and Medeiros (2008); Liu, Patton, and Sheppard (2012).

DATA

There are three primary sources of data for this project: announcement dates for USDA WASDE, crop production, grain stock, prospective planting, and acreage reports, Kansas City Board of Trade transaction data, and KCBT daily price data. All other covariates are constructed from these. This chapter describes each data set and the variables contructed from it.

Data Sources

The dates on which USDA reports are released are collected from the USDA archive website hosted by the Albert R. Mann Library of Cornell University.[1] In all, there are 75 unique report dates from the beginning of 2008 and the end of 2012. The WASDE and crop production reports are always released together at the same date and time. The prospective planting and acreage reports are published at end of March and June respectively resulting in only five dates each which tend to coincide with one of the four grain stock reports each year. Generally, there are fifteen unique report dates per calendar year, twelve from WASDE releases, and three for the acreage, prospective planting, and grain stocks.

Only the report dates are collected, not the informational content of the reports. The WASDE report contains domestic new wheat crop survey data from NASS during the growing season of each year. The first report to contain the initial information is in May and is updated each month through August, with the numbers finalized in October.

[1] www.usda.mannlib.cornell.edu

Transaction price data for the KCBT HRW futures contract are taken from a tick data set posted on their website prior to the merger of KCBT with CME. Only a five year period of 2008 to 2012 is used where this data set contains the most reliably collected information from the CME Globex platform. The data contain all transactions as well as bid, ask, pre and post market opening bid and ask, open and closing price ranges. All other quotes except for actual matched trades are trimmed away in order to study the actual transactions themselves. By doing so, this study can focus on the exchange behavior of the market rather than looking at pre and post market reactions or bid-ask spreads. Additionally, while the transaction data is complete, the quoting of asks and bids is sporadic, undermining the integrity of any analysis.

To supplement the transaction data, daily price data for the KCBT HRW wheat futures contract are downloaded from the Quandl website covering the same sample period, 2008 to 2012. This series provides the daily open and settle prices, open interest and total volume for that day. The daily high and low prices are also recorded. These data are useful for several reasons. First, the daily open and closing prices are not the same as the first and last transaction of each day. The opening price is often changed by the overnight session, and the settle price is calculated from a range of prices near the end of the trading day. Using the settle price from the previous, I can determine if the price move limit is reached during a given trading day. When the limit is settled at, the price limit increases on the ensuing trading session. Thus, days on which the price discovery mechanism of the market is restricted can be identified. This process of price move limits is fully explained in chapter 2.

During the sample period no major changes take place regarding day trading hours, trading platform, or report releases.[2] All the transactions of interest trade during the day session from Monday to Friday, opening at 9:30 am and closing at 1:15 pm CST. The sample also coincides with the period that KCBT HRW futures first traded exclusively on CME Group's Globex electronic platform. USDA reports are made available at 7:30 am Central Time on release dates, prior to the market opening. This way I can evaluate the effect of reports on the same trading session as the report release.

The final data set contains only transactions from the most actively traded nearby contract, which are rolled from one contract to the next on the first trading day of the contract delivery month. Each contract is the nearby for approximately 60 to 90 days. Only the day trading session is considered. The day session uses side-by-side trading in which both electronically- matched trades are recorded and traditional pit, or open outcry, trades. Open Outcry trades are distinguishable from electronic trades. The evening or overnight electronic trading session is not considered due to the sparseness of observations. There are 4,531,673 observations or transactions for the nearby contract from 2008 until 2012 for only the day session. The daily price data are trimmed to nearby contract in the same manner to match the transaction data. The nearby contract is rolled over on the first of every delivery month to the next contract nearest to deliver. Overall, the full sample period covers 1261 total trading days.

Prior to any analysis, the transaction and daily price data are natural logged to reduce the weight of outliers and deal with likely heteroskedasticity. Several sam-

[2]With the exception of the change in the closing of the night session from 6 am to 7:15 am in 2009. This change does not affect the analysis since the night session is not used except for the one noted occasion.

pling schemes are employed. Using standard calendar time I calculate daily 5 minute returns using the “previous tick” method of interpolation, filling in for missing observations. Transaction time is considered for the unsampled price series transaction by transaction. Also, RV for various chunks of interest inside a day can be calculated, like the first 15 minutes.

After selecting a sampling scheme, the series is differenced to calculate period-to-period returns from which realized volatility estimates can be made. For analysis and replication of previous studies, Lehecka, Wang, and Garcia (e.g. 2014); McNew and Espinosa (e.g. 1994), an 11-day event window is constructed. With five days before and after each report plus the 75 announcement days themselves the event window should contain 825 days total. But report release date windows frequently overlap resulting in truncated intervals. To deal with this, days that are overlapped are counted twice, providing an observation for each event window. This way the full 825 day sample is still intact with equal number of days for each USDA publication.

Summary Statistics

Agricultural commodities exhibited remarkable price variability during the time period in consideration. The majority of this volatility falls in the year 2008, which is commonly held as the height of a global food crisis that began in 2007. Table 1 shows basic summary statistics for the total price series, transaction by transaction. While the first row gives statistics for the entire sample period, the second two compares 2008 to the rest of the sample. 2008 has a much larger price range and variability than the 2009 to 2012 period and drives much of the data variation. This will be explicitly controlled for when appropriate.

Table 1: Summary Statistics for KCBT Transaction Price Series

Transaction by Transaction Data					
Years	N	Mean	Std Dev	Minimum	Maximum
2008 - 2012	4531673	732.20	148.90	456.25	1384.75
2008	479380	811.79	171.79	501.00	1384.75
2009 - 2012	4052293	722.78	143.09	456.25	990.50

Figure 5 helps visualize the path of prices of the sample period, where 2008 has noticeably high prices and greater changes in prices. This is the only time that the price series will be considered in levels.

Figure 5: Daily KCBT Settle Price Series 2008 to 2012

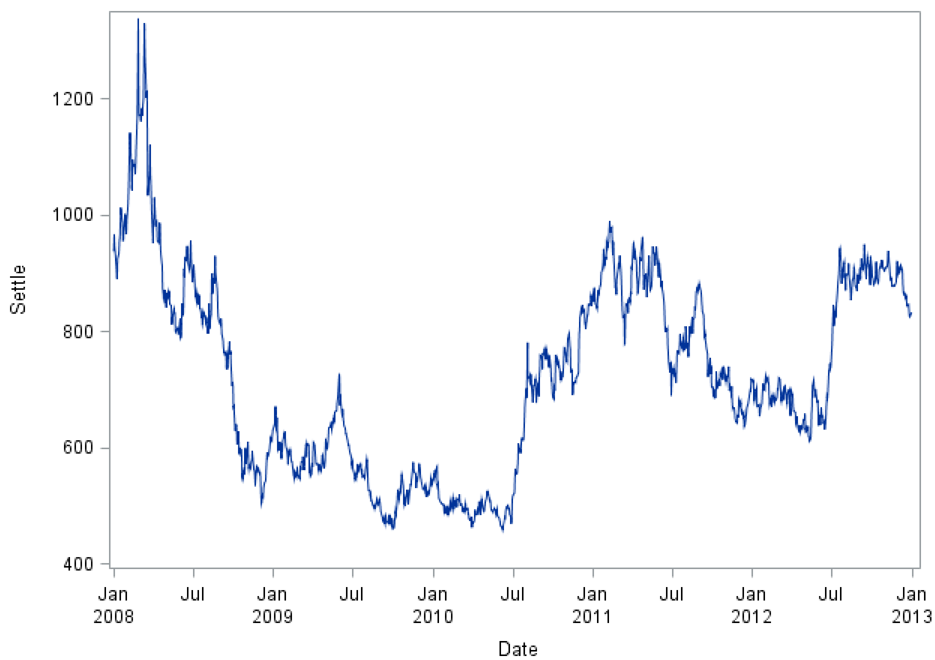


Table 2 displays summary statics for daily realized volatility estimates for two different sampling methods using two different methods of RV. The first is 5 minute calendar time sampling. The second is the unsampled transaction flow returns. Noticeably, the minimum is always 0 on a day due to the fact that price limits during the most volatile period of early 2008 prevented any trading from taking place for several days.

Table 2: Summary Statics of Daily RV Estimates

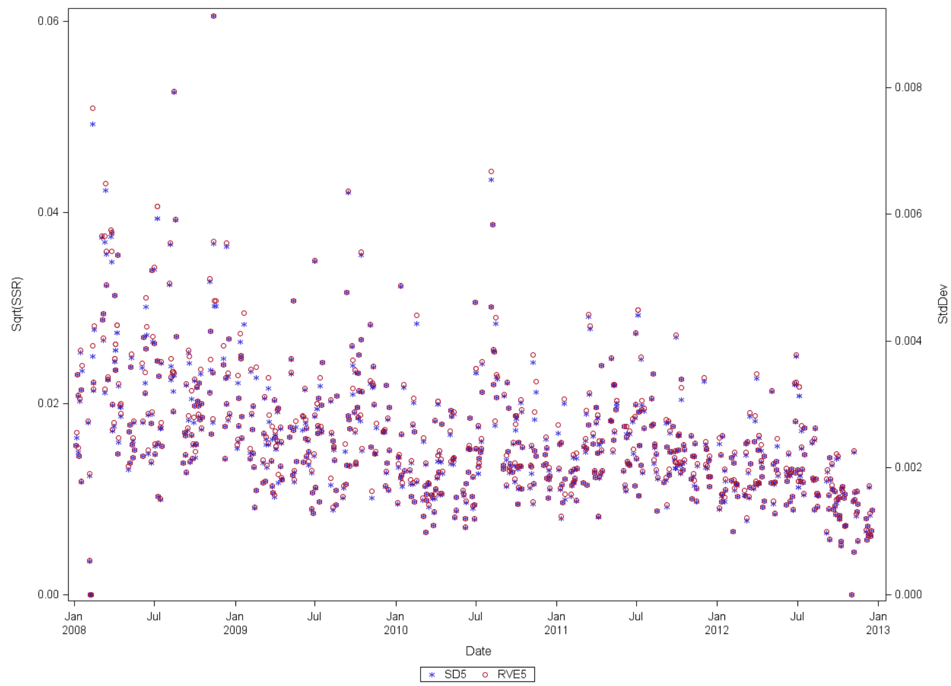
$Variable_{sampling}$	N	Mean	Std Dev	Minimum	Maximum
$\sqrt{SSR}_{5min}$	825	0.016585	0.006802	0	0.060532
SD_{5min}	825	0.002471	0.001013	0	0.009124
$\sqrt{SSR}_{tick}$	825	0.030695	0.022028	0	0.33079
SD_{tick}	825	0.000548	0.000446	0	0.005257

The two measure of volatility are presented for each sampling method for comparison. The square root of the sum of squared returns is noticeably larger in all case merely because it is not divided by, n, the number of observations sampled. As a results, moving down the table to more frequent sampling methods produces a larger $\sqrt{SSR}$ because there are more returns added into the calculation. The opposite effect appears for the SD measure. This is because as sampling frequency increases price moves get smaller between transactions, hence returns shrink. Also n, the divisor, increases proportionally as frequency increases.

