Imperial College London

Department of Mathematics

A study of price impact in commodity futures markets

Author: Diego SAVERYS (CID: 02475053)

A thesis submitted for the degree of

MSc in Mathematics and Finance, 2023-2024

Declaration

The work contained in this thesis is my own work unless otherwise stated.

Acknowledgements

I would like to thank my internal supervisor, Johaness Muhle-Karbe, for his ideas, comments and feedback during this project. His expertise on the topic brought fresh perspectives to the research and was extremely helpful. Additionally, I would like to thank Velador Associates Ltd, and Stuart Jamieson in particular, for providing me with the opportunity to work on this project, and giving me access to relevant data and interesting research ideas. Finally, I would also like to thank my parents for supporting me through all my academic years, and providing me with the opportunity to pursue my ambitions.

Abstract

Commodity futures contracts, with their frequent expiration periods, require investors to regularly roll over their positions to maintain exposure. This rolling over process can create significant price impacts in the market. This thesis explores these price impacts using a non-linear propagator model developed by Muhle-Karbe et al. [25] and identifies the most suitable functional forms for calibrating price impact models across different commodity futures. Our findings indicate that price impacts in these markets are typically non-linear, with an optimal concavity coefficient smaller than the standard square-root law traditionally applied in price impact modeling. Furthermore, given that the timing of futures contract rollovers for major indices, such as the SP-GSCI, is generally known, the thesis investigates the potential for designing simple anticipatory arbitrage strategies to exploit these predictable price movements. While we demonstrate that such strategies were profitable until 2010, their effectiveness has markedly declined in recent years. This decline is likely due to the transition to electronic trading and increased market liquidity, which have reduced price impacts around rolling events.

Contents

List of Figures

List of Tables

Introduction

Since the early 2000s, commodity futures markets have experienced a remarkable rise in investment and participation. Commodities have become increasingly popular, offering diversification and protection against inflation, especially among hedgers and speculators. The primary method for engaging with commodities is through futures markets, which allow investors to replicate the returns of physical commodities without the hassle of owning them. Since futures contracts come with regular expiration dates, investors typically have to sell the contracts that are about to expire, and buy new ones with longer maturities, in a bid to maintain their exposure. This phenomenon is described as the 'rolling forward' of futures contract. Major indices typically adhere to strict rules, determining the timing of the rolls well in advance and making this information public. To reduce tracking errors, commodity investors typically execute their roll trades within the index's designated roll window. This makes the timing of rolling activity highly predictable. Moreover, since rolling involves the substantial sale of nearby contracts and the purchase of further maturity contracts, basic supply and demand principles suggest that nearby prices will drop and prices for the next contracts will rise. Therefore, the rolling forward of futures contracts should create significant and predictable market anomalies, presenting opportunities for investors to use simple anticipatory strategies capitalizing on these temporary price impacts.

While recent literature has closely examined the effects of futures contracts rolling, the results remain mixed and it's unclear if this predictability translates into clear and predictable price impacts. On the one hand, Mou [24] found that the rolling activity of commodity indices in futures markets created statistically and economically significant price impacts, leading to profitable trading opportunities through anticipatory arbitrage. He also showed that the rolling forward of futures contracts created significant order flow cost for index investors. Similarly, other studies [1, 7, 30] reported significant price impacts and trading opportunities in the crude oil futures market due to the roll. On the other hand, a more recent part of the literature [4, 15, 20, 27] examined the price impact surrounding roll events across different data sets and commodities, finding little to no impact of futures rolling on commodity futures prices.

These studies differ in terms of data frequency, data availability, commodities considered, and the specific rolls studied. An important aspect shared by all these studies is the methodology used to detect potential price impacts. Specifically, they all employ a similar approach: simple linear regression of the return differential between the next and deferred contracts, based on the number of contracts traded as part of the roll. While linear regressions may seem suitable for examining the impact of increased investment on commodity futures prices, these models might not be well-suited for studying price impacts in limit order books. Indeed, the price impact of trades is a concave function of trade size, which then gradually decays over time [17]. As noted in [3], large trades tend to have a non-linear impact, suggesting the need to explore potential non-linearity in price impacts and their evolution over time.

This lack of focus on price impact of futures rolling specifically in limit order books, and the oversight of potential non-linear effects is a significant gap. This problem is crucial for stakeholders, including institutional investors, hedge funds, and market regulators. Indeed, understanding the precise dynamics of price impacts in limit order books can help institutional investors and hedge funds develop more effective trading strategies that capitalize on or mitigate these impacts. For market regulators, insights into non-linear price impacts are essential for ensuring market stability and preventing market manipulation. Finally, for index investors, understanding how price impact of trades propagates throughout time allows them to mitigate the significant costs associated with the rolling forward of their futures contracts, and to execute their according trade optimally. Addressing this gap in the literature will lead to more accurate price impact models for commodity futures markets, and a better understanding of market behavior.

The first and main goal of this thesis is to understand how price impact propagates throughout time in commodity indices, including during rolling periods. Our approach involves using trade imbalance to accurately explain and predict price changes. Previous studies have shown that calibrating trade imbalances with the propagator model effectively captures key stylized facts of market impact, such as intraday liquidity shifts, impact propagation, and highly concave instantaneous impact [25, 18]. Therefore, to specifically capture the dynamics of price impact within commodity futures markets, we follow Muhle-Karbe et al's framework for using the non-linear propagator model [25], and focus on finding the most adequate functional form to calibrate the model to the different commodities, as detailed in Chapter 4.

Additionally, after understanding the price impact dynamics, a second goal will be to design simple anticipatory trading strategies to try and take advantage of the market anomalies during futures rolling. Specifically, we build on Mou [24] and implement three simple anticipatory trading strategies. We then evaluate the performance of these strategies over time. Notably, the data considered in the design of the trading strategies in the literature is limited to the period until 2010. Therefore, evaluating the performance of these strategies on post-2010 data will shed light on whether anomalies still exist and whether it is still profitable to trade surrounding the rolling events.

To the best of my knowledge, this thesis represents one of the first significant studies focused on the specific price impact dynamics in commodity futures markets. While it is widely recognized that price impact exhibits concavity as a function of trade size, there is an ongoing debate regarding the most accurate functional form to model this relationship in practice. Existing literature which also examine trades in commodity futures, suggests that a square root function often provides a robust fit for metaorders. However, when examining public trades at the fill level, the optimal functional form remains less clear, with any degree of concavity being beneficial, though a square root may not necessarily be the best choice. Additionally, it also appears to be the first study considering trading strategies surrounding rolling events post-2010, leading to more accurate and relevant findings on the potential profits from simple anticipatory arbitrage strategies in commodity futures markets.

This thesis is structured as follows. We start by introducing futures markets in Chapter 1, specifically focusing on commodity index futures and the motivation for investors to trade them. Next, in Chapter 2, we discuss the state of the literature on the price impact of futures contracts rolling, before detailing the data used in the analysis in Chapter 3. Chapter 4 introduces the price impact model used in the analysis and how it is calibrated. It also presents the results of our calibration, including the optimal scale and concavity of market impact. In Chapter 5, we present the different strategies that will be used to anticipate the price impact, how to evaluate them and present their performance. Finally, we offer concluding remarks, and discuss potential limitations and avenues for further research.

Chapter 1

Commodity Index Futures and the Background

1.1 Futures contracts

A futures contract is an agreement to buy or sell a specific underlying asset at a predetermined price on a future date [10]. In general, the contracts are made by producers and suppliers to avoid market volatility. For example, consider an airline and a fuel supplier. If an airline buys jet fuel at £2.50 per gallon from a supplier and sells tickets based on that fuel cost, both parties may wish to secure that price. Engaging in a futures contract eliminates the uncertainty of fluctuating fuel prices in the future, providing them with financial stability.

1.1.1 Evolution of futures markets

The first futures trading exchange was the Dojima Rice Exchange, established in Japan in 1730 for trading rice futures [31]. Subsequently, futures markets gained popularity with the opening of the New York Cotton Exchange (NYCE) in 1870 and the London Metals and Market Exchange in 1877. From the 1970s onwards, the futures trading markets experienced significant expansion. The Chicago Mercantile Exchange (CME) began offering futures on foreign currencies, the Chicago Board of Trade (CBOT) introduced T-bond futures, and the New York Mercantile Exchange (NYMEX) added financial futures, including crude oil and gas. Additionally, the Commodities Exchange (COMEX) started offering metal futures. Today, futures contracts are available not only for commodities but also for other asset classes, including foreign exchange, interest rates, and equity indices.

1.1.2 Underlying mechanisms

Most futures markets are registered with the Commodity Futures Trading Commission (CFTC), which regulates them. This regulation ensures that futures contracts are highly standardized, with specific quantities and set expiration dates. For instance, WTI Crude Oil futures expire every month of the year. Futures tickers consist of three parts. For example, CLF24 represents a WTI Crude Oil contract (CL) expiring in January (F) 2024 (24). The letters corresponding to each expiry month are provided in Table 1.1.

Historically, upon expiration of a futures contract, the agreed-upon cash amount was exchanged for the physical good. Nowadays, it is more common to settle contracts with cash, exchanging the difference between the agreed price and the current spot price to simplify the settlement process.

Key	Month
F	January
G	February
H	March
J	April
K	May
M	June
N	July
Q	August
U	September
V	October
X	November
Z	December

Table 1.1: Futures Expiry Months Keys.

1.1.3 Futures' prices

Let $F(t, T)$ denote the price of a futures contract at time t with expiry T. Typically, this price is determined through a replication strategy, where the futures' price is closely linked to the spot price of the underlying asset by constructing a portfolio that mimics the payoff of the futures contract. However, in practice, several factors can influence the price, causing deviations from the spot price. For example, the replicator might incur extra costs such as storage and insurance, and there can be advantages to holding the underlying assets, such as dividends.

The futures curve, often referred to as the forward curve, illustrates the relationship between futures contract prices and their time to maturity. While both terms are sometimes used interchangeably, it's important to note the distinction between them. The futures curve represents the prices of standardized futures contracts traded on an exchange, where these prices are marked to market daily. In contrast, the forward curve represents prices for forward contracts, which are customizable agreements typically traded over-the-counter (OTC) and settled at the end of the contract. Despite these differences, both curves provide a snapshot of the current value of the underlying asset based on anticipated future buy or sell transactions. The shape of the futures curve is crucial for market participants as it influences trading strategies and market expectations. When the futures curve slopes upwards, the spot price is lower than the futures' price, and the market is described as being in *contango*. In this situation, the maturing contract is less expensive $(F(t, T_1) < F(t, T_2))$. The opposite condition, where the spot price is above the futures' price, is described as *backwardation*, and the maturing contract is more expensive $(F(t, T_1) > F(t, T_2))$, as illustrated in the case of WTI Crude Oil in Figure 1.1. Typically, backwardation is favorable for investors with long positions, as they benefit from the futures' prices rising to meet the spot price. As the time to maturity decreases, the futures' price tends to converge to the spot price of the underlying asset, expressed as $F(T, T) = S_T$, with S_t the spot price at time t.

1.1.4 Futures Rolling and Open interest

Due to the expiring nature of futures contracts, indexes or investors typically need to maintain exposure to the underlying asset by selling the near-to-maturity contract and purchasing a later-maturity contract in the same underlying asset. This process is known as futures rolling. The expiring contract, often referred to as the front or expiring contract,

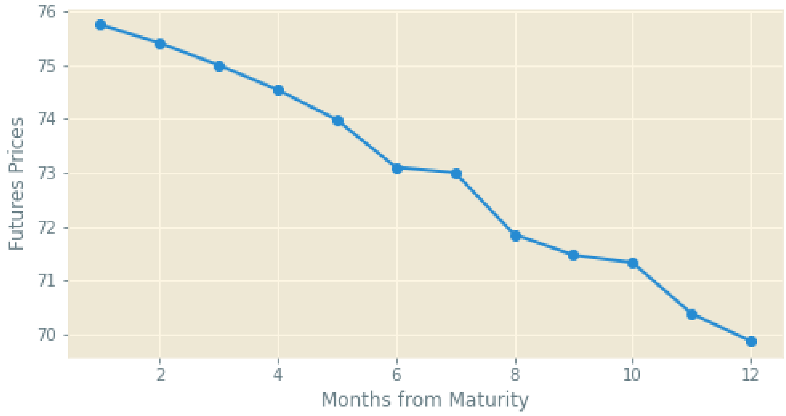


Figure 1.1: Futures curve for WTI Crude Oil (CL) on 2022-01-02.

is the one being rolled from, while the new contract with a later maturity, referred to as the deferred contract, is the one being rolled into. While futures contracts allow investors to gain exposure to an underlying asset, it's important to note that this exposure is not static. Because the futures contracts must be rolled over periodically, the price exposure is effectively reset with each roll. This periodic resetting means that while the futures contracts track the underlying asset, the rolling process introduces a form of revaluation that can reduce direct exposure to short-term volatility of the underlying asset. However, this also means that investors do not lock in a fixed price for the duration of their exposure, but instead continuously adjust to market conditions as each contract expires and a new one is entered. This characteristic can reduce the impact of sudden price swings but also means the position is subject to the ongoing influence of market dynamics at each roll.

The roll yield refers to the return gained from rolling over a futures position from a near-to-maturity contract to a longer-dated contract. Specifically, Mou [24] defines it as "the difference between the log price of the expiring contract that investors roll out of and the log price of the deferred contract they roll into".

A crucial concept in the futures market is open interest, representing the number of contracts that have been traded (opened) but not yet liquidated by an offsetting trade or delivery. Information on open interest is valuable for estimating the number of contracts that might be rolled during a rolling period. Changes in open interest can significantly affect the liquidity of a futures contract. An increase in open interest indicates that new money is flowing into the market, enhancing liquidity and making it easier to enter or exit positions [13]. The CFTC publishes regular Commitment of Traders (COT) reports providing open data on open interest, broken down by trader type and market position, allowing for a good estimation of the rolling behavior of different market participants.

1.2 Commodity Index Markets

In this paper, we focus our analysis on commodity futures markets, with particular emphasis on commodity indices. In the following section, we provide a concise overview of the appeal of commodity index investing, identifying the key participants, highlighting the most significant indices, and examining the evolution of commodity index investing over time.

1.2.1 Who invests and why ?

The most significant theory of portfolio management was developed by Markowitz [22]. According to his pioneering work, investors should construct portfolios to maximize expected returns for a given level of risk. Specifically, an investor should first determine an acceptable level of risk and then adjust their portfolio to match this risk tolerance. Physical commodities, with returns that are relatively uncorrelated with those of more traditional asset classes, have grown in popularity over the years. In addition to their attractive returns, this lack of correlation helps reduce portfolio risk, partly because commodities serve as an inflation hedge [30]. However, physical commodities were rarely used in practice due to the high costs of purchasing and storing items like grain or crude oil. This is where commodity futures became useful. Through futures contracts, investors can achieve the returns of physical commodities synthetically, without holding the actual commodities. Since in an efficient market, the return and risk of the futures position should mirror those of the underlying commodity [30], futures contracts have bridged the gap and become extremely popular over the years.

Typically, futures traders are categorized into two types: hedgers and speculators. As the names suggest, hedgers aim to mitigate risk to which they are exposed, usually holding long positions in physical commodities and short positions in futures markets to offset their price risk exposure. They trade potential financial upside for certainty. Speculators, on the other hand, take on such risk, seeking to profit from price changes. They have a directional view of the markets and take long or short positions based on their predictions of future price movements.

The weekly COT reports provided by the CFTC include aggregate long and short positions of participants in the commodity futures markets, classifying participants into three categories: commercials, non-commercials, and non-reportables. Commercials are associated with hedgers, and non-commercials with speculators. The third category typically includes commodity index traders. These traders are a more recent classification, lacking a physical presence in commodity markets and not taking a directional view on futures' prices. Their primary motivation is to leverage the risk-reducing benefits of futures contracts through buy-and-hold strategies, requiring them to roll their futures positions regularly to maintain this exposure.

1.2.2 Financialisation and the most important indices

The process by which commodity futures have become a popular asset class for portfolio investors is referred to as 'financialization', as described by Cheng and Xiong [11]. The trend in commodity index investment can be divided into four financialization stages [20]: the pre-financialization period (1980 to 1990), the early stage (1991 to 2003), the growth stage (2004 to 2011), and the post-financialization period (post-2012). During the first period, investments in commodities were typically very difficult and rare, as it predates the development of major commodity indexes. Starting in 1991, the development of major commodity indexes attracted significant interest, leading to a surge in investment in commodity futures markets during the growth stage. This boost was partly driven by the conclusion of Gorton and Rouwenhorst [14] that commodities could provide equity-like returns and diversification. After 2012, investment stabilized, maintaining historical highs despite not continuing to grow as rapidly as before.

The two most important commodity indexes are the Standard & Poor's Goldman Sachs Commodity Index (SP-GSCI) and the Bloomberg Commodity Index (BCOM). Introduced in 1991, the SP-GSCI was the first major investable commodity index and includes the most liquid commodity futures. It currently holds 24 commodities from all sectors: energy, industrial metals, precious metals, grains, softs, and livestock. The index follows strict rules, is rebalanced annually, and is rolled monthly from the 5th to the 9th business day, typically referred to as 'Goldman Roll'. It is particularly popular among hedge funds and other traders for portfolio diversification, although it is heavily weighted towards the energy sector, with about 60-70% of its weight dependent on this sector. The BCOM also offers broad-based exposure to commodities but ensures no single commodity or sector dominates the index, aiming to reduce volatility compared to non-diversified commodity investments. Created by AIG in 1998, it was later acquired by UBS in 2009 and Bloomberg in 2020. The BCOM represents all commodity sectors and its roll period lasts five days, from the 6th to the 10th business day of the month.

Over time, many other commodity futures indexes have been developed to meet the diverse needs of market participants. Notable examples include the United States Oil Fund (USO), the UBS Bloomberg Constant Maturity Commodity Index (CMCI), the JP Morgan Commodity Curve Index (JPM CCI), and the DB Liquid Commodity Index. Each of these indexes follows specific rules but must adhere to a rolling schedule to maintain exposure to commodity futures contracts as they expire.

1.3 Arbitrage and the limits to arbitrage

As defined in the lectures, an arbitrage is a portfolio (a.k.a. a trading strategy, or an investing strategy) L that, starting with no initial capital to invest, and without taking any risk, makes money at some point; i.e. an arbitrage is a portfolio L with zero initial capital and with final value V_T^L which satisfies $V_T^L < 0$ with zero probability (i.e. $P(V_T^L < 0) = 0$) and $V_T^L > 0$ with strictly positive probability (i.e. $P(V_T^L > 0) > 0$). Arbitrage theory is a cornerstone of financial economics. Malkiel [21] argues that a market is considered efficient if it incorporates all relevant information completely and accurately into prices. This concept, that he defined as the Efficient Market Hypothesis (EMH), states that all available information are reflected into share prices, making it impossible to consistently achieve higher returns than the overall market on a risk-adjusted basis. Arbitrage opportunities theoretically align with the EMH, as rational arbitrageurs should eliminate price discrepancies quickly. However there exists several limitations to this theory in practice. Indeed, the limits to arbitrage theory suggests that market inefficiencies can persist due to capital available for arbitrage activities [29]. In particular, the effectiveness of arbitrage diminishes when the size of index investments declines and arbitrage capital increases. As highlighted by Mou [24], empirical evidence indicates that anomalies can persist due to slow-moving arbitrage capital and market frictions. In particular, commodities futures markets, which do not have short-selling constraints and offer high leverage, still exhibit anomalies. Such anomalies can be exploited by statistical arbitrage strategies for instance, relying on statistical models to identify and exploit short-term price deviations.

Chapter 2

Literature Review of Commodity Index Futures Price Impact

2.1 Origins

Studying price impact around rolling events in commodity futures markets has gathered significant attention, particularly following the influential 2008 testimony by Michael Masters. Employed by a large hedge fund at the time, Masters asserted that the surge in commodity index investing was driving up futures prices for commodities. This assertion, now known as the "Masters Hypothesis" as termed by Irwin and Sanders [27], suggested that the intense buying pressure from index investors was creating a substantial bubble in commodity futures prices. This bubble, through arbitrage linkages between futures and spot prices, was subsequently affecting spot prices as well. Since the congressional hearing, numerous studies have delved into whether this effect truly exists, focusing not only on the general impact of commodity index investing on futures prices but also examining the specific effects of futures rolling on futures prices.

2.2 Price impact around futures rolls

2.2.1 Mechanism

The main mechanism through which there might be a price impact surrounding futures rolling operates as follows: during a roll month, traders sell nearby futures contracts and purchase second nearby contracts. Figure 2.1 illustrate this process for the case of WTI Crude Oil. It can be seen that, every month, one contract is being traded much more than the other ones. Given that many traders need to sell expiring contracts and buy deferred contracts during the rolling period, it is anticipated that the price of the maturing contract will decrease, while the price of the deferred contract will increase. Specifically, nearby prices are pushed lower than they would be otherwise, and deferred contract prices are pushed higher. This results in a decrease in the price spread between the nearby and deferred futures contracts, often referred to as the order flow cost of roll trades.

Additionally, a large amount of trading activity is expected around roll dates, making these trades and their directions predictable. This predictability can lead to predatory trading, as described by Brunnermeier and Pedersen [8]. Predatory traders may become aware of the need for large-scale liquidations or acquisitions by other traders and exploit this information to their advantage, exacerbating price movements. This phenomenon can contribute to increased volatility and market inefficiencies during roll periods.

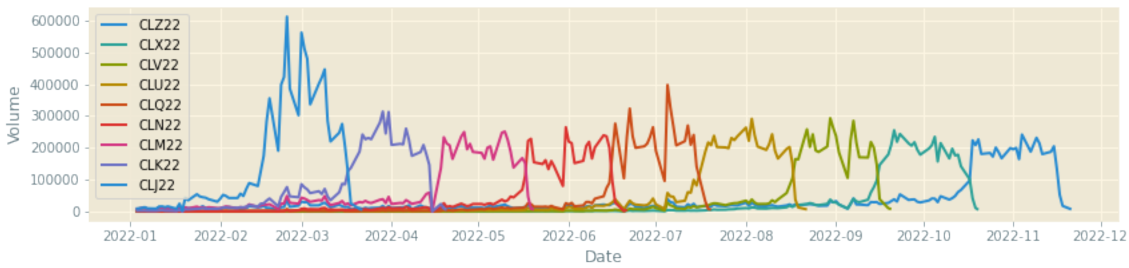


Figure 2.1: Volume trends for WTI Crude Oil (CL) futures.

2.2.2 Related Literature

As hinted at in the introduction, a wide range of studies specifically studied the effects of commodity index investing on commodity futures prices, and all covered at least partially an analysis of price impact surrounding rollover periods ([1, 4, 7, 15, 20, 24, 27, 30]). The methodologies employed by various papers to capture the price impact around futures rolls focused on the relationship between index investments and futures prices through regression analyses, spread calculations, and causality tests.

Mou [24] provides decisive conclusions on the potential price impact during the rolling period. By analyzing data from 19 commodities traded in the SP-GSCI, 17 additional commodities not part of the index, and weekly index trader positions from the COT reports from January 1980 to March 2010, he examines roll yields and spreads through a panel regression of contract returns on the amount of index investment, a dummy variable for index inclusion, and control variables. Mou's findings indicate significant and variable price pressure on the relative prices of the involved contracts during the Goldman roll, and he suggests trading strategies to exploit these opportunities for substantial returns. Building on Mou's methodology but extending the dataset to 2019, Irwin et al. [20] observed that roll impacts were present during the early and growth stages of financialization but disappeared in the post-financialization period. They attribute the disappearance of these opportunities to increased capital investment and arbitrage mechanisms reducing arbitrage possibilities as awareness grew.

Hamilton and Wu [15], using weekly positions from the COT reports and multiple linear regressions of returns against past returns and the amount of index investing, found minimal impact of index fund investing on futures prices, although crude oil futures prices may have been affected. Similarly, Bessembinder et al. [4], employing regression analysis with higher frequency data on oil futures prices, determined that the temporary price impact of an order imbalance on roll days was nearly entirely reversed within one minute in the expiring contract and within three minutes in the second contract.

Stoll and Whaley [30], acknowledging some impact on the crude oil futures market during the index roll period, concluded that commodity index rolls generally have minimal impact on futures prices. Through linear regression of the return differential between nearby and deferred contracts on the number of futures contracts rolled, they found that futures markets could absorb commodity index roll activity without significant price effects. Likewise, Sanders and Irwin [27], by calculating spreads and their changes for a single fund's roll transaction across energy futures markets from 2007 to 2012, found little evidence that rolling activity impacted spreads in the energy futures markets.

Finally, both [1] and [7] studied futures rolling price impact using daily changes in trader position, available in the Large Trader Reporting System (LTRS) report from the CTFC. The former focused on agricultural futures markets from 2003 to 2012 and used GARCH models to examine the spread between daily settlement prices of deferred and nearby contracts, concluding that CITs temporarily increase futures prices in specific maturities, indicating trading opportunities. The latter employed bivariate Granger causality tests with CIT positions, exploring the lead-lag dynamics between financial index trader positions and daily futures returns. Still focusing on agricultural commodities, they found that CIT positions impacted daily returns during the roll period in 5 out of 12 markets. However, contrary to expectations, the cumulative impacts were negative, suggesting that CIT rolling activity simultaneously pressured nearby prices upward and first deferred prices downward.

Related to the predictable futures index rolling, a similar body of research exists around index rebalancing in equities, which examines the price impact and potential arbitrage opportunities associated with predictable changes in index composition. Notably, Shleifer [28] explored the price effects of inclusion in the S&P 500, finding that stocks added to the index experienced a significant price increase, which he attributed to increased demand from index funds. Harris and Gurel [16] provided further evidence of this effect, suggesting that the price impact is largely temporary and driven by liquidity imbalances. More recently, Wurgler and Zhuravskaya [34] examined the phenomenon of downward-sloping demand curves for stocks and how index changes can lead to non-fundamental price movements, reinforcing the idea that index rebalancing created predictable price shifts that could be exploited by arbitrageurs. Adding to this literature, Wang et al. [32] highlighted how the predictable rebalancing of ETFs could be anticipated and exploited by hedge funds, leading to significant price impacts around the rebalancing dates. These studies collectively highlight that predictable changes in index composition, whether in equity or futures markets, create opportunities for market participants to anticipate and profit from the resulting price movements, raising important considerations about market efficiency and the role of index-linked investing.

2.2.3 Limitations

The literature on the price impact of commodity futures rolling presents contrasting results. While there is some indication of potential price effects, it remains unclear whether these impacts are consistent across all markets, persist over time, and are significant enough to exploit through anticipatory arbitrage. Previous studies have predominantly focused on weekly or daily changes in open interest positions, which might miss higher-frequency dynamics where price impacts are more likely to manifest and subsequently reverse. Performing regression of price changes on the changes in positions captured from open interest limits the ability to properly capture a price impact. This limitation arises from the availability of data, as high-frequency trading data was not commonly accessible.

In this study, I have access to limit order book data, including volume traded at a higher frequency, to conduct a more granular analysis. This allows to observe the immediate price changes associated with rolling periods and the subsequent reversals that are not detectable with lower-frequency data. Additionally, the literature has not yet reached a definitive conclusion on the exact impact structure and the order flow cost of rolling large commodity indexes such as the SP-GSCI. By employing a non-linear propagator model, I aim to capture the precise non-linear dynamics of price impacts in the limit order book. Recognizing that price impact is typically a concave, non-linear function of trade size, this study is expected to provide an accurate fit and offer valuable insights into the behavior of price impact in futures markets for both roll and non-roll days.

Chapter 3

Data Description

In this section, we describe the data that is used to perform our analysis. With the objective of analysing the price impact surrounding futures rolling in commodities index, we gather data about the volumes traded and prices of the different commodities included in the SP-GSCI index. Velador Associates provided me with futures data for various commodities sourced from Tick Data[1]. The data files followed the format *XXMYY.csv*, where XX represents the ticker, M denotes the expiry month, and YY indicates the expiry year. Each futures contract's file included the following columns: Symbol, Date, Time, Open, High, Low, Close, Volume, Tick Count, Date, and Time. The data was available at a minute-by-minute frequency throughout the trading day. For most commodities, the contracts were available from 1990 to the end of 2023. However, *Volume* data was mainly available from 2008 onwards. Note that the direction of each trade was not indicated, hence we need to estimate them. We show how this is done in Section 3.3.2 below.

3.1 Commodities in the SP-GSCI Index

We focus our analysis on the commodities included in the Standard and Poor's Goldman Sachs Commodity Index due to its position as the main commodity futures index, and the predictability of its rolling period. Indeed, the SP-GSCI rolls contracts forwards from the 5th to the 9th business day of each month where the contracts are expiring. Additionally, the composition of the index is known, as well as the rolling schedule for each commodity, making it easy to predict which contracts will be rolled each month. The SP-GSCI contains 24 commodities, however, due to data availability, we discard 5 commodities and focus on the 19 commodities for which we have futures data available. They are listed in Table 3.1 and span four sectors: agriculture, livestock, energy, and metals.

Each commodity has unique expiry months and trading frequencies. For example, Crude Oil (WTI), with the ticker *CL*, has a new contract expiring every month. In contrast, Corn, with the ticker *CN*, only has contracts expiring in February, April, June, August, and December, as illustrated in Figure 3.1.

3.2 Descriptive Statistics

Table 3.2 presents descriptive statistics for the key variables associated with each commodity. Panel A details the daily closing prices, while Panel B provides information on the daily traded volumes. The data reveals considerable fluctuations in closing prices across all commodities, with standard deviations notably large relative to the mean prices. Among the commodities, *CC* (Cocoa) emerges as the most expensive, while *HO* (Heating Oil) is

[1] Available at `https://www.tickdata.com/`.

Commodity	Sector	Trading Facility	Ticker	Maturity of contracts at Month Begin
Chicago Wheat	Agriculture	CBT	WC	H, H, K, K, N, N, U, U, Z, Z, Z, H
Kansas Wheat	Agriculture	KBT	KW	H, H, K, K, N, N, U, U, Z, Z, Z, H
Corn	Agriculture	CBT	CN	H, H, K, K, N, N, U, U, Z, Z, Z, H
Soybeans	Agriculture	CBT	SY	H, H, K, K, N, N, X, X, X, X, F, F
Coffee	Agriculture	ICE-US	KC	H, H, K, K, N, N, U, U, Z, Z, Z, H
Sugar	Agriculture	ICE-US	SB	H, H, K, K, N, N, V, V, V, H, H, H
Cocoa	Agriculture	ICE-US	CC	H, H, K, K, N, N, U, U, Z, Z, Z, H
Cotton	Agriculture	ICE-US	CT	H, H, K, K, N, N, Z, Z, Z, Z, Z, H
Lean Hogs	Livestock	CME	LH	G, J, J, M, M, N, Q, V, V, Z, Z, G
Live Cattle	Livestock	CME	LC	G, J, J, M, M, Q, Q, V, V, Z, Z, G
Feeder Cattle	Livestock	CME	FC	H, H, J, K, Q, Q, Q, U, V, X, F, F
WTI Crude Oil	Energy	NYM/ICE	CL	G, H, J, K, M, N, Q, U, V, X, Z, F
Heating Oil	Energy	NYM	HO	G, H, J, K, M, N, Q, U, V, X, Z, F
RBOB Gasoline	Energy	NYM	XB	G, H, J, K, M, N, Q, U, V, X, Z, F
Brent Crude Oil	Energy	ICE-UK	CO	H, J, K, M, N, Q, U, V, X, Z, F, G
Gasoil	Energy	ICE-UK	GO	G, H, J, K, M, N, Q, U, V, X, Z, F
Natural Gas	Energy	NYM/ICE	NG	G, H, J, K, M, N, Q, U, V, X, Z, F
Gold	Metals	CMX	GC	G, J, J, M, M, Q, Q, Z, Z, Z, Z, G
Silver	Metals	CMX	SV	H, H, K, K, N, N, U, U, Z, Z, Z, H

Table 3.1: Commodity Futures in the SP-GSCI Index.

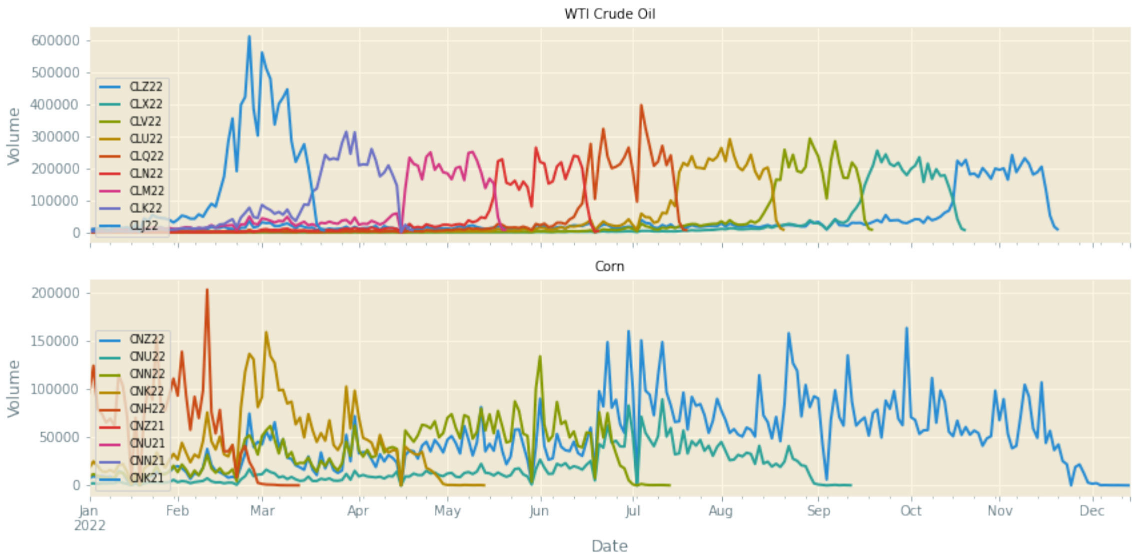


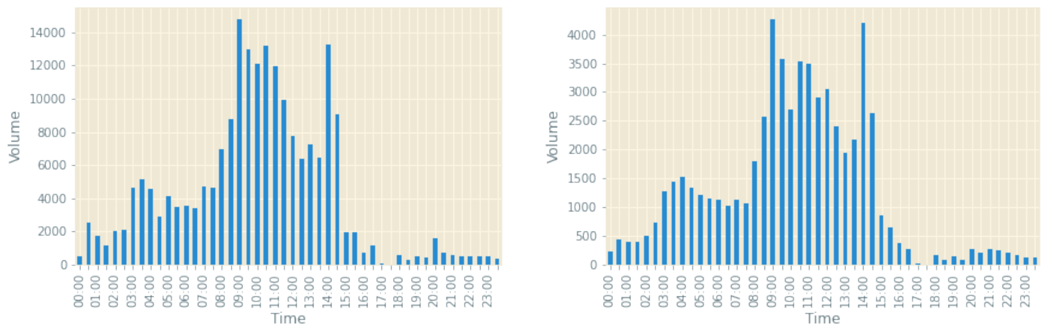
Figure 3.1: Volume trends for WTI Crude Oil (CL) and Corn (CN) futures.

the least costly future contract. Regarding traded volumes, there is significant variation, including days with no trading activity for all contracts, likely due to data limitations during the early periods of the analysis, alongside notably high maximums. Predictably, the CL contracts are the most actively traded, whereas *FC* (Feeder Cattle) appears to attract less interest from investors. The number of contracts analyzed for each commodity varies, reflecting the differing numbers of contracts traded annually.

Table 3.2: Descripive Statistics of commodity futures in the SP-GSCI index.

Ticker	mean	std	min	max	contracts
		Panel A: Daily Closing Prices			
WC	480.080	182.818	230.750	1326.500	174
KW	593.359	177.076	368.250	1340.750	54
CN	363.434	153.394	174.250	831.500	173
SY	880.909	334.291	411.500	1786.250	174
KC	126.115	49.514	42.000	314.000	174
SB	13.434	5.529	4.280	35.280	140
CC	2049.022	1087.143	680.000	12250.000	174
CT	71.556	21.684	28.520	219.700	139
LH	66.613	18.058	27.725	132.350	243
LC	97.708	30.294	54.800	189.850	207
FC	117.735	44.620	47.650	264.400	276
CL	50.716	29.419	1.430	145.320	413
HO	1.545	0.961	0.296	4.944	413
XB	2.157	0.638	0.443	4.308	212
CO	74.069	25.624	19.500	146.610	229
GO	700.095	218.759	193.000	1360.250	175
NG	3.973	2.199	1.300	15.491	377
GC	886.577	580.549	253.800	2441.300	173
SV	1306.521	886.266	351.000	4843.000	173
		Panel B: Daily Traded Volumes			
WC	12446.761	19575.821	0.000	178753.000	174
KW	8012.654	7523.945	0.000	48243.000	54
CN	29541.307	46852.840	0.000	411359.000	173
SY	22675.067	34383.585	0.000	245468.000	174
KC	3282.438	5873.848	0.000	40899.000	174
SB	11224.377	18336.053	0.000	156111.000	140
CC	2909.225	5602.919	0.000	36913.000	174
CT	3391.068	5797.971	0.000	48494.000	139
LH	3680.257	5968.874	0.000	34239.000	243
LC	4433.376	7344.732	0.000	48038.000	207
FC	943.513	1727.536	0.000	11396.000	276
CL	88132.313	154171.639	0.000	1164773.000	413
HO	6452.978	10814.775	0.000	123633.000	413
XB	10766.742	11008.086	0.000	73068.000	212
CO	68846.423	75586.300	0.000	475019.000	229
GO	24737.454	20278.522	0.000	137358.000	175
NG	29402.402	42968.415	0.000	333756.000	377
GC	54790.087	86390.351	0.000	618848.000	173
SV	16305.212	26784.280	0.000	337765.000	173

In Tables A.1 and A.2 in the appendix, where descriptive statistics are separately calculated for Goldman roll days and non-roll days, a significant difference in daily traded volumes between these two periods is evident. Notably, trading activity intensifies during roll days, as exemplified by the CL contract, which averages around 152,585 contracts traded during rolling periods, compared to only 67,264 contracts on non-rolling days. Figure 3.2 further illustrates this by showing the variation in hourly traded volumes for a single CL contract on roll versus non-roll days. This observation reinforces the notion that the rolling over of futures contracts could have a price impact, as higher trading volumes typically contribute to a more substantial price impact.



(a) Roll days.

(b) Other days

Figure 3.2: Comparison of volume traded for CLZ22 during roll and non-roll days..

3.3 Data Processing

It was necessary to pre-process the data in several ways to prepare it for price impact analysis using the non-linear propagator model and for designing trading strategies. The various steps included combining different futures contracts, reconstructing the order sign from the traded volume, normalizing the data for comparability across stocks, and resampling the data to trade at various time granularities.

3.3.1 Futures rolling

Given the nature of futures contract, and their frequent expiration, there is a need to combine different contracts together in order to perform the price impact analysis as described in Section 4.1.2. Specifically, we need to obtain a single continuous database of prices and volumes over the whole time frame to be able to use continuous training and testing sample. As shown in Table 3.1, the traded contracts are known for each commodity and every month, and we also know when the contracts are being rolled over in the SP-GSCI index. Therefore, to combine the different contracts together, and to be able to fit the model to the actively traded contracts that are held by the SP-GSCI, we decide to only keep the active contracts as described in the table, and switch from one contract to the next at the end of the rolling period of the SP-GSCI index. The resulting 'synthetic' continuous contract is illustrated in Figure 3.3 for the case of HO (Heating Oil) and SY (Soybeans). It can be seen that the technique produces consistently the most traded contract, allowing for an adequate study of price impact of futures contracts trading activity.

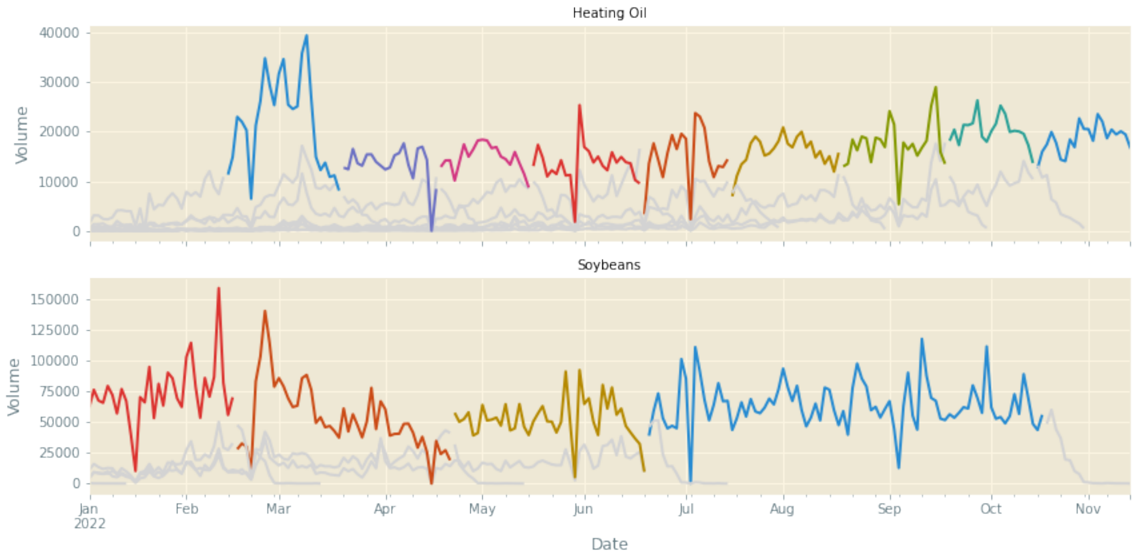


Figure 3.3: Synthetic contracts for HO and SY futures.

3.3.2 Order Sign

In order to properly calibrate the price impact models, it is important to know if the trade was a buy or a sell order, namely, it is important to get the sign of the trade volume. Unfortunately, the data files provided only contains volume data without specifying the order of the operation. Hence, without access to specific limit order book data indicating whether the transaction was buyer or seller initiated, an algorithm is needed to reconstruct the signed order flow. For this purpose, we rely on the tick classification rule introduced by Holthausen and Leftwich [19, Section 3.3, page 244]. Specifically, they assume that transactions are seller-initiated if they trade on a downtick, and buyer-initiated if they trade on an uptick. In particular this means that if the change in price from the prior price is negative, the transaction is classified as seller-initiated, whereas if it is positive, it is classified as buyer-initiated.

3.3.3 Normalisation

To ensure comparability of the impact coefficients across all commodity futures, a normalization of the data is performed following [25]. Specifically, we normalize the data using a daily volatility estimate for the future's price, σ, and an estimate of the daily traded volume of the future, ADV. Both these variables are computed using a rolling window and are defined as

$$ADV_i(k) = \frac{1}{k}\sum_{j=1}^{k} v_{i-j},$$

$$\sigma_i(k) = \frac{1}{k}\sum_{j=1}^{k} \sigma_{i-j},$$

for day i, with $k = 20$ (rolling window over the past month, ≈ 20 days), v_i the signed volume traded on day i, and σ_i the standard deviation of returns on day i.

3.3.4 Resampling

The data used in our analysis is available at a minute-by-minute frequency. We chose to use the *Close* column as the true price within each minute because it represents the last available price in the one-minute binned data. For our trading strategy, which operates on a daily frequency, we resample the data by using the last available price for each day as the daily closing price. For the opening price of each day, we use the last available price from the previous day. However, it is important to note that for trading decisions at the start of each bin (e.g., each day), we do not have knowledge of the closing price at the end of that bin. Therefore, in our strategy, the previous bin's closing price is used as the opening price for the next bin, acknowledging that this introduces a lag in the price used for decision-making. This approach ensures that our strategy is based on information that would realistically be available at the time of execution.

Chapter 4

Price Impact Analysis

4.1 Methods

Price impact is a reaction to order-flow imbalance. Indeed, trading of a stock causes price moves for the stock that otherwise would not have happened [33]. Over time, a vast amount of models have been developed to best capture this price impact. In this section, we describe the framework for using the non-linear propagator model in a discrete setting as proposed by Muhle-Karbe et al. [25] based on the original model developed by Bouchaud et al. [6], which describes how each trade's price impact propagates across time.

4.1.1 Propagator Models

Let Q be our position and dQ our trades, S be the price if we didn't trade, $Q = 0$. Then the price when we trade is $P = S + I(Q)$ where $I(Q)$ is the price impact of Q. This is the actual price which is shifted relative to its fundamental level S by the order flow.

In the propagator model, each trade has an immediate impact described by a function $g(\cdot)$ which gradually diminishes over time according to a decay kernel. In particular, the dynamics of the price impact for an asset are [25, Definition 2.1, page 4]:

$$\Delta I_n = -\beta I_{n-1}\Delta t + g(\Delta Q_n)$$

where $\beta > 0$ and $g \in C^1$ is some odd function that is concave on $[0, \infty]$. The decay parameter β measures the timescale of price impact's reversal. Hence, the term $-\beta I_{n-1}\Delta t$ allows the price shift to dissipate over time. The price impact of buy and sell orders is symmetric, and concavity implies that trading twice the amount at most doubles the price impact [25].

As seen in the lectures, it is possible to implement this specification incrementally to reduce computational cost, by using the recursive relationship existing between impacts:

$$I_n = (1 - \beta\Delta t)I_{n-1} + g(\Delta Q_n).$$

Several versions of the functional forms have been used in the literature. We briefly cover a few of them below, before detailing the functional form that will be used in this analysis.

Obizhaeva-Wang Model

The simplest case is the original OW model, developed in [26] where each trade has linear impact that decays exponentially, i.e. $f(x) \propto x$:

$$I_n - I_{n-1} = -\beta I_{n-1}\Delta t + \lambda\Delta Q_n.$$

This baseline model is an exponential moving average (EMA) on $\lambda\Delta Q_n$, and β, λ are the only parameters to estimate. λ measures the sensitivity of prices to order flow, and is often referred to as 'Kyle's lambda'. Note that a smaller λ indicates smaller market impact.

Non-linear Propagator Model

In the original propagator model, Bouchaud et al. [6] suggested the use of a functional form such as $f(x) \propto log(x)$. However, after reviewing empirical studies of concave propagator models, and to improve the model's fit by passing to a concave parametric class of functions, Bouchaud et al. [5] suggested the use of $f(x) \propto x^p$ with $p \in [0.2, 0.5]$ to capture the non-linear behavior of price impact. The most well known model specification here is the square root rule, where the impact decays as follows:

$$I_n - I_{n-1} = -\beta I_{n-1}\Delta t + \lambda sign(\Delta Q_n)\sqrt{|\Delta Q_n|}$$

This is an EMA on $\lambda sign(\Delta Q_n)\sqrt{|\Delta Q_n|}$. The advantage of this specification is that the model allows to capture the non-linearity in Q, but is still linear in λ. Therefore, one still fits the model via linear regression on the non-linear features.

Reduced Form Model

Muhle-Karbe et al. [25] also suggest the use of reduced form version of the propagator model where the square-root impact is proxied by a stochastic liquidity parameter $\lambda_t \propto 1/\sqrt{v_t}$, leading the corresponding linear price impact model as:

$$I_n - I_{n-1} = -\beta I_{n-1}\Delta t + \lambda \frac{\Delta Q_n}{\sqrt{v_t}}.$$

This is a parametric form of the extended OW model of Fruth et al. [12] and is motivated by the lack of tractability of non-linear propagator models in general. Indeed, Muhle-Karbe et al. [25] argue that this approach provides a bridge between linear price impacts and non-linear propagator models by approximating non-linear discrete-time models with linear continuous-time models, thereby linking the unobservable illiquidity parameter to measurable market activity as proxied by moving averages of trading volumes.

4.1.2 Our Model

To identify the most accurate functional form for modeling price impact across different commodities, we need a function that facilitates testing multiple forms in a straightforward manner. The propagator model, with a functional form $g(x) \propto x^c$, is a natural choice due to its simplicity and widespread use in the literature. To explore a broad range of functional forms and achieve the best fit, we employ the function $g(x) = \lambda\,\mathrm{sign}(x)\,|x|^c$, where $c \in [0.1, 1]$. This approach allows us to determine the optimal concavity coefficient for each commodity future.

Specifically, an EMA is applied to the expression $\lambda sign(\Delta Q_n)|\Delta Q_n|^c$, preserving the linearity with respect to λ while introducing non-linearity in the relationship between order flow Q and price impact. This approach allows the model to be fitted using linear regression on these non-linear features, thereby enabling the determination of the optimal concavity for the price impact. The resulting price impact model is given by:

$$I_n - I_{n-1} = -\beta I_{n-1}\Delta t + \lambda sign(\Delta Q_n)|\Delta Q_n|^c. \tag{4.1.1}$$

By fitting this model, we can assess how changes in trading volume affect prices and adjust the concavity coefficient c to achieve the most accurate representation of price impact in the market.

Choice of parameters

Given the specification 4.1.1, several key parameters still need to be determined: the bin size Δt, the decay parameter β, and the prediction horizon h. While a grid search for each of these parameters could have been conducted, the associated computational cost would have been substantial. Moreover, existing literature suggests that the model fit is not significantly impacted by these specific parameters. Therefore, we are confident that the reduction in computational cost achieved by fixing these parameters outweighs the potential drawbacks of not optimizing each parameter individually.

In their study, Muhle-Karbe, Wang, and Webster [25] grouped data into 10-second bins to reduce data size and standardize the number of data points across different days and stocks. However, due to the granularity of our data, which is available only on a minute-by-minute basis, we use 60-second bins instead. This choice does not significantly impact the accuracy of our model fitting, as previous research found negligible differences in R^2 when using bins ranging from 10 to 300 seconds [23, 9].

Next, we need to determine the decay parameter β. In their work, Muhle-Karbe et al. [25] set β to 0.7, which corresponds to a price impact half-life of $log(2)/\beta = 60$ minutes. They also demonstrate that performing a grid search over β has minimal effect on the results for prediction horizons ranging from 1 to 120 minutes. Therefore, as long as the prediction horizon remains within a reasonable range, the model fit should not be significantly impacted. Following the approach discussed in the lectures, we set $\beta = log(2)/3600$ to achieve a half-life of 1 hour.

Lastly, the behavior of price impact varies across different time horizons. In [25], the authors examine prediction horizons of $h = 1$, 15, and 60 minutes. For a given horizon h, they define the horizon-specific return as

$$\Delta_t^h P = (P_{t+h} - P_t)/P_t,$$

and the impact return as

$$\Delta_t^h I = I_{t+h} - I_t.$$

They then perform simple linear regressions for each horizon using the model

$$\Delta^h P = \Delta^h I(\lambda) + \epsilon$$

and evaluate the R^2 of the regression. As expected, shorter time horizons yielded higher in- and out-of-sample R^2 values. Therefore, to better understand the price change over a specific time interval due to the volume traded within that interval and to optimize the model fit, we set $h = \Delta t$, corresponding to 60 seconds.

Estimation

With the set-up clearly defined and the remaining parameters fixed, it is now possible to estimate the model. We are interested in the price impact's magnitude, as estimated by λ, with a higher value indicating a higher impact. To ensure comparability of the impact coefficients across all commodities, a normalization of the data is performed as described in 3.3.3. The resulting impacts are

$$\Delta I_n = -\beta \Delta t I_{n-1} + \sigma \lambda sign\left(\frac{\Delta Q_n}{ADV}\right)\left|\frac{\Delta Q_n}{ADV}\right|^c. \tag{4.1.2}$$

For each $c \in \{0.1, 0.15, ..., 0.95, 1\}$, we originally pre-compute the impacts by setting $\lambda = 1$ in 4.1.2. Next, we estimate λ using linear regression in

$$\Delta^h P = \lambda \Delta^h I(1) + \epsilon,$$

taking advantage of the linearity of the model in λ. Note that λ is fitted for every hour of each trading day. This is motivated by the fact that liquidity varies across the day with both trading volumes and price volatility higher at the start and at the end of the day. Indeed, Muhle-Karbe et al. [25] state that "Refiting λ to account for time-of-day effects significantly improves the model", and "the R^2 of the non-linear propagator model benefits from the refitting". Consequently, for every value of c, we obtain a corresponding value of λ for each hour during the trading day.

Finally, to identify the optimal combinations of parameters that best fit the data, we adopt the methodology from [25]. This involves splitting the sample into in-sample and out-of-sample parts. The model parameters are fitted using the in-sample data, and their performance is then evaluated on the out-of-sample data. In this study, we use a moving window approach with 6 months of training data and 3 months of testing data.

Evaluation

All statistical estimations are performed on a commodity-by-commodity basis using a 6-month training sample, followed by a subsequent 3-month validation sample. To evaluate the performance of each model, we provide performance statistics averaged across all regressions. Specifically, we examine the in-sample (IS) and out-of-sample (OOS) R^2 to determine which pair best fits the data in both contexts. The in-sample R^2 shows how well the model fits the data with a poor IS R^2 indicating an implementation bug or poor model features. The out-of-sample R^2 indicates how well the model works out of sample, with a poor OOS R^2 indicating overfitting compared to IS R^2.

Additionally, following [25], to assess the model's robustness across the universe of commodity futures and the time period, we consider the estimator's t-statistic:

$$\frac{\mathrm{mean}(\hat{\lambda})}{\mathrm{std}(\hat{\lambda})}.$$

The mean and standard deviation are computed across all λ from different (commodity, sample) pairs and help assess how stable the model is across all pairs. It can be generalized over any dimension that one wants λ stable over and a poor t-stat indicates that the model is incorrectly normalized or does not fit well. A higher t-stat, all else being equal, implies a more stable model. Comparability across commodities is achieved through normalizations, which make λ unitless and comparable across the cross-section.

4.2 Results of the calibration

We now present the results of our price impact analysis on commodity futures. We begin by discussing the overall fit for all commodities, then move on to a comparative analysis by sector. Finally, we conduct an in-depth examination of the energy sector. Due to the limited availability in volume data before 2010, we restrict the analysis to the period from 2010 onwards for the purpose of the calibration.

4.2.1 Overall results

General Analysis

Table 4.1 provides a summary and comparison of each model's performance at a 1-minute horizon ($h = 1$ minute). The performance metrics for the different impact models are averaged across all commodities and years since 2010, revealing a clear hierarchy among the models. Notably, the model with a concavity coefficient of $c = 0.1$ shows strong performance, achieving a training R^2 of 0.650 and a test R^2 of 0.646. This indicates that the price impact model with a strictly concave coefficient effectively captures the dynamics of price impact in commodity futures markets. The model's out-of-sample fit improves as the concavity coefficient increases to 0.25, reaching its peak performance at $c^* = 0.25$, where it achieves an out-of-sample R^2 of 0.665 with a t-statistic of 3.353. A higher t-statistic suggests greater model stability, reinforcing that the model with this optimal concavity coefficient outperforms the others in terms of fit and stability. Hence, the model with $c^* = 0.25$ performs well in terms of goodness of fit, and seems to be a good choice for generalizing across different commodities and years. Beyond this point, the model's fit deteriorates linearly as the concavity coefficient increases further, as illustrated in Figure 4.1. This is an important results as it confirms that the lower concavity worked better than the square root ($c = 0.5$ with out-of-sample R^2 of 0.621) and the out-of-sample R^2 seems to be the largest for the most extreme concavities. The contrast becomes even more evident when comparing the optimal model to the commonly used linear impact model with a coefficient of $c = 1$. The linear model achieves an in-sample R^2 of 0.399 and an out-of-sample R^2 of 0.350. Although these results are reasonable on their own, the optimal model delivers significant improvements, with approximately 85% better out-of-sample fit. Additionally, the linear model demonstrates lower stability across commodities and time, as indicated by its t-statistic of 1.161. Therefore, the optimal price impact model performs very well in-sample, out-of-sample and its model shows great stability across all commodities and years.

impact c	Train R^2	Test R^2	T-Stat
0.10	0.650	0.646	3.097
0.15	0.660	0.656	3.291
0.20	0.666	0.663	3.381
0.25	0.669	0.665	3.353
0.30	0.668	0.664	3.223
0.35	0.663	0.658	3.028
0.40	0.655	0.644	2.807
0.45	0.643	0.637	2.583
0.50	0.629	0.621	2.371
0.55	0.612	0.602	2.178
0.60	0.592	0.581	2.005
0.65	0.571	0.557	1.852
0.70	0.548	0.531	1.716
0.75	0.524	0.504	1.596
0.80	0.499	0.475	1.490
0.85	0.474	0.445	1.394
0.90	0.449	0.414	1.309
0.95	0.424	0.382	1.231
1.00	0.399	0.350	1.161

Table 4.1: Performance of various price impact models for all commodities.

To determine whether the results would be similar when fitting the model on data from a single year, we present a similar analysis, but solely focused on data from the year 2023. The results are shown in Table 4.2. From this table, it is visible that the model fit improves slightly for all concavity coefficients, and the models exhibit greater stability across commodities. Once again, a clear ranking of the models emerges, with the model fit improving from $c = 0.1$ to the optimal value of $c^* = 0.25$, where the out(in)-of-sample R^2 reaches 0.687(0.681). Additionally, the t-statistic (3.724) confirms that the model with

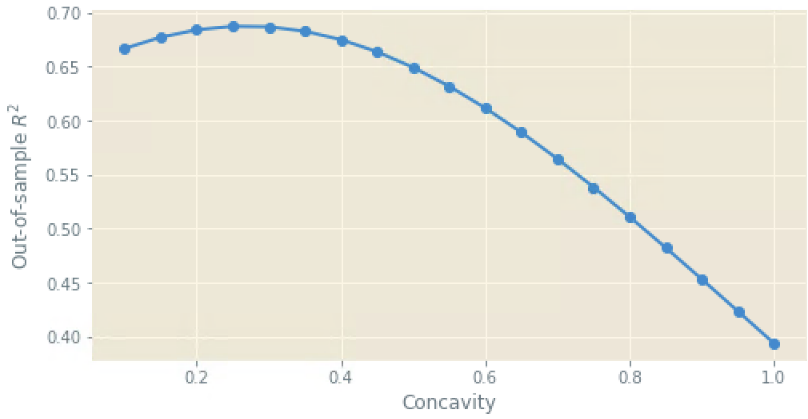


Figure 4.1: Out-of-sample R^2 against coefficient of concavity c for all commodities since 2010-01-01.

$c^* = 0.25$ is a strong candidate for generalizing across commodities. As observed in the broader dataset analysis, the fit deteriorates linearly as the concavity coefficient increases beyond $c^* = 0.25$. The improvement in fit compared to the linear impact model with $c = 1$ is significant, showing a 60% better in-sample fit and a 74% better out-of-sample fit. Moreover, the linear model is less stable, as indicated by a t-statistic of 1.407.

impact c	Train R^2	Test R^2	T-Stat
0.10	0.659	0.667	3.185
0.15	0.670	0.677	3.506
0.20	0.677	0.684	3.706
0.25	0.681	0.687	3.724
0.30	0.681	0.687	3.574
0.35	0.678	0.683	3.323
0.40	0.671	0.675	3.041
0.45	0.661	0.664	2.773
0.50	0.648	0.649	2.536
0.55	0.632	0.632	2.335
0.60	0.614	0.612	2.167
0.65	0.594	0.589	2.027
0.70	0.571	0.565	1.909
0.75	0.548	0.538	1.808
0.80	0.523	0.511	1.718
0.85	0.498	0.482	1.635
0.90	0.472	0.453	1.556
0.95	0.446	0.424	1.480
1.00	0.421	0.394	1.407

Table 4.2: Performance of various price impact models for all commodities since 2023.

In general, these results confirm that it is indeed possible to generalize the model across commodities and years without significantly deteriorating the performance of the model. As expected, the model performs better out-of-sample when it is fitted on more recent data. Indeed, commodity futures markets have evolved significantly over time. Changes in market structure, regulatory environments, and the increasing role of algorithmic trading and institutional investment have altered the dynamics of price impact. For instance, research indicates that the financialization of commodity markets in the early 2000s, driven by the influx of institutional investors, fundamentally changed the behavior of these markets [11]. Similarly, shifts in market liquidity, trading volume, and the emergence of new trading strategies over the years can lead to variations in price impact that a model might not fully capture when applied across different time periods. However, while the model shows strong performance within a single year like 2023, it is a good sign that it is able to generalize relatively well across multiple years. This suggests that, while models could be periodically recalibrated or adapted to account for these evolving market conditions, this is not a necessity and it is possible to maintain a great accuracy and stability when fitting the model over multiple years.

Optimal Impact Analysis

After having analyzed the performance averaged across all years and commodities, we now turn to the analysis of each commodity separately. Table 4.3 summarizes the R^2 results for all commodity futures analyzed. These R^2 values correspond to the optimal coefficients of impact and concavity, averaged over the period since 2010. For each commodity, the table presents the optimal concavity and impact coefficients, along with the R^2 values and the t-statistic, which helps assess the stability of the model's fit over time.

	Optimal Impact	λ	Train R^2	Test R^2	T-stat
WC	0.20	6.245	0.749	0.750	7.024
KW	0.20	6.931	0.716	0.719	5.962
CN	0.10	3.448	0.800	0.796	5.189
SY	0.25	8.117	0.728	0.728	6.335
KC	0.25	10.788	0.694	0.692	8.945
SB	0.20	6.618	0.766	0.767	8.746
CC	0.25	8.460	0.701	0.697	8.188
CT	0.30	11.184	0.586	0.586	4.926
LH	0.25	10.462	0.656	0.655	3.750
LC	0.25	7.756	0.663	0.655	4.126
FC	0.25	10.163	0.605	0.564	3.269
CL	0.45	28.045	0.675	0.661	3.718
HO	0.25	7.147	0.566	0.560	3.896
XB	0.20	7.399	0.552	0.544	3.686
CO	0.25	6.696	0.618	0.617	4.213
GO	0.15	4.305	0.729	0.728	3.542
NG	0.25	17.530	0.691	0.693	2.902
GC	0.50	35.899	0.690	0.686	4.754
SV	0.35	13.848	0.719	0.717	5.448

Table 4.3: Average R^2 results for best model for each commodity.

A key observation is that the optimal concavity coefficient for all commodities lies within the range of $[0.1, 0.5]$. Interestingly, with the exception of *Gold (GC)* and *Crude Oil (CL)*, the concavity coefficient does not exceed 0.3. This suggests that strictly concave impact models are particularly effective in capturing the dynamics of price impact over time, highlighting the importance of incorporating this concavity in modeling. Given that the scale of c reflects certain characteristics of the order book for a particular future, the optimal values indicate that even a relatively small amount of trading can significantly impact prices, yet the difference in impact between small and large trades is minimal. This trend of low c values across most commodities suggests that it doesn't take much trading to generate price impacts in commodity futures markets.

Due to the data normalization, differences in the λ coefficients across commodities can be interpreted meaningfully. A higher coefficient indicates a greater price impact relative to the volume traded as a percentage of the total daily volume. The highest impact coefficient is observed for *Gold (GC)* with a value of 35.9, followed by *Crude Oil (CL)* at 28 and *Natural Gas (NG)* at 17.5. These findings are particularly surprising given that Gold, Crude Oil, and Natural Gas are typically considered highly liquid assets. In general, more liquid assets have lower price impacts because their high trading volumes can absorb large trades without causing substantial price movements. The expectation would be for these commodities to exhibit lower price impact coefficients, reflecting their liquidity. However, the observed higher coefficients suggest that even small trades can significantly affect the prices of these commodities. This could be due to various factors, such as market structure, the nature of trading in these specific commodities, or periods of heightened market sensitivity. Another possible explanation lies in the rolling of futures contracts, which can lead to increased trading volume and frequency. If there is a clear imbalance in the direction of these trades, it could result in a more significant price impact.

The in-sample and out-of-sample R^2 values provide valuable insight into the model's performance for all commodities over time. The in-sample R^2 values consistently exceed 0.5 for all commodities and are generally within the range of $[0.6, 0.75]$. This indicates that the model fits the historical data well. Notably, for all commodities, the performance remains robust when considering the out-of-sample R^2, suggesting that the model is effective in predicting price impact on unseen data. For example, *Corn (CN)* demonstrates an in-sample R^2 of 0.8 and an out-of-sample R^2 of 0.796, indicating that the model's predictive power is retained even when applied to new data. Additionally, the t-statistic is 5.189, signaling exceptional model stability over time. The strong performance of the model across all commodities, particularly its stability and predictive power, indicates that the underlying price impact mechanisms are consistent over time and can be effectively captured by the model.

4.2.2 Analysis by sector

Next, we compare results across different sectors, with a particular focus on the energy sector due to its significant influence on the SP-GSCI index. Table 4.4 presents the optimal concavity coefficient, as well as the average impact coefficient and the average R^2 across all commodities and all years within each sector.

	Optimal Impact	λ	Train R^2	Test R^2	T-stat
Agriculture	0.2	6.289	0.715	0.714	5.348
Energy	0.30	10.824	0.633	0.628	2.962
Livestock	0.2	7.617	0.641	0.624	3.854
Metals	0.45	26.472	0.700	0.697	4.559

Table 4.4: Average R^2 results for the best model for each sector.

