Does Speculation Make Commodity Spot Markets More Volatile?

– An empirical approach testing the Masters hypothesis with new evidence

Senior Thesis

Yixiao (Carl) Cao

Vassar College, Department of Economics

May 2022

Abstract

This paper examines how increased speculator participation in the commodity futures market affects spot price volatility. Consistent with results from the majority of existing literature on this subject, I find no evidence that speculators destabilize the commodity spot market. More importantly, I include data from 2012 to 2021, a period that is rarely covered or discussed by existing work, and compare the result with that obtained from pre-2012 data. My findings show that as speculator share across commodity markets increases to a certain extent, they become less stabilizing compared to pre-2012 periods. My analysis suggests that while speculators continue to reduce spot price volatility overall in the status quo, their excessive share can potentially accentuate spot price movements in some cases. This implies that the well-known Masters hypothesis, which recommends an ideal range of speculator share, could be valid under certain circumstances.

Acknowledgements

I would like to thank my thesis advisor Professor Sarah Pearlman for her help, guidance, and feedback throughout this year-long effort, my academic advisor Professor Esteban Argudo for embarking me on the exciting and rewarding journey of economics research and equipping me with essential skills in data science, Professor Paul Johnson for his helpful consultation on the econometric details in this paper, as well as the entire Economics faculty at Vassar College for pushing me to grow intellectually the whole time.

I would also like to thank my parents for their support to my academic and career pursuits, as well as my housemates Yuchen, Kailai, and Weining for the friendship and wonderful time spent together throughout my senior year.

1 Introduction

The popular assertion that derivatives tend to destabilize the underlying primary markets has persisted for more than three centuries (Mayhew, 2000). For instance, Lowenstein (1997) commented in the Wall Street Journal after a market decline triggered by the 1997 market mini-crash (known as the “Asian contagion” or the “Tom Yum Goong crisis”) that:

> When the Nobel Prize was awarded to the inventors of the formula for pricing stock options, the formula’s practical utility was widely noted. Last Monday, when the market cracked, we saw how practical indeed.
>
> To start at the conclusion, the vast growth in the market for options unleashed tremendous selling pressure in the stock market and accentuated Monday’s decline. Previously, the very same options had encouraged stock prices to rise above the level justified by business valuations.

In short, derivatives have been blamed for guiding asset prices away from their intrinsic values and accentuating volatile movements in either direction. But to what extent is this notion true, and to what extent has it been supported by empirical evidence? This question has attracted considerable academic scrutiny, particularly after financial crises and recessions when prices of financial assets experienced high volatility and drastic fluctuations. The purpose of this paper is to present a comprehensive review of the theoretical and empirical literature on this issue and test similar hypotheses by constructing empirical models on more recent periods, especially after the year of 2012, which most of the existing literature was unable to capture. Furthermore, this paper primarily focuses on the relationship between commodity futures speculation and commodity spot price. The results suggest that futures speculation continues to stabilize the spot market overall, but to a lesser extent more recently as speculator participation continues to soar across commodity markets. In some rare instances, speculation starts to be destabilizing. Towards the end of this paper, I discuss some possible explanations for the underlying mechanism of the results as well as some potential directions of future work.

While it's overall important to study the extent to which various types of financial derivatives such as stock options impact their underlying assets, there are at least three characteristics of the commodity futures market that make it distinctively interesting. First, commodities are essential to our daily lives. They are continuously produced and consumed on a massive scale, closely tied to day-to-day agricultural and industrial production, and therefore can influence almost every aspect of our society. For this reason, stability in commodity prices is particularly important: large fluctuations and high volatility in commodity prices can cause tremendous disruptions to industrial, agricultural, and societal functionalities. In the most extreme cases, instability in price might lead to detrimental outcomes such as power shortage, famine, supply chain crises, and so on. In contrast, this is usually not the case in other primary and derivative markets such as stocks and their corresponding option markets: highly disruptive systematic shocks in the stock market are generally rare and are typically blamed for factors such as macroeconomic conditions, wars, and pandemics, not options trading. Furthermore, idiosyncratic shocks on individual companies can have a relatively trivial and indirect impact on the overall functionality of our society compared to commodities. This is primarily why a non-finance person should also care about commodity futures. Second, commodity futures are complex instruments: their prices are sensitive to a wide range of fundamental factors including production level, consumption behavior, seasonality, constraints in natural resources, governmental policies, and international relations, and thus it is a challenging and rewarding task for scholars and financial researchers to study the interaction between these factors. In particular, two unique characteristics of commodities are worth noting: storage and futures basis. Since the production and consumption of commodities are highly dynamic, peculation in commodity markets can be achieved in the form of storage. If the product is storable, as most commodities are, inventories and expectations on future shortage or surplus can affect the demand function of a particular commodity by adding the factor of speculative demand, driving anticipations of price to increase or decrease. This is not the case for the stock or options market, as the total amount of stock shares is relatively fixed without new issues, and "hoarding" stocks is rarely meaningful or feasible. The determination of stock prices and their corresponding option prices is predominately derived from financial analyses and the industry outlook of individual companies. In addition, the price gap between the futures and the spot markets, measured by the futures basis (spot price minus future price), provides a linkage between the two markets and helps investors decide the best time to buy and sell and hedgers decide when to hedge a purchase or sale. The consideration of historical and seasonal bases adds great complexities to the analysis of commodity prices, while the stock market usually does not exhibit this feature. For these reasons, stock pricing is greatly different from, and arguably less complex than, how commodity prices are determined. Third, commodity futures are uniquely important to producers and consumers of the underlying commodity good. Access to the futures market allows different parts of the supply chain to manage their risks by taking a futures position and locking a desirable price. Futures price also directly influences the behavior of different parties. With the introduction of the futures market, a producer might change how much to produce and what production techniques to use (Mayhew, 2000). Similarly, merchants and end-users might adjust their consumption patterns according to the futures market dynamics. For example, observing a spike in futures price of corn, large-scale, long-term purchasers such as chain restaurants might adjust their menus and look for substitute grains that are potentially cheaper. For all of these reasons, this paper chooses to delve deeper into commodity futures as opposed to other types of derivatives.

There are generally two types of participants in the commodity futures market: speculators and hedgers. Speculators are individuals and firms who take risk in order to make profit from price movements. According to the Chicago Mercantile Exchange (CME) group, speculators in the futures market are broadly comprised of individual traders, proprietary trading firms ("prop shops"), portfolio managers, hedge funds, and market makers. They have little interest in the underlying physical commodity and typically buy and sell the futures contracts as a financial product. On the other hand, hedgers are individuals and firms who buy and sell the actual physical, deliverable commodity. They can be producers, wholesalers, retailers or manufacturers who have an invested interest in the underlying asset and whose businesses are affected by changes in commodity prices. The primary motive for hedgers to participate in futures trading is to manage and offset risk, as they can "lock" the price at which they wish to buy or sell their physical commodities in case of unexpected price shocks.

The role of commodity futures speculation has been widely debated. The traditional theory in defence of speculation by Friedman (1953) argues that profitable speculation must involve buying when price is low and selling when price is high (both can happen at any time in the case of commodity futures when traders can take both long and short positions). The arbitrage effect causes prices in commodity markets to converge to their fundamentals, thus speculation eliminates mispricing and contributes to price stabilization. Furthermore, it is commonly believed that speculators provide market liquidity, as they facilitate trade between all market participants, and providing liquidity is a crucial market function that enables individuals to easily enter or exit the market. This is particularly important in commodity markets, where the number of hedgers is limited and trading between them is generally inactive, speculators serve as risk takers whom the hedgers can trade with and unload their risks to.

The concern about speculative trading, on the other hand, finds support in theoretical models where noise traders, speculative bubbles, and herding can drive prices away from fundamental values and thus destabilize markets (Brunetti, Buyuksahin, & Harris, 2016).

The perception that speculation drives commodity price fluctuations has led to policy proposals aiming to stabilize the commodity market by restricting speculative trading in the futures market. The most well-known proponent of such restriction, hedge fund manager Michael W. Masters, claims that excessive speculation leads to excessive volume and volatility, thus an aggregate speculative position limit should be imposed. Masters argues that the ideal range of speculator share should be $25\% - 35\%$ of total open interest (Masters, 2009). The proposed range was consistent with our data before 2012 (see Table 6). Between 1992 to 2012, all commodity markets within the scope of this paper had speculators sharing less than 35% of the market, and markets such as sugar and cocoa were to some extent "under-speculated", having less than 25% of the market being speculators. However, speculator's market share has grown tremendously in the past decade and has exceeded almost double of Masters' proposed range in 2021 (see Table 6 and Figure 1). From Figure 1, we see that the increase of the share is rather continuous, without identifiable discrete jumps or spikes. When we break down to individual commodities, a typical example of such increase is crude oil, as illustrated by Figure 2.