This table high-lights the primary difference between $\sqrt{SSR}$ and SD RV estimates. The former gets larger as more returns are sampled while the latter gets smaller. Furthermore, under different sampling frequencies, $\sqrt{SSR}$ is not comparable. Since different frequencies contain different numbers of returns summed, the more frequent one will necessarily be larger. Hence, one must be careful when interpreting the magnitude of $\sqrt{SSR}$ estimates. On the other hand, SD measures are normalized by the number of transactions making them comparable over different timescales.

Figure 6 helps visualize the path of RV over the event window period. This figure also exemplifies the fact that once adjusted for scale, $\sqrt{SSR}$ and SD are nearly identical and yield identical statistical results. This justifies their interchangeability.

Figure 6: Realized Volatility Estimates: SD vs. $\sqrt{SSR}$



As a result, in reporting analysis in which both measures are feasible, justification will be given for choosing one measure over the other leaving the alternative's results suppressed for the convenience of the reader.

In order to perform analysis on the most relevant range of the trading as determined by previous research, RV estimates are calculated for the first 15 minutes of each days when prices have been shown to measurably shift. Summary statics for these measures are displayed in Table 3. Again the difference in magnitude is noticeable but what is more important is that the average of all $\sqrt{SSR}$ is not a useful measure. This is because observations at the tick level are bound to be different for first fifteen minutes of each trading. Hence, that RV estimates are entirely incomparable. Since much of the analysis in this work uses transaction level data, the $\sqrt{SSR}$ RV estimator is not used or addressed because it can not make useful comparisons. This decision is confirmed by the results of Figure 6 which show how show similar the SD measure is to the $\sqrt{SSR}$.

Table 3: Summary Statics for First 15 Minute RV

Variable	N	Mean	Std Dev	Minimum	Maximum
SD_{15tick}	825	0.000649	0.000486	0	0.005698
$\sqrt{SSR}_{15tick}$	825	0.015428	0.009822	0	0.141879

RESULTS

Objectives

My objective is to measure if and in what manner USDA crop reports change realized volatility measures of uncertainty in the HRW wheat futures market. My primary measure, RV, is calculated from transaction level data rather than interpolated from models providing me with the ability to study the real time shifts in the distribution of prices. Using this new measure of uncertainty for the wheat market allows for comparison with familiar techniques in addition to analysis at a finer level of granularity.

I begin by testing the hypotheses considered by Lehecka, Wang, and Garcia (2014) and McNew and Espinosa (1994) using their methods. Tests inspired by Lehecka, Wang, and Garcia (2014) confirm that there is a price impact at the minute by minute level on wheat futures when relevant USDA information is made publicly available. Tests similar to McNew and Espinosa (1994) identify the distributional influence of USDA reports on daily wheat RV rather than using implied volatility. This enables me to test their primary hypothesis of the USDA reports reducing uncertainty using this new data. This test shows, however, that RV is significantly larger on report days than non-report days.

The final and unique contribution of this work exploits the nature of high frequency data to measure the effect of USDA reports on RV calculated for smaller scales inside of a day rather than entire days. The replication work provides the preliminary and baseline results which confirm and motivate the final component of this work. Building on these, I perform robustness check with regression and statistical analysis to possibly identify trends hidden by lower frequency observation.

Replication of Lehecka, Wang, and Garcia (2014)

Lehecka, Wang, and Garcia (2014) consider the average absolute price change on report and non-report days for high frequency corn data. Table 4 presents a replication of their primary results using high frequency wheat data. It compares the average absolute deviations from the median return for each minute of report days and pre and post report days. This uses the same 5 day before and after the release of a USDA WASDE, crop production, grain stock, prospective planting or acreage report. The only first 15 minutes of the day trading session are reported because it is assumed the reaction is instantaneous and any reaction after this period may not be attributed to the report's release.

The first column in Table 4 presents the average absolute deviations from the median return for that minute only on days that a USDA report is released. The next column reports the same statistic but for only the 5 days before and after a report day. These are calculated using Equation 1. The third column shows that report days consistently have a larger average absolute deviation from the median return.

The F-stat is an equal variance test on absolute returns, showing that the variance of absolute returns is significantly larger on report days at the 99% level for the first 5 minutes with sporadic significance for the next 10 minutes. Due to the known non-normality of high frequency returns as well as large skewness and kurtosis, the Kruskal-Wallis rank-sum non-parametric test looks for differences in absolute returns on report and non-report days. This test is useful in identifying location differences in distributions know to have non-normal properties, like high-frequency transaction data. It is reported as a χ^2 statistic. The Kruskal-Wallis shows similar results, with significance at the 99% for the first 4 minutes and even more sporadic significance trailing off over the next 11 minutes.

Table 4: Average Absolute Deviations from the Median per Minute of Report and Non-Report Days

t	AAD_r	AAD_{nr}	$DAAD$	F_{stat}	F_{pv}	KW_{stat}	KW_{pv}
9:30	2.026	0.497	1.530	6.225	0.000	67.733	0.000
9:31	0.452	0.197	0.255	5.689	0.000	8.406	0.004
9:32	0.305	0.165	0.141	2.266	0.000	7.456	0.006
9:33	0.244	0.151	0.093	2.666	0.000	10.311	0.001
9:34	0.189	0.138	0.051	2.525	0.000	2.777	0.096
9:35	0.189	0.143	0.046	1.199	0.168	8.051	0.005
9:36	0.147	0.140	0.007	1.109	0.296	2.768	0.096
9:37	0.140	0.128	0.012	1.185	0.184	7.386	0.007
9:38	0.130	0.120	0.011	1.654	0.005	0.011	0.915
9:39	0.125	0.120	0.005	1.235	0.134	0.655	0.419
9:40	0.115	0.129	-0.013	1.018	0.479	0.040	0.842
9:41	0.159	0.124	0.035	1.128	0.268	3.519	0.061
9:42	0.190	0.119	0.071	2.173	0.000	9.163	0.002
9:43	0.140	0.117	0.023	1.193	0.181	9.099	0.003
9:44	0.144	0.108	0.036	2.438	0.000	0.477	0.490

The full sample of KCBT transaction level data from 2008 to 2012 is used instead of the July 2009 to May 2012 of Lehecka, Wang, and Garcia (2014), showing the impact of USDA reports is significant over the entire sample in question. This calls for one minor correction for when the overnight electronic trading session changes from closing at 6 to 7:15 am on July 1, 2009. As has been specified, there are 75 unique report dates during the 2008 to 2012 dates. This means that there should be 75 observations for every minute of report days and 750 for every non-report day. This is not the case, however, due to sparse trading during certain days. Nevertheless, there are between 550 and 650 observations per non-report minute and 70 per report minute.

Only one other deviation is made from Lehecka, Wang, and Garcia (2014)'s approach which is that because contracts are so thinly traded in the night session, the previous tick method is used to insure there is a transaction price in the last minute of every morning session. This guarantees a representative first minute return can always be calculated. The first minute, as evidenced by Table 4 is always the largest, hence it is important to provide that this return can be computed in the same manner as their work. All other missing values are left as such, per Lehecka, Wang, and Garcia (2014)'s methods.

Overall, the results using KCBT transaction wheat data are consistent with the findings of Lehecka, Wang, and Garcia (2014). There is a statistically significant increase in the variability of returns on report days compared to the five days before and after a USDA publication. This impact appears to be even more fleeting in the wheat market than in the corn market. These preliminary results confirm that USDA reports do influence the wheat market and provide a basis for further investigation of the hypothesis of whether USDA reports effect volatility in the market. This is the only analysis done where days are pooled to calculate volatility, all other measures will be realized volatility computed from prices within a day.

Replication of McNew and Espinosa (1994)

The metric of interest varies from the variance of returns used in the F and Kruskal-Wallis tests conducted by Lehecka, Wang, and Garcia (2014) necessitating a different approach to testing. I then turn to the work of McNew and Espinosa (1994) to construct a simple manner to compare the daily volatility of the event window around USDA reports in the wheat market. Table 5 presents the average daily RV measure of each day across months and years of the event window. Following the methods of McNew and Espinosa (1994), I perform tests comparing days of the event window.

These numbers are normalized by the volatility of the day before a report release. This helps to remove seasonal or time fixed effects that might be present and invalidate testing (McNew and Espinosa, 1994, p. 484). For that reason, the day before the report or day -1 is not reported because its average is 1 by construction. The literature standard of 5 minute calendar time returns are used to calculate RV. In this case, daily standard deviations are used to make the analysis more comparable to previous work with implied volatility. Each day then has 75 observations.

Table 5: Average Daily Wheat RV Relative to Day Before Release

t	-5	-4	-3	-2	0	+1	+2	+3	+4	+5
$\sigma_t \backslash \sigma_{-1}$	1.108	1.154	1.139	1.062	1.398	1.141	1.161	1.139	1.160	1.116

The first striking element of Table 5 is that the first day the wheat market trades on the new USDA information is much larger than any other day on average in the report window. Furthermore, every day on average is above one. This means that every day in the event window is on average more volatile than the day before a USDA report is released.

Table 6 considers differences in RV on report and non-report days for specific months in which reports are released. In this table each day has only 5 observations except for March, June, and September which have 10 because two reports are released in those months. These averages show very little difference in volatilities across all days with the noticeable exception of the January. This is not surprising because January's release date contains not just the WASDE and crop production but also the 4th quarter Grain Stocks report from the previous year. September and June are also noticeably larger than other months. September includes both the monthly WASDE/crop production as well as the quarter grain stock. June also sees the release of a WASDE and crop production report at the beginning of those months followed by the acreage report at the end of the month. We would expect the additional information content of these reports to further influence average volatility.

Table 6: Average Wheat RV Per Month Relative to Day Before Release

t	-5	-4	-3	-2	0	+1	+2	+3	+4	+5
Month										
Jan.	1.209	1.127	1.315	1.077	1.688	1.242	1.221	1.294	1.169	1.238
Feb.	0.968	0.944	0.974	0.991	1.208	1.083	1.018	0.758	1.432	0.860
Mar.	1.181	1.192	1.338	1.135	1.373	1.255	1.170	1.510	1.444	1.355
Apr.	1.140	1.068	1.179	1.126	1.206	1.166	1.132	1.096	1.055	1.105
May	1.045	1.216	1.294	1.179	1.381	1.101	0.952	0.938	1.097	0.938
Jun.	0.866	0.943	0.827	0.903	1.520	1.139	0.973	1.097	0.993	1.043
Jul.	1.333	1.606	1.350	1.046	1.337	1.198	1.114	1.089	1.319	1.195
Aug.	1.671	1.506	1.469	1.494	1.362	1.191	1.483	1.076	1.077	1.321
Sep.	1.021	1.189	1.162	1.028	1.627	1.165	1.728	1.476	1.441	1.181
Oct.	0.826	0.842	0.834	0.741	1.067	1.037	0.974	1.008	0.948	0.899
Nov.	1.096	1.188	1.001	1.226	1.438	1.029	1.410	1.575	1.092	1.376
Dec.	1.168	1.131	0.995	0.913	1.210	1.098	1.004	0.912	1.068	0.990

Extending the analysis of Table 5, Kruskal-Wallis non-parametric tests are performed to compare realized volatility of all the observations of each day in the event window to one another. These tests are performed on the volatility relative to Day -1, the day before the report. The p-values of the test between each day of the event window are reported in Table 7. The null hypothesis is RV is the same on each day of the event window. Stated formally - $H_0 : \sigma_i = \sigma_j$. The alternative is stated at the top of the table. The table is symmetric so the top right numbers are not displayed.

The results show that on the first day where the KCBT trades on USDA report information, Day 0, has statistically different price volatility than nearly every other

Table 7: Probability Values for Kruskal-Wallis Test Statistic on Daily Wheat RV

	H$_a$: $\sigma_i \neq \sigma_j$									
j	-5	-4	-3	-2	-1	0	+1	+2	+3	+4
i										
-4	0.613									
-3	0.821	0.762								
-2	0.497	0.187	0.476							
-1	0.069	0.034	0.363	0.879						
0	0.003	0.007	0.005	0.000	0.000					
+1	0.505	0.842	0.645	0.204	0.071	0.008				
+2	0.480	0.818	0.637	0.158	0.006	0.016	0.985			
+3	0.994	0.478	0.960	0.430	0.762	0.002	0.469	0.443		
+4	0.632	0.878	0.794	0.248	0.363	0.009	0.842	0.848	0.651	
+5	0.972	0.628	0.869	0.384	0.069	0.003	0.634	0.514	0.847	0.689

day in the event window at the 99% level. This is except for Day +1, for which the difference is only significant at the 95% level, showing there may be some residual volatility impact on that day following the report. The averages of daily RV suggests that the day before the report is also different than all other days in the event window. These results only provide marginal significance for that hypothesis. No other pattern stands out in the table suggesting limited difference between volatility on most days.

These results are constructed in a manner very similar to McNew and Espinosa (1994) in order to make them comparable. What stands out is that the results are in direct contrast to the findings of McNew and Espinosa (1994) and Isengildina-Massa et al. (2008b). Both of those work find that USDA reports reduce futures market uncertainty as measured by implied volatility at the daily level. These basic results suggests that USDA reports instead increase market uncertainty as measured by the primary statistic, RV, challenging the hypothesis that USDA reports reduce uncertainty and facilitate price consensus. This also raises questions about differences between RV and implied volatility in measuring uncertainty in futures markets.

While inspired by McNew and Espinosa (1994)'s main results, Table 7 departs slightly from their methods. First, the Kruskal-Wallis rank sum test is reported to be consistent with analysis from the previous section. McNew and Espinosa (1994) perform this same test as well as the Van der Waerden test, but only report the latter because they yield identical results. Both tests are performed here as well confirming identical results. In order to be consistent with analysis in the previous section, the Kruskal-Wallis p-values are reported here instead.

Furthermore, Table 7 presents the results of a two-tailed test, not the one tailed tests performed by McNew and Espinosa (1994). This is because, as suggested by the averages of Table 5, the volatility is actually greater on report days then non-report. This means the directional tests implemented in McNew and Espinosa (1994) would return no results, because on average reports have more rather than less volatility relative to other days. Nor do the volatility measures lend themselves to the pre-report gradual increase in volatility that their hypothesis suggests. Instead the tests of Table 7 show a possible lull before the report release.

While Table 7 gives no directional tests, it is a fact that Day 0 is significantly larger than all other report days. The day before the report release, Day -1, is moderately smaller compared to sporadic days. As a two tailed test, the null hypothesis rejection region is split in half make the results even more robust than one tailed tests.

Intraminute Average Realized Volatility

On the surface it would appear that the replication of Lehecka, Wang, and Garcia (2014) in Table 4 accomplished the task of studying intraminute volatility. While informative, their work does not provide the type of intraday analysis that realized volatility estimators can. Lehecka, Wang, and Garcia's methods differ significantly from RV in its calculation of price variability because it treats all days the same except for the class distinction between report and pre/post report days. They begin with merely an average of returns for each minute, that is report days are pooled together to compare. They then calculate a variance of returns across the same minute of multiple days for their F test, and look at only the returns to compute the Kruskal-Wallis statistic.

These methods are not the same as calculating a variance (or standard deviation) of returns inside of a day. Lehecka, Wang, and Garcia (2014) samples in a manner that produces one observation per minute to study minute to minute price changes, while I instead utilize all the observations from a given minute to calculate a distinct volatility measure for each minute of each day. These individual RV estimates can then be analyzed for their measure of the spread or distributive properties of prices rather than mean price changes.

Because the initial results suggest that uncertainty, as measured by RV, actually increases following USDA report releases, moving to intraday analysis may explain how RV is reacting to this new information. An alternative hypothesis is that there is an initial adjustment period or shock which increases uncertainty following a report release, something possibly driven by different and conflicting interpretation of the USDA data. Following the resolution of these difference of opinion possibly uncertainty reduces later in the day. This can only be identified by looking at RV over intervals inside the trading day.

To better assess this possible change in RV, I disaggregate non-report days into the 5 days before the report, the pre-report days, and the five days after, the post-report days. By doing so, I can then compare days in two dimensions. First, compare across days to see how the distribution of RV changes from the days leading up to and the report date is self. I can also compare the report day to the days after the report.

This comparison across days is similar to previous studies comparing implied volatility across days, but this analysis adds an additional dimension by comparing RV of particular trading minutes inside of a day. This allows the original hypothesis to be tested with a closer observation of the reaction but also allows for potentially identifying results consistent with the alternative hypothesis suggested.

In order to take full advantage of the new time scale added, the full transaction-by-transaction sample is used to exploit its greater number of observations. The higher frequency of the data provides at times over 200 transactions in a single minute. The ensuing tests use the standard deviation measure of RV because the sum of squared returns cannot be compared for different number of transactions per minute as well as to remain consistent with tests from previous sections.

At this level of granularity the previous discussion of microstructure noise in chapter 4 is immediately relevant. Looking at standard deviations inside of each minute of each day is likely to contain a large amount of microstructure effects which will provide the appearance of fundamental price volatility when they are instead frictions of the trading process. As a result, at this level of frequency I cannot test any hypothesis about the fundamental value of wheat. Nevertheless, I posit that distortions like microstructure noise are interesting in themselves because they reflect trading behavior in the wheat market.

Hence, detecting changes in the RV due to shifts in the fundamental value, the microstructure noise component, or both is equally demonstrative of the impact of USDA reports on trading in the market as well as uncertainty.

Table 8 show the average RV per minute for the first 30 minutes of the day trading session on report and pre-report days. A longer period of the opening trading day

Table 8: Intraminute RV on Report and Pre-Report Days

t	ARV_r	ARV_{pr}	$DARV$	KW_{stat}	KW_{pv}
9:30	0.0011	0.0008	0.0003	15.1526	0.0001
9:31	0.0008	0.0005	0.0003	10.7153	0.0011
9:32	0.0007	0.0005	0.0002	3.6153	0.0573
9:33	0.0006	0.0005	0.0002	4.5700	0.0325
9:34	0.0008	0.0005	0.0003	7.7528	0.0054
9:35	0.0006	0.0005	0.0001	6.3544	0.0117
9:36	0.0006	0.0005	0.0001	1.4420	0.2298
9:37	0.0006	0.0005	0.0001	6.4947	0.0108
9:38	0.0005	0.0005	0.0001	0.0040	0.9495
9:39	0.0006	0.0005	0.0002	0.7613	0.3829
9:40	0.0008	0.0005	0.0003	0.5347	0.4646
9:41	0.0006	0.0005	0.0001	0.0178	0.8937
9:42	0.0008	0.0004	0.0003	2.4979	0.1140
9:43	0.0007	0.0005	0.0002	0.0412	0.8391
9:44	0.0008	0.0005	0.0003	0.3353	0.5626
9:45	0.0007	0.0005	0.0002	2.8102	0.0937
9:46	0.0008	0.0005	0.0003	5.3881	0.0203
9:47	0.0007	0.0005	0.0002	1.4123	0.2347
9:48	0.0008	0.0004	0.0004	15.9685	0.0001
9:49	0.0006	0.0005	0.0002	5.5055	0.0190
9:50	0.0007	0.0005	0.0002	13.1480	0.0003
9:51	0.0006	0.0004	0.0002	1.4053	0.2358
9:52	0.0007	0.0005	0.0002	0.1987	0.6558
9:53	0.0007	0.0005	0.0002	2.0913	0.1481
9:54	0.0007	0.0005	0.0002	2.0322	0.1540
9:55	0.0007	0.0005	0.0003	4.2273	0.0398
9:56	0.0007	0.0004	0.0003	2.0651	0.1507
9:57	0.0006	0.0004	0.0002	2.6192	0.1056
9:58	0.0005	0.0004	0.0001	1.3684	0.2421
9:59	0.0006	0.0005	0.0002	1.0588	0.3035

is shown to consider the possibility of a longer more gradual adjustment in volatility through time. The first column presents the per minute averages for days on which reports are released. There are typically 70 report days for each minute. The second column reports the same for the five days before the report release of the event window for which there are always at least 330 days. The third displays the difference between the first two columns.

The difference column shows RV on average is larger on the day USDA wheat information is released than on the five previous trading days. Again, the Kruskal-Wallis tests for locational differences in minute-by-minute RV, comparing report and pre-report days. The test statistic is significant at least at the 95% level for about the first 7 minutes, very similar to Lehecka, Wang, and Garcia (2014)'s pooled volatility results. What is more interesting is that following this there is no major trend of difference in RV between pre-report days and the report date. This means that following the initial jump in volatility, it returns to the same level as the previous days if not occasionally higher. While not show here for brevity, this same finding plays out through the entire trading day.

Table 9 makes a similar comparison of intraminute RV on report days and days following the report. Here average RV is again larger on report days than on the days following a USDA publication, however, the difference is not as large. Here, the Kruskal-Wallis test provides a significant difference at the 95% level in the distribution of minute-by-minute RV estimates for only the first 5 minutes. Thus, in minutes of the days following the release of reports RV is smaller over all on the post report days but not significantly so. This is partially consistent with implied volatility results in that volatility is not significantly different on post report days than report days. Here the tendency is that RV is largest on report days and slightly smaller on the days following rather than report days being the smallest with gradual decreases afterwards.

The result of higher RV on report days could mask a gradual lagged reduction in the volatility over the few days following the report. This would in some ways confirm the initial hypothesis just over a more generous interval of time. However, it rules out entirely that in later minutes in the day volatility is significantly reduced. In order to considered a longer more gradual reduction in uncertainty I can compare the change from the days before the report to the days after. Hence, this comparison could capture an overall reduction in the magnitude of volatility following the report day's increase.

Table 10 presents the results of comparing the 5 trading days prior to report day to the 5 trading days after the USDA report's release. Column three displays the difference of post-report average RV from pre-report days' average RV. Surprisingly, the majority of minutes of the post-report days have a slightly higher average volatility than the days prior, as evidenced by the negative sign. This finding is not unanimous, since on some days pre-report days are slightly larger. Nevertheless, as the Kruskal-Wallis test results show, the spread of RV minutes of pre and post report days is not enough to suggest the distribution of observed RV for each minute is different.

These results are interesting in that they provide strong evidence that the release of USDA information does not reduce uncertainty in the wheat futures market either in the immediate day following the release or over the 5 days following its release. Considering all the tests, it is clear that on the day of the reports' release locational difference exist in the RV for the first 10 minutes roughly from all other days in the event window. While little statical significance results in any other minute or day comparison, RV in magnitude is still larger on average on the report day than any other day. More interestingly, average RV on the days following the report is still

Table 9: Intraminute RV on Report and Post-Report Days

t	ARV_r	ARV_{po}	$DARV$	KW_{stat}	KW_{pv}
9:30	0.0011	0.0009	0.0002	7.2318	0.0072
9:31	0.0008	0.0005	0.0002	5.2044	0.0225
9:32	0.0007	0.0005	0.0002	2.6616	0.1028
9:33	0.0006	0.0005	0.0001	3.5602	0.0592
9:34	0.0008	0.0005	0.0003	6.7996	0.0091
9:35	0.0006	0.0005	0.0000	2.1214	0.1453
9:36	0.0006	0.0005	0.0001	1.1763	0.2781
9:37	0.0006	0.0005	0.0000	3.9934	0.0457
9:38	0.0005	0.0005	0.0000	0.0238	0.8775
9:39	0.0006	0.0006	0.0000	0.1966	0.6575
9:40	0.0008	0.0005	0.0002	0.0476	0.8273
9:41	0.0006	0.0006	0.0000	0.0305	0.8614
9:42	0.0008	0.0006	0.0002	0.8396	0.3595
9:43	0.0007	0.0006	0.0001	0.5703	0.4502
9:44	0.0008	0.0006	0.0002	0.0440	0.8339
9:45	0.0007	0.0006	0.0002	3.0206	0.0822
9:46	0.0008	0.0005	0.0002	4.8854	0.0271
9:47	0.0007	0.0005	0.0002	1.0724	0.3004
9:48	0.0008	0.0005	0.0003	9.9339	0.0016
9:49	0.0006	0.0005	0.0002	4.2133	0.0401
9:50	0.0007	0.0005	0.0002	7.2682	0.0070
9:51	0.0006	0.0005	0.0001	0.3799	0.5377
9:52	0.0007	0.0005	0.0002	0.3117	0.5767
9:53	0.0007	0.0005	0.0002	2.9996	0.0833
9:54	0.0007	0.0005	0.0002	2.0783	0.1494
9:55	0.0007	0.0006	0.0002	1.8924	0.1689
9:56	0.0007	0.0006	0.0002	1.8519	0.1736
9:57	0.0006	0.0005	0.0001	0.6934	0.4050
9:58	0.0005	0.0005	0.0001	0.5607	0.4540
9:59	0.0006	0.0005	0.0002	1.3473	0.2458

Table 10: Intraminute RV on Pre and Post-Report Days

t	ARV_{pr}	ARV_{po}	$DARV$	KW_{stat}	KW_{pv}
9:30	0.0008	0.0009	-0.0001	4.9314	0.0264
9:31	0.0005	0.0005	0.0000	2.3345	0.1265
9:32	0.0005	0.0005	0.0000	0.1126	0.7372
9:33	0.0005	0.0005	0.0000	0.1349	0.7135
9:34	0.0005	0.0005	0.0000	0.0352	0.8512
9:35	0.0005	0.0005	-0.0001	2.5924	0.1074
9:36	0.0005	0.0005	0.0000	0.0098	0.9212
9:37	0.0005	0.0005	-0.0001	0.4867	0.4854
9:38	0.0005	0.0005	0.0000	0.1875	0.6650
9:39	0.0005	0.0006	-0.0001	0.6585	0.4171
9:40	0.0005	0.0005	-0.0001	0.7974	0.3719
9:41	0.0005	0.0006	-0.0001	0.3025	0.5823
9:42	0.0004	0.0006	-0.0002	1.1629	0.2809
9:43	0.0005	0.0006	-0.0001	3.1325	0.0768
9:44	0.0005	0.0006	-0.0001	0.3622	0.5473
9:45	0.0005	0.0006	-0.0001	0.0786	0.7792
9:46	0.0005	0.0005	-0.0001	0.0035	0.9531
9:47	0.0005	0.0005	-0.0001	0.0109	0.9168
9:48	0.0004	0.0005	-0.0001	1.9277	0.1650
9:49	0.0005	0.0005	0.0000	0.2293	0.6321
9:50	0.0005	0.0005	0.0000	2.4289	0.1191
9:51	0.0004	0.0005	-0.0001	0.8916	0.3450
9:52	0.0005	0.0005	0.0000	0.1436	0.7047
9:53	0.0005	0.0005	0.0000	0.2718	0.6021
9:54	0.0005	0.0005	0.0000	0.0369	0.8477
9:55	0.0005	0.0006	-0.0001	1.3121	0.2520
9:56	0.0004	0.0006	-0.0001	0.0154	0.9012
9:57	0.0004	0.0005	-0.0001	1.7197	0.1897
9:58	0.0004	0.0005	-0.0001	0.4431	0.5056
9:59	0.0005	0.0005	0.0000	0.0522	0.8193

slightly larger than the days prior. These results present contradictory evidence to previous studies on USDA's impact on uncertainty in agricultural futures markets. Though one cannot conclude that USDA reports directly lead to increased volatility.

Robustness Checks Using Regression Analysis

Baseline Model Controls

Having found that realized volatility does not decline following the release of USDA reports, I move to regression analysis to consider whether other factors may explain observed changes in RV. This analysis does not seek to establish causality but rather check if the results are robust to the consideration of other technical factors that theory would predict to influence RV. By controlling for different factors, I can better identify the effect of USDA reports on market volatility and price uncertainty.

The primary control I introduce is the change in the fundamental value of wheat. Recall from the discussion in chapter 4, the fundamental value of futures is unobservable. But with the arrival of relevant USDA data, one would expect a shift in market participants' grasp of the fundamental value of their contracts. In its simplest form, one can think of changes in the fundamental value of wheat brought on by USDA information as a change in price. Thus I use the move in the level of prices over a given interval of time as a proxy for the change in the fundamental value of wheat.

Controlling for the move in prices is important for to two reasons. First, not all USDA reports are created equal. Some reports contain more information than others e.g. the January report. While some report days have a greater price impact than others, e.g. the September or October report. Secondly, the absolute size of the overall price move affects RV. It is a mathematical necessity that as the price move gets larger over a given range the RV over that same range will be larger as well. This is due to the fact that as the movement of prices is larger, the range of prices is larger for a calculation of volatility thus increasing the magnitude of the measure.

In order to capture the magnitude of the shift itself the absolute price move is calculated to approximate the change in fundamental value. In addition to controlling for the absolute price move, controlling for the number of contracts traded in a day is also indicative of how active as well as volatile the market is. The daily measure of market volume controls for this.

I begin with a very simple model with only the variable of interest and a single control then build gradually to the full slate of controls in order to study the impact of their inclusion. In all models the dependent variable consists of RV for the first 15 minutes of each trading day. The period of observation is limited to the first 15 minutes since the previous analysis suggested a short-lived impact from USDA reports.

The first model includes only the dummy for report days and the volume of thousands of contracts traded that day. The previous sections have already established that RV is higher on report days, thus I expect this coefficient on that estimate to be positive. Conversely, the more contracts are traded the more liquid the market is. I expect this to then reduce volatility since market participants can more easily exchange contracts reducing frictions in the market and large price jumps. This model is presented in Equation 11.

$$\boldsymbol{RV}_{15,i} = \boldsymbol{\beta}_0 + \boldsymbol{\beta}_1 \boldsymbol{ReportDate}_i + \boldsymbol{\beta}_2 \boldsymbol{Volume}_i + \boldsymbol{\varepsilon}_i \tag{11}$$

In this equation $\boldsymbol{RV}_{15,i}$ is the dependent variable and represents the standard deviation of natural log transaction to transaction returns for the first 15 minutes of each day $i = 1...825$ for all the days of the event window 5 year sample. The $\boldsymbol{\beta}_i$'s represent the estimated coefficients. $\boldsymbol{ReportDate}_i$ is the variable of interest which is merely an indicator variable which turns on for each of the 75 report dates, using non-report days as the contrast. The first control is then added, $\boldsymbol{Volume}_i$, which is the daily volume of contracts exchanged for each day i. Naturally, an error term is included.

To then see how the addition of the absolute price move influences the coefficient estimate of the report day indicator, I then add the price move variable to the equation as shown in Equation 12. Basic theory predicts that the coefficient of the absolute price move should be positive since as the price change increases the size of the volatility estimate for that same time also increases. I predict the addition of the price move might mitigate some of the impact of report days on RV.

$$\boldsymbol{RV}_{15,i} = \boldsymbol{\beta}_0 + \boldsymbol{\beta}_1 \boldsymbol{ReportDate}_i + \boldsymbol{\beta}_2 |\boldsymbol{PriceMove}|_{15,i} + \boldsymbol{\beta}_3 \boldsymbol{Volume}_i + \boldsymbol{\varepsilon}_i \quad (12)$$

Two more controls are made based our initial price move control, presented in Equation 13. Here an indicator for when the price move is negative in direction is added in order to capture possible differences in the magnitude of the market reaction. In an efficient market, one would anticipate market reactions that drive prices down or up to be symmetrical because one does not contain more information than the other. Thus, I predict the coefficient on this to be economically and statistically insignificant. This variable is titled $\boldsymbol{NegPM}_i$.

$$\boldsymbol{RV}_{15,i} = \boldsymbol{\beta}_0 + \boldsymbol{\beta}_1 \boldsymbol{ReportDate}_i + \boldsymbol{\beta}_2 |\boldsymbol{PriceMove}|_{15,i} + \boldsymbol{\beta}_3 \boldsymbol{Volume}_i + \boldsymbol{\beta}_4 \boldsymbol{NegPM}_i + \boldsymbol{\alpha}_5 (\boldsymbol{PM}_{15,i} \times \boldsymbol{RD}_i) + \boldsymbol{\varepsilon}_i \quad (13)$$

Then interaction term between the Report Day indicator and the absolute price move variable is included to attempt to control for a conditional impact of USDA reports on wheat realized volatility based on the level of the absolute price move. It is shown in Equation 13 as $(\boldsymbol{PM}_{15,i} \times \boldsymbol{RD}_i)$. I expect this coefficient to be positive and significant. It is difficult to predict exactly how is will impact the magnitude of the price move variable and more so the report day dummy variable.

Two more models are specified adding two different types of controls. First, I include a dummy to indicator weather a the price move limit was struck on a particular day. That means did the wheat price hit the 60 cent up or down limit that KCBT places on trading.[1] Controlling for these days is important because RV is likely to be restricted as evidenced by several dates in early February 2008 where no trading took place because of the price move limit. Because of the limiting effect on prices by this rule, the next trading day might see additional correction or be exceptionally volatile. Thus an additional indicator is added to see if the day following a limit move day sees greater volatility. These variables are indicated by the $\boldsymbol{Limit}_i$ and $\boldsymbol{Limit}_{i-1}$ symbols respectively. I expect them to both be positive since if the limit is struck we expect prices to be very volatile on that day.

Finally, crop year indicators are included for the the 5 crop years covered in the sample. The new wheat crop begins in July then runs through June of the next year. As a result 2007-2008 is the first crop year. The most recent season of the sample, the 2012-2013 year, is used as the baseline so that the exceptionally volatile years of 2007 and 2008 are not used as the comparison. The matrix of crop year indicator is represented by $\boldsymbol{CropYear}_i$ for each day i. The fully specified model is displayed in

[1]See chapter 2 for a full discussion of this institutional detail.

Equation 14.

$$\boldsymbol{RV}_{15,i} = \boldsymbol{\beta}_0 + \boldsymbol{\beta}_1 \boldsymbol{ReportDate}_i + \boldsymbol{\beta}_2 |\boldsymbol{PriceMove}|_{15,i} + \boldsymbol{\beta}_3 \boldsymbol{Volume}_i + \boldsymbol{\beta}_4 \boldsymbol{NegPM}_i$$
$$+ \boldsymbol{\alpha}_1 (\boldsymbol{PM}_{15,i} \times \boldsymbol{RD}_i) + \boldsymbol{\beta}_5 \boldsymbol{Limit}_i + \boldsymbol{\beta}_6 \boldsymbol{Limit}_{i-1} + \boldsymbol{\gamma}_1 \boldsymbol{CropYear}_i + \boldsymbol{\varepsilon}_i \quad (14)$$

To estimate this equation, Ordinary Least Squares is used with Newey-West standard errors to correct for autocorrelation and heteroskedasticity. Again, the unsampled tick-to-tick returns are used to calculate the RV of the first 15 minutes of each trading day, i, in the event window in order to isolate the report effect as well as capture all market activity. The price move variable is calculated as the absolute change in log price over the 15 minute period following the opening of trading. Also, volume is reported in 1000's of contracts traded.

Table 11 presents the coefficient estimates and standard errors for Equation 11, Equation 12, Equation 13 in the first 4 columns. Equation 14 is shown in the last two columns with the limit indicators and then crop years added. Post estimation standardized coefficients for the two continuous variables are provided underneath their unstandardized estimates. The conventional transformation of the estimates that I use is shown in Equation 15 (Bring, 1994). Given the variety of scale and units of measurement for volume and the price move, the standardized coefficients assist in evaluating the magnitude of impact on the dependent variable. These are titled "Beta" per the literature. Following convention dummy variables and interaction terms are left unaltered (Friedrich, 1982).

$$\beta_k^* = \beta_k \frac{s_{x_k}}{s_y} \quad (15)$$

Table 11: Report and Non-report Day Model with Controls

	Dependent Variable for OLS using Newey-West Std. Er.					
Variable	$RV_{First15}$	$RV_{First15}$	$RV_{First15}$	$RV_{First15}$	$RV_{First15}$	$RV_{First15}$
Intercept	0.000874***	0.000721***	0.000747***	0.000781***	0.000753***	0.000291***
	(0.000074)	(0.000065)	(0.000067)	(0.000063)	(0.000063)	(0.000074)
Report Date	0.000371***	0.000283***	0.000287***	0.000027	0.000018	-0.000044
	('0.000094)	(0.000078)	(0.000079)	(0.000116)	(0.000120)	(0.000121)
Volume	-0.000024***	-0.000025***	-0.000026***	-0.000024***	-0.000024***	-0.000005
Beta Volume	-0.1975	- 0.2057	-0.2140	-0.1975	-0.1975	-0.0411
	(0.000006)	(0.000005)	(0.000006)	(0.000006)	(0.000007)	(0.000007)
|Price Move|		0.025500***	0.025800***	0.018800***	0.018200***	0.011500***
Beta |Price Move|		0.2970	0.3005	0.2190	0.2120	0.1339
		(0.004006)	(0.004014)	(0.003252)	(0.003786)	(0.002904)
Neg PM			-0.000049	-0.000056*	-0.000042	-0.000033
			(0.000030)	(0.000030)	(0.000027)	0.0000229
|Price Move| x Report Date				0.028600***	0.029900**	0.035400***
				(0.012900)	(0.012600)	(0.012700)
Limit					0.000033	-0.000114
					(0.000287)	(0.000306)
Lag Limit					0.000599**	0.000340
					(0.000317)	(0.000333)
Crop Years						Yes
R^2	0.0792	0.1644	0.1669	0.1867	0.222	0.3635

Notes: Standard Errors are shown in parentheses. A single (*) denotes statistical significance at the 10%, a double asterisk (**) indicates significance at the 5%, and a triple asterisk (***) means statistical significants at the 1% level.

The results for the basic regression in column one are exactly as predicted. The coefficient on Report Date is positive and significant. It can be interpreted as report days being .000371 more volatile than non-report days. While the magnitude of this estimate seems minuscule, keep in mind that the average RV is about .00065 on all trading days. So a .000371 increase represents an increase of almost 60% in RV on average for report days compared to non-report days. Volume's coefficient is negative and significant but very small, indicating it does not have a large economic impact on RV. The standardized coefficient for volume so that a one standard deviation increase in the volume of contracts produces a .1975 standard deviation in realized volatility.

Adding the price move variable produces results consistent with the predictions, shown in the second column. The coefficient estimate is also positive and significant at the 99% level with a much larger magnitude. This means that for a 1% increase in the absolute price change, there is a .0255 increase in the magnitude of RV. This confirms the theory that as the price move increase so does RV. The coefficient on report dates adjusts downward to .000283, a drop of about 15% in magnitude. This means that as expected some portion of volatility is driven by the overall change in price not just report releases, however the coefficient is still positive and significant. That implies that controlling for the size of the price move which necessarily drives up RV and is likely to be larger on report days, reports days are still more volatile than non-report days.

The standardized coefficients or "betas" for this regression shows the volume estimates consistent at .2075 while the coefficient for the absolute price move to be .297. This means a one standardization increase in the absolute price move results in about a .3 standard deviation increase in the realized volatility. This shows that in magnitude the price move has a greater impact on realized volatility. This conforms to the expectation that the price move would have a greater bearing on volatility than volume.

The coefficient on volume is unchanged by the addition of the price move. The third column includes an indicator variable for the direction of the price move over the first first 15 minutes of each day, the same range for which the RV estimated. As predicted the magnitude is very small and statistically insignificant. Furthermore it's inclusion has no major impact on the other estimates.

The fourth column includes a interaction term between report days and the absolute price move. This is to capture a possible conditional effect that the magnitude of the price move on report days may differ from that on non-report days. As is turns out it is significant and positive as I anticipated.

The interpretation of the interaction term is slightly more subtle. In order to get the conditional impact, we must sum the coefficient of the report day dummy variable with the multiple of the coefficient of the interaction term and the appropriate price move. This means for a one percentage increase in the price move variable, a $.000027 + .028600(1)$ increase in volatility results. Interestingly, the inclusion of the interaction term wipes out both the magnitude and significance of the report day indicator, implying that most of the difference in RV on report and non-report days is driven by the magnitude of the price move. It also reduces the magnitude of the the price move variable itself by about 30%.

This conforms to the theory I set forth because we would expect larger moves in the fundamental value of wheat to have a large impact on volatility. Furthermore, this also suggest that there is little difference between RV on report and non-report days unless there is large change in the understanding of market fundamentals, proxied by the price move variable. Also, the indicator for the price move direction gains significance, suggesting that price moves do not have as symmetric an effect on RV as thought. It is very small and thus possibly economically insignificant but it could suggest that market participants interpret "bad" news more noisily than good news. As a result, negative price moves increase volatility more than positive price moves but only slightly so.

The final two columns of Table 11 include first the indicator variables for whether or not a trading day hit the price move limit set by KCBT. Both coefficients are positive as predicted but small in magnitude. Only the lagged variable is significant implying that it appears the day after the market is restricted by the limit price move does seem some residual price volatility. However, the final regression with the full set of crop year dummies removes the significance of the lagged limit variable. While not presented, the five crop year dummies are all significant and positive, meaning that all crop years prior to the 2012-2013 year were more volatile on average.

It is also interest to note how this time fixed effects impact the other estimates. First, one should notice that the statistical significance of the volume estimate is completely eliminated and its coefficient drops by an order of magnitude. While the price move is still significant its coefficient is reduced by nearly 50%. The report day indicator is still small and statistically insignificant. Interestingly, the interaction term gains in magnitude and remains significant, meaning that controlling for different volatilities in different crop cycles the conditional impact of a price move on a report day is greater. Lastly, the annual fixed effects reduce the estimates of the limit move indicators making them insignificant.

Overall these result show the necessity of controlling for annual fixed effects but also evince that when a large shift in prices occurs, which I posit represents an arrival of more information valuable to the market, report days exhibit greater volatility than no report days. Also, there does not seem to be a economically or statistically significant difference in negative versus positive price moves which conforms with an efficient market.

Individual Report Model with Controls

Up until now, this study has only considered report and non-report days with no distinction between which reports are released. It is possible that different reports under consideration contain different levels of useful information to the wheat futures market and as a result will impact the market differently. Unfortunately, of the five reports studied here, the WASDE, crop production, prospective planting, grain stocks, and acreage reports, there is significant overlap between all there release dates. As was described in greater detail in the institutional background, the WASDE and crop production are always released in tandem so disentangling their individual announcement effects is not possible.

Nevertheless, with only intermittent overlap, a simple model can attempt to capture some of what is driving the reports in general to be more volatile than non-report days. This model will include dummy variables for each of the reports except for crop production then interaction terms will be added for each report with the price move. The full slate of controls will be added leading to the final model shown in Equation 16.

$$\begin{aligned}\boldsymbol{RV}_{15,i} &= \boldsymbol{\beta}_0 + \boldsymbol{\beta}_1\boldsymbol{Volume}_i + \boldsymbol{\beta}_2|\boldsymbol{PriceMove}|_{15,i} + \boldsymbol{\beta}_3\boldsymbol{NegPM}_i + \boldsymbol{\gamma}_1\boldsymbol{Report}_i \\ &+ \boldsymbol{\gamma}_2(\boldsymbol{PM}_{15,i} \times \boldsymbol{Report}_i) + \boldsymbol{\beta}_5\boldsymbol{Limit}_i + \boldsymbol{\beta}_6\boldsymbol{Limit}_{i-1} + \boldsymbol{\gamma}_3\boldsymbol{CropYear}_i + \boldsymbol{\varepsilon}_i \quad (16)\end{aligned}$$

In this equation, the majority of the terms are recognized. However the model includes $\boldsymbol{Report}_i$ to represent the matrix of dummies for the WASDE, grain stock, prospective planting, and acreage reports. Also the term $\boldsymbol{PM}_{15,i} \times \boldsymbol{Report}_i$ captures the matrix of interaction terms between each report and the price move. The full model includes the limit move indicators and crop year fixed effects.

Here the γ's represent the coefficients for the interaction matrices. In expectation, the predictions for the individual report should not differ from the report dummy sign and magnitude as well as the interaction terms. Of the reports, I expect the WASDE as the most frequent of the reports, to be the most significant and possibly largest of all the reports. Due to the scarcity of observations, especially for the prospective planting and acreage reports, I expect high standard errors to mitigate any impact they could have.

Table 12 displays the results for the above model. The first column begins with just the individual report indicators. The volume, price move, and negative move variables carry similar signs as the analogous regression without individual reports. As expected the WASDE is significant at the 99% level and about half the magnitude of the previous report indicator implying that some of that coefficients impact was coming from some of the other reports. Surprisingly, the acreage report indicator is an order of magnitude larger than the WASDE and also significant at the 99% level.

The next column presents the inclusion of interaction terms between the each report indicator and the absolute price move. This produces interesting results in that it wipes out the magnitude and significance of the WASDE report indicator. The acreage report is still significant at the highest level and has even increased slightly in magnitude but curiously its interaction term with the absolute price move is insignificant and the opposite sign. This is very contradictory evidence and could very well be a result of the lack of data since the acreage report only has 5 observations.

Only the prospective planting interaction term is significant but no others are. The inclusion of the limit move indicators has not major impact on any other estimates except for washing out the significance of the negative price move indicator which

Table 12: Individual Reports Model with Controls

Dependent Variable for OLS using Newey-West Std. Er.				
$Variable$	$RV_{First15}$	$RV_{First15}$	$RV_{First15}$	$RV_{First15}$
Intercept	0.000738***	0.000750***	0.000723***	0.000252***
	(0.000063)	(0.000063)	(0.000063)	(0.000067)
Volume	-0.000022***	-0.000022***	-0.000021***	-0.000001
Beta Volume	-0.1811	-0.1811	-0.1728	-0.0082
	(0.000007)	(0.000006)	(0.000005)	(0.000006)
|Price Move|	0.021500***	0.018700***	0.018200***	0.011500***
Beta PM	0.2504	0.2178	0.2120	0.1339
	(0.003346)	(0.003301)	(0.003817)	(0.002900)
Neg PM	-0.000053*	-0.000058**	-0.000045	-0.000035
	(0.000030)	(0.000030)	(0.000027)	(0.000027)
WASDE	0.000176***	0.000049	0.000032	0.000009
	(0.000063)	(0.000117)	(0.000120)	(0.000109)
Acreage	0.001722***	0.002050***	0.002027***	0.002030
	(0.000626)	(0.000761)	(0.000767)	(0.001042)
Grain Stock	0.000134	-0.000021	0.000010	-0.000125
	(0.000157)	(0.000143)	(0.000168)	(0.000155)
Prospective P	0.000195	-0.000298	-0.000328	-0.000133
	(0.000249)	(0.000254)	(0.000355)	(0.000362)
WASDExPM		0.015500	0.018000	0.018800
		(0.016200)	(0.016400)	(0.015100)
AcrexPM		-0.011700	-0.008815	-0.004721
		(0.034300)	(0.034300)	(0.040200)
GSxPM		0.005933	0.003647	0.011200
		(0.019200)	(0.020300)	(0.018400)
PPxPM		0.044900**	0.048300	0.034100
		(0.020800)	(0.030400)	(0.028900)
Limit			-0.000009	-0.000133
			(0.000275)	(0.000286)
Lag Limit			0.000614**	0.000351
			(0.000311)	(0.000329)
Crop Years				Yes
R^2	0.2325	0.2406	0.2759	0.3824

Notes: Standard Errors are shown in parentheses. A single (*) denotes statistical significance at the 10%, a double asterisk (**) indicates significance at the 5%, and a triple asterisk (***) means statistical significants at the 1% level.

had crept up in the last regression. The final model with the crop year fix effects is very telling that it completely eliminates any significance in the individual reports or their interaction terms. All that is left is the absolute price move variable significant and at a magnitude comparable to same model without individual reports.

This implies that individually no report is systematically different than one another enough to statistically effect realized volatility once controlling for crop years. Part of the messy estimates and standard errors is likely driven by the lack of distinctness between reports since often a grain stock report comes out with the acreage and prospective planting. Once or twice a year the two also overlap with the WASDE and crop production. In conclusions it seems that try to break out individual reports is asking for more detail than the data can actually provide since there is no way to decipher which price movements are due to one report or another on the same day.

Event Window Model

In my final model I break out the individual days of the event window. In this manner I can analyze the event window similarly to the tabular analysis in previous sections but with greater ability to control for confounding factors. This means that indicator variables are added for each day of the event window so individual coefficient can be estimated for it in addition to the report date. I begin with the two original controls for volume and absolute price move, making the analysis similar to the first but with the addition of the event window dummies. Then the limit move and crop year fixed effects are added.

The individual report and interaction with price move analysis is not added to prevent the tedious study of variables of no interest like the interaction of a grain stock report the - 5 day of the even window. Furthermore, the loss of power and degrees of freedom would only be exacerbated by the lack of observations for particular reports. Sticking to a simple framework, the new model with all the controls is displayed in Equation 17.

$$\begin{aligned}\boldsymbol{RV}_{15,i} = \boldsymbol{\beta}_0 + \boldsymbol{\beta}_1 \boldsymbol{Volume}_i + \boldsymbol{\beta}_2 |\boldsymbol{PriceMove}|_{15,i} + \boldsymbol{\beta}_3 \boldsymbol{Limit}_i + \boldsymbol{\beta}_4 \boldsymbol{Limit}_{i-1} \\ + \boldsymbol{\gamma}_1 \boldsymbol{EVENT}_i + \boldsymbol{\gamma}_2 \boldsymbol{CropYear}_i + \boldsymbol{\varepsilon}_i \quad (17)\end{aligned}$$

In this equation, $\boldsymbol{EVENT}_i$ represents a matrix of dummy variables for 10 days of the event window excluding day - 1, the date before a USDA report. This way, all event window day estimate coefficients, represented by the matrix $\boldsymbol{\gamma}_1$ are calculated relative to the day before a report like the previous tabular analysis. Consistent with previous analysis, a positive coefficient is expected for each day of the event window with the largest and most significant for the report day.

I expect similar signs and magnitudes for the price move and volume variables. Theory predicts a positive sign for both $\boldsymbol{Limit}$ variables because clearly these days are likely to have more trading however this could be washed out by days in which the limit was too restrictive leading to minimal trading. Finally, I expect a positive sign for most if not all of the crop year dummies since level graphs show significant price variability in years early in the sample compared to the most recent. That matrix of dummies is represented by $\boldsymbol{CropYear}_i$ with a matrix of estimates similar to the event window, $\boldsymbol{\gamma}_2$.

Table 13 presents the incremental results for Equation 17. The first column contains the simplest form of this regression using only the price move and volume controls but adding the even window matrix instead of merely the report day indicator. The estimated coefficients for this equation are mostly as predicted. The price move and volume coefficients barely change which is expected since all that has been done is parsing out the coefficient of the report day into its various parts. Most of the days in the event window are positive but only the report day and the day after the report are the largest. Days -2, -3, and -5 turn out to have negative but small and insignificant coefficients meaning that RV is slightly lower on those days relative to the day before a report.

The report date is still significant at the 99% and almost the same magnitude as previous regressions. It is deceptive to claim it is identical although because Table 13 present coefficients relative to the day before any report rather than merely report relative to non-report days. Hence, the coefficient does not have the same meaning. Interestingly, the day after a report is significant at 95% and positive showing a residual impact on realized volatility the day following a USDA report release.

The second regression includes the price limit indicator and the limit indicator for the previous trading day. Neither variables are significant and have no major impact on any other estimates, suggesting that price limits do not influence intra-minute price volatility. The final column includes all previous controls and controls for all crop years relative the most recent, 2012-2013 crop year. While not printed all crop year dummies are significant at the 99% level. Their addition makes noticeable adjustments to several of the variables of interest.

The inclusion of crop year fixed effects completely eliminates the significance of the volume coefficient, suggesting that all impact on RV due to the level of volume is instead driven by attributes of seasonality. The absolute price move and report date coefficients remain nearly identical and of equal significance. The crop years are important controls since they pick up the volatility due to the food crisis of 2007 and 2008 previously mentioned, suggesting that the results are not driven by extraneous market supply and demand factors.

Table 13: Event Window Model with Controls

Dependent Variable for OLS using Newey-West Std. Er.			
Variable	$RV_{First15}$	$RV_{First15}$	$RV_{First15}$
Volume	-0.224***	-0.215***	-0.087
	(0.047)	(0.046)	(0.056)
|Price Move|	0.299***	0.296***	0.256***
	(0.046)	(0.048)	(0.045)
-5	-0.143**	-0.157**	-0.417**
	(0.063)	(0.065)	(0.087)
-4	-0.072	-0.075	-0.348***
	(0.084)	(0.090)	(0.095)
-3	-0.196**	-0.242***	-0.483***
	(0.061)	(0.063)	(0.082)
-2	-0.132*	-0.148**	-0.423***
	(0.073)	(0.075)	(0.086)
Report Date	0.534***	0.510***	0.226
	(0.165)	(0.157)	(0.170)
+1	0.201*	0.108	-0.148
	(0.124)	(0.114)	(0.121)
+2	0.093	0.045	-0.216
	(0.157)	(0.128)	(0.136)
+3	-0.058	-0.073	-0.334***
	(0.112)	(0.110)	(0.111)
+4	-0.018	-0.019	-0.270***
	(0.084)	(0.087)	(0.099)
+5	-0.126	-0.156*	-0.379***
	(0.079)	(0.080)	(0.095)
Limit	No	0.078	-0.214
		(0.629)	(0.685)
Lag Limit	No	1.165*	0.674
		(0.662)	(0.696)
Crop Years	No	No	Yes
R^2	0.175	0.207	0.314

Notes: Standard Errors are shown in parentheses. A single (*) denotes statistical significance at the 10%, a double asterisk (**) indicates significance at the 5%, and a triple asterisk (***) means statistical significants at the 1% level.

CONCLUSION AND DISCUSSION

This work uses realized volatility of transaction-by-transaction Kansas City Board of Trade's Hard Red Winter wheat futures data to test the hypothesis that the release of relevant USDA crop reports reduces uncertainty in the futures market. This hypothesis was developed and tested by McNew and Espinosa (1994) and Isengildina-Massa et al. (2008b) using implied volatility from options pricing models. Both of these papers show strong evidence of reduced futures price volatility following the release of relevant USDA crop reports.

I consider this hypothesis by stating that the supply and demand information contained in USDA reports, especially the WASDE, could help equilibrate beliefs about the fundamental value of wheat contracts leading to price consensus observed by a measurable reduction in the intraday spread of prices following the information's release. The primary contribution is the use of a new tick data set of HRW wheat contracts from the KCBT website from which volatility can be calculated from observed prices rather than backed out of options pricing models.

Computing RV from actual transaction prices allows me to investigate the intraday effects of USDA information. Breaking the day into smaller intervals of time I estimate RV to try to identify if price consensus is reached by measuring the magnitude of RV throughout the day. The results of various tests resoundingly reject this hypothesis and instead suggest that RV is, in fact, higher following report releases than lower. I perform robustness checks to ensure that the increase in RV is not a result of mechanical process of calculating volatility, institutional trading limits, or seasonality. These regressions confirm that RV is still higher on USDA report release dates than days before or after.

Frictions in the trading process, called microstructure noise, are prevalent in high frequency data like that employed in this study. This means that volatility estimates contain both the variability in the fundamental value and microstructure noise. At the highest level of frequency, which I use for much of the analysis, previous research indicates that this noise component is prominent meaning that at no point can I make any claims about the volatility of fundamental values, which are unobservable.

The results are compelling because even if fundamental price variability can not be distinguished, the noise term is the result of behavior in the market. If this noise term increases or changes then clearly behavior is changing in the market, likely as a result of the new information. Although I cannot parse whether the variation observed is due to fundamental value change, microstructure noise, or both, it is clear that I have identified a salient feature of observed wheat futures volatility - the variability of intraday wheat prices does not decrease following the release of USDA crop information.

As I proposed the hypothesis, an increase in RV suggests an increase in uncertainty following new crop information. This result is in direct contrast to the literature that has studied this same hypothesis. Nevertheless, the results provided by RV raise serious questions. Are realized and implied volatility actually measuring the same thing? If not, then what is realized volatility capturing that is different than implied volatility?

Some of the basic differences between Implied volatility and RV can easily be identified. The RV estimates used in this paper provide a snap shot look in time at the tangled variability of fundamental wheat futures value and microstructure noise. It provides an instantaneous view of the distribution of prices. Implied volatility on the other hand is more forward looking, giving an expected variability of the underlying futures price over the life of an option for that future. Implied volatility is predictive given a set of parameters in contrast to the hardened snap shot nature of RV.

The striking difference in results matched with the difference in underlying computation suggests that realized volatility is not a measure of uncertainty but instead provides a glance at the spread of expected prices. While implied volatility is forward looking and thus a better measure of forthcoming price risk, RV measures rather the spread of disparate expectations at the time of the report. This dispersion of expectations is the same in essence as the difference of opinion previously discuss. Hence it seems reasonable that RV is instead a measure of differences of opinions in the market.

With such a short lived impact the question arises as to why this matters. First, the quickness with which the market reacts to the new information suggests a market that is very efficient, meaning prices in the wheat futures market reflect fundamentals accurately. Secondly, this means that the wheat futures market equilibrates quickly to new information. The time that these differences of opinion are shorted out is similar to the moving of a supply or demand schedule to new information. Thus wheat futures prices quickly return to long run equilibrium preventing disequilibrium markets leading to resource missallocation.

One would expect implied volatility and RV to be correlated to some degree, however the mechanical nature of the Black-Scholes option pricing model means that prices can hypothetically change a lot and not move implied volatility in the same direction. Thus it is entirely possible that RV is increases while IV is decreases. This raises another important question on which one can only speculate.

Is there some feature of USDA information that drives up volatility? I hazarded a guess at this possible result when discussing the difference in opinion literature. It seems plausible that traders develop private heterogeneous beliefs about the underlying value of wheat prior to any USDA publication. Following a crop report release, the authoritative nature of the report may challenge traders expectations prompting them to agree to disagree on the interpretation of the data leading to an increase in the distribution of prices as they trade to profit on their disparate opinions. The data suggests that if this occurs the process happens quickly, but never reduced volatility below its original level.

Popular criticism may implicate high-frequency trading as the culprit of increased volatility in the wheat market. With such a hypothesis one would expect to see activity often called "banging the beehive" around the release of government information. In such a case, high frequency traders attempt overwhelm the market with orders prior to the information release in order to move the price in the opposite direction of the anticipated price move, as described in the the Wall Street Journal article on the natural gas futures market (Dicolo and Rogow, 2012). By creating larger price spreads they can then make greater profits. In the case of the wheat market these HFT might attempt to use the frictions of the market to make profits. However, results suggesting pre-release market moves are never observed. Furthermore, the presents of HFT would not explain the sudden decline and return to normal volatility.

What is most clear from this analysis is that a simplistic model of information fails to describes how the market reacts to and incorporates information. The analysis using RV confirms that more nuanced tools are needed in order to continue to further understand how price discovery manifests itself in the market.

REFERENCES CITED

Adjemian, M.K. 2012. “Quantifying the WASDE Announcement Effect.” *American Journal of Agricultural Economics* 94:238–256.

Allen, R. 2007. “Safeguarding America’s Agricultural Statistics: A Century of Successful and Secure Procedures, 1905-2005.” *U.S. Department of Agriculture, National Agricultural Statistics Service*, pp. 1–60.

Andersen, T.G., and T. Bollerslev. 1998. “Answering the Skeptics: Yes, Standard Volatility Models do Provide Accurate Forecasts.” *International Economic Review* 39:885–905.

—. 1997. “Intraday Periodicity and Volatility Persistence in Financial Markets.” *Journal of Empirical Finance* 4:115–158.

Andersen, T.G., T. Bollerslev, F.X. Diebold, and H. Ebens. 2001. “The Distribution of Realized Stock Return Volatility.” *Journal of Financial Economics* 61:43–76.

Andersen, T.G., T. Bollerslev, F.X. Diebold, and P. Labys. 2000a. “Exchange Rate Returns Standardized by Realized Volatility are (Nearly) Gaussian.” *National Bureau of Economic Research Working Paper Series* No. 7488.

—. 2000b. “Great Realizations.” *Risk* 13:105–108.

—. 2003. “Modeling and Forecasting Realized Volatility.” *Econometrica* 71:579–625.

Arrow, K. 1962. “Economic Welfare and the Allocation of Resources for Invention.” In *The Rate and Direction of Inventive Activity: Economic and Social Factors*. Princeton University Press, pp. 609–626.

Bai, X., J.R. Russell, and G.C. Tiao. 2001. “Beyond Merton’s Utopia (I): Effects of Non-Normality and Dependence on the Precision of Variance Estimates Using High-Frequency Financial Data.” Unpublished, Working Paper, University of Chicago.

Bandi, F.M., and J.R. Russell. 2008. “Microstructure Noise, Realized Variance, and Optimal Sampling.” *The Review of Economic Studies* 75:339–369.

—. 2005. “Realized Covariation, Realized Data, and Microstructure Noise.” Unpublished, Working Paper, University of Chicago.

Bandi, F.M., J.R. Russell, and Y. Zhu. 2008. “Using High-Frequency Data in Dynamic Portfolio Choice.” *Econometric Reviews* 27:163–198.

Banerjee, S. 2011. “Learning from Prices and the Dispersion in Beliefs.” *Review of Financial Studies*, pp. 3025–3068.

Banerjee, S., R. Kaniel, and I. Kremer. 2009. "Price Drift as an Outcome of Differences in Higher-Order Beliefs." *The Review of Financial Studies* 22:3707–3734.

Banerjee, S., and I. Kremer. 2010. "Disagreement and Learning: Dynamic Patterns of Trade." *The Journal of Finance* 65:1269–1302.

Barndorff-Nielsen, O.E., and N. Shephard. 2004. "Econometric Analysis of Realized Covariation: High Frequency Based Covariance, Regression, and Correlation in Financial Economics." *Econometrica* 72:885–925.

Bring, J. 1994. "How to Standardize Regression Coefficients." *The American Statistician* 48:pp. 209–213.

Carter, C.A. 2012. *Futures and Options Markets: An Introduction*. Davis, CA: RebelText.

Colling, P.L., and S.H. Irwin. 1990. "The Reaction of Live Hog Futures Prices to USDA Hogs and Pigs Reports." *American Journal of Agricultural Economics* 72:84–94.

de Pooter, M., M. Martens, and D. van Dijk. 2008. "Predicting the Daily Covariance Matrix for S&P 100 Stocks Using Intraday Data - But Which Frequency to Use?" *Econometric Reviews* 27:199–229.

Dicolo, J.A., and G. Rogow. 2012. *Wall Street Journal*. : "Gas Market Stung by Rapid Traders", retrieved online: www.wjs.com, October, 30th.

Easley, D., and M. O'Hara. 1992. "Time and the Process of Security Price Adjustment." *The Journal of Finance* 47:577–605.

Ederington, L.H., and J.H. Lee. 1995. "The Short-Run Dynamics of the Price Adjustment to New Information." *Journal of Financial and Quantitative Analysis* 30:117–134.

Egelkraut, T.M., P. Garcia, S.H. Irwin, and D.L. Good. 2003. "An Evaluation of Crop Forecast Accuracy for Corn and Soybeans: USDA and Private Information Services." *Journal of Agricultural and Applied Economics* 35:79–95.

Engle, R.F. 2000. "The Econometrics of Ultra-High-Frequency Data." *Econometrica* 68:1–22.

Epps, T.W. 1979. "Comovements in Stock Prices in the Very Short Run." *Journal of the American Statistical Association* 74:291–298.

Falk, B., and P.F. Orazem. 1985. "A Theory of Future's Market Responses to Government Crop Forecasts." Unpublished, Staff Paper Series No. 150, Department of Economics, Iowa State Univseriity.

Fama, E.F. 1970. "Efficient Capital Markets: A Review of Theory and Empirical Work." *The Journal of Finance* 25:383–417.

—. 1991. "Efficient Capital Markets: II." *The Journal of Finance* 46:1575–1617.

Fishe, R.P., J.P. Janzen, and A. Smith. 2014. "Hedging and Speculative Trading in Agricultural Futures Markets." *American Journal of Agricultural Economics* 96:542–556.

Fortenberry, T., and D. Sumner. 1993. "The Effects of USDA Reports on Futures and Options Markets." *Journal of Futures Markets* 13:157–173.

French, K.R., and R. Roll. 1986. "Stock Return Variances: The Arrival of Information and the Reaction of Traders." *Journal of Financial Economics* 17:5–26.

Friedrich, R.J. 1982. "In Defense of Multiplicative Terms in Multiple Regression Equations." *American Journal of Political Science* 26:pp. 797–833.

Garcia, P., S.H. Irwin, R.M. Leuthold, and L. Yang. 1997. "The Value of Public Information in Commodity Futures Markets." *Journal of Economic Behavior & Organization* 32:559–570.

Good, D.L., and S.H. Irwin. 2006. "Understanding USDA Corn and Soybean Production Forecasts: Methods, Performance and Market Impacts over 1970-2004." Department of Agricultural and Consumer Economics University of Illinois Urbana-Champaign.

Goodhart, C.A., and M. O'Hara. 1997. "High Frequency Data in Financial Markets: Issues and Applications." *Journal of Empirical Finance* 4:73–114.

Grossman, S.J., and J.E. Stiglitz. 1980. "On the Impossibility of Informationally Efficient Markets." *The American Economic Review* 70:393–408.

Hansen, P.R., and A. Lunde. 2006. "Realized Variance and Market Microstructure Noise." *Journal of Business & Economic Statistics* 24:127–161.

Hasbrouck, J. 1993. "Assessing the Quality of a Security Market: A New Approach to Transaction-Cost Measurement." *Review of Financial Studies* 6:191–212.

—. 2007. *Emprical Market Microstructure: The Insitutions, Economics, and Econometrics of Securities Trading*. New York: Oxford University Press.

Hayek, F.A. 1945. "The Use of Knowledge in Society." *The American Economic Review* 35:519–530.

Hirshleifer, J. 1971. "The Private and Social Value of Information and the Reward to Inventive Activity." *The American Economic Review* 61:561–574.

Isengildina-Massa, O., S.H. Irwin, and D.L. Good. 2006. “The Value of USDA Situation and Outlook Information in Hog and Cattle Markets.” *Journal of Agricultural and Resource Economics* 61:262–282.

Isengildina-Massa, O., S.H. Irwin, D.L. Good, and J.K. Gomez. 2008a. “The Impact of Situation and Outlook Information in Corn and Soybean Futures Markets: Evidence from WASDE Reports.” *Journal of Agricultural and Applied Economics,* 40:89–103.

—. 2008b. “The Impact of WASDE Reports on Implied Volatility in Corn and Soybean Markets.” *Agribusiness* 24:473–490.

Janzen, J.P., A.D. Smith, and C.A. Carter. 2014. “The Quality of Price Discovery Under Electronic Trading: The Case of Cotton Futures.” Unpublished, Working Paper, U.C. Davis.

Just, R.E. 1983. “The Impact of Less Data on the Agricultural Economy and Society.” *American Journal of Agricultural Economics* 65:872–881.

KCBT. 2015. “Kansas City Board of Trade Archived Website: http://web.archive.org/web/20080315044939/http://www.kcbt.com/about_kcbt.htm.”

Lehecka, G.V. 2014. “The Value of USDA Crop Progress and Condition Information: Reactions of Corn and Soybean Futures Markets.” *Journal of Agricultural and Resource Economics* 39:88–105.

Lehecka, G.V., X. Wang, and P. Garcia. 2014. “Gone in Ten Minutes: Intraday Evidence of Announcement Effects in the Electronic Corn Futures Market.” *Applied Economic Perspectives and Policy* 36:504–526.

Liu, L.Y., A.J. Patton, and K. Sheppard. 2012. “Does Anything Beat 5-minute RV? A Comparison of Realized Measures Across Multiple Asset Classes.” Unpublished, Working Paper, Duke University.

MacKinlay, A.C. 1997. “Event Studies in Economics and Finance.” *Journal of Economic Literature* 35:13–39.

Madhavan, A. 2000. “Market Microstructure: A Survey.” *Journal of Financial Markets* 3:205–258.

Malinowska, M. 2010. “Price and Trading Response to Public Information.” Unpublished, Working Paper Series No. 1177, European Central Bank.

Marone, H. 2008. “How do Wheat Prices React to USDA Reports.” Unpublished, UNDP/ODS Working paper, United Nations Development Programme, Office of Development Studies, New York.

McAleer, M., and M.C. Medeiros. 2008. “Realized Volatility: A Review.” *Econometric Reviews* 27:10–45.

McNew, K.P., and J.A. Espinosa. 1994. “The Informational Content of USDA Crop Reports: Impacts on Uncertainty and Expectations in Grain Futures Markets.” *Journal of Futures Markets* 14:475 – 492.

Miller, S. 1979. “The Response of Futures Prices to New Market Information: The Case of Live Hogs.” *Southern Journal of Agricultural Economics* 2:67–70.

Morris, S., and H.S. Shin. 2002. “Social Value of Public Information.” *The American Economic Review* 95:1521–1534.

Moschini, G., and D.A. Hennessy. 2001. “Uncertainty, Risk Aversion, and Risk Management for Agricultural Producers.” *Handbook of Agricultural Economics* 1:88–153.

Oomen, R.C. 2005. “Properties of Bias-Corrected Realized Variance Under Alternative Sampling Schemes.” *Journal of Financial Econometrics* 3:555–577.

Pagel, I., P. De Jongh, and J. Venter. 2007. “An Introduction to Realized Volatility.” *Investment Analysts Journal* 65:47–57.

Russell, J., and R. Engle. 2010. “Analysis of High Frequency Financial Data.” In Y. Ait-Sahalia and L. Peter, eds. *Handbook of Financial Econometrics*. Elsevier, vol. Volume 1 - Tools and Techniques, pp. 384–424.

Sumner, D.A., and R.A.E. Mueller. 1989. “Are Harvest Forecasts News? USDA Announcements and Futures Market Reactions.” *American Journal of Agricultural Economics* 71:1–8.

USDA. 2014. “Economic Research Service: Wheat Overview Website http://www.ers.usda.gov/topics/crops/wheat.aspx.”

Vogel, F.A., and G.A. Bange. 1999. “Understanding USDA Crop Forescasts.” U.S. Department of Agriculture, National Agricultural Statistics Service and Office of the Chief Economist, World Agricultural Outlook Board.

Wasserfallen, W., and H. Zimmermann. 1985. “The Behavior of Intra-Daily Exchange Rates.” *Journal of Banking and Finance* 9:55–72.

Williams, J.C. 2001. “Chapter 13 Commodity futures and options.” In B. L. Gardner and G. C. Rausser, eds. *Marketing, Distribution and Consumers*. Elsevier, vol. 1, Part B of *Handbook of Agricultural Economics*, pp. 745 – 816.

Zhang, L., P.A. Mykland, and Y. Ait-Sahalia. 2011. “Edgeworth Expansions for Realized Volatility and Related Estimators.” *Journal of Econometrics* 160:190–203.

Zhou, B. 1996. "High-Frequency Data and Volatility in Foreign-Exchange Rates." *Journal of Business & Economic Statistics* 14:45–52.

Zulauf, C.R., and S.H. Irwin. 1998. "Market Efficiency and Marketing to Enhance Income of Crop Producers." *Review of Agricultural Economics* 20:308–331.