The table highlights several important findings related to the performance and stability of the optimal price impact model across different commodity sectors. Firstly, the optimal impact coefficient (ranging from 0.2 to 0.45) confirms the superiority of a strictly concave, non-linear model over a linear one when analyzing price impact in commodity futures markets. This non-linearity captures the nuanced price dynamics more effectively, as demonstrated by the results across all sectors. Next, the metals sector stands out for its stability, with a high T-stat of 4.559, indicating robust model fit both in- and out-of-sample. This stability is likely due to the fact that this sector includes only two commodities, Gold (GC) and Silver (SV), which share similar characteristics, making the generalization easier and the model more consistent across time. Additionally, commodities in the metals sector suffer the highest price impact ($\lambda = 26.472$), largely due to the high impact found for GC. The Energy and Livestock sectors show similar patterns, with relatively high λ values (10.824 and 7.617, respectively), suggesting significant price impact within these sectors. The reasonable stability and predictive performance in both training and test data (R^2 around 0.63) indicate that the models capture the price impact dynamics well for these commodities. Finally, the Agriculture sector has the lowest optimal concavity coefficient and impact λ, and it achieves the best performance in terms of model fit. While it achieves a good in-sample R^2 of 0.715, its out-of-sample performance remains impressive at 0.714, reflecting great generalization. Unsurprisingly, this sector also shows the higher stability (T-stat of 5.348), indicating that the model's fit generalizes well across different time periods and commodities within the Agriculture sector.

Energy Sector

As mentioned in Section 1.2.2, about 60 to 70% of the weight of the SP-GSCI index is dependent on the energy sector. For this reason, we present a deeper dive into this sector, and present the performance of the different models, an analysis of intraday impact of the trades, and the evolution of the performance of the different models over time. The corresponding figures and tables are available in the appendix for the other sectors.

Table 4.5 provides a summary and comparison of each model's performance at a 1-minute horizon averaged across all commodities in the energy sector and years, revealing again a clear hierarchy among the models. Notably, it can be seen that the model fit improves when increasing the concavity coefficient to reach optimality at $c^* = \{0.25, 0.30\}$ with in sample R^2's of 0.633, and out-of-sample R^2's of 0.628. Additionnaly, the t-statistic of 3.250 (for $c = 0.25$) implies that the model is relatively stable over commodities and time. The performance of the model then deteriorates when increasing the concavity coefficient. Most importantly, the difference is again striking compared to the linear model with $c = 1$. Indeed, the linear model achieves R^2's of 0.394 (training) and 0.351 (testing), implying an improvement of about 61% and 79% in and out-of-sample respectively. Similar to most commodities, the same conclusion is obtained that not much trading is required to have an impact, and there is little difference in impact between a small trade and a big trade.

Intraday Effect

To dive deeper into the intraday effects of trades, Figure 4.2 provides the fit results for the commodities in the energy sector with the test period starting on 2024-01-01. From the figure, the in- and out-of-sample R^2 are plotted, as well as the coefficient of impact λ. The plots are shown for the coefficient of concavity that had the best average in-sample R^2.

The figure shows that for all commodities, the model exhibits strong consistency between in-sample and out-of-sample performance, as indicated by the close alignment of the blue and green curves. This suggests that the model generalizes well to unseen data across different times of the day. The patterns of price impact differ across commodities. For Brent Crude Oil (CO) and Gasoil (GO), the impact remains relatively stable throughout the early hours of trading but experiences a sharp increase towards the end of the trading day. This could be due to end-of-day adjustments or the closing of positions, leading to heightened activity and volatility. Conversely, Crude Oil (CL), RBOB Gasoline (XB), and Natural Gas (NG) display a different pattern where the price impact is highest shortly after the trading session begins. This initial spike may be driven by the rush of early trades as market participants react to overnight news or market conditions. The impact

impact c	Train R^2	Test R^2	T-Stat
0.10	0.613	0.609	3.756
0.15	0.622	0.619	3.723
0.20	0.629	0.625	3.527
0.25	0.633	0.628	3.250
0.30	0.633	0.628	2.962
0.35	0.630	0.625	2.698
0.40	0.624	0.618	2.470
0.45	0.615	0.608	2.278
0.50	0.603	0.595	2.118
0.55	0.588	0.579	1.983
0.60	0.571	0.561	1.869
0.65	0.553	0.540	1.770
0.70	0.532	0.518	1.684
0.75	0.511	0.493	1.608
0.80	0.488	0.467	1.538
0.85	0.465	0.440	1.474
0.90	0.441	0.412	1.413
0.95	0.418	0.382	1.355
1.00	0.394	0.351	1.300

Table 4.5: Performance of various price impact models for commodities in the energy sector.

Figure 4.2: Evolution of model fit for the energy sector with training start on 2014-01-01.

then decreases during the middle of the trading day, a period typically characterized by lower volatility and more balanced supply and demand dynamics. Finally, there is another increase in impact towards the end of the day, possibly as traders execute final transactions before the market closes, which might occur during periods of lower liquidity.

In general, the higher price impact during out-of-market hours for most commodities suggests that reduced trading volume and liquidity at these times make the market more sensitive to individual trades, leading to more significant price movements.

Evolution over time

As discussed, the performance of a model tends to evolve over time, specifically due to the evolving nature of commodity futures markets. To understand the evolution of the performance of each model further, Figure 4.3 provides the test R^2 values for every backtest window for the range of different c for the energy futures contracts. For each future, we are interested in finding if there is any c that is consistently optimal, if there are distinct regime changes in optimal c, or if there are frequent changes.

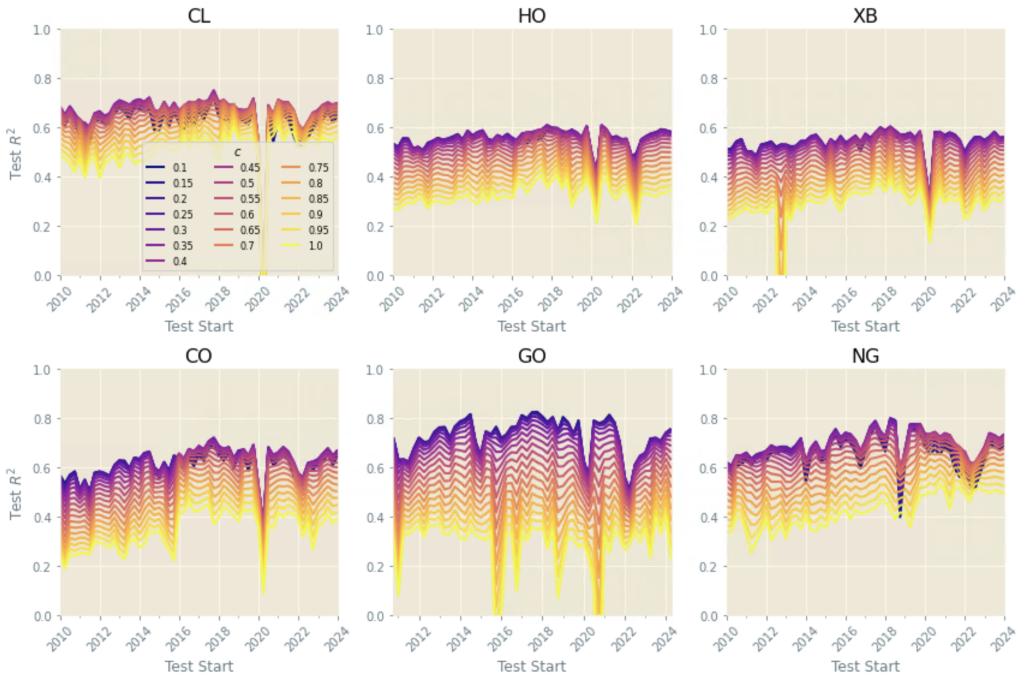


Figure 4.3: Evolution of the out-of-sample model fit for the energy sector over time.

Across all energy futures, there is a slight upward trend in test R^2 values, indicating an overall improvement in model performance over time. However, this trend is occasionally interrupted by periods of decline, suggesting that the model's effectiveness varies with changing market conditions. For WTI Crude Oil (CL), the optimal c fluctuates but generally remains within the range of [0.2,0.5], reflecting the need for regular adjustments in the concavity of the impact function to maintain optimal performance. Typically, the fit of the optimal model remains above 0.4 and often surpasses 0.7.

Similarly, for Natural Gas (NG), the model's fit fluctuates over time, with lower coefficients being optimal in the early years, eventually shifting to less concave models, with an optimal c of around 0.45 by the end of the sample. Heating Oil (HO) and RBOB Gasoline (XB) exhibit relatively stable patterns over time, with few shifts in the optimal concavity coefficient. The out-of-sample R^2 for the optimal model ranges between 0.4 and 0.6, with a low c value of 0.2 or 0.25 proving adequate throughout the entire timeframe.

Gasoil (GO) shows a similar pattern, with little change in the optimal concavity coefficient over time. However, as with Brent Crude Oil (CO), there is more variation in model performance depending on the chosen c value, highlighting the importance of selecting the appropriate concavity for these commodities. For CO, the out-of-sample fit ranges from around 0 to above 0.6 during most of the period, depending on the selected coefficient.

The optimal coefficient for CO remained around 0.15 until 2015, after which it shifted to a higher value of about 0.35.

Importantly, across all commodities, the linear model with $c = 1$ consistently performs the worst over the entire timeframe. This result reaffirms the findings presented in earlier sections and underscores the importance of calibrating price impact models in commodity futures markets with strictly concave functions to capture the inherent non-linearities in limit order books. All models across all commodities show a noticeable decline in performance around 2020, aligning with the onset of the COVID-19 pandemic. This effect is particularly significant for Crude Oil but is also visible across all commodities, reflecting the market disruptions and heightened volume and volatility of that period. Since the models are trained in 6-month windows and tested in 3-month windows, the decline in out-of-sample performance can be attributed to the fact that the models were trained on data from different market conditions and then evaluated during the unique circumstances of the pandemic.

4.2.3 Seasonality Analysis

To complement the price impact analysis, we introduce a preliminary exploration into seasonality patterns specific to the commodity futures market. In particular, beyond the commonly considered factors such as the time of day, the time remaining until the next roll date could potentially exert a significant influence on price dynamics. Instead of re-estimating our model with an additional seasonality parameter, we focus on analyzing stylized facts using data from the *Crude Oil (CL)* futures market, while leaving room for further research to build on these findings.

Seasonality around roll dates may manifest in several ways, including increased volatility, heightened trading volumes (suggesting enhanced liquidity), and imbalances in the direction of trades. Understanding these patterns is important because they can provide insights into market behavior that are distinct from the more generic patterns observed in other markets.

Specifically, three main effects might be expected around roll dates. First, increased volatility: as the roll date approaches, uncertainty over the roll process, combined with adjustments in positions by various market participants, may lead to increased volatility. Understanding how volatility changes as a function of time to the roll date can help market participants better manage risk and anticipate periods of heightened price movements. Second, increased trading volume: a rise in trading volume around roll dates might indicate greater liquidity. Analyzing trading volume as a function of time to roll dates allows us to identify potential periods of heightened market activity and assess whether such patterns are systematic. High volumes can also signal periods when market orders are more likely to be executed without significant price impact, which is critical for both liquidity providers and takers. Finally, imbalances in trade direction: if trading flows around roll dates are highly directional (e.g., more buying than selling or vice versa), this could have a notable impact on price dynamics. For instance, in the FX market (FX fixings), it is observed that price impact tends to increase during periods of directional flow, even when volumes are high. It remains an open question whether similar patterns hold for commodity futures around roll dates. Identifying any directional bias in trade flow could have important implications for trading strategies, hedging approaches, and market-making activities.

To illustrate the potential analysis of seasonality effects, Figure 4.4 presents insights into the average trading volume as a function to time to roll dates for the case of *Crude Oil (CL)*. Specifically, for both the maturing and deferred contracts, the figure highlights two aspects: the average absolute volume as a function of the number of days until the next roll event (left), providing an understanding of how liquidity evolves in both contracts around these events, and the average signed volume as a function of number of days until the next roll event (right), offering insights into any potential seasonality in the direction of the trades near roll dates.

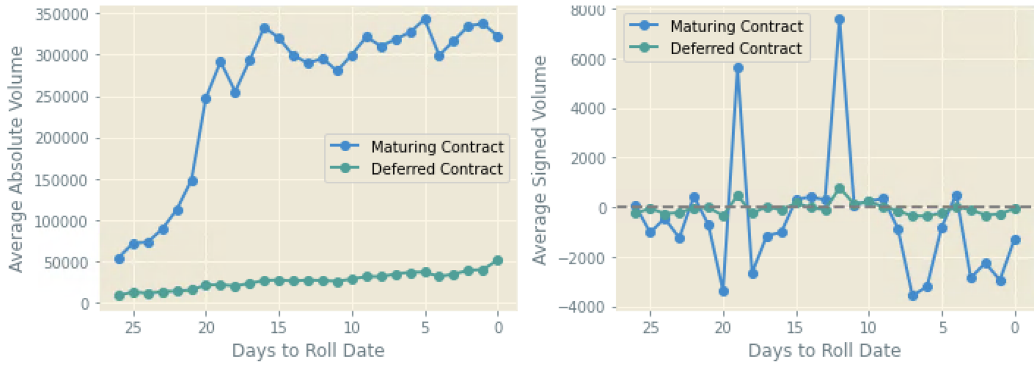


Figure 4.4: Average trading volume as a function of days to roll dates for Crude Oil.

Starting with the average absolute volume as a function of time to roll dates, it can be seen that there is a sharp increase in volume traded in the maturing contract from 25 to 20 days before the next roll period. For the case of Crude Oil, this corresponds to the last rolling period. This is expected as during the previous rolling period, we expect investors to start taking position in the (then) deferred contract, which is the maturing contract in our case, explaining this sharp increase in absolute trading volume. From days 20 to 0, a slow upward trend is visible with some fluctuations, suggesting that the maturing contracts remains the most liquid one during this period. Additionally, there is a slight upward trend in the absolute volume traded for the deferred contracts from day 25 to 0, indicating that investors are gradually taking interest in the deferred contract, which is expected to see a significant rise in trading activity at the start of the next rolling period. Note that very similar patterns have been found for most other commodities as illustrated in Appendix B.

These findings imply a clear transition of liquidity from the maturing contract to the deferred contract as the rolling period starts. The observed increase in trading volume for the maturing contract, followed by a gradual rise for the deferred contract, suggests that market participants systematically adjust their positions during the roll period. This pattern of liquidity transition could have several implications. For instance, it highlights predictable periods of heightened market activity that could be exploited for optimizing trade execution strategies.

Turning to the analysis of the average signed trading volume as a function of days to roll dates, the pattern is less clear. There are significant fluctuations in the directions of the net trade volume for both the deferred and maturing contracts. On certain days, there is a noticeable positive imbalance, with more contracts being bought (such as on days 19 and 12 for instance), while other days show more selling activity. Still, from days 5 to 0 before the next rolling period, there appears to be a slight negative imbalance for the maturing contracts, suggesting that investors may be unwinding their positions in anticipation of the roll.

To understand these results, it is crucial to note that the averages are calculated from the sum of the signed trading volumes for each day across multiple years. This aggregation means that the data reflect a mix of different types of investors and trading behaviors, which may not fully represent those investors who need to roll over their positions regularly. Additionally, the signs of the trades were estimated, which introduces potential discrepancies compared to the actual market data. Despite these limitations, the figures provide valuable insights into the behavior around roll dates. They suggest that while there is no consistent directional pattern in signed volumes well before the roll, there is some evidence of position unwinding in the days immediately leading up to the roll period. This observation could inform future studies on the behavior of various market participants around roll dates, helping to identify specific conditions under which these imbalances become more predictable or pronounced.

Investigating these potential seasonality effects provides a foundation for understanding the unique characteristics of the commodity futures market around roll dates. Such analysis could serve as a basis for further research, particularly in refining models of price impact to account for seasonal patterns, developing trading strategies that take advantage of predictable volume and volatility changes, and enhancing market efficiency by providing more transparent information on trading behavior around key dates.

Chapter 5

Trading Strategies

5.1 Background

To explore trading strategies designed to exploit potential temporary price movements caused by the Goldman Roll, we build on the two initial strategies proposed by Mou [24] and introduce a third strategy that aligns with the same underlying concept.

These strategies are rooted in the observation that the process of rolling over futures contracts in commodity markets typically results in the underpricing of contracts nearing maturity and the overpricing of deferred contracts. By leveraging this phenomenon, straightforward long-short positions can be established to capitalize on the resulting price discrepancies.

Let $F(t, T)$ represent the price of a futures contract at time t with maturity T. Our focus is on the spread between the near-maturing contract with maturity T_1 and the deferred contract with maturity T_2, defined as:

$$SP_t(T_1, T_2) = F(t, T_1) - F(t, T_2).$$

As noted by Mou (2010) [24], the spread serves as a measure of the potential profit or loss per unit of the commodity when futures contracts are rolled forward. It represents the value associated with a calendar spread strategy, which involves taking a short position in the maturing contract while simultaneously holding a long position in a deferred contract. This strategy aims to limit exposure to changes in the absolute price level, thereby allowing traders to fully capture the effects of price pressures resulting from the futures roll. In a market with minimal impact, the spread is generally expected to remain relatively stable over short periods. However, when market impact is significant, the spread tends to narrow during the roll period. This is due to the temporary downward pressure on the price of the maturing contract, $F(t, T_1)$, and the upward pressure on the deferred contract price, $F(t, T_2)$.

5.2 Strategies

There are two primary approaches to capturing the anticipated drop in the spread. The first approach involves positioning ahead of the roll by establishing spread positions before the actual rolling period begins. Specifically, this involves shorting the maturing contracts and going long on the deferred contracts before the roll, with positions unwound precisely when the index rolls its futures forward. Strategies 1 and 2 will focus on this approach. The second approach, known as "back-running," involves taking opposite positions concurrently with the Goldman Roll and unwinding them a few days after the temporary price impact dissipates. Strategy 3 will explore this method.

5.2.1 Strategy 1

The first strategy is structured as follows. Each month, the contracts that the SP-GSCI will roll forward and the timing of the roll are publicly available. The SP-GSCI conducts its futures roll from the 5th to the 9th business day of each month. For this strategy, beginning 10 business days before the first roll date and continuing until 6 days prior, we short the maturing contracts and long the deferred contracts that the index will roll into. We then unwind these positions during the Goldman Roll, allowing us to capitalize on the price impact generated during the roll by anticipating it 10 business days in advance.

5.2.2 Strategy 2

Strategy 2 closely mirrors Strategy 1 but with a shorter lead time. Positions are established only 5 days before the Goldman Roll. From 5 days before the first roll day until the day before the roll begins, we short the maturing contracts and purchase the deferred contracts. These positions are then unwound during the roll, aiming to capture the price impact with a shorter anticipatory window.

5.2.3 Strategy 3

Strategy 3 takes a different approach by aiming to profit from the price impact after it has occurred. During the Goldman Roll, this strategy involves buying the maturing contracts and shorting the deferred contracts, effectively taking the opposite position to the SP-GSCI index. The positions are then unwound during the 5 days following the roll dates, allowing the strategy to benefit as the temporary price impact fades.

5.3 Performance Evaluation

Given that these strategies are implemented over very short timeframes, it is reasonable to assume that any substantial abnormal excess returns are primarily a result of the rolling over of futures contracts included in the SP-GSCI index, rather than other factors. To assess the performance of the strategies, we focus on their excess returns. For strategies $j \in \{1, 2\}$, the excess return on day t is defined as

$$r_t = \frac{SP_{t_j}(T_1, T_2) - SP_t(T_1, T_2)}{(F(t_j, T_1) + F(t_j, T_2))/2} = \frac{(F(t_j, T_1) - F(t_j, T_2)) - (F(t, T_1) - F(t, T_2))}{(F(t_j, T_1) + F(t_j, T_2))/2} \quad (5.3.1)$$

where t is the day when the strategy is unwound, which occurs on each day i of the rolling period, where $i = 1, ..., 5$. t_j represents the day the strategy is initiated, specifically $t_j = t - 10$ for strategy 1, and $t_j = t - 5$ for strategy 2.

For strategy 3, the calculation is slightly different due to the reversal of positions, leading to:

$$r_t = \frac{SP_t(T_1, T_2) - SP_{t_j}(T_1, T_2)}{(F(t, T_1) + F(t, T_2))/2} = \frac{(F(t, T_1) - F(t, T_2)) - (F(t_j, T_1) - F(t_j, T_2))}{(F(t, T_1) + F(t, T_2))/2}. \quad (5.3.2)$$

It is important to highlight that these are excess returns, and the monthly excess return from investing in commodity i using Strategy j is calculated as the 5-day average of $r_t^{i,j}$ when the commodity is rolled forward within the month, and zero otherwise.

Performance Indicators

For performance evaluation, we follow the methodology of [24], grouping the commodities by sector to construct equally weighted portfolios, as well as an aggregate portfolio that includes all commodities. Each month, the portfolio return is computed as the average return of the commodities within the portfolio. Additionally, we adhere to the standard metrics commonly used in trading strategy assessments. Specifically, for monthly excess returns, we report the mean and standard deviation of returns, the t-statistic, skewness, kurtosis, the minimum and maximum returns, and two key risk measures: the Sharpe Ratio and the Maximum Drawdown.

The Sharpe Ratio is defined as

$$SR = \frac{R_p - R_f}{\sigma_p},$$

where R_p is the portfolio return, R_f the risk-free rate, and σ_p represents the standard deviation of the portfolio's excess returns. The Maximum Drawdown is a critical risk metric that measures the largest peak-to-trough decline in the value of a portfolio over a specified period, before a new peak is achieved. It is defined as $\max(DD_t)$ where

$$DD_t = \frac{max_{s \leq t} V_s - V_t}{max_{s \leq t} V_s},$$

and V_t represents the portfolio value at time t.

By reporting these measures, we aim to provide a comprehensive overview of the performance and risk characteristics of the strategies.

5.4 Strategies' Performance

We now turn to the analysis of the performance of the three strategies across sectors and time. The full sample period is divided into three sub-period: 1999-2000, 2000-2010, and 2010-2024, and the commodities are grouped into sectors portfolios.

5.4.1 Strategy 1

We begin by presenting the results obtained from applying the first strategy. To recap, this strategy entails shorting the contracts approaching maturity and going long on deferred contracts 10 business days prior to the rolling period, with the positions being closed out during the rolling period. Table 5.1 provides a summary of the monthly excess return statistics for the portfolios.

When examining the mean monthly excess returns of Strategy 1, we observe that during the first sample period (1990-2000), the monthly returns were generally low but positive across most sectors, except for Metals, with values ranging from 0.051% to 0.176%. However, these returns saw a significant increase in the following decade (2000-2010), with mean excess returns ranging from 0.108% in Agriculture to 0.252% in Livestock, and even turning positive for the Metals sector (0.017%). Unfortunately, this improvement did not extend into the final sub-period (post-2010), where mean returns dropped sharply, becoming negative for the Agriculture and Livestock portfolios, and remaining modest for the Energy and Metals sectors.

A similar trend is evident when analyzing the risk-adjusted returns using the monthly Sharpe Ratio. The pre-2000 period exhibited modest but positive Sharpe Ratios for all sectors except Metals, followed by a substantial increase during 2000-2010, with ratios ranging from 0.258 (Metals) to 0.557 (Energy). This indicates not only higher returns but also a more favorable return-to-risk profile during that period. However, this positive

	Agriculture	Livestock	Energy	Metals	Total
		Panel A: 1990 - 2000			
Mean	0.051	0.176	0.141	-0.001	0.046
T-Stat	2.477	3.194	3.464	-0.160	3.192
Std	0.224	0.601	0.446	0.047	0.156
Skewness	1.285	0.031	-0.332	2.235	1.099
Kurtosis	7.972	2.160	5.385	14.885	8.404
Min	-0.825	-1.917	-1.562	-0.167	-0.618
Max	1.223	2.161	2.048	0.268	0.806
Sharpe Ratio	0.228	0.293	0.316	-0.015	0.294
Max Drawdown	1.238	2.739	2.492	0.412	0.828
# of obs	118	119	120	119	118
		Panel B: 2000 - 2010			
Mean	0.108	0.260	0.240	0.017	0.088
T-Stat	5.014	4.802	6.180	2.849	5.942
Std	0.237	0.591	0.427	0.068	0.162
Skewness	1.586	1.259	1.766	6.399	1.936
Kurtosis	3.776	2.989	5.981	53.496	4.769
Min	-0.376	-0.909	-0.610	-0.083	-0.241
Max	1.182	2.911	2.273	0.620	0.837
Sharpe Ratio	0.456	0.428	0.562	0.261	0.540
Max Drawdown	1.134	3.797	0.688	0.099	0.241
# of obs	121	121	122	121	121
		Panel C: 2010 - 2024			
Mean	-0.013	-0.049	0.003	0.002	0.018
T-Stat	-0.725	-0.950	0.065	0.954	1.222
Std	0.243	0.678	0.681	0.031	0.193
Skewness	-0.042	0.204	-3.262	5.499	2.083
Kurtosis	8.043	2.604	31.310	50.496	14.676
Min	-1.248	-2.229	-5.668	-0.079	-0.644
Max	1.126	2.774	2.905	0.294	1.307
Sharpe Ratio	-0.055	-0.073	0.005	0.073	0.093
Max Drawdown	4.206	14.119	7.878	0.157	1.464
# of obs	171	171	170	169	171

Table 5.1: Performance of Strategy 1 during the three sub-periods.

trend reversed post-2010, with Sharpe Ratios turning negative for the Agriculture and Livestock portfolios, and a sharp decline in the Energy portfolio's ratio to just 0.005. In terms of risk, the maximum drawdown remained relatively low across all portfolios and sub-periods, suggesting that these strategies generally carried low risk.

Overall, the Energy sector consistently outperformed the other sectors across all sub-periods, with mean monthly excess returns of 0.141% in the pre-2000 period, rising to 0.237% in 2000-2010, and continuing to achieve positive returns in the final sub-period. Similarly, the Energy sector exhibited the highest Sharpe Ratios across all periods. The Livestock portfolio performed relatively well during 1990-2000, but the Agriculture portfolio performed slightly better in terms of risk-adjusted returns in the 2000-2010 period. The Metals portfolios underperformed in the first sample period, but later managed to achieve positive mean monthly excess returns and Sharpe Ratios of 0.261 (2000-2010) and 0.073 (2010-2024).

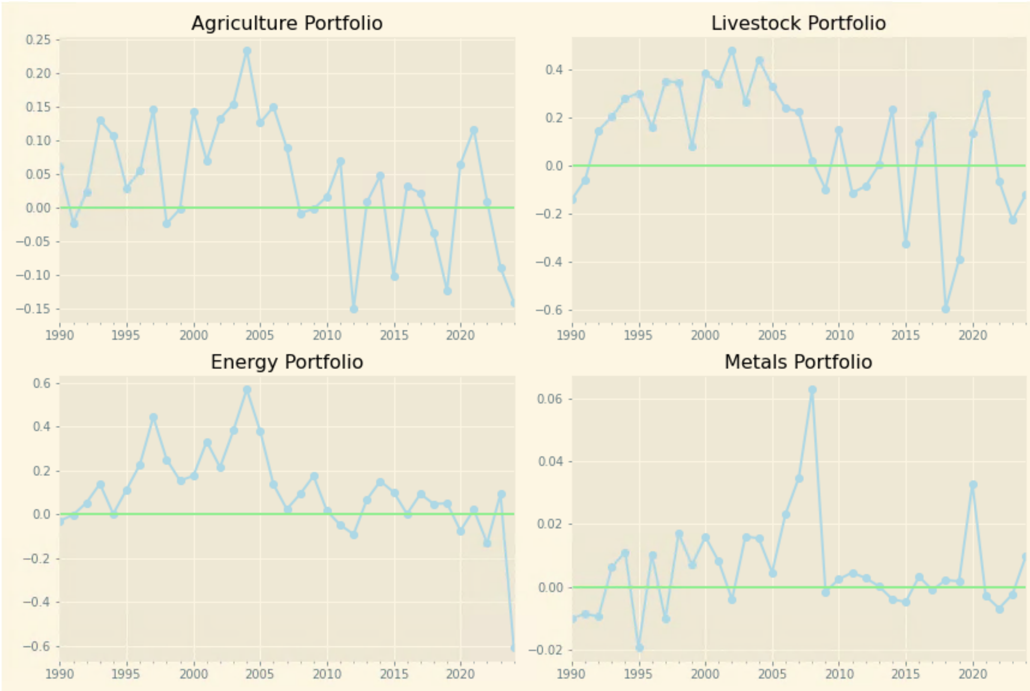


Figure 5.1: Average Monthly Excess Returns of the Four Sector Portfolios with Strategy 1

Figure 5.1 illustrates the excess returns year by year for the 4 sector portfolios. From the Figure, it can be seen that the Energy and Livestock portfolios had consistently positive returns until the year 2010, before oscilating between the negative and the positive thereafter. The pattern is relatively similar for the Agriculture portfolio, with slightly more variations in the period until 2010. Finally, the Metals portfolio had very irregular monthly excess returns over time, oscillating between the positive and the negative for most years, except for 2008 and 2020, where it performed particularly well.

5.4.2 Strategy 2

Next, the second strategy is examined, which involded establishing the 'roll' position 5 business days before the Goldman Roll. The results are detailed in Table A.6.

When examining the mean monthly excess returns of Strategy 2, we observe again that during the first period, the returns were generally low but positive across most sectors, except for Metals, with values ranging from 0.066% to 0.111%. However, contrary to Strategy 1, these returns saw a slight decrease in the following decade (2000-2010) for the Agriculture and Livestock portfolio, while the mean excess returns of the Energy and Metals sectors rose to respectively 0.130% and 0.007%. As in Strategy 1, the mean excess returns were lowest in the final sub-period, becoming negative for the Agriculture and Livestock portfolios, and remaining modest for the Energy and Metals sectors.

When analyzing the risk-adjusted returns using the monthly Sharpe Ratio, we see that the strategy performed better than Strategy 1 in the first sample period, before being slightly less performing in the second sample period with monthly Sharpe Ratios in the range 0.120 (Metals) to 0.327 (Energy). Eventually, the positive trend reversed post-2010, with Sharpe Ratios turning negative for the Agriculture and Livestock portfolios, and a sharp decline in the Energy portfolio's ratio to just 0.033. In terms of risk, the maximum drawdown remained relatively low across all portfolios and sub-periods, suggesting that these strategies generally carried low risk.

Overall, the Energy sector consistently outperformed the other sectors across all sub-periods, with mean monthly excess returns of 0.111% in the pre-2000 period, rising to 0.130% in 2000-2010, and continuing to achieve positive returns in the final sub-period. Similarly, the Energy sector exhibited the highest Sharpe Ratios across most sub-periods. The Livestock and Agriculture portfolio performed relatively well during 1990-2000, before losing out the lead to Energy in terms of risk-adjusted returns in the 2000-2010 period. The Metals portfolios underperformed until 2010, but later managed to achieve positive mean monthly excess returns and a monthly Sharpe Ratio of 0.066.

Lastly, Figure B.7 illustrates the excess returns year by year for the 4 sector portfolios. Similar to Strategy 1, it can be seen that the Agriculture, Energy and Livestock portfolios had relatively positive returns until the year 2008, before oscillating between the negative and the positive thereafter. Once again, the Metals portfolio performed particularly well in 2008. In terms of volatility, it appears that the Livestock portfolio moved the most over time.

5.4.3 Strategy 3

Finally, we analyze the performance of Strategy 3, which involved taking the opposite position during the Goldman Roll and closing the trades 5 business days later. The results of this strategy are presented in Table A.7.

When examining the mean monthly excess returns of the strategy, we observe again that during the first period, the returns were generally low but positive across most sectors, except for Metals, with values ranging from 0.066% to 0.111%. Then, similar to Strategy 1, the returns increased in the following decade (2000-2010) for the Agriculture and Energy portfolio to respectively 0.045% and 0.124, while the they slightly decreased for the Livestock portfolio, and barely became positive for Metals. As in Strategy 1 and 2, the mean excess returns decreased in the final sub-period, with the slight difference that they remained positive and modest for all portfolios.

The same trend is evident when analyzing the risk-adjusted returns using the monthly Sharpe Ratio. The strategy exhibited relatively similar Sharpe Ratios across the first two sample periods, with Sharpe Ratios in the range 0.205 (Agriculture) to 0.383 (Energy) in the period 2000 to 2010. Note that the Sharpe Ratio remained very modest for Metals throughout the entire period under study. Eventually, as in Strategy 1 and 2, the monthly Sharpe Ratio decreased in the final subperiod, with the notable difference that they remained positive for the Agriculture, Livestock and Energy portfolios. In terms of risk carried during the strategy, the maximum drawdowns were very low across all portfolios and sub-periods, suggesting that these strategies generally carried low risk.

Overall, the Livestock portfolio performed the best in the first sample period with a monthly Sharpe Ratio of 0.485, before leaving the best performance to the Energy portfolio for the remaining sub-periods with a Sharpe Ratio of 0.383 in 2000-2010. The Agriculture had positive performance measures, but performed slightly less than the other two portfolios. Similar to the Strategy 1 and 2, the Metals portfolios did not perform well in the first sample period, however, here, the performance did not increase significantly over the time period. Finally, Figure B.8 illustrates the average monthly excess returns year by year for the four sector portfolios. Again, the Energy, Agriculture and Livestock portfolios had similar patterns over time, with some oscillation between the positive and the negative, even though they still managed to maintain mostly positive returns throughout. Notably, the livestock portfolio achieved roughly 0.5% monthly excess returns on average in 2020. Contrary to Strategies 1 and 2, the Metal portfolio did not have peaks of performance in years 2008 and 2020, and exhibited mostly negative mean excess returns over the whole sample period.

5.4.4 Strategies' comparison

In this section, we provide a direct comparison of the strategies' performance using their annualized Sharpe Ratios. Table 5.2 presents the annualized excess returns for each strategy across all subperiods. Following the methodology of Mou [24], we annualize the excess returns of Strategy 1 based on 10-day returns and those of Strategies 2 and 3 based on 5-day returns. Accordingly, we annualize the returns of Strategy 1 using a factor of 252/10, and those of the other strategies using a factor of 252/5. The annualized Sharpe Ratios facilitate the comparison of similar portfolios. Generally, a Sharpe Ratio greater than 1 is considered acceptable, while a ratio above 2 is deemed very good. Conversely, a ratio below 1 is regarded as sub-optimal.

	Agriculture	Livestock	Energy	Metals	Total
Strategy 1					
Pre-2000	1.145	1.470	1.587	-0.074	1.475
2000-2010	2.288	2.210	2.820	1.311	2.712
Post-2010	-0.278	-0.365	0.025	0.369	0.469
Strategy 2					
Pre-2000	2.760	1.491	2.438	-0.289	2.642
2000-2010	1.607	1.255	2.323	0.858	2.766
Post-2010	-0.767	-0.454	0.234	0.469	0.274
Strategy 3					
Pre-2000	0.590	3.445	1.542	-0.295	1.087
2000-2010	1.453	2.211	2.719	0.043	2.438
Post-2010	0.823	0.244	0.586	0.286	0.427

Table 5.2: Annualized Sharpe Ratios by Trading Days for each strategy.

In the first sample period, from 1990 to 2000, Strategy 1 displayed Sharpe Ratios ranging from 1.145 in Agriculture to 1.587 in Energy. Strategy 2 outperformed during this period, with Sharpe Ratios ranging from 1.491 in Livestock to 2.760 in Agriculture. Conversely, Strategy 3 performed poorly in Agriculture, with a Sharpe Ratio of 0.590, but excelled in Livestock, achieving an impressive annualized Sharpe Ratio of 3.445. It is noteworthy that all portfolios within the Metals sector exhibited negative Sharpe Ratios during this period. In the subsequent period from 2000 to 2010, the performance of Strategy 1 improved significantly, with all Sharpe Ratios exceeding 2, except for Metals, which had a Sharpe Ratio of 1.311. Strategy 3 also experienced a substantial performance increase, with Sharpe Ratios ranging from 1.453 to 2.719, except for Metals, which had a Sharpe Ratio of 0.043. Surprisingly, Strategy 2 did not see a significant performance boost during this period, though it still performed relatively well, with Sharpe Ratios ranging from 1.255 to 2.323. Finally, in the last sub-sample period, the performance of all strategies declined markedly across most sectors. Strategies 1 and 2 exhibited negative Sharpe Ratios in the Agriculture and Livestock portfolios, while the Energy and Metals portfolios managed to maintain positive but modest performance. Although Strategy 3 did not have any negative Sharpe Ratios, all its ratios were below 1, indicating that the strategy generated limited excess returns relative to the risk taken.

5.5 Discussion

Finally, we dive deeper into the key insights gained from the analysis and explore the reasons behind the performance of these strategies. The key takeaway from these results is the significant variation in strategy performance across different sectors and time periods. Specifically, the strategies exhibited strong performance in the early periods, with Strategies 1 and 3 showing substantial improvement during the 2000-2010 period, while Strategy 2 also performed relatively well. However, in the final period, all strategies experienced a notable decline in performance. When examining sector-specific strengths and weaknesses, the Energy sector consistently performed well across all strategies, particularly in the early periods. Strategy 3 demonstrated exceptional performance in the Livestock sector but faced challenges in Agriculture, while Strategy 2 performed relatively well in the Agriculture sector during the early period. Conversely, the Metals sector generally underperformed in the early periods but showed a notable improvement in the final sub-period, overtaking other sectors for Strategies 1 and 2.

To fully understand these results, it's crucial to examine the average changes in spreads over the different time periods considered. As indicated by Equations 5.3.1 and 5.3.2, the monthly excess returns are largely driven by the changes in spreads between expiring and deferred contracts. The strategies are specifically designed to exploit the anticipated decrease in spreads during the rolling period, driven by the tendency for the price of the maturing contract to be lower and the deferred contract price to be higher during the Goldman Roll. Figure 5.2 illustrates the average change in spreads between the nearby and first deferred contracts for index commodities across all sub-periods, broken down into 5-day intervals, providing insights into the evolving performance of these strategies.

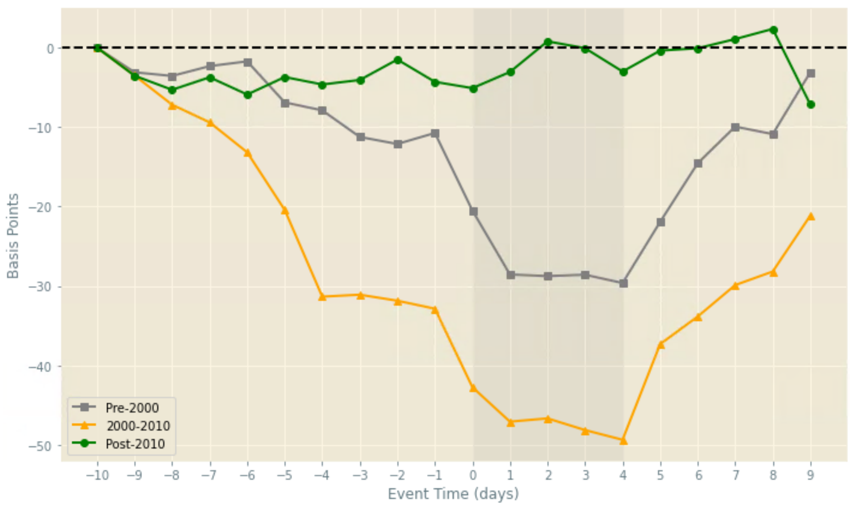


Figure 5.2: Average spreads between the nearby and first deferred contracts of commodities around the GR in three time periods.

The figure highlights the differences in strategy performance over time. In the earlier periods, a clear V-shaped pattern in spreads is visible, with spreads declining leading up to the Goldman Roll (visible in grey shades in the figure), stabilizing during the roll, and then increasing afterward. This pattern was most pronounced in the 2000-2010 period, where spreads declined by nearly 50 basis points during the roll, supporting the notion of front-running impacts as discussed by Mou [24]. Strategy 1, which involved entering positions earlier days -10 to -6, benefited more during this period due to a more significant decline in spreads earlier on, explaining its relative outperformance in 2000-2010. Conversely, in the pre-2000 period, Strategy 2 (which enters positions later in the event period, days -5 to -1) performed well, as the difference in spread changes between the two periods was less pronounced. Additionally, the widening of spreads post-Goldman Roll helps explain the positive performance of Strategy 3 during these periods.

The most significant insight from this figure is the disappearance of the V-shaped pattern in the post-2010 period (green line). During this period, the average spreads remain near zero across all days, indicating that the price impact or opportunities to capitalize on spread changes during the roll have largely vanished, which explains the overall poor performance of these strategies after 2010. This finding aligns with the work of Irwin et al. [20], who found that spreads between nearby and deferred futures prices decreased significantly during the early (1991-2003) and growth (2004-2011) stages of financialization, with a substantial decline in impact during the post-financialization period (2012-2019). This suggests that while price impacts from futures rolling were significant in earlier periods, they have largely dissipated in more recent years, consistent with the findings of Bessembinder et al. [4], Hamilton and Wu [15], and Sanders et al. [27].

Since Mou's original paper [24] identified price impacts in futures rolling and the resulting arbitrage opportunities, the field has attracted considerable attention. This increased attention likely led to greater capital flows into these strategies, raising awareness and reducing arbitrage opportunities. As more market participants sought to exploit these opportunities, the market became more efficient, diminishing the potential for excess returns. This outcome is consistent with the work of Cheng and Xiong [11], who examined the effects of financialization on commodity markets and found that as more capital entered these markets, the potential for excess returns diminished. By 2010, increased participation and awareness among market participants had reduced the impact of the Goldman Roll on spreads, explaining the diminished performance of these strategies in the last subsample period. Another reason for the reduction in the price impact surrounding rolling events is brought by Irwin et al [20]. They suggest that the reduction in price impact during rolling events can be largely attributed to the significant increase in liquidity supply in commodity futures markets, which has been driven mainly by the shift to electronic trading.

Next, to better understand the variation in strategy performance across different sectors, Figure 5.3 presents the average spreads during the event days across three time periods, broken down by sector.

The figure reveals that the Energy portfolio exhibits a clear V-shaped pattern in the earlier periods, with spreads decreasing leading up to the Goldman Roll, stabilizing during the roll, and increasing afterward. This pattern is consistent with findings in the literature, such as those by Mou [24], who documented the significant price impact during the Goldman Roll, particularly in energy markets. The pronounced V-shape in the Energy sector for the earlier periods highlights the potential for high excess returns during these times. However, in the period after 2010, this pattern disappears, with spreads forming a relatively straight line, indicating a significant reduction in the opportunities to exploit price impacts. This shift helps explain the reversal in strategy performance after 2010, as the market became more efficient and the potential for excess returns diminished.

In both the Agriculture and Livestock portfolios, similar patterns are observed, though with some notable differences. In the pre-2000 period, the Agriculture portfolio shows a more linear decrease in spreads, while in the Livestock portfolio, spreads tend to increase during the Goldman Roll. This difference is crucial in understanding the sector-specific performance: Strategy 2 performed best in the Agriculture portfolio during this period, likely because the most significant decrease in spreads occurred between days -5 to -1, aligning with Strategy 2's timing. Conversely, Strategy 1 excelled in the Livestock portfolio, where the sharpest spread decrease occurred earlier in the event period (days -10 to -5). This observation is in line with studies such as Hamilton and Wu [15], which noted

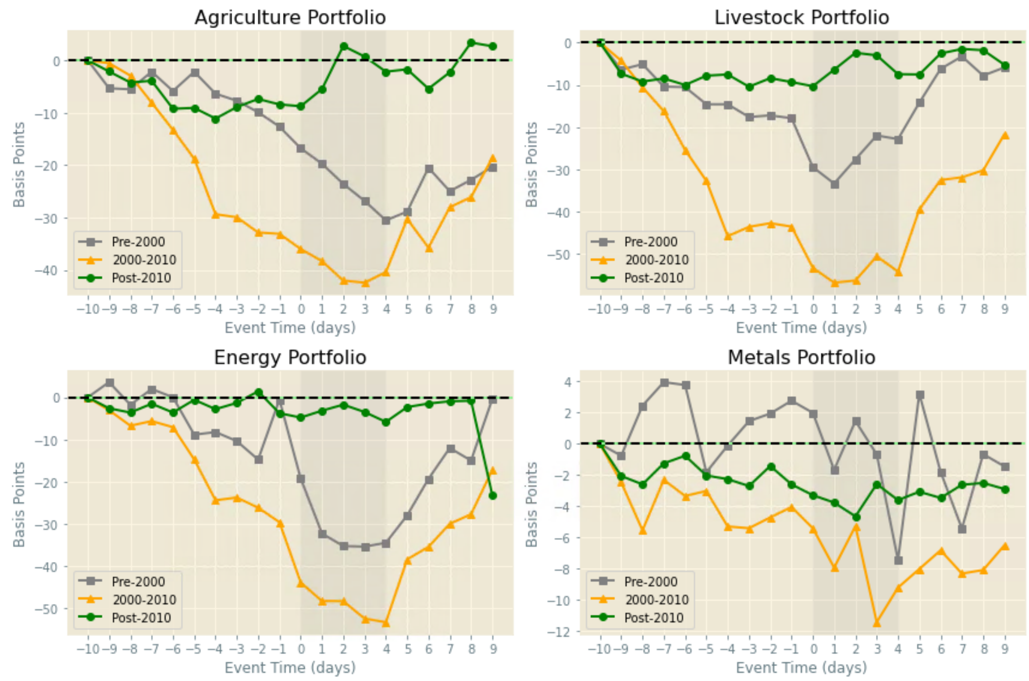


Figure 5.3: Average spreads between the nearby and first deferred contracts of commodities around the GR in three time periods per sectors.

sector-specific differences in how futures markets respond to rolling activities, particularly in agricultural markets where liquidity dynamics can vary significantly. In the 2000-2010 period, both the Agriculture and Livestock portfolios performed better with Strategy 1, which aligns with the sharper decrease in spreads observed during this period—reaching 40 to 50 basis points. This sharp decline underscores the effectiveness of Strategy 1 in capturing early spread decreases, as supported by Irwin and Sanders [27], who emphasized the increased influence of financial traders in this period, leading to more pronounced price movements around rolling activities.

The Metals portfolio presents a unique case. Unlike other sectors, it does not exhibit a strong or consistent pattern in spread changes. During the pre-2000 period, average spreads in the Metals sector tended to increase initially, then oscillate between positive and negative values, leading to poor performance across all strategies. This lack of a clear pattern in spreads, and thus a predictable price impact, aligns with findings by Bessembinder et al. [4], who observed that metals markets often behave differently from other commodities, with less pronounced reactions to rolling activities due to different market structures and participant behaviors. In the 2000-2010 period, the Metals portfolio showed some improvement in spread behavior, with a more noticeable decline, which partially explains the modest improvement in strategy performance. However, this decline was not as pronounced or consistent as in other portfolios, leading to less significant gains. Finally, in the post-financialization period (2010 onwards), the Metals portfolio exhibited a small decline in spreads, reaching only -4 basis points at most. This modest decline explains why the Metals portfolio performed relatively better than others in the last sub-period, despite the general decline in strategy effectiveness. This outcome is consistent with research by Cheng and Xiong [11], which suggests that as markets mature and more participants engage in arbitrage strategies, the opportunities for significant price impacts, and thus excess returns, diminish.

Finally, in the analysis of the results, we saw that the metals portfolios performed particularly well in 2008 and 2020, as visible in Figures 5.1 and B.7. The strong performance of the metals portfolio, including silver and gold, in 2008 and 2020 can be attributed to several factors, particularly the role of gold as a safe-haven asset during periods of financial crisis. Indeed, during both the 2008 global financial crisis and the 2020 COVID-19 pandemic, investor demand for safe-haven assets surged, with gold often serving as a primary refuge. As market uncertainty increased, investors flocked to gold, driving up its price and causing significant movements in the spreads between nearby and deferred futures contracts. This heightened volatility and the associated price impacts during the Goldman Roll likely created greater opportunities for the strategies to capitalize on these changes in spreads, leading to higher excess returns. The phenomenon of gold acting as a safe haven during times of economic turmoil is well-documented in the literature. For instance, Baur and Lucey [2] found that gold often acts as a hedge and a safe haven against stocks, particularly during extreme market conditions. Moreover, the financial crises in 2008 and 2020 led to increased market liquidity in safe-haven assets and the implementation of aggressive monetary policies, including low-interest rates and quantitative easing. These factors contributed to higher demand for precious metals, not only as safe havens but also as hedges against inflation and currency depreciation. This increased demand further amplified the price impacts during futures rolls, enhancing the effectiveness of strategies based on exploiting changes in spreads.

Conclusion

The first and main objective of this thesis was to understand how price impact propagates over time in commodity indices, including during rolling periods. Using the nonlinear propagator model framework developed by Muhle-Karbe et al. [25], we focused on identifying the most suitable functional form for calibrating the model to different commodity futures. Specifically, we experimented with a range of concavity coefficients, c, in the expression 4.1.2, aiming to find the optimal coefficient, c^*, that would provide the best model fit and stability. Additionally, we calibrated the model for every hour of the trading day to gain insights into intraday patterns of price impact around rolling events and assessed its performance across time and different sectors.

Our findings indicate that, across all commodities and time periods, the optimal concavity coefficient, $c^* = 0.25$, yielded the best performance. Notably, this coefficient significantly improved model fit and stability compared to the linear impact model with $c = 1$, which has been typically used to study price impacts around roll dates in commodity futures contracts. Furthermore, the optimal model outperformed the commonly used square-root law ($c = 0.5$), which is typically applied to account for the concavity of price impact in limit order books. This suggests that relatively little trading is needed to create a price impact and that there is minimal difference in impact between small and large trades. Additionally, we observed that the model's performance was highly robust across both time and commodities, with the highest impacts found for Gold, Crude Oil, and Natural Gas. We also found that price impacts were generally higher during out-of-market hours and that the model's effectiveness varied under changing market conditions. For example, during the onset of the COVID-19 pandemic, the model showed much poorer out-of-sample fit for most commodity futures when trained on pre-pandemic data, which did not account for this unpredictable change in market conditions.

After completing the price impact analysis, the second goal of this thesis was to design simple anticipatory trading strategies to exploit market anomalies during the rolling over of futures contracts. Specifically, building on the concept that the rolling over of futures contracts involves large-scale selling of maturing contracts and buying of deferred contracts, we replicated two strategies developed by Mou [24] and implemented a third simple strategy to evaluate their performance after 2010. The objective was to determine whether such anomalies could still be profitable over time. Strategies 1 and 2 anticipated the rolling over of futures contracts by establishing positions in advance, while Strategy 3 essentially took the opposite approach.

Regarding the results of these strategies, we found outcomes similar to those reported by Mou [24] up to 2010. Both Strategies 1 and 2 showed low but positive monthly Sharpe Ratios and monthly excess returns until 2000, followed by a significant performance increase with the financialization of the commodity futures market, achieving annualized Sharpe Ratios as high as 2.8 for the Energy sector during the 2000-2010 period. The third strategy exhibited a similar pattern, with modest performance in the initial period, followed by substantial improvement from 2000 to 2010, reaching annualized Sharpe Ratios of about 2.7 for the Energy portfolio. Overall, the Energy portfolio performed the best across all strategies, while the Metals sector generally underperformed in the earlier periods.

More importantly, the performance of all strategies declined significantly post-2010, with Strategies 1 and 2 showing negative returns for the Agriculture and Livestock portfolios and only modest returns for other sectors and Strategy 3. During this period, average spreads hovered around zero across all days, indicating that opportunities to profit from spread changes during the rolling period had largely disappeared. This aligns with findings by [20], who observed a substantial decline in impact during the post-financialization period. The decline in strategy performance can be attributed to the increased number of market participants attempting to exploit the potential anomaly since the publication of Mou's paper, leading to greater arbitrage capital and investor participation, which in turn made the market more efficient and reduced the potential for excess returns. Additionally, the shift to electronic trading boosted liquidity in the commodity futures markets, resulting in lower price impacts surrounding rolling events.

Finally, we discuss certain limitations of this thesis and open the discussion for potential future research. First, the order sign data had to be estimated rather than directly observed; while the estimation process was reasonably robust, it may not be fully accurate, potentially affecting the precision of the results. Additionally, the analysis relied on data from the full order book of commodity futures, which did not specifically focus on the trades of institutional investors that regularly roll over their contracts. As a result, the volume analyzed might include transactions from hedgers or other investors with different objectives, not necessarily linked to rolling over trades. Furthermore, the trading strategies were entirely focused on the SP-GSCI index, the largest and most liquid commodity index, where the rolling period is well-known. However, its popularity and liquidity may have diminished the opportunities for arbitrage over time. There could be other opportunities in less liquid contracts if it were possible to estimate when institutional investors might roll them over.

In conclusion, several avenues for future research could build upon the findings of this thesis. First, while we introduced the potential seasonality in terms of time to the next roll event, incorporating this seasonality into the modeling of price impact could provide significant insights. This approach may reveal how price impacts vary when trading flows are highly directional around these times and whether the expected effects are observed when flow is indeed directional. Second, based on our understanding of the dynamics of price impact, optimal trading strategies could be developed to enhance trading execution on specific days. For example, institutional investors needing to roll over a certain volume on a particular day could use the estimated price impact to decide how and when to execute their trades optimally throughout the day, thereby reducing order costs and increasing efficiency. Finally, a promising area for future research is exploring the cross-impact that trading one futures contract has on others that are highly correlated. Understanding this cross-impact could lead to a more comprehensive and accurate representation of price impact and help clarify the exact effects of trading in commodity futures contracts.

Appendix A

Additional Tables

This section provides the additional tables that are useful in our analysis. Tables A.1 and A.2 provide descriptive statistics for daily futures prices and daily traded volumes separated between roll and non-roll days of the SP-GSCI index. Tables A.3, A.4 and A.5 present the performance of the various price impact models for commodities in the Agriculture, Metals and Livestock sectors, and Tables A.6 and A.7 present the performance of Strategies 2 and 3 during the three sub-sample periods.

Ticker	mean	std	min	max	contracts
		Panel A: Daily Closing Prices			
WC	477.080	182.674	233.500	1119.750	174
KW	595.957	179.109	388.250	1181.500	54
CN	448.890	122.076	307.750	786.500	173
SY	880.126	335.281	429.000	1767.500	174
KC	125.879	49.292	43.000	279.550	174
SB	13.365	5.549	5.140	32.580	140
CC	2047.172	1183.847	710.000	10582.000	174
CT	72.908	25.201	32.150	212.060	139
LH	67.275	18.305	32.550	131.825	243
LC	97.747	30.277	58.300	189.850	207
FC	117.745	44.584	53.100	261.225	276
CL	50.660	29.428	10.740	144.410	413
HO	1.540	0.958	0.302	4.610	413
XB	2.150	0.638	0.658	4.263	212
CO	78.274	24.623	29.370	145.200	229
GO	698.573	218.156	241.750	1360.250	175
NG	3.391	1.252	1.629	9.347	377
GC	936.685	585.933	256.500	2366.900	173
SV	1311.369	891.505	364.000	4093.000	173
		Panel B: Daily Traded Volumes			
WC	19372.963	21879.686	0.000	94819.000	174
KW	11690.713	5307.038	0.000	32920.000	54
CN	96710.035	47294.792	0.000	265502.000	173
SY	34141.292	37838.582	0.000	205432.000	174
KC	4846.990	5649.447	0.000	35461.000	174
SB	16492.314	18336.410	0.000	71035.000	140
CC	2780.106	3891.645	0.000	31812.000	174
CT	4948.788	5717.171	0.000	23168.000	139
LH	6205.520	7226.771	0.000	31804.000	243
LC	7277.889	8673.768	0.000	36432.000	207
FC	1451.900	1972.473	0.000	9554.000	276
CL	152585.057	187849.973	0.000	1164773.000	413
HO	13180.785	13989.126	0.000	123633.000	413
XB	22179.302	7616.352	0.000	52751.000	212
CO	122046.837	70942.267	0.000	475019.000	229
GO	44297.046	13413.481	0.000	137358.000	175
NG	93080.942	32147.198	0.000	257477.000	377
GC	92740.987	102772.867	0.000	580884.000	173
SV	24737.236	33696.572	0.000	337765.000	173

Table A.1: Descripive Statistics of commodity futures during the GR in the SP-GSCI index.

Ticker	mean	std	min	max	contracts
		Panel A: Daily Closing Prices			
WC	480.300	182.835	230.750	1326.500	174
KW	593.168	176.949	368.250	1340.750	54
CN	361.628	153.479	174.250	831.500	173
SY	880.972	334.227	411.500	1786.250	174
KC	126.132	49.533	42.000	314.000	174
SB	13.439	5.528	4.280	35.280	140
CC	2049.158	1079.744	680.000	12250.000	174
CT	71.478	21.461	28.520	219.700	139
LH	66.544	18.031	27.725	132.350	243
LC	97.705	30.297	54.800	188.500	207
FC	117.734	44.628	47.650	264.400	276
CL	50.735	29.418	1.430	145.320	413
HO	1.546	0.961	0.296	4.944	413
XB	2.158	0.638	0.443	4.308	212
CO	73.384	25.720	19.500	146.610	229
GO	700.390	218.898	193.000	1326.000	175
NG	4.028	2.260	1.300	15.491	377
GC	883.216	580.056	253.800	2441.300	173
SV	1306.165	885.917	351.000	4843.000	173
		Panel B: Daily Traded Volumes			
WC	11940.641	19300.565	0.000	178753.000	174
KW	7743.032	7592.011	0.000	48243.000	54
CN	28122.403	45794.806	0.000	411359.000	173
SY	21757.684	33926.932	0.000	245468.000	174
KC	3167.407	5873.817	0.000	40899.000	174
SB	10850.745	18279.391	0.000	156111.000	140
CC	2918.721	5708.561	0.000	36913.000	174
CT	3300.587	5790.003	0.000	48494.000	139
LH	3413.893	5756.508	0.000	34239.000	243
LC	4178.102	7158.879	0.000	48038.000	207
FC	874.798	1680.060	0.000	11396.000	276
CL	67264.499	135143.273	0.000	920643.000	413
HO	5103.495	9501.708	0.000	60729.000	413
XB	8481.212	10125.637	0.000	73068.000	212
CO	60173.415	72721.182	0.000	468292.000	229
GO	20941.393	19169.739	0.000	118795.000	175
NG	23439.863	38823.094	0.000	333756.000	377
GC	52244.560	84576.696	0.000	618848.000	173
SV	15685.846	26099.370	0.000	251964.000	173

Table A.2: Descripive Statistics of commodity futures outside of the GR in the SP-GSCI index.

impact c	Train R^2	Test R^2	T-Stat
0.10	0.674	0.164	0.108
0.15	0.681	0.148	0.136
0.20	0.683	0.124	0.167
0.25	0.680	0.098	0.203
0.30	0.673	0.075	0.244
0.35	0.661	0.054	0.293
0.40	0.645	0.036	0.353
0.45	0.626	0.020	0.427
0.50	0.605	0.006	0.513
0.55	0.580	-0.009	0.604
0.60	0.555	-0.023	0.695
0.65	0.528	-0.039	0.782
0.70	0.500	-0.055	0.862
0.75	0.472	-0.072	0.931
0.80	0.444	-0.089	0.986
0.85	0.417	-0.106	1.024
0.90	0.391	-0.123	1.045
0.95	0.365	-0.138	1.050
1.00	0.341	-0.155	1.040

Table A.3: Performance of various price impact models for commodities in the agriculture sector.

impact c	Train R^2	Test R^2	T-Stat
0.10	0.628	0.629	3.477
0.15	0.640	0.641	4.247
0.20	0.649	0.650	5.041
0.25	0.655	0.656	5.445
0.30	0.658	0.658	5.153
0.35	0.658	0.658	4.470
0.40	0.656	0.654	3.792
0.45	0.650	0.648	3.248
0.50	0.642	0.638	2.834
0.55	0.631	0.626	2.521
0.60	0.618	0.611	2.281
0.65	0.603	0.595	2.091
0.70	0.585	0.576	1.938
0.75	0.566	0.554	1.812
0.80	0.545	0.531	1.706
0.85	0.523	0.506	1.614
0.90	0.499	0.480	1.533
0.95	0.475	0.453	1.461
1.00	0.451	0.424	1.396

Table A.4: Performance of various price impact models for commodities in the metal sector.

impact c	Train R^2	Test R^2	T-Stat
0.10	0.573	0.545	2.046
0.15	0.577	0.551	2.459
0.20	0.579	0.554	2.675
0.25	0.577	0.552	2.692
0.30	0.572	0.546	2.585
0.35	0.563	0.536	2.425
0.40	0.551	0.521	2.254
0.45	0.537	0.503	2.089
0.50	0.519	0.480	1.938
0.55	0.500	0.454	1.800
0.60	0.479	0.424	1.675
0.65	0.457	0.391	1.560
0.70	0.434	0.355	1.454
0.75	0.411	0.315	1.355
0.80	0.388	0.273	1.263
0.85	0.365	0.228	1.178
0.90	0.343	0.179	1.097
0.95	0.322	0.126	1.021
1.00	0.302	0.067	0.949

Table A.5: Performance of various price impact models for commodities in the livestock sector.

	Agriculture	Livestock	Energy	Metals	Total
		Panel A: 1990 - 2000			
Mean	0.066	0.080	0.111	-0.002	0.045
T-Stat	4.225	2.291	3.763	-0.444	4.043
Std	0.170	0.382	0.325	0.049	0.120
Skewness	2.624	-0.417	-0.478	-0.604	0.663
Kurtosis	11.554	2.724	4.881	9.510	9.617
Min	-0.331	-1.430	-1.368	-0.219	-0.563
Max	1.107	1.016	1.303	1.155	0.567
Sharpe Ratio	0.389	0.209	0.343	-0.041	0.372
Max Drawdown	0.607	2.850	3.271	0.487	0.701
# of obs	118	119	120	119	118
		Panel B: 2000 - 2010			
Mean	0.034	0.070	0.130	0.007	0.037
T-Stat	2.490	1.928	3.599	1.318	4.286
Std	0.148	0.394	0.396	0.064	0.096
Skewness	0.441	0.487	3.320	7.982	1.478
Kurtosis	4.152	3.773	22.440	77.607	3.829
Min	-0.500	-1.244	-0.866	-0.086	-0.217
Max	0.636	1.666	2.986	0.597	0.426
Sharpe Ratio	0.227	0.177	0.327	0.121	0.390
Max Drawdown	1.679	2.998	2.236	0.220	0.403
# of obs	121	121	122	121	121
		Panel C: 2010 - 2024			
Mean	-0.022	-0.031	0.012	0.001	0.005
T-Stat	-1.412	-0.836	0.429	0.858	0.504
Std	0.209	0.478	0.360	0.022	0.146
Skewness	-2.117	0.066	0.405	3.505	0.848
Kurtosis	19.549	3.419	11.559	32.673	13.502
Min	-1.559	-1.884	-1.754	-0.084	-0.749
Max	0.905	1.615	2.174	0.186	0.919
Sharpe Ratio	-0.108	-0.064	0.033	0.066	0.039
Max Drawdown	4.423	11.662	3.869	0.206	1.460
# of obs	171	171	170	169	171

Table A.6: Performance of Strategy 2 during the three sub-periods.

	Agriculture	Livestock	Energy	Metals	Total
Panel A: 1990 - 2000					
Mean	0.012	0.174	0.097	-0.002	0.017
T-Stat	0.902	5.338	2.389	-0.457	1.663
Std	0.149	0.359	0.447	0.048	0.106
Skewness	-0.343	1.397	1.857	4.398	1.020
Kurtosis	4.578	2.403	11.935	41.566	3.767
Min	-0.610	-0.438	-1.629	-0.144	-0.297
Max	0.485	1.671	2.703	0.396	0.422
Sharpe Ratio	0.083	0.485	0.217	-0.042	0.153
Max Drawdown	0.972	0.563	1.630	0.612	0.902
# of obs	118	119	120	119	118
Panel B: 2000 - 2010					
Mean	0.045	0.128	0.124	0.000	0.045
T-Stat	2.252	3.398	4.212	0.066	3.778
Std	0.222	0.409	0.324	0.034	0.132
Skewness	3.104	0.127	0.988	0.308	2.659
Kurtosis	14.926	2.538	5.429	14.493	9.935
Min	-0.444	-1.310	-0.971	-0.175	-0.289
Max	1.437	1.467	1.560	0.185	0.764
Sharpe Ratio	0.205	0.311	0.383	0.006	0.343
Max Drawdown	1.020	2.434	1.063	0.423	0.213
# of obs	121	121	122	121	121
Panel C: 2010 - 2024					
Mean	0.023	0.018	0.034	0.001	0.010
T-Stat	1.515	0.451	1.080	0.527	0.784
Std	0.201	0.513	0.407	0.016	0.170
Skewness	-1.090	-0.543	-0.703	-1.923	-3.367
Kurtosis	19.808	9.011	22.340	22.812	34.740
Min	-1.446	-3.013	-2.929	-0.124	-1.453
Max	0.932	1.996	2.220	0.077	0.678
Sharpe Ratio	0.116	0.034	0.083	0.040	0.060
Max Drawdown	1.767	7.110	3.457	0.203	1.547
# of obs	171	171	170	169	171

Table A.7: Performance of Strategy 3 during the three sub-periods.

Appendix B

Additional Plots

This section provides the additional figures that are useful in our analysis. Figures B.1 to B.6 present the evolution of the model fits over one day for the Agriculture, Metals and Livestock sectors, as well as the evolution of the fit of the models over the years. Next, Figures B.7 and B.8 present the average monthly excess returns of the four sector portfolios with strategies 2 and 3. Finally, Figures B.9 to B.26 present the average trading volumes as a function of days to roll dates for all commodities.

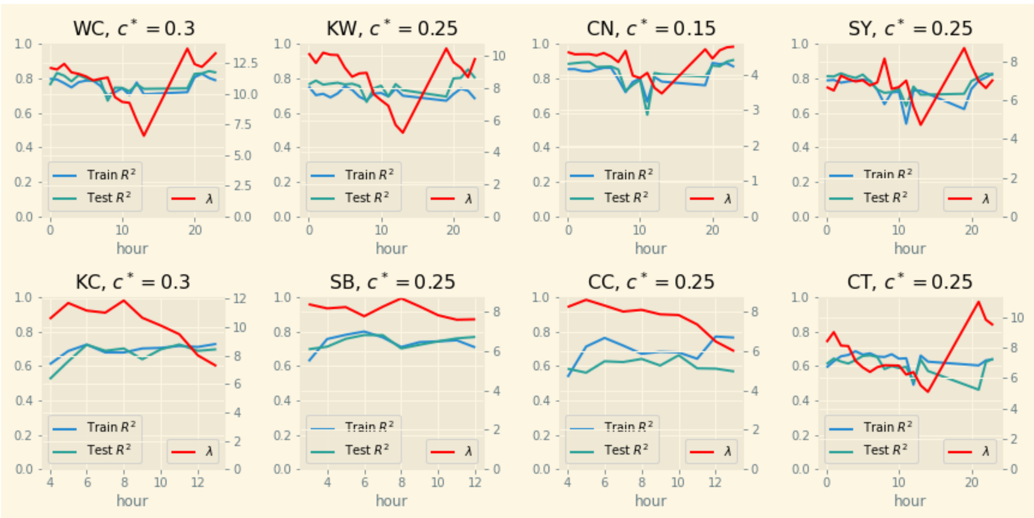


Figure B.1: Evolution of model fit for the agriculture sector with training start on 2024-01-01.

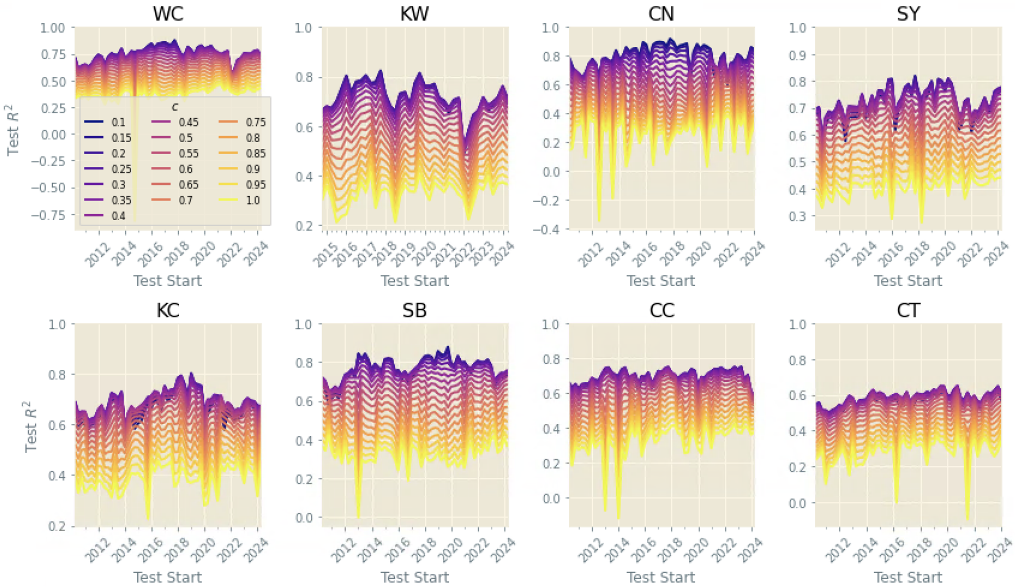


Figure B.2: Evolution of model fit for the agriculture sector over time.

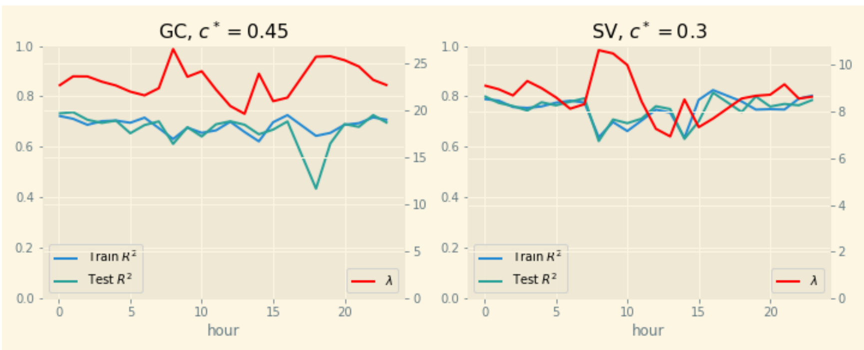


Figure B.3: Evolution of model fit for the metal sector with training start on 2024-01-01.

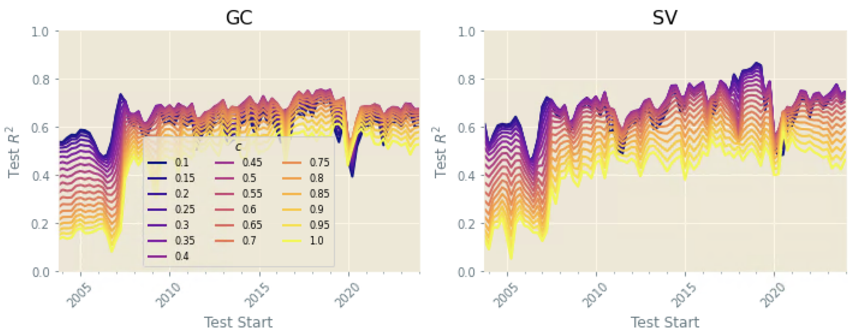


Figure B.4: Evolution of model fit for the metal sector over time.

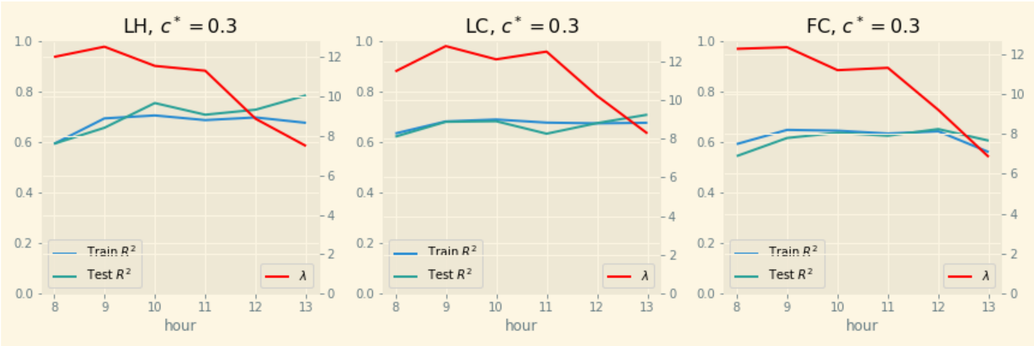


Figure B.5: Evolution of model fit for the livestock sector with training start on 2024-01-01.

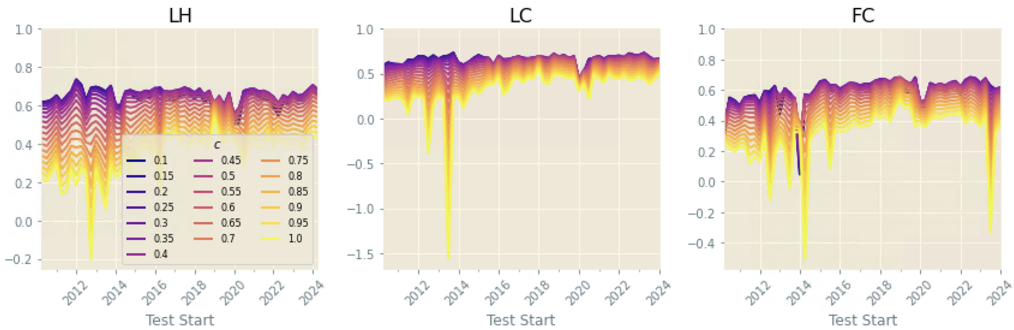


Figure B.6: Evolution of model fit for the livestock sector over time.

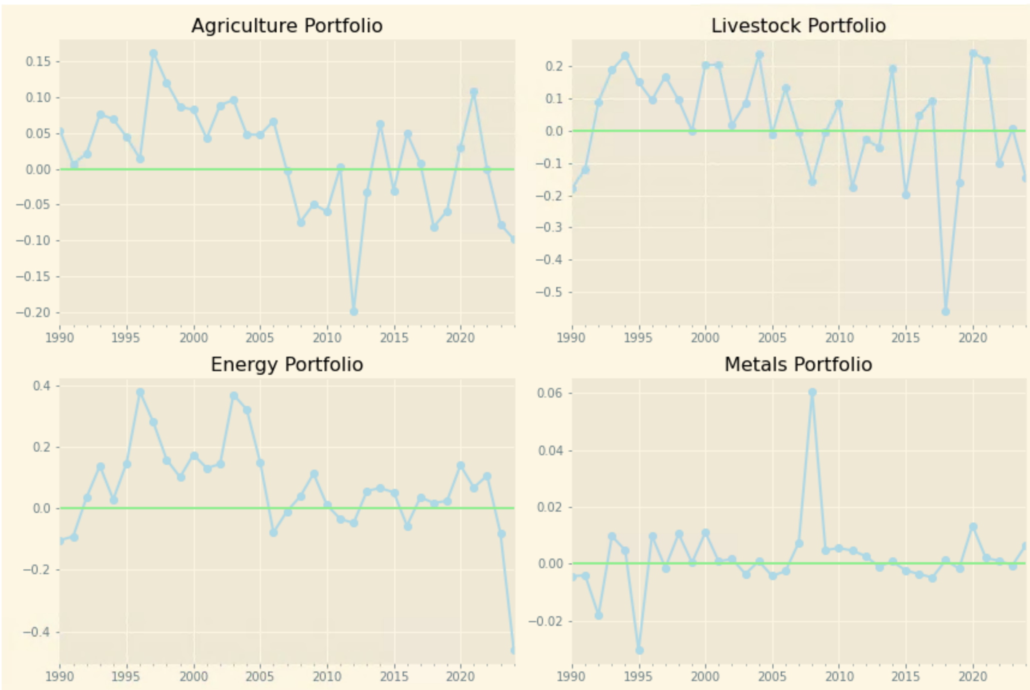


Figure B.7: Average Monthly Excess Returns of the Four Sector Portfolios with Strategy 2

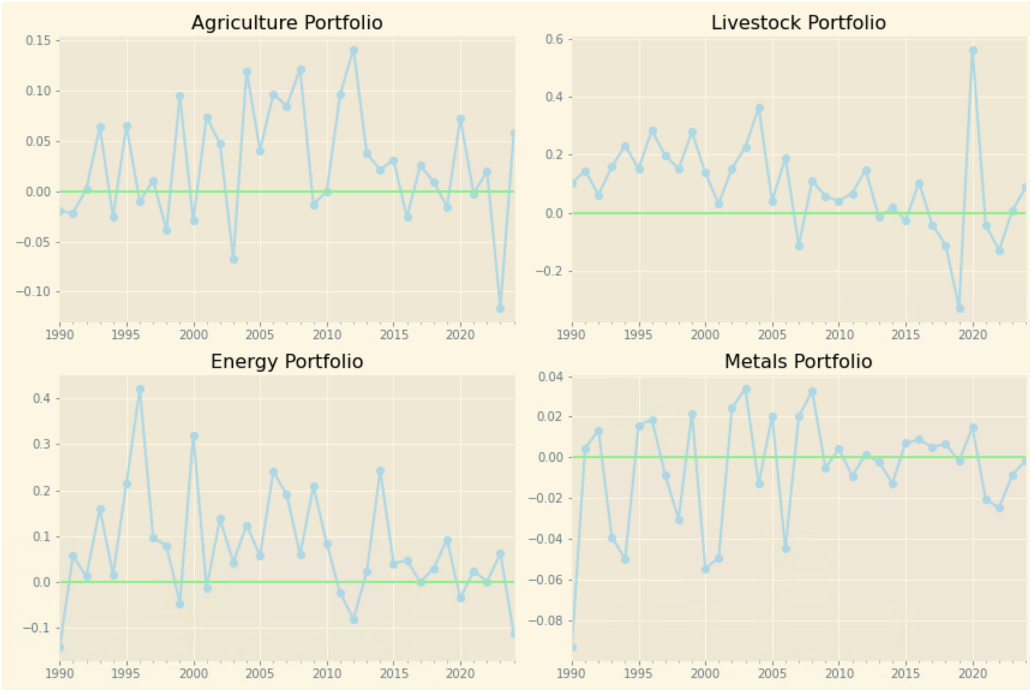


Figure B.8: Average Monthly Excess Returns of the Four Sector Portfolios with Strategy 3

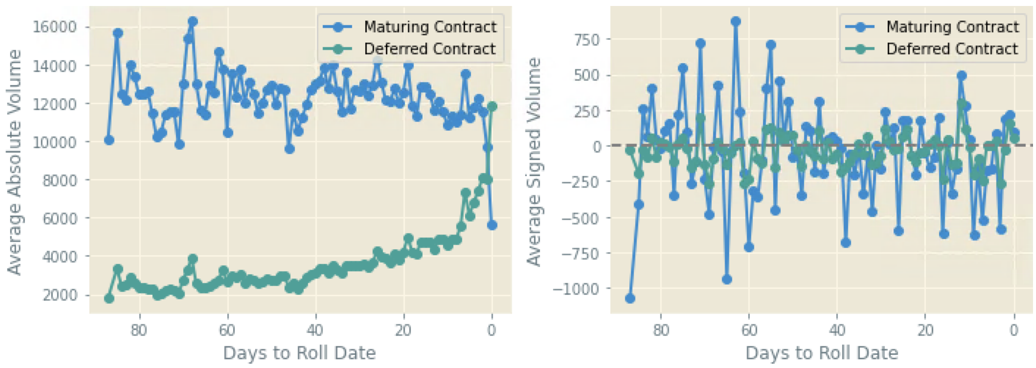


Figure B.9: Average trading volume as a function of days to roll dates for CC.

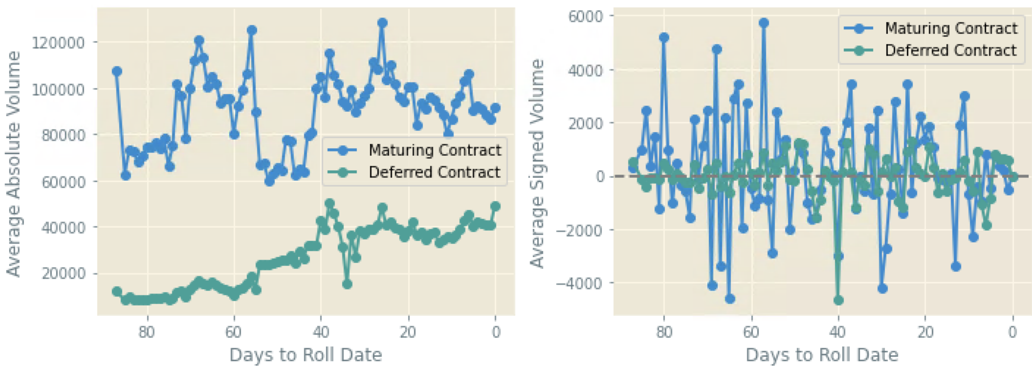


Figure B.10: Average trading volume as a function of days to roll dates for CN.

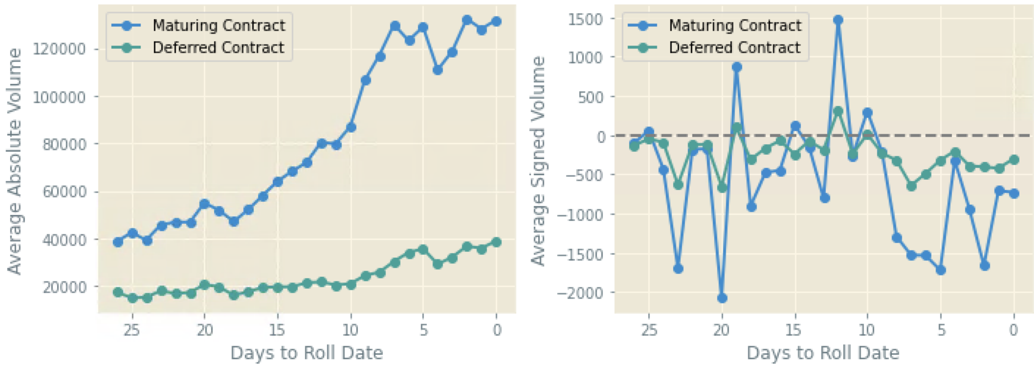


Figure B.11: Average trading volume as a function of days to roll dates for CO.

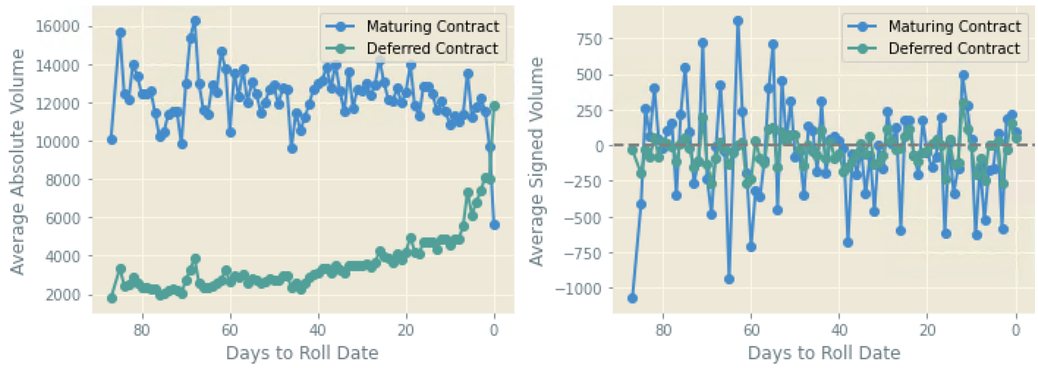


Figure B.12: Average trading volume as a function of days to roll dates for CT.

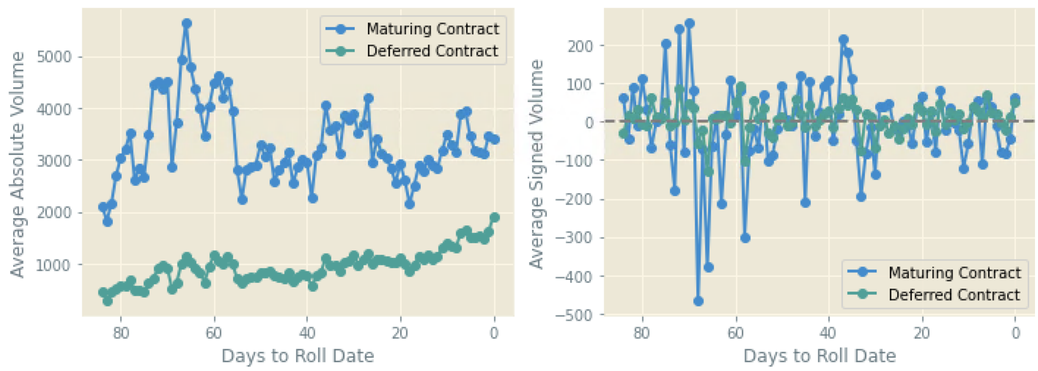


Figure B.13: Average trading volume as a function of days to roll dates for FC.

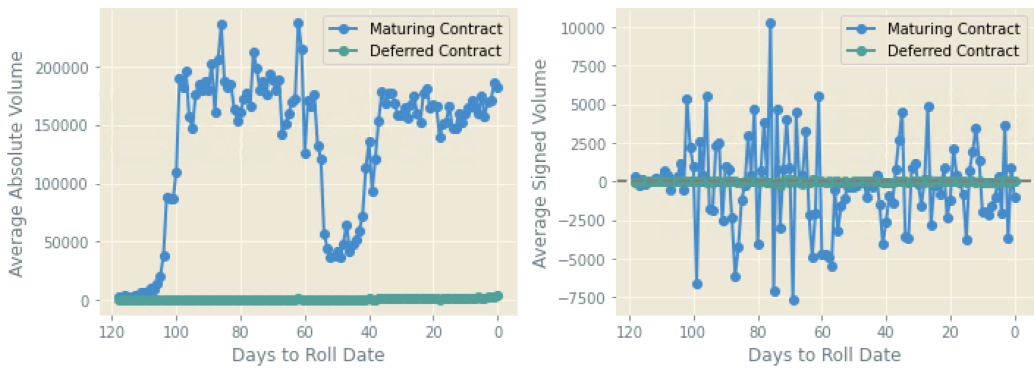


Figure B.14: Average trading volume as a function of days to roll dates for GC.

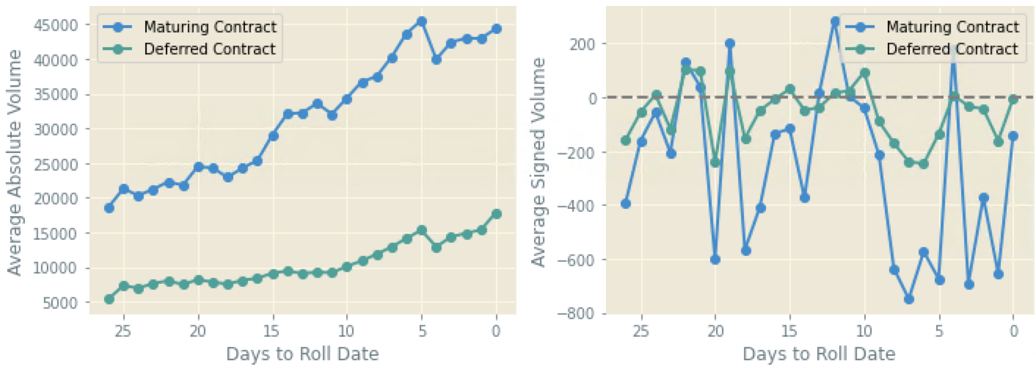


Figure B.15: Average trading volume as a function of days to roll dates for GO.

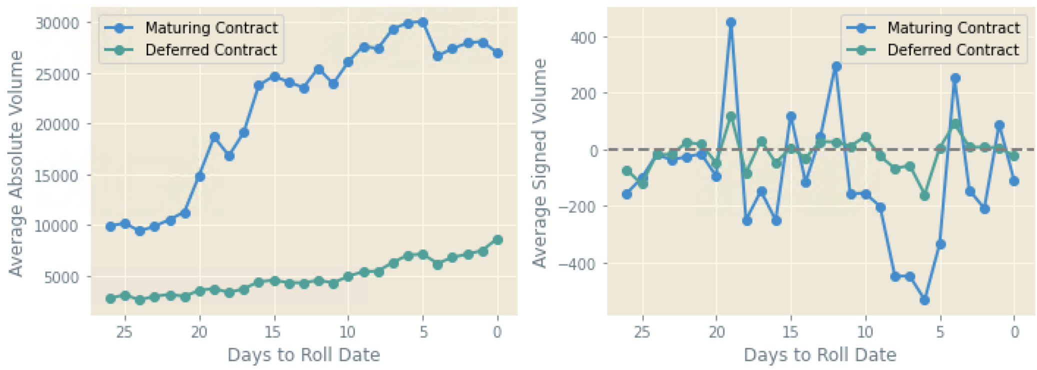


Figure B.16: Average trading volume as a function of days to roll dates for HO.

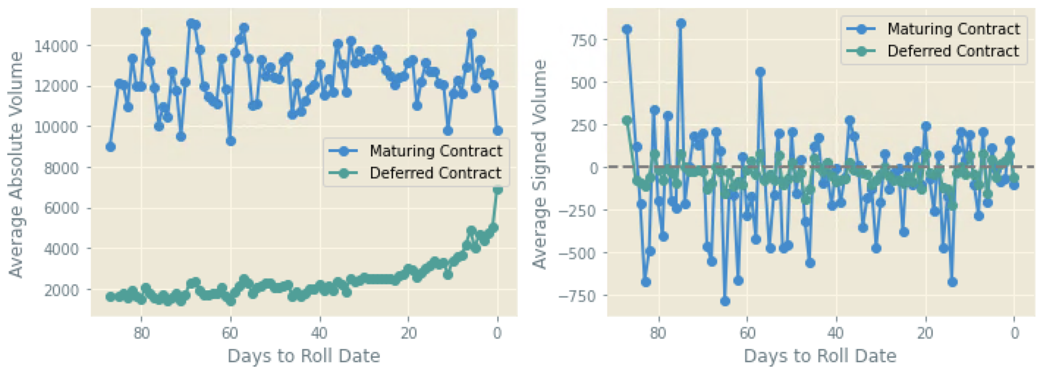


Figure B.17: Average trading volume as a function of days to roll dates for KC.

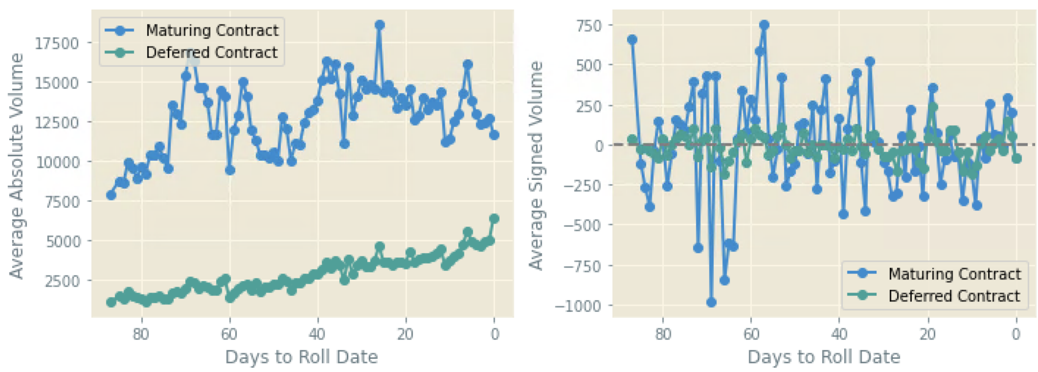


Figure B.18: Average trading volume as a function of days to roll dates for KW.

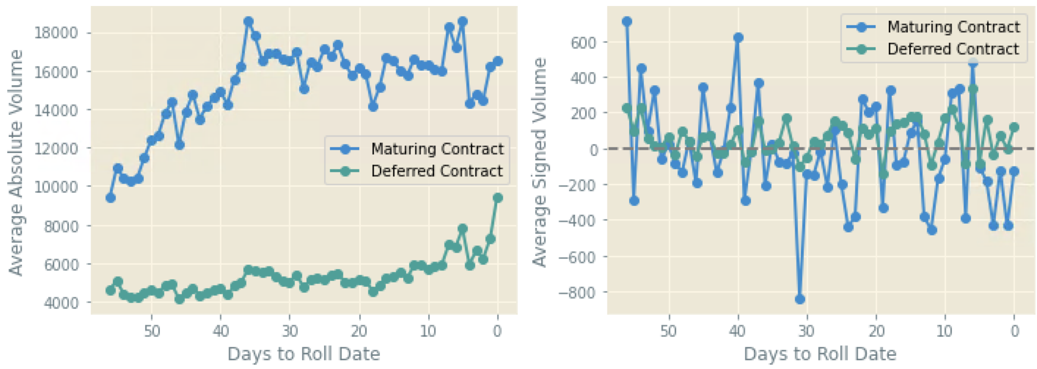


Figure B.19: Average trading volume as a function of days to roll dates for LC.

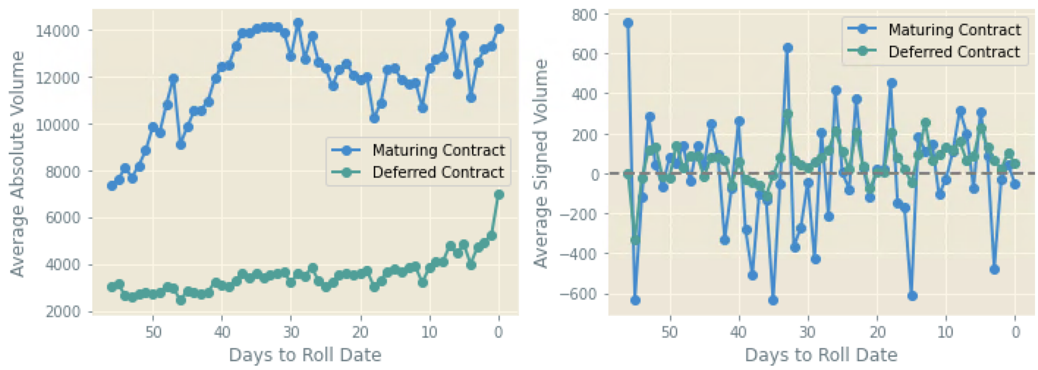


Figure B.20: Average trading volume as a function of days to roll dates for LH.

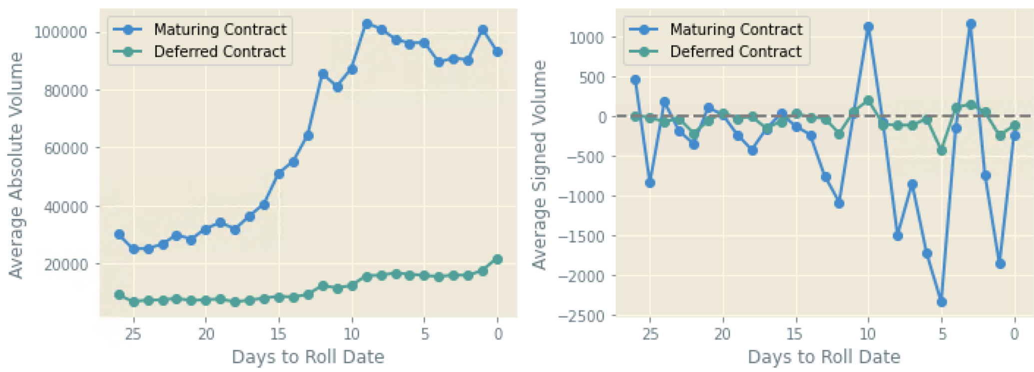


Figure B.21: Average trading volume as a function of days to roll dates for NG.

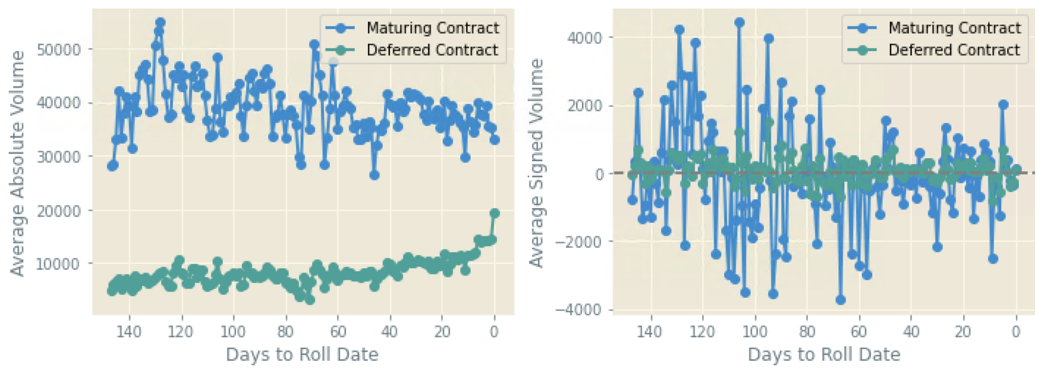


Figure B.22: Average trading volume as a function of days to roll dates for SB.

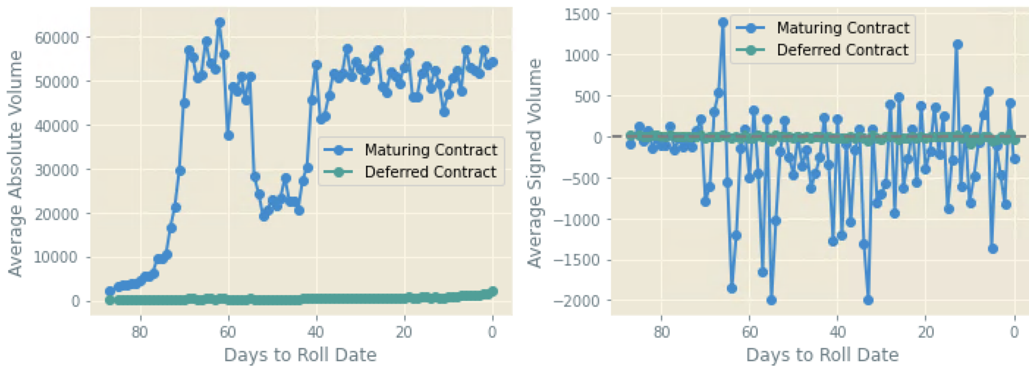


Figure B.23: Average trading volume as a function of days to roll dates for SV.

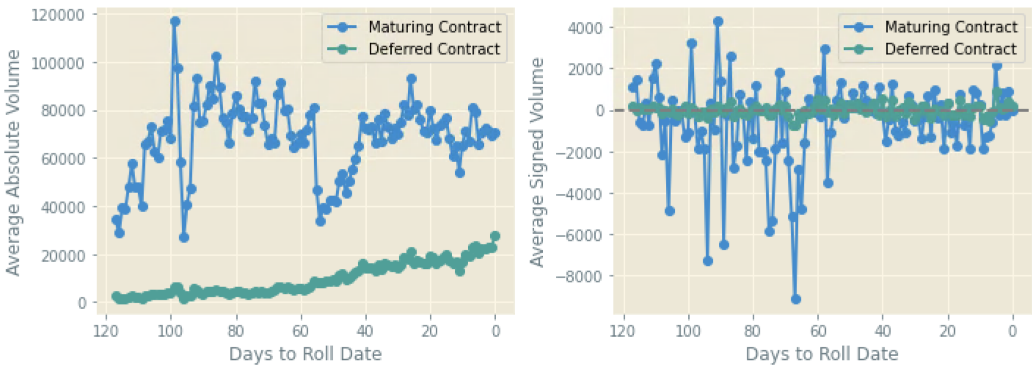


Figure B.24: Average trading volume as a function of days to roll dates for SY.

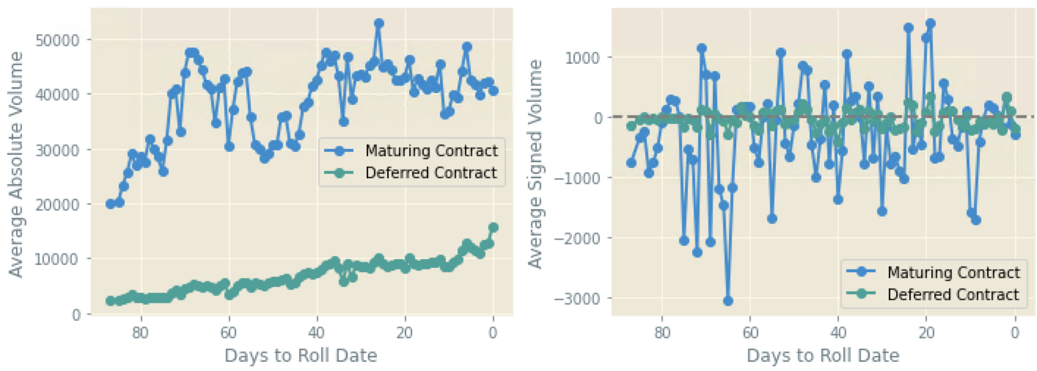


Figure B.25: Average trading volume as a function of days to roll dates for WC.

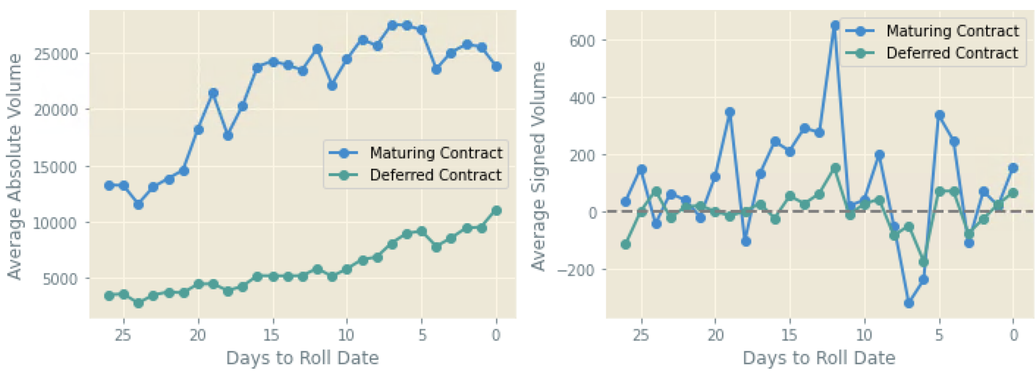


Figure B.26: Average trading volume as a function of days to roll dates for XB.

Bibliography

[1] N. M. Aulerich, S. H. Irwin, and P. Garcia, *Bubbles, food prices, and speculation: evidence from the cftc's daily large trader data files*, in The economics of food price volatility, University of Chicago Press, 2014, pp. 211–253.

[2] D. G. Baur and B. M. Lucey, *Is gold a hedge or a safe haven? an analysis of stocks, bonds and gold*, Financial review, 45 (2010), pp. 217–229.

[3] N. Bershova and D. Rakhlin, *The non-linear market impact of large trades: Evidence from buy-side order flow*, Quantitative finance, 13 (2013), pp. 1759–1778.

[4] H. Bessembinder, A. Carrion, L. Tuttle, and K. Venkataraman, *Liquidity, resiliency and market quality around predictable trades: Theory and evidence*, Journal of Financial economics, 121 (2016), pp. 142–166.

[5] J.-P. Bouchaud, J. D. Farmer, and F. Lillo, *How markets slowly digest changes in supply and demand*, in Handbook of financial markets: dynamics and evolution, Elsevier, 2009, pp. 57–160.

[6] J.-P. Bouchaud, Y. Gefen, M. Potters, and M. Wyart, *Fluctuations and response in financial markets: thesubtle nature ofrandom'price changes*, Quantitative finance, 4 (2003), p. 176.

[7] C. Brunetti and D. Reiffen, *Commodity index trading and hedging costs*, Journal of Financial Markets, 21 (2014), pp. 153–180.

[8] M. K. Brunnermeier and L. H. Pedersen, *Predatory trading*, The Journal of Finance, 60 (2005), pp. 1825–1863.

[9] E. Busseti and F. Lillo, *Calibration of optimal execution of financial transactions in the presence of transient market impact*, Journal of Statistical Mechanics: Theory and Experiment, 2012 (2012), p. P09010.

[10] J. Chen, *Futures market*, May 2022.

[11] I.-H. Cheng and W. Xiong, *Financialization of commodity markets*, Annu. Rev. Financ. Econ., 6 (2014), pp. 419–441.

[12] A. Fruth, T. Schöneborn, and M. Urusov, *Optimal trade execution and price manipulation in order books with time-varying liquidity*, Mathematical Finance, 24 (2014), pp. 651–695.

[13] A. Ganti, *Open interest: Definition, how it works, and example*, February 2024.

[14] G. Gorton and K. G. Rouwenhorst, *Facts and fantasies about commodity futures*, Financial Analysts Journal, 62 (2006), pp. 47–68.

[15] J. D. Hamilton and J. C. Wu, *Effects of index-fund investing on commodity futures prices*, International economic review, 56 (2015), pp. 187–205.

[16] L. Harris and E. Gurel, *Price and volume effects associated with changes in the s&p 500 list: New evidence for the existence of price pressures*, the Journal of Finance, 41 (1986), pp. 815–829.

[17] J. Hasbrouck and D. J. Seppi, *Common factors in prices, order flows, and liquidity*, Journal of financial Economics, 59 (2001), pp. 383–411.

[18] N. Hey, J.-P. Bouchaud, I. Mastromatteo, J. Muhle-Karbe, and K. Webster, *The cost of misspecifying price impact*, arXiv preprint arXiv:2306.00599, (2023).

[19] R. W. Holthausen, R. W. Leftwich, and D. Mayers, *The effect of large block transactions on security prices: A cross-sectional analysis*, Journal of Financial Economics, 19 (1987), pp. 237–267.

[20] S. H. Irwin, D. R. Sanders, and L. Yan, *The order flow cost of index rolling in commodity futures markets*, Applied Economic Perspectives and Policy, 45 (2023), pp. 1025–1050.

[21] B. G. Malkiel, *Efficient market hypothesis*, in Finance, Springer, 1989, pp. 127–134.

[22] H. Markowitz, *The utility of wealth*, Journal of political Economy, 60 (1952), pp. 151–158.

[23] L. P. Mertens, A. Ciacci, F. Lillo, and G. Livieri, *Liquidity fluctuations and the latent dynamics of price impact*, Quantitative Finance, 22 (2022), pp. 149–169.

[24] Y. Mou, *Limits to arbitrage and commodity index investment: front-running the goldman roll*, Available at SSRN 1716841, (2010).

[25] J. Muhle-Karbe, Z. Wang, and K. Webster, *Stochastic liquidity as a proxy for nonlinear price impact*, Operations Research, 72 (2024), pp. 444–458.

[26] A. A. Obizhaeva and J. Wang, *Optimal trading strategy and supply/demand dynamics*, Journal of Financial markets, 16 (2013), pp. 1–32.

[27] D. R. Sanders and S. H. Irwin, *Energy futures prices and commodity index investment: New evidence from firm-level position data*, Energy Economics, 46 (2014), pp. S57–S68.

[28] A. Shleifer, *Do demand curves for stocks slope down?*, The Journal of Finance, 41 (1986), pp. 579–590.

[29] A. Shleifer and R. W. Vishny, *The limits of arbitrage*, The Journal of finance, 52 (1997), pp. 35–55.

[30] H. R. Stoll and R. Whaley, *Commodity index investing and commodity futures prices*, Journal of Applied Finance (Formerly Financial Practice and Education), 20 (2010).

[31] T. I. Team, *What is the history of futures?*, October 2022.

[32] G. J. Wang, C. Y. Yao, and A. Yelekenova, *Etf rebalancing, hedge fund trades, and capital market*, Available at SSRN 4324054, (2023).

[33] K. T. Webster, *Handbook of price impact modeling*, Chapman and Hall/CRC, 2023.

[34] J. Wurgler and E. Zhuravskaya, *Does arbitrage flatten demand curves for stocks?*, The Journal of Business, 75 (2002), pp. 583–608.