The fright of large scale price instability in commodity markets has propelled regulations targeting speculations in the futures market: the governing bodies of different commodity futures markets across the financial world have implemented policies such as speculation limits to manage the potential of "over speculation". However, there has been tremendous difficulty in reaching a consensus of how speculation should be regulated and how much speculation is "excessive". More recently, in March 2021, the U.S Commodity Futures Trading Commission (CFTC) amended its rules ("Final Rule"), limiting speculative positions that market participants may take in certain commodity contracts, especially during the delivery month. The Commission argues that the limits are imposed in order to protect futures markets from excessive speculation that can cause unreasonable or unwarranted price fluctuations. Such regulatory choice makes the research topic of this paper highly relevant and meaningful: is it really necessary to regulate speculation if they weren't so harmful to price stability? An inquiry of the relationship between speculation and price stability (and possibly other market qualities such as efficiency and liquidity) is evidently critical to this policy-making process.

This paper is motivated by the large and constant growth of speculator's participation in the commodity futures market and the revitalizing interest in futures market regulations targeting on speculations. Specifically, what is driving the growth and what is the impact of such growth? Relating back to Friedman's proposition, profitable speculation stabilizes the market, and unprofitable speculation will be driven out in the long run. The continuous upward trend of speculator's share over a long time frame implies that speculation in commodity futures market has been overall profitable, and profitable speculation should continue to contribute to stability and liquidity of commodity markets. This paper examines whether this is the case, or whether such relationship has changed in today's context.

2 Literature Review

2.1 Theory on Speculation

A variety of theoretical arguments have been advanced over the years to explain why speculative trading in the derivatives market can influence price stability in the underlying primary market. Friedman (1953)'s suggestion that profitable speculation ought to stabilize prices led to a wave of research that explores the contrary, that is, how speculative trading can be simultaneously profitable and destabilizing. For example, Baumol (1957) provides a formal model in which speculative trading is destabilizing. The model assumes predictable seasonal fluctuations in price where the introduction of speculation accelerates price movement. In contrary to Friedman (1953)'s proposition, Baumol (1957) argues that speculators, even though they might be profitable, cannot foretell the price of the asset, and therefore only buy the asset after the bottom and sell the asset after the peak, contributing to the upswing or downswing of the price. The model captures the popular intuition that speculators are momentum traders, that is, they follow the trends of the price movements and amplify the magnitude of the volatility. However, the model has a few major drawbacks: there's no random component to prices, speculative demand is not derived from utility maximization, and stability is measured by frequency and amplitude of oscillations instead of variance (Lowenstein, 1997).

Deeming Baumol (1957)'s theory as unsatisfactory, Telser (1959) constructs a general model with a profit-maximizing monopolist speculator, showing that speculation unambiguously stabilizes prices. The model measures stability by the variance and assumes the speculator's foresight is imperfect. The solution to the model suggests that the stabilizing effect exists regardless of whether the speculation is profitable. When the speculator realizes their maximum profit, the presence of the speculator reduces the variance of price to a quarter of the variance in absence of speculation.

To the extent that speculators can inaccurately predict price, their behavior following a false prediction can also drive market fluctuation. Hart and Kreps (1986) consider a market with constant supply where demand is subject to large but rare shocks. There always will be prior signals for high demand periods, and sometimes will be signals for low demand periods. In so far as high demand periods are very rare, the signals are still mostly wrong. When speculators enter the market at a positive signal by buying underlying assets in anticipation of a positive demand shock, they would be stabilizing the market only if the signal is accurate. Therefore, in most occasions, price of the subsequent period depresses even more than they would have been, as speculators dump their shares after their chance of price appreciation fades.

The various models discussed above are mostly theories on speculation in general, with little or no contextualization in any specific asset or derivative market. Although some of these papers derived their models in the context of commodity futures, the models' settings and formulations are not predicated on any inherent characteristics of commodity futures or any other financial derivatives. The theories are rather inconclusive about whether speculative trading stabilizes or destabilizes price, and the competing arguments mainly diverge from the assumptions about speculator behavior, i.e. when and how would a speculator enter and exit the market. Therefore, it is important to provide an accurate characterization of speculator behavior in the market, and it would reasonable to believe that the conclusion might differ across various types of markets. Therefore, the below sections will discuss more recent work on speculation in derivatives markets in general and speculation in commodity futures in particular.

2.2 Speculation in Derivatives Market

Derivatives, as the quip goes, "are too complicated to explain but too important to ignore". Since the financialization period when a great variety of derivatives such as options and futures were introduced to the market on an unprecedented scale, there has been a wide range of literature studying their effects. Danthine (1978) argues that the introduction of futures trading improves the depth of the spot equity market and reduces its volatility because the cost to informed traders responding to mis-pricing will be reduced. In contrast, Stein (1987) finds that introducing more speculators into the market for a given commodity inflicts an externality on those traders already in the market and affects their ability to make inferences based on current prices. Sometimes the externality is negative and lowers the informativeness of the price, leading to destablization and welfare reduction. The above competing theories mainly treat informativeness of price as a mechanism towards stability, and evaluates the effect of speculators on information.

There has also been a great amount of empirical work testing the effect of derivatives on the cash market. Figlewski (1981) finds that trading in the Governmental National Mortage Association (GNMA) futures market has increased volatility of the cash market. The hypothesized mechanism is similar to that proposed by Stein (1987) that the existence of the futures market has led to trading by new classes of investors who do not have access to as good information as traders in the existing cash market. The story behind the phenomenon that less informed traders are entering the derivatives market can be related to the varying barriers to entry across each market, and it is therefore important to view and analyze such phenomenon in different contexts. Harris (1989) finds a statistically significant increase in the volatility of S&P 500 stocks after the introduction of trading in S&P index futures and options. While the paper finds an empirical relationship between the introduction of trading in index derivative and volatility in the cash stock market, there has been no clear causation identified. Another study on the stock index derivatives but in a different context is Wang, Li, and Cheng (2008), who find that the introduction of Hong Kong Hang Seng Chinese Enterprise Stock Index (H-share Index) futures induces additional speculative trading in the underlying equities and leads to an increase in volatility. However, the subsequent introduction of H-share index options increased the level of informed trading and opens up opportunities for arbitrage activities using futures against options, thus leading to a significant decline in spot market volatility as futures and options trading grows more efficient. Focusing on the impact of trading in index options on the underlying assets, Chatrath, Ramchander, and Song (1995) find that an increase in the intraday and interday volatility in the cash market is followed by an increase in the level of options trading, and an increase in options trading is followed by a decline in cash market variability. The findings are overall consistent with the notion that options trading has a stabilizing influence on cash markets. In comparison to the wide range of studies on derivatives for indices, bonds, and commodities, the research on individual stock options has been relatively rare since those usually are evaluated on a case by case basis.

While the literature on speculation in derivatives markets provides us some empirical results and a basic understanding of some potential mechanisms at work such as informativeness, it is important to recognize that there are some key differences between the structure of options or index futures and commodity futures markets. For instance, commodity futures were invented for institutional buyers who intend to actually take possession of commodities and manage their risks, thus market participants essentially have different incentives to trade, and this could in turn impact the effect of speculative trading. Therefore, the next section will provide a similar literature review in the context of commodity futures specifically.

2.3 Speculation in Commodity Futures

The fluctuation of commodity prices before and after the 2008 financial crisis, accompanied by a rapid growth in the financialization of commodity markets , led to considerable academic interest in the impact of speculation in the futures market on commodity prices. The literature can be broadly divided into two strands: one studies the effect of futures speculation on the futures market, the other looks at the effect on the spot market. However, through the process of calendar spread arbitrage (entering a long and short position on the same futures contract but with different delivery dates), the futures market will be closely linked to the spot market for the same underlying commodity and the spot and futures prices must be highly correlated. Thus the results from one strand could be informative to the other. In the first strand, Irwin and Sanders (2012) look at the effect of index traders on futures price and find very little evidence that index positions influence return or volatility in commodity futures markets using Fama-MacBeth tests. Using highly disaggregated data of individual trader positions, Brunetti and Buyuksahin (2009) show that speculators do not cause price movement or volatility in the futures market. Focusing on the financial crisis period, Brunetti et al. (2016) use an instrumental variable approach to show that hedge fund position changes consistently cause lower volatility and do not destabilize futures markets. In the second strand, Kim (2015) conducts a cross-sectional study by pooling large price changes together and finds that speculator participation in the futures market stabilizes price and lowers price volatility in the spot market.

Researchers have developed theoretical frameworks to study the relationship between futures trading and spot market prices. Sockin and Xiong (2015) develop a model to analyze information aggregation in commodity markets and highlight the important feedback effects of informational noise originating from both supply shocks and futures market trading on commodity spot price. Furthermore, they point out that consumers are not necessarily able to recognize commodity price distortion originated from futures trading due to informational frictions, thus they won't necessarily reduce consumption even if spot price is distorted. Knittel and Pindyck (2013) also provide a theoretical framework of supply and demand to show that speculation had little effect on crude oil prices. While the theories oftentimes disagree with each other, the the existing empirical literature analyzing the effect of speculation on price change and volatility has reached a predominant conclusion that futures speculation has a stabilizing effect on price.

The existing literature primarily focuses on data before 2012 in order to examine the cause of the spikes in commodity prices around the financial crisis period. However, the literature on more recent data is largely missing. With the expansion of high-frequency, algorithmic trading hedge funds and their active participation in the commodity futures market, as discussed in the introduction section, the share of speculators in the market has greatly increased. Furthermore, global events such as the Covid-19 pandemic and Russia's war on Ukraine imposed drastic changes and uncertainties on the global commodities market, thus the relationship between futures speculation and price stabilization should be re-evaluated. This paper revisits the topic and examines whether the conclusions drawn from the stream of literature surrounding the 2008 financial crisis period still hold in today's context.

3 Data

I use 12 commodity futures that are traded in the U.S futures market, and obtain the corresponding daily spot prices and futures trading volume from the Commodity Research Bureau (now a part of barchart.com). The commodities cover energy, grain, soft, and livestock sectors, including: crude oil, gasoline, natural gas, wheat, soybean, corn, cotton, cocoa, coffee, sugar, lean hogs, and live cattle. These are the most actively traded commodities by volume in their respective sectors[1], thus the sample is representative of the commodity futures market as a whole. For futures volume, when multiple contracts are considered at a time, I only consider the daily nearest contract (see Glossary) so that the futures data for each commodity forms a continuous time series. I divide my analysis into two periods: October 1992 to July 2012, August 2012 to October 2021. The rationale is to partially replicate the result from previous literature, take a closer look at more recent data, and observe any changes in market dynamics through time.

The mean and standard deviation of nominal spot prices by commodity are summarized in Table 4. By just observing the statistics, I find that for most commodities, the price has increased significantly while the standard deviation, i.e. volatility, has decreased. This certainly does not bring us to any conclusion since the 1992-2012 period contains some of the most volatile times for commodity prices, but the observation gives us some intuition that maybe price volatility goes down as speculation grows. Some of the commodities have

[1] Top contracts by volume in the U.S agriculture market: corn, soybean, wheat, and sugar; in the U.S energy market: crude oil, natural gas, gasoline, and heating oil.

been traded in different exchanges throughout the history. The details on the exchanges and data availability are summarized in Table 5.

To find the positions of speculators, I use the Commitment of Traders (COT) report from the U.S Commodity Futures Trading Commission (CFTC). I use the legacy version of it and follow the procedure from Irwin and Sanders (2012) to compute the speculator market share. The data provides weekly reports of total open interest, and separates traders into commercial (hedgers) and non-commercial (speculators) types. The speculator market share is given by:

$$S = \frac{(NCL + NCS + NCSP)}{2 \cdot TOI}$$

where S is speculator market share, NCL is non-commercial long position, NCS is non-commercial short position, $NCSP$ is non-commercial spread (see Glossary), and TOI is total open interest. The open interest reflects Tuesday's closing positions for a given market and is aggregated across all contract expiration months (Irwin & Sanders, 2012). Non-commercial traders, i.e. speculators, can hold positions on both long and short sides, and thus the spread is included in the calculation (note that for commercial traders, theoretically they only take a one-sided position in the futures markets to hedge against a specific unfavorable price risk because they hold the opposite position in the spot market).

There have been ongoing complaints that the legacy COT trader designations may be inaccurate. For example, a corporation can participate in the market as both speculators and hedgers and their positions might be mis-categorised. In addition, speculators may have an incentive to self-classify their activity as commercial hedging to circumvent speculative position limits. In response to such concerns, the CFTC added the disaggregated version of the COT to provide more detailed categorizations of the traders. However, the disaggregated version is only available after 2009. Since our analysis covers a much longer period, for data consistency, the original legacy version of the report is used in this paper.

After collecting and combining the data described above, I generate the main dependent and independent variables: spot price volatility and speculator market share. To construct weekly volatility, I compute daily return as the first order difference between the logarithm of daily spot price and then compute the standard deviation of the daily return within a week. The rationale for constructing a series in weekly frequency is to match the frequency of the COT report, which comes out every Tuesday. The speculator market share variable requires some additional processing before being used in the regression analysis. From Figure 1 and Figure 2, we can tell that the series exhibits a clear upward trend, which will be problematic to use in regressions (see further discussions on this issue in section 4). To remove the trend factor from the data, I regress the series on a timestamp index:

$$\text{speculator share}_t = \text{time index}_t + u_t$$

where the residuals u_t are stored as the "de-trended" speculator share.

To add control variables for some of the regression analysis, I compute the weekly percentage change in total open interest. I also obtain data for quarterly gross domestic product (GDP) growth rate, monthly percentage change in industrial production index and monthly inflation rate from the Federal Reserve Economic Data (FRED) of the Federal Reserve Bank of St.Louis. These data represent macroeconomic factors that should be controlled for in regression analysis of commodity futures volatility, as suggested by previous literature such as Kim (2015).

The choice of control variables is limited to data availability. I choose variables such as open interest change and GDP growth to control for the macroeconomic and market conditions that can be confounding predictors of volatility. However, factors such as inventory are also widely noted in previous literature as important control variables. Due to limited access to data, those variables are not included in my analysis.

After matching all the data by the corresponding date of observation, I construct a main data set which will be used for most of the empirical analysis. A detailed description of the variables can be found in Table 7.

4 Empirical Strategy

One of most common empirical strategies to test association and potentially causation between two variables is Ordinary Least Squares (OLS) regression. However, it is important to remain cautious when running regressions in a cross-sectional manner on time-series data. Due to the difference in the data generating process, i.e. independent sampling versus time-series, we are looking at a slightly different version of the standard assumptions of the OLS regression. Specifically, in the time-series context, the following assumptions have to hold in order to obtain consistent estimators from an OLS regression:

1. The stochastic process $\{y_t, x_{1t}, x_{2t}, ..., x_{kt}\}_{t=0}^{\infty}$ follows a linear model, and is stationary and weakly dependent. In this case, the linear relationship is assumed, and we need to check stationarity and weak dependency. Here we focus on checking the stationarity of the main variable of interest, i.e., the speculator market share variable. The straightforward way to check stationarity without any formal statistical testing is through visualization. Suppose we arbitrarily choose to look at the time-series of a few commodities such as wheat, coffee, and live cattle, and create plots such as Figure 3. The figure shows that speculator market share exhibits a clear upward trend. Our summary statistics in Table 6 provides a more clear and general picture: the summary statistics partition the data of each commodity into two disjoint periods, and we can observe that for all commodities in our sample, the mean of the corresponding speculator market share has increased (thus is not constant across the two periods).

   Therefore, to satisfy the stationarity assumption, it is important to use the de-trended version of the speculator share series. The process is described in Section 3. As shown in Figure 4, the de-trended speculator share series is now approximately stationary.

2. The independent variables should not be a constant or have exact co-linearity. Our data generating process already guarantees this (and our statistical program will automatically omit the variable and report a warning if such problem arises).

3. $E[\epsilon_t|x_t] = 0$, that is, the error term is unrelated to the independent variable of the same period, usually referred to as "contemporaneous exogeneity" (as opposed to "strict exogeneity" which requires the error term to be uncorrelated with the independent variable at all periods). This assumption is difficult to test because the error term is something that we don't observe. Strict exogeneity doesn't seem to hold in our case due to a potential "reversed causality" issue. Since our main independent variable is speculator share and our dependent variable is volatility, a larger error term implies more unexpected shocks in volatility, and that would attract more speculation, causing speculator share to rise. So the error term will be correlated with the main independent variable at certain periods. However, such correlation might not be contemporaneous: current shocks in volatility aren't necessarily correlated with speculation in the same period, and even if it is, one way to take such possibility into consideration is to use lagged speculator share as an alternate predictor. Therefore, I choose to run the regression on lagged series of speculator market share from 0 up to 5 lags.

By satisfying the above assumptions, an OLS regression will produce consistent estimators. To produce statistics such as standard errors that are asymptotically valid (valid when sample size gets large), two more assumptions are important to consider: contemporaneous homoskedasticity and no serial correlation. The former assumes $\text{Var}(\epsilon_t|x_t)$ to be constant, and the latter assumes $\text{Cov}(\epsilon_t\epsilon_s|x_tx_s) = 0$ for all $t \neq s$. However, after some preliminary regressional tests, it turns out that the residuals of our OLS model are both heteroskedastic and auto-correlated. The residual versus predicted values plot exhibits a cone shape, and the Breusch-Godfrey test for serial correlation shows that the autocorrelation of the residuals are statistically significant up to 24 lags, which is approximately 6 month. The evidence from the residuals suggests that the error terms in the model violate the standard OLS assumptions.

To address these two violations of the assumptions, I adopt two strategies. First, I use the heteroskedasticity and autocorrelation consistent (HAC) standard error method to produce Newey-West standard errors in an OLS regression. The Newey-West estimation is used to provide an estimate of the covariance matrix of the parameters when a regression model is applied in situations where the standard assumptions do not apply. Specifically, it can be used to overcome serial correlation and heteroskedasticity in the error term, which is indeed the case in our model. The truncation parameter is set by the "rule of thumb", $m = 0.75T^{1/3}$, where $T = 18457$ is the full sample size. So all of my OLS regression models are run by the Newey-West method with parameter value $m = 11$. The regressions not only produce consistent estimators of β_1, but also asymptotically valid standard errors, which can be used to find the p-values and statistical significance of the coefficients.

Second, I use Generalized Auto-Regressive Conditional Heteroskedasticity (GARCH), which is a time-series framework specifically designed to model volatility. The primary motive for using GARCH is that the contemporaneous correlation between error term and the explanatory variable in the OLS model hasn't been fully addressed by using lagged variables. In the presence of auto-correlation between the lagged terms, if the error term is correlated with the speculator share variable, then it also could be correlated with the lagged terms unless we lag significantly many periods to reduce the auto-correlation. This process would essentially become a black box in the sense that we wouldn't know how many lags to use to address the contemporaneous correlation while still meaningfully interpret the result. On the other hand, time series approaches like GARCH are more structural frameworks to understand financial data such as return and volatility. The GARCH model is more commonly adopted by existing literature on similar topics so that it allows me to compare my estimation to previous results by running the model on the same periods.

4.1 OLS Regression with Newey-West Standard Errors

The hypothesized relationship between spot price volatility and futures speculator market share can be tested by an Ordinary Least Squares (OLS) regression specified as the following:

$$\begin{aligned}\text{weekly\_volatility}_{it} = \beta_0 &+ \beta_1\text{detrended\_share}_{it} + \beta_2\text{inflation}_{it} + \beta_3\text{gdp\_pch}_{it} \\ &+ \beta_5\text{production}_{it} + \beta_6\%\Delta\text{open\_interest}_{it} + \text{commodity fixed effects}_i \\ &+ \text{seasonal fixed effects} + \epsilon_{it}\end{aligned} \tag{1}$$

The regression in equation (1) tests the relationship between speculator market share and spot price volatility over the full sample period (1992-2021) and is estimated using the heteroskedasticity and autocorrelation consistent (HAC) standard error method. The de-trended share variable, associated with the coefficient β_1, is the main variable of interest and the rest are control variables. Under Newey-West estimation, the result shows that β_1 is statistically significant ($p = 0$) and is negative ($\hat{\beta_1} = -0.01905$), implying that volatility is negatively associated with speculator market share.

Reverse causality would be a reasonable challenge to the above model: speculators typically expand their share during periods when market volatility is high, thus the volatility might in turn impact speculation level. To partly mitigate such challenge, we can run the same regression on lagged speculator share to test the relationship between past speculation and future market volatility. As previously mentioned, running lagged regressions also takes into account the possibility of contemporaneous endogeneity. These regressions are specified in the following equation:

$$\begin{aligned} \text{weekly\_volatility}_{it} = \beta_0 &+ \beta_1 \text{detrended\_share}_{it-k} + \beta_2 \text{inflation}_{it} + \beta_3 \text{gdp\_pch}_{it} \\ &+ \beta_5 \text{production}_{it} + \beta_6 \%\Delta \text{open\_interest}_{it} + \text{commodity fixed effects}_i \\ &+ \text{seasonal fixed effects} + \epsilon_{it} \end{aligned} \tag{2}$$

The index $t-k$ of the detrended share variable represents k lags (in terms of weeks) of the original series. I run the regression for $k = 1, 2, 3, 4, 5$ and the coefficients on the main variable continues to be statistically significant, while the magnitude of the coefficient diminishes. The detailed results are reported and discussed in the result section.

In order to test the change of market dynamics in more recent periods, i.e. has such relationship changed after 2012, I run the same regression with post2012, a dummy variable indicating whether the data is before or after 2012, being interacted with every independent variable. The reason for choosing 2012 as a cut-off point is to align with my observation that most existing literature analyzes pre-2012 markets. The formulation of the model with the interaction terms is given by:

$$\begin{aligned}\text{weekly\_volatility}_{it} = {} & \beta_0 + \beta_1 \text{detrened\_share}_{it} + \delta_0 \text{post2012} \\ & + \delta_1 \text{detrended\_share*post2012}_{it} + \beta_2 \text{inflation}_{it} \\ & + \delta_2 \text{inflation*post2012}_{it} + \beta_3 \text{gdp\_pch}_{it} \\ & + \delta_3 \text{gdp\_pch*post2012}_{it} + \beta_4 \text{production}_{it} \\ & + \delta_4 \text{production*post2012}_{it} + \beta_5 \%\Delta \text{open interest}_{it} \\ & + \delta_5 \%\Delta \text{open interest*post2012}_{it} + \text{commodity fixed effects}_i + \\ & + \text{commodity fixed effects*post2012}_{it} + \text{seasonal fixed effects} \\ & + \text{seasonal fixed effects*post2012} + \epsilon_{it}\end{aligned} \tag{3}$$

The regression in equation (3) allows us to understand the differences between new and old evidence (before and after 2012). The coefficient δ_0 is statistically significant and negative, suggesting that the market volatility did experience a decline after 2012. However, the estimated coefficient on the interaction term δ_1 is not statistically significant ($p = 0.32$) and is positive (0.0112), suggesting that there's little evidence of how markets after 2012 have changed in terms of the relationship between futures speculation and spot price volatility.

4.2 GARCH Model of Volatility

I adopt a Generalized Auto-Regressive Conditional Heteroskedasticity (GARCH) model to investigate the relationship between futures speculation and spot price volatility. GARCH is a common statistical technique used to help model and predict the volatility of returns on financial assets. The stochastic model having time-varying conditional variances can be defined by decomposing the time-series y_t into a mean model and a prediction error:

$$y_t = \text{mean model} + \epsilon_t$$

while the prediction error is written as a product of a distribution U_t and the conditional standard deviation σ_t, that is,

$$\epsilon_t = U_t \sigma_t$$

and the conditional variance σ_t^2, under GARCH (p,q) model, can be generated as

$$\sigma_t^2 = \omega + \sum_{i=1}^{p} \alpha_i \epsilon_{t-i}^2 + \sum_{j=1}^{q} \beta_j \sigma_{t-j}^2$$

The GARCH specification in this paper follows Kim (2015): the conditional mean is constructed as a first order auto-regressive process (AR(1)) with control variables including percentage changes in open interest, inflation, GDP growth, and production growth, and then the analysis on spot volatility is based on the following GARCH (1,1) model:

$$\begin{aligned} r_{it} &= a_0 + a_1 r_{it-1} + c_0 \%\Delta \text{open\_interest}_{it} + c_1 \text{inflation}_t + c_2 \text{gdp\_pch}_t + c_3 \text{production}_t + \epsilon_{it} \\ \epsilon_t &= z_t \sigma_t \\ \sigma_t^2 &= \omega + \alpha \epsilon_{t-1}^2 + \beta \sigma_{t-1}^2 + \theta_1 \text{detrended\_share}_{it} + \lambda_1 \text{inflation}_t + \lambda_2 \text{gdp\_pch}_t + \lambda_3 \text{production} \\ &\quad + \sum_{i-1}^{3} \phi_i s_{it} \end{aligned} \tag{4}$$

Equation (4) is estimated using the maximum likelihood method with robust standard errors. The distribution model is assumed to be a standard normal z. r_{it} is the weekly return variable and s_i is a seasonal dummy variable. The main coefficient of interest is θ_1, which measures the association between speculator market share and volatility. Since different commodities exhibit different market behaviors, the model is estimated individually for each commodity. The negative or insignificant coefficient implies that the speculative positions does not increase conditional spot volatility for most commodities. It also supports the stabilizing theory that speculators' trading in the futures market lowers price volatility. A detailed discussion of the results can be found in the next section.

5 Result

5.1 Results of the OLS Regressions

I examine whether spot price volatility across markets is related to speculation in the futures market. Table 1 reports the estimation results using the OLS regressions described in equation (1) and (2). I test whether spot price volatility is associated with the speculator market share of the same period, and of the previous periods. The regression shows that controlling for some other relevant variables, a unit increase (in our case, an increase of 1 in the de-trended series of speculator market share) in speculator market share is related to 0.0191 reduction in weekly volatility, and such relationship is statistically significant at $p = 0.001$ level. While the nominal value of the predicted coefficient seems small, notice that weekly volatility is a very small measure as well. In fact, for the full sample, the average weekly volatility is 0.0162. So a unit increase in speculator market share is, on average, associated with a more than 118% decrease in spot price volatility. To be more clear, since the speculator share variable is also on a percentage scale, we can say that a percentage point increase in speculation is associated with 1.18% decrease in spot price volatility. Such relationship is not only statistically significant, but also impressive in terms of its magnitude.

Regression (2) is estimated to evaluate the relationship between volatility and past speculation, and the underlying intuition is that if speculation does cause volatility to change in any direction, there might be a lagged effect. The coefficient on speculator share lagged one period is -0.0176 and on lagged two periods is -0.0161, etc., indicating a statistically significant relationship between lagged speculator share and volatility. In fact, when we continue such experiment with data lagged 5 periods, such relationship will lose its statistical significance, and the coefficient continues to go down as we lag more periods, indicating that the magnitude of such relationship diminishes through time.

Table 1: Results from Regression (1) to (2)

	Dependent variable: weekly volatility					
Lags (L) on detrended_share variable	L=0	L=1	L=2	L=3	L=4	L=5
detrended_share	-0.0191***	-0.0176***	-0.0161**	-0.0151**	-0.0132*	-0.0127*
(de-trended speculator market share)	(0.00536)	(0.00531)	(0.00535)	(0.00535)	(0.00528)	(0.00525)
inflation	-0.000385	-0.000329	-0.000286	-0.000241	-0.000142	-0.000118
	(0.00147)	(0.00146)	(0.00146)	(0.00146)	(0.00146)	(0.00147)
gdp_pch	-0.000690***	-0.000687***	-0.000686***	-0.000688***	-0.000692***	-0.000695***
(quarterly GDP growth rate)	(0.000178)	(0.000178)	(0.000178)	(0.000178)	(0.000178)	(0.000179)
production	-0.000845**	-0.000848**	-0.000850**	-0.000855**	-0.000859**	-0.000865**
	(0.000304)	(0.000304)	(0.000305)	(0.000305)	(0.000306)	(0.000307)
%Δ open_interest	-0.0102**	-0.0123***	-0.0129***	-0.0128***	-0.0127***	-0.0125***
(percentage change in total open interest)	(0.00315)	(0.00300)	(0.00298)	(0.00298)	(0.00299)	(0.00300)
_cons	0.0162***	0.0162***	0.0162***	0.0162***	0.0162***	0.0162***
	(0.000546)	(0.000546)	(0.000546)	(0.000547)	(0.000548)	(0.000549)
N	18522	18522	18522	18522	18522	18522

Newey-West Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Regression (3) is estimated to inquire the characteristics of the new evidence (after 2012) in particular. The regression tests whether post2012 is a significant predictor of spot price volatility, and whether the evidence after 2012 provides a stronger support for the hypothesized relationship between spot price volatility and futures speculation. The result, reported in Table 2, indicates that spot price volatility did experience a significant reduction after 2012 (coefficient = -0.005 and p-value < 0.001), but the post 2012 data does not exhibit any stronger or weaker evidence of our hypothesis: the coefficient on the interaction term is positive and not significant. While this is to some extent disappointing, it is also reasonable since I pick 2012 as the cut-off year quite arbitrarily (based on the periods that similar literature has covered or not covered). It would still be interesting to use other methods that do not rely on such arbitrary time cut-offs to learn more about how the association between spot price volatility and futures speculation has grown stronger (or weaker).

Table 2: Results from Regression (3)

	(1)
	weekly_volatility
detrended_share	-0.0257***
	(0.007)
post2012	-0.00526***
	(0.001)
detrended_share_post	0.0112
	(0.011)
N	18522

Newey-West Std. Err. in parenthesis

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

5.2 Results from GARCH Models

I assess whether speculators' trading share in the futures market stabilizes the spot market using the GARCH model described in Equation (4). The estimations are run separately on pre-2012 and post-2012 data, and then on the full sample. The results of the model estimations are summarized in Table 3 below. To better understand the main research question and to conserve space, I only present the estimated coefficients and p-values of the detrended speculator share variable in the conditional variance function for each commodity type.

Table 3: The Effect of Future Speculation on Spot Volatility (GARCH estimation)

	Full Sample		Pre-2012		Post-2012	
	Coefficient	p-value	Coefficient	p-value	Coefficient	p-value
Crude oil	-4.190***	0.005	-6.189***	0.001	-3.181	0.359
Gasoline	10.108	0.291	-8.409**	0.023	8.576	0.172
Natural gas	-1.846*	0.051	-2.841**	0.049	0.091	0.946
Wheat	-0.821	0.709	-3.894*	0.057	7.939*	0.064
Soybean	-1.811	0.707	-6.998	0.241	2.173	0.588
Corn	1.086	0.749	0.610	0.921	-0.404	0.858
Cotton	-0.580	0.792	0.048	0.985	1.386	0.490
Cocoa	5.324	0.411	5.096*	0.095	-	-
Coffee	-2.779	0.389	-4.473	0.118	-5.667***	0.003
Sugar	1.865	0.751	8.040	0.762	-1.526	0.435
Lean hogs	3.738	0.245	2.257	0.669	0.810	0.735
Feeder cattle	4.419	0.274	-	-	3.577	0.301
Live cattle	7.473	0.348	13.251	0.525	5.132	0.124

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

The coefficients and p-values reported in Table 3 indicate there is no evidence that futures speculation destabilizes the spot price. The omitted entries are incidents of failed estimation: the maximum likelihood method, in those cases, cannot move out a corner area or does not find a convergence due to a variety of reasons such as model complexity.

For the full sample, all of the positive coefficients are not statistically significant, suggesting very little evidence of the destabilizing story of futures speculation. In addition, the coefficient estimates are negative and statistically significant in crude oil and natural gas, suggesting that speculation is negatively associated with spot price volatility in these two markets, supporting the stabilizing story of futures speculation.

As suggested in the introduction, the full sample spans a very long time period, and there could be a lot of changes in the market dynamics during the 30-year evolvement of the financial market. Indeed, some interesting comparisons can be made between the pre-2012 and post-2012 results. In the pre-2012 markets, there seems to be a stronger stabilizing effect of futures speculation: for crude oil, gasoline, natural gas, wheat, and coffee, we see a negative coefficient estimate with considerably small p-value, indicating a negative association between futures speculation and spot volatility. The finding is consistent with the literature studying pre-2012 periods such as Kim (2015). However, in the post-2012 markets, we see some of these stabilizing relationships diminish or gone. In the crude oil market, for example, the coefficient is still negative, but not statistically significant anymore. A more stark contrast can be found in the wheat market, where the coefficient "flipped" from a significant negative to a significant positive. Similar phenomenon occurred in gasoline and natural gas markets. Besides, we also see coffee as an interesting case, as speculation continues to be a negative predictor of volatility, the statistical significance of the coefficient grows in post-2012 periods. Overall, we see that speculation is almost never destabilizing, and most of the time stabilizing or having no effect in the spot price. In some markets, the pre-2012 and post-2012 dynamics might have experienced changes such that the negative relationship becomes weaker.

6 Discussion

In early March, 2022, just days after Russia invaded Ukraine, a strange thing happened in the market for wheat futures: the prices of wheat future May contracts spiked up to the limit, halting trading of the day. A New York Times article described this event as a "wheat whale" (Flitter, 2022), and analysts have attributed it to retail investors congregating on Reddit piling into an agricultural ETF called the Teucrium Wheat Fund. As a result of expectations of an increased wheat price due to the war, the fund had experienced more than five-fold increase in trading volume since March, 2022.

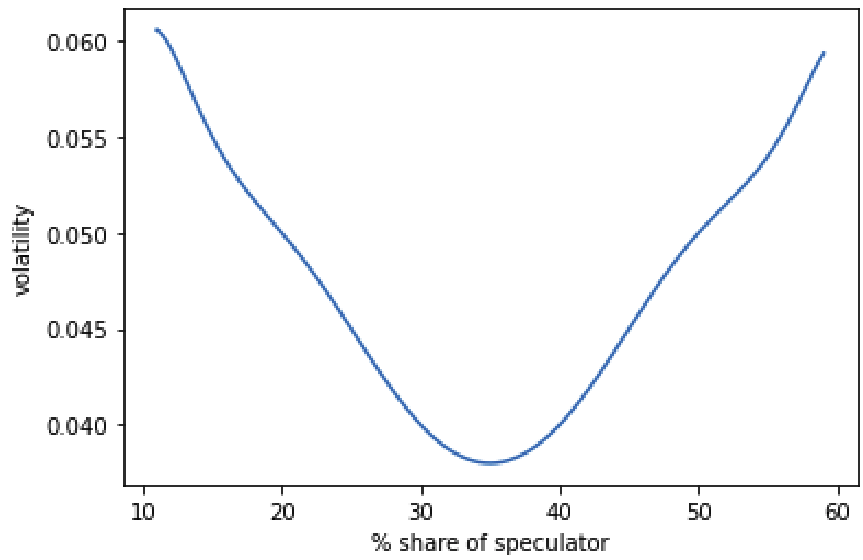
While the "wheat whale" might have been just another strange and interesting finance story to tell, it sheds lights on at least two things as far as this paper is concerned. First, why does speculation increase so much through time, particularly after 2012? As discussed in the introduction, the increase in speculator market share follows a steady upward trend without any identifiable jumps, so it's unlikely that the increase has been caused by any specific event or policy. The wheat whale story provides a possible explanation: the steady increase in speculation in commodity futures might have been caused by the introduction of new financial instruments such as ETFs, as well as the increased number of online brokers who offer investors zero commission fee. These advancements are regarded as the democratization of finance, as they make the financial market more accessible to retail investors. These instruments and platforms not only allow investors to trade at no cost, but also allow them to access asset classes like commodity futures through ETFs, something that they wouldn't be able to trade otherwise. In addition, the increase in speculation can also be attributed to an increase in the number of hedge funds and other investment companies (the number of hedge funds in the U.S increases by 3% per year), as well as a wave of innovation in financial technology and new trading ideas or methods emerged in more recent years. These funds and trading technologies have been well-accepted by more wealthy investors or institutional investors. The approach of quantitative and algorithmic trading, for example, had been able to generate high risk adjusted returns in the past two decades and had therefore received much popularity among investors. The unique features of commodity futures, such as low correlation with other asset classes and the ability to find various types of arbitrage opportunities, have led to the success of a significant number of quantitative trading strategies.

The second thing that the wheat whale story tells us is how investors can be very uninformed and yet drive price towards any direction dramatically. This actually isn't the first time when amateur investors engage in social media driven, coordinated buying regimes. Speculators on Reddit have once successfully bid up the price and volume of the stock of GameStop, an event now known as the Gamestop mania. While these can be considered as rare instances in which some investors create a huge momentum to wield in financial markets when equipped with tools like social media, it wouldn't be hard to imagine how the long-term presence of uninformed investors can consistently trade in a manner that makes the market more volatile. This section discusses how our empirical results match with the theories of information, and how future work can be developed to formally test these theories.

6.1 Masters hypothesis, informed investors, and my analysis

The results from previous sections, especially from the GARCH model, provide us a much clearer picture of the story told by the Masters hypothesis and previous literature. The Masters hypothesis (Masters, 2009) argues that "excessive speculation leads to excessive volatility" and proposes an optimal range of speculator share: 25% to 35%. However, literature such as Irwin and Sanders (2012) and Kim (2015) quickly rebuts the hypothesis by presenting empirical results from pre-2012 data, showing that speculation has not been destabilizing. The potential relationship between speculation and high market volatility, once attracted substantial academic interest, then became a lingering debate. But what if both sides are correct? The opponents of Masters' theory might have made their conclusions too hastily: the speculator shares in all of the commodities covered in this and previous studies never exceeded 35% before 2012 (see Table 6). So the Masters hypothesis wasn't necessarily about discouraging speculation at the time it was made, but more about calling for precautions when speculation becomes "excessive". This is why our new evidence becomes important. After 2012, speculation in most commodity markets has grown tremendously: all of the markets have more than 35% of participants being speculators, and in the natural gas market, speculators occupy a startling 60% share. So it is only meaningful to evaluate the Masters hypothesis when we now have data of periods when speculator share goes way beyond the proposed range, and this paper fills the gap by comparing results from more recent periods to those from earlier conclusions. Although the overall "not destabilizing" story of speculation still holds in most commodities, we do see changes in some markets. One prominent example is the wheat market: before 2012, speculation was significantly stabilizing; after 2012, speculation has been significantly destabilizing ("significant" in a statistical context). Such patterns in our empirical results make us wonder whether we have reached the point at which Masters' concern becomes real.

Here I hypothesize a rudimentary framework to explain the change in the pre-and-post-2012 market dynamics, as roughly sketched by the diagram below:



The diagram is obtained by a polynomial interpolation over a set of possible values of weekly volatility and speculator market share in the wheat market. The idea demonstrated in the diagram is that speculators start to stabilize the market as they enter, and continues to reduce volatility as they expand their shares in the market to a point at which they occupy too much of the trades and begin to destabilize the market. This is no doubt a very rough guess and one of the many ways to explain what's going on in our data. However, several previously established theoretical and empirical work have provided support to this hypothesis. The underlying story of why the market behaves this way is closely related to the types of speculators in the market. For example, Figlewski (1981), Stein (1987), and Wang et al. (2008) attribute the increase or decrease in volatility to the level of information in the market and how informed the investors are. In particular, they propose that the existence of futures trading (or any type of derivative instrument in general) might lead to trading by new classes of investors who do not have access to as good information as traders in the existing market. In addition, the introduction of new derivatives such as options can increase the level of informed trading and open up arbitrage opportunities to reduce spot market volatility. Combining these theories with our empirical results, one can arrive at the following explanation: traders in pre-2012 markets could be primarily more informed, and as the informed traders enter the market, volatility decreases due to the increased informativeness of price. In post-2012 periods, there could be more and more uninformed traders entering the market, leading to more volatile spot prices. The mechanism could be a mixture of many factors. For example, these less informed new entrants can impose a negative externality to the existing traders as their actions reduce the informativeness of price. In addition, there could be a threshold beyond which the traders entering the market are less informed. While the efficient market hypothesis usually assumes that asset prices reflect all available information, it is commonly known that the assumption does not hold in reality: even mature, institutional investors could be misinformed or uninformed in some ways, not to say individual investors who have little access to information. The information in commodity markets is complex and multi-dimensional, including supply and demand of a particular good, climate, policy, and macroeconomic factors. Therefore, it is reasonable to assume that only a limited number of investors in the futures market can be thought of as adequately informed, usually those who are merchants or producers themselves (hedgers), or experienced institutional investors. These investors possess information that is difficult for other investors to access, such as first-hand knowledge about agricultural production, expert research in macroeconomics, advanced machines and algorithm that can dig out information from price, and etc. If this assumption holds, then the Masters hypothesis implies that only fewer than 25% to 35% of the investors in commodity futures markets can be considered as informed.

The analysis above not only provides a framework to explain and understand our empirical results, but also offers a way to maintain the stabilizing effect of speculation even when their share expands significantly. Our analysis suggests that the problem with speculation is not that it's excessive, but that it's uninformed. That is to say, the optimal threshold share of speculators can be increased if information becomes more accessible to investors. From a policy-making perspective, we should be concerned with the amount of information barrier and information friction investors face, rather than the nominal share of speculators in the market.

6.2 Future work

Research in this direction can be extended in two ways. First, one can develop theoretical models that take information barriers and information friction into consideration. One notable example of this stream of research is Sockin and Xiong (2015), which highlights the feedback effects of informational noise from supply shocks and futures market trading on commodity demand and spot prices. Second, one can continue to explore empirical methods that better captures the causal relationship between speculation, information, and volatility. In order to do this, one would have to come up ways to formally quantify the amount of information that a particular type or group of investors possess, and that would be without doubt a challenging task.

7 Conclusion

In this paper, I assess whether futures speculation destabilizes the commodity market and compare the results obtained from different periods to analyze the change in market dynamics. I use an OLS regression and a GARCH model to assess the impact of speculator share on spot price weekly volatility, and study the different results obtained from pre-2012 and post-2012 data. My analysis indicates that weekly volatility, measured by the standard deviation of daily returns within a week, is negatively associated with speculator market share. The spot price volatility measured by conditional variance in the GARCH model is either negatively associated or not associated with speculator market share. Speculators in the futures market overall contribute to reducing spot price volatility.

I also show that the stabilizing effect of speculation is diminished in post-2012 periods, and this is particularly the case in a few commodity markets where speculation begins to destabilize spot price. Through an in-depth literature review, I connect the new findings with existing theoretical framework and empirical evidence, providing a possible explanation for such change.

Finally, I propose a potential reason why speculation has steadily increased, and how policy makers can shift their focus from regulating speculation to reducing information barrier and information friction in the market. In future research, I plan to extend my analysis to investigate the relationship between information and commodity price fluctuations.

Tables

Table 4: Summary Statistics of Commodity Spot Price

Commodity	Full Sample		1992-2012		2012-2021	
	Mean	Stdev	Mean	Std	Mean	Std
Crude Oil	50.14	28.87	44.0	29.7	63.53	21.59
Gasoline	1.48	0.84	1.24	0.83	1.99	0.61
Natural Gas	3.92	2.18	4.35	2.48	3.03	0.82
Wheat	451.72	161.74	404.39	156.68	553.96	119.78
Soybean	846.61	313.64	738.38	278.94	1080.44	250.66
Corn	357.55	150.16	318.93	142.28	441.07	131.66
Cocoa	2224.7	763.09	1967.09	741.7	2781.48	445.53
Coffee	140.96	53.68	137.68	61.74	148.03	28.38
Cotton	66.94	21.22	64.84	24.49	71.47	9.93
Sugar	13.82	5.76	13.16	6.56	15.23	2.96
Feeder cattle	116.76	36.04	98.0	19.85	156.92	29.51
Lean hogs	66.05	18.06	61.28	15.58	76.54	18.68
Live cattle	94.47	26.55	80.08	15.62	125.53	16.94

Table 5: Information on Commodity Futures and Data Availability

Commodity	Contract Size	Exchange	From	To	Obs
Cocoa	10 METRIC TONS	CSCE	1986-01-15	2004-12-28	799
		NYBT	2005-01-04	2008-07-08	184
		ICUS	2008-07-15	2021-10-05	691
Coffee	37,500 POUNDS	CSCE	1986-01-15	2004-12-28	799
		NYBT	2005-01-04	2008-07-08	184
		ICUS	2008-07-15	2021-10-05	691
Cotton	50,500 POUNDS	NYCE	1986-01-15	2004-12-28	800
		NYBT	2005-01-04	2008-07-08	184
		ICUS	2008-07-15	2021-10-05	691
Corn	1,000 BUSHELS	CBT	1986-01-15	1997-12-30	437
	5,000 BUSHELS	CBT	1998-01-06	2021-10-05	1240
Crude oil	1,000 BARRELS	NYME	1986-01-15	2021-10-05	1675
Feeder cattle	44,000 POUNDS	CME	1986-01-15	2015-06-09	1347
	50,000 POUNDS	CME	2015-06-16	2021-10-05	330
Gasoline	42,000 GALLONS	NYME	2006-02-14	2021-10-05	817
Lean hogs	40,000 POUNDS	CME	1996-04-02	2021-10-05	1328
Live cattle	40,000 POUNDS	CME	1986-01-15	2021-10-05	1677
Natural gas	10,000 MMBTU'S	NYME	1990-04-12	2021-10-05	1573
Soybean	1,000 BUSHELS	CBT	1986-01-15	1997-12-30	437
	5,000 BUSHELS	CBT	1998-01-06	2021-10-05	1240
Sugar	112,000 POUNDS	CSCE	1986-01-15	2004-12-28	799
		NYBT	2005-01-04	2008-07-08	184
		ICUS	2008-07-15	2021-10-05	691
Wheat	1,000 BUSHELS	CBT	1986-01-15	1997-12-30	437
	5,000 BUSHELS	CBT	1998-01-06	2021-10-05	1240

Table 6: Summary of Speculator Market Share

	1992-2012	**2012-2021**	**Growth**
Speculator market share and growth			
Cocoa	0.23	0.41	78.26%
Coffee	0.28	0.42	50.0%
Corn	0.25	0.4	60.0%
Crude oil	0.24	0.51	112.5%
Feeder cattle	0.32	0.48	50.0%
Gasoline	0.24	0.39	62.5%
Lean hogs	0.35	0.48	37.14%
Live cattle	0.31	0.43	38.71%
Natural gas	0.29	0.60	106.9%
Soybean	0.29	0.38	31.03%
Sugar	0.20	0.35	75.0%
Wheat	0.32	0.54	68.75%

Table 7: Full Dataset Variables Description

Variable	Type	Information
commodity	categorical	commodity name
date	datetime	date of observation
speculatorshare	numerical	share of speculation
detrended_share	numerical	share of speculation moving the trend
inflation	numerical	monthly inflation rate
gdp_pch	numerical	quarterly GDP % change
production	numerical	monthly industrial production % change
close	numerical	close spot price of the Tuesday
log_price	numerical	logarithm of price
weekly_return	numerical	Tuesday-to-Tuesday difference in log_price
weekly_volatility	numerical	weekly standard deviation of daily return
post2012	dummy	whether it's after 2012
%Δopen_interest	numerical	weekly percentage changes in futures open interest

Figures

Figure 1: Speculator Market Share (Full Sample)

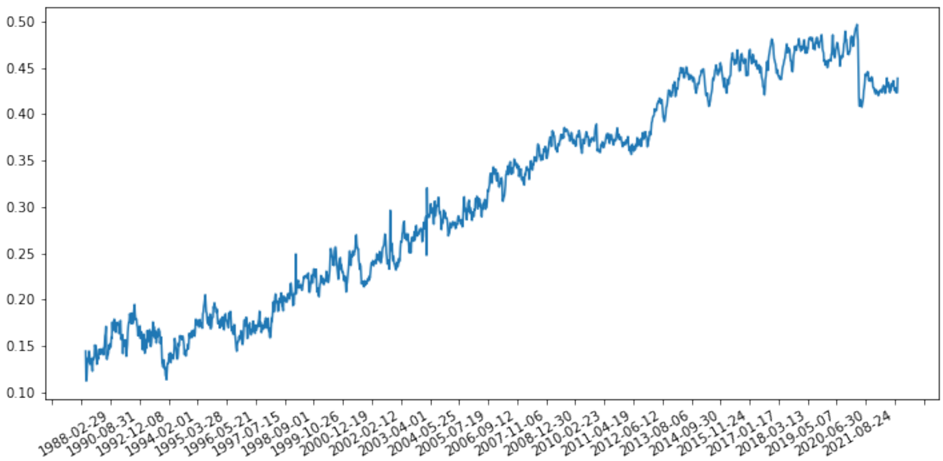


Figure 2: Speculator Market Share (Crude Oil)

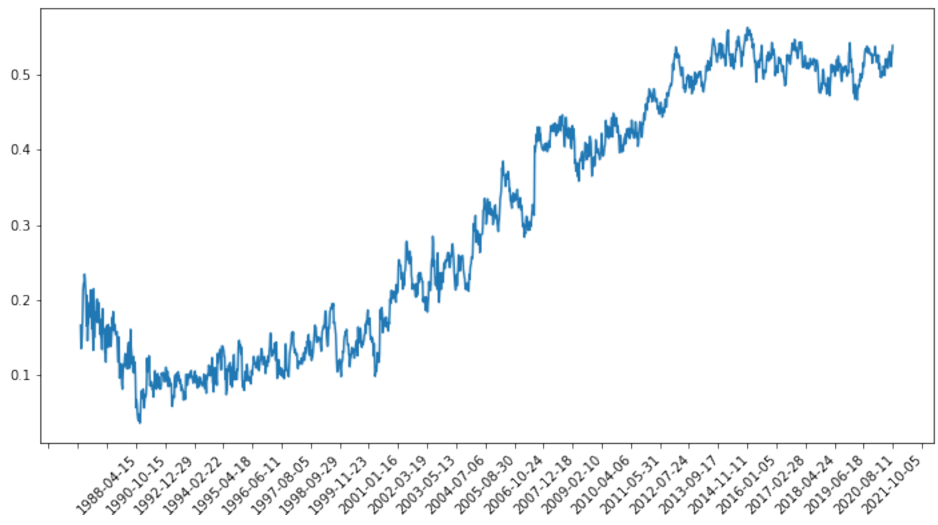


Figure 3: Time-series Stationarity Check for Wheat, Sugar, and Live Cattle (with trends)

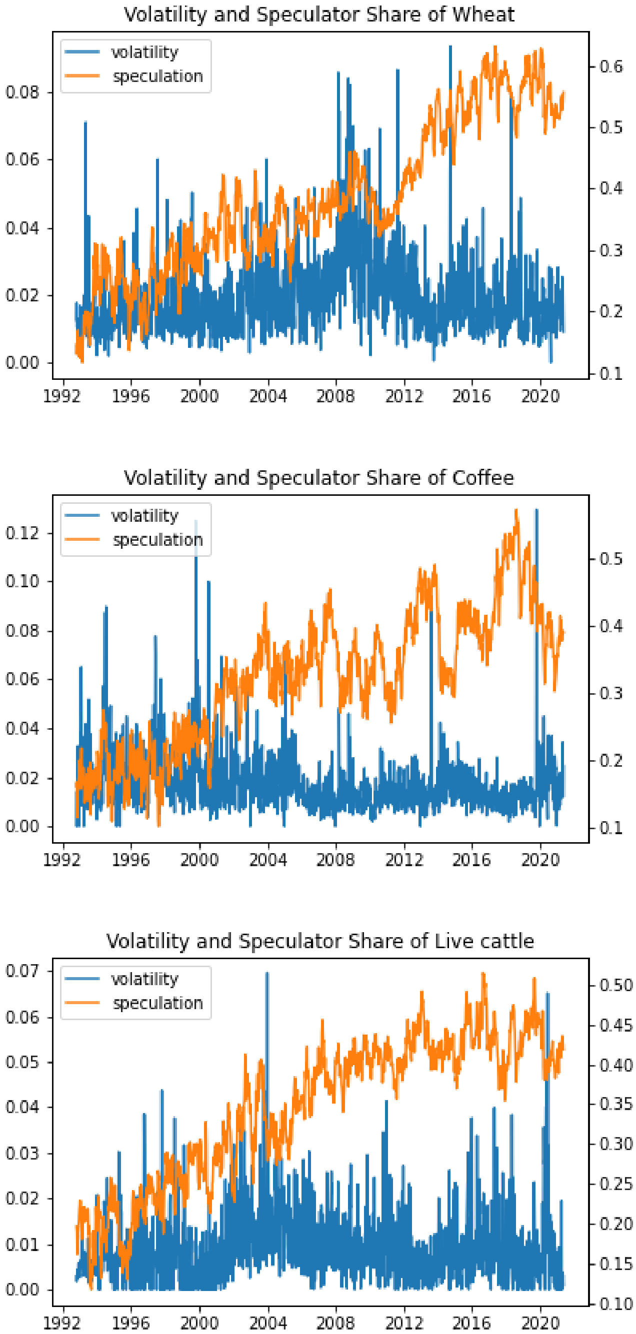
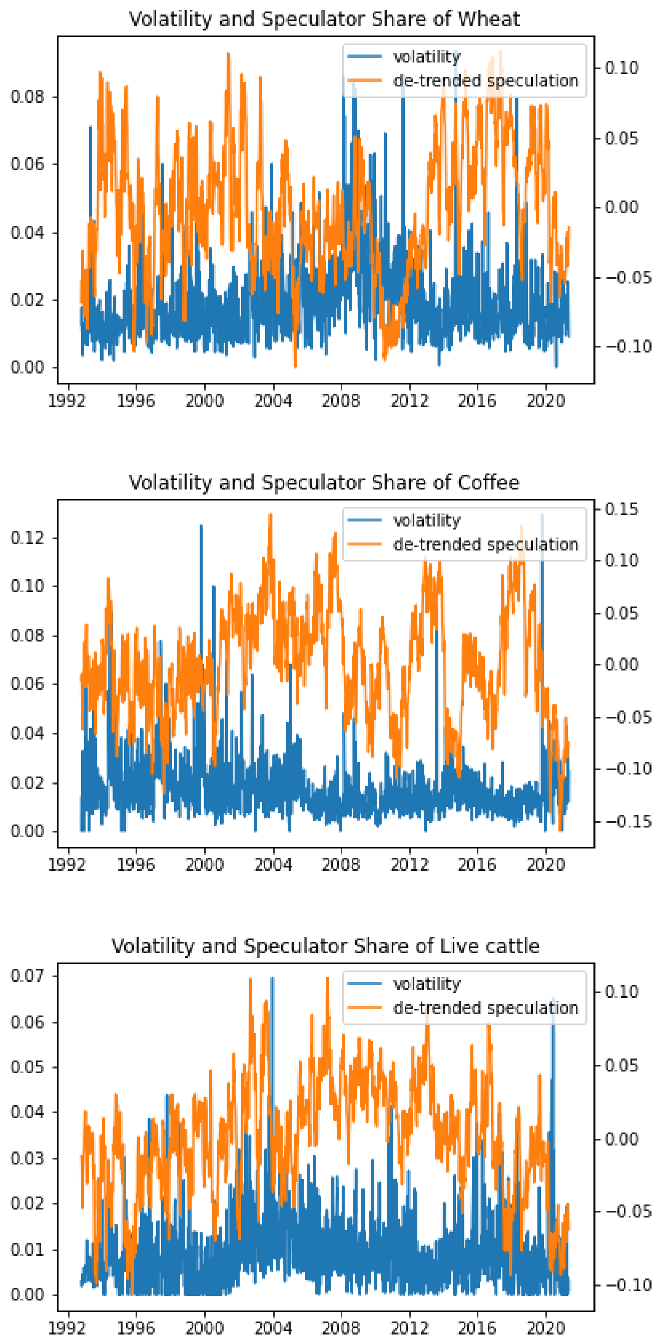


Figure 4: Time-series Stationarity Check for Wheat, Sugar, and Live Cattle (trends removed)



Glossary

daily nearest contract is the closest unexpired contract for the underlying commodity. 14

herding behavior is when investors follow the crowd instead of their own analysis. 6

noise traders are overconfident and misinformed traders who typically trade securities without the support of professional advise or advanced analysis. 6

non-commercial spread is a measurement of the extent to which each non-commercial trader holds equal long and short futures positions. 15

speculative bubbles are spikes in asset prices caused by irrational speculative activity that is not supported by the fundamentals. 6

Tom Yum Goong crisis The crisis started in Thailand (known in Thailand as the Tom Yam Kung crisis) on 2 July, with the financial collapse of the Thai baht after the Thai government was forced to float the baht due to lack of foreign currency to support its currency peg to the U.S. dollar.. 3

References

Baumol, W. J. (1957). Speculation, profitability, and stability. *The Review of Economics and Statistics*, *39*(3), 263. doi: 10.2307/1926042

Brunetti, C., & Buyuksahin, B. (2009). Is speculation destabilizing? *SSRN Electronic Journal*. doi: 10.2139/ssrn.1393524

Brunetti, C., Buyuksahin, B., & Harris, J. H. (2016). Speculators, prices, and market volatility. *Journal of Financial and Quantitative Analysis*, *51*(5), 1545-1574. doi: 10.1017/s0022109016000569

Chatrath, A., Ramchander, S., & Song, F. (1995). Does options trading lead to greater cash market volatility? *Journal of Futures Markets*, *15*(7), 785-803. doi: 10.1002/fut.3990150704

Danthine, J.-P. (1978). Information, futures prices, and stabilizing speculation. *Journal of Economic Theory*, *17*(1), 79-98. doi: 10.1016/0022-0531(78)90124-2

Figlewski, S. (1981). Futures trading and volatility in the gnma market. *The Journal of Finance*, *36*(2), 445-456. doi: 10.1111/j.1540-6261.1981.tb00461.x

Flitter, E. (2022, Apr). *A market mystery: The 'wheat whale' that came out of nowhere.* The New York Times. Retrieved from https://www.nytimes.com/2022/04/08/business/wheat-market-russia-ukraine-reddit.html

Friedman, M. (1953). *Essays in positive economics.* The University of Chicago Press.

Harris, L. (1989). Samp;p 500 cash stock price volatilities. *The Journal of Finance*, *44*(5), 1155. doi: 10.2307/2328637

Hart, O. D., & Kreps, D. M. (1986). Price destabilizing speculation. *Journal of Political Economy*, *94*(5), 927.

Irwin, S. H., & Sanders, D. R. (2012). Testing the masters hypothesis in commodity futures markets. *Energy Economics*, *34*(1), 256-269. doi: 10.1016/j.eneco.2011.10.008

Kim, A. (2015). Does futures speculation destabilize commodity markets? *Journal of Futures Markets*, *35*(8), 696-714. doi: 10.1002/fut.21716

Knittel, C., & Pindyck, R. (2013). The simple economics of commodity price

speculation.
doi: 10.3386/w18951

Lowenstein, R. (1997, Nov). Why stock options are really dynamite. *The Wall Street Journal*. Retrieved from https://webreprints.djreprints.com/4401390655727.html

Masters, M. W. (2009, Aug). Testimony before the commodity futures trading commission.

Mayhew, S. (2000, Feb). The impact of derivatives on cash markets: What have we learned? *Unpublished manuscript, University of Georgia*.

Sockin, M., & Xiong, W. (2015). Informational frictions and commodity markets. *The Journal of Finance*, *70*(5), 2063-2098. doi: 10.1111/jofi.12261

Stein, J. C. (1987). Informational externalities and welfare-reducing speculation. *Journal of Political Economy*, *95*(6), 1123-1145. doi: 10.1086/261508

Telser, L. G. (1959). A theory of speculation relating profitability and stability. *The Review of Economics and Statistics*, *41*(3), 295. doi: 10.2307/1927455

Wang, S. S., Li, W., & Cheng, L. T. (2008). The impact of h-share derivatives on the underlying equity market. *SSRN Electronic Journal*. doi: 10.2139/ssrn.1145306