UNIVERSITY OF TURKU
Turku School of Economics

Commodity Futures Term Structure and Single-Commodity ETP Returns

Evidence from 2015–2025

Accounting and Finance,
Department of Accounting and Finance
Bachelor's thesis

Author:
Niki Nieminen

Supervisor:
Md Khaled Hossain Rafi

23.4.2026
Turku

Student’s statement regarding the use of Artificial Intelligence (AI) for preparing and/or writing this thesis:

☐ **I have not used any AI-based tools.**

☒ **I have used AI-based tools**. Their use is documented in the Appendix. The AI tools were used in a way that complies with academic integrity guidelines.

The originality of this thesis has been checked in accordance with the University of Turku quality assurance system using the Turnitin Originality Check service.

Bachelor's thesis

Subject: Accounting and Finance
Author: Niki Nieminen
Title: Commodity Futures Term Structure and Single-Commodity ETP Returns: Evidence from 2015–2025
Supervisor: Md Khaled Hossain Rafi
Number of pages: 48 pages (+ appendices 5 pages)
Date: 23.4.2026

This thesis examines whether the term structure of commodity futures helps explain the returns of futures-based single-commodity exchange-traded products (ETPs) and whether this relationship changed during the COVID-19 period. The empirical analysis is based on weekly panel data for 16 non-leveraged single-commodity ETPs from the energy, metals, and agriculture sectors over the period from 2015 to 2025. The data is obtained from LSEG Workspace. The main explanatory variable is a commodity-specific measure of futures term structure constructed from front-month and second-month futures prices. Product returns are analyzed using fixed-effects panel regressions with product-level clustered standard errors, complemented by robustness checks using alternative covariance estimators and week fixed effects.

The results show that, in the full sample baseline setting, a steeper contango is associated with lower ETP returns. This indicates that futures curve conditions are relevant for explaining the returns of futures-based commodity ETPs beyond movements in the underlying commodity price alone. However, the evidence does not support the view that the negative contango effect became uniformly stronger during the COVID-19 period. Instead, the relationship varied across pandemic subperiods: it weakened after the initial shock, while the recovery-period estimate in the full sample became positive but remained only weakly significant. Subsample and robustness analyses further show that the stronger time variation is driven mainly by energy products, whereas the baseline negative relationship is not limited to the energy sector.

The findings imply that investors in futures-based commodity ETPs should consider not only expected commodity price movements but also the slope of the futures curve, especially in energy-related products. Future research could extend the analysis by using larger product samples, more product-specific roll measures, and alternative approaches to modeling common time effects.

Keywords: commodity ETPs, futures term structure, contango, backwardation, COVID-19, panel data

Kandidaatintutkielma

Oppiaine: Laskentatoimi ja rahoitus
Tekijä: Niki Nieminen
Otsikko: Hyödykefutuurien termirakenne ja yksittäisiin hyödykkeisiin sijoittavien ETP-tuotteiden tuotot: Evidenssiä vuosilta 2015–2025
Ohjaaja: Md Khaled Hossain Rafi
Sivumäärä: 48 sivua (+ liitteet 5 sivua)
Päivämäärä: 23.4.2026

Tiivistelmä

Tässä tutkielmassa tarkastellaan, selittääkö hyödykefutuurien termirakenne futuuripohjaisten yksittäisiin hyödykkeisiin sijoittavien pörssilistattujen tuotteiden (ETP) tuottoja, ja muuttuiko tämä yhteys COVID-19-pandemian aikana. Empiirinen analyysi perustuu viikoittaiseen paneeliaineistoon, joka kattaa 16 vivuttamatonta yksittäisiin hyödykkeisiin sijoittavaa ETP-tuotetta energia-, metalli- ja maataloussektoreilta vuosilta 2015–2025. Aineisto on hankittu LSEG Workspace -palvelusta. Keskeinen selittävä muuttuja on hyödykekohtainen futuurien termirakenteen mitta, joka on laskettu lähimmän ja toiseksi lähimmän futuurikuukauden hinnoista. Tuottoja analysoidaan kiinteiden vaikutusten paneeliregressioilla tuotetason klusteroiduilla keskivirheillä, ja lisäksi tehdään robustisuustarkistuksia vaihtoehtoisilla kovarianssiestimointimenetelmillä sekä viikkotason kiinteillä vaikutuksilla.

Tulokset osoittavat, että koko otoksen perusmallissa jyrkempi contango liittyy alhaisempiin ETP-tuottoihin. Tämä viittaa siihen, että futuurikäyrän rakenne on merkityksellinen futuuripohjaisten hyödyke-ETP-tuotteiden tuottojen selittäjä pelkän kohde-etuuden hintakehityksen lisäksi. Evidenssi ei kuitenkaan tue näkemystä, jonka mukaan negatiivinen contango-vaikutus olisi yhdenmukaisesti voimistunut COVID-19-pandemian aikana. Sen sijaan yhteys vaihteli pandemian eri alavaiheissa: se heikkeni alkusokin jälkeen, kun taas elpymisvaiheen estimaatti koko otoksessa muuttui positiiviseksi, mutta jäi vain heikosti merkitseväksi. Osa-aineisto- ja robustisuusanalyysit osoittavat lisäksi, että voimakkaampi ajallinen vaihtelu on pääosin energiatuotteiden aiheuttamaa, kun taas perusmallissa havaittu negatiivinen yhteys ei rajoitu pelkästään energiasektoriin.

Löydökset implikoivat, että futuuripohjaisten hyödyke-ETP-tuotteiden sijoittajien tulisi ottaa huomioon odotettujen hintamuutosten lisäksi myös futuurikäyrän kaltevuus, erityisesti energiaan liittyvissä tuotteissa. Jatkotutkimuksessa analyysiä voitaisiin laajentaa suuremmilla tuoteotoksilla, tuotekohtaisemmilla rullaamismitoilla sekä vaihtoehtoisilla tavoilla mallintaa yhteisiä aikaefektejä.

Avainsanat: Hyödyke-ETP-tuotteet, futuurien termirakenne, contango, backwardation, COVID-19, paneeliregressio

TABLE OF CONTENTS

FIGURES

TABLES

1 Introduction

Commodity exchange-traded products have made commodity investing more accessible to a wider range of investors. However, the performance of futures-based commodity products is not determined by movements in the underlying commodity price alone. Because these products maintain exposure through futures contracts, investor returns also depend on the structure of the underlying futures market. When the futures curve is in contango, rolling exposure forward is generally associated with less favourable return outcomes, whereas backwardation is associated with more favourable conditions for maintaining exposure. (Erb & Harvey, 2006; Fama & French, 1987)

This issue is economically important because the shape of the commodity futures curve reflects underlying market conditions such as inventories, convenience yield, and risk premia. These factors vary across commodities and over time, which implies that the return implications of contango and backwardation are unlikely to be constant (Gorton et al., 2013; Working, 1949). At the same time, commodity exposure is increasingly accessed through exchange-traded products rather than through direct participation in futures markets. This makes the relationship between futures curve conditions and realized product returns especially relevant for investors in futures-based single-commodity ETPs.

Existing research provides a strong foundation for understanding commodity futures pricing, roll yield, and the performance of commodity investment products. At the same time, less attention has been paid to this relationship in a unified panel setting covering multiple non-leveraged single-commodity ETPs across different commodity groups and market phases. This thesis addresses that issue by examining whether a market-level measure of futures term structure helps explain the returns of futures-based single-commodity ETPs and whether this relationship changed during the COVID-19 period.

The COVID-19 crisis provides a particularly useful setting for this analysis. The pandemic disrupted commodity markets through demand shocks, inventory pressures, storage constraints, and elevated uncertainty, with especially strong effects in energy markets (Borgards et al., 2021; Szczygielski et al., 2022). These conditions make it possible to examine not only whether futures term structure matters for ETP returns on average, but also whether its importance differed across distinct market phases.

The empirical analysis uses weekly panel data for 16 non-leveraged single-commodity ETPs over the period 2015–2025. The sample includes products from the energy, metals, and agriculture sectors. By combining insights from commodity futures pricing and commodity ETP performance in one empirical framework, this thesis contributes to a more precise understanding of how futures curve conditions are reflected in realized product returns.

1.1 Research Question

The main research question of this thesis is whether the term structure of commodity futures helps explain the returns of futures-based single-commodity ETPs and whether this relationship changed during the COVID-19 period. To address this broader question, the study examines the following more specific research questions:

- Does the shape of the futures curve help explain the returns of futures-based single-commodity ETPs?
- Did this relationship change during the COVID-19 period?

These questions are relevant from both academic and practical perspectives. From an academic perspective, the thesis connects the literature on commodity futures pricing with the literature on commodity exchange-traded products. From a practical perspective, the results help clarify why returns on futures-based commodity ETPs may differ from movements in the underlying commodity price alone and why futures curve conditions may matter for investors.

1.2 Structure

This thesis is structured as follows. Chapter 2 presents the theoretical background, covering commodity futures pricing, the term structure of futures prices, the characteristics of commodity markets, the structure of futures-based commodity ETPs, and commodity market conditions during the COVID-19 period. Chapter 3 describes the empirical methodology, including the data, sample selection, variable construction, model specification, subsample analyses, and robustness considerations. Chapter 4 reports the empirical results for the full sample, the commodity-group subsamples, and the robustness checks. Chapter 5 concludes by summarizing the main findings, discussing their implications, and outlining the main limitations of the study.

2 Theoretical background

2.1 Futures contracts and pricing mechanisms

2.1.1 Spot and Futures Prices

Futures contracts are exchange-traded derivative instruments that specify the purchase or delivery of an underlying asset at a predetermined price on a future date. In commodity markets, futures contracts have traditionally facilitated risk transfer between commercial participants and financial investors. They therefore serve not only as instruments for price exposure, but also as mechanisms through which market participants share risk over time. (Dusak, 1973)

The relationship between spot and futures prices is commonly described through the cost-of-carry framework. In its simplest form, the futures price reflects the current spot price adjusted for carrying costs, such as financing and storage. In commodity markets, however, this relationship cannot be understood solely in mechanical terms. Because commodities are physical goods, the pricing of futures contracts is also influenced by inventory conditions and the economic benefits of holding the physical commodity.

Working (1949) argues that the spread between spot and futures prices should be understood through the "price of storage," which includes both observable storage costs and the convenience yield associated with holding inventories. This means that the difference between spot and futures prices reflects an equilibrium outcome in the underlying commodity market rather than merely a forecast of future spot prices. Fama and French (1987) similarly show that commodity futures prices contain information about both expected spot-price movements and risk premia, indicating that futures prices reflect broader economic fundamentals rather than simple price expectations alone.

This is particularly important in commodity markets, where production, transportation, storage, and consumption conditions directly affect price formation. As a result, the relationship between spot and futures prices is shaped by both financial and physical market forces. This provides the foundation for understanding why the futures curve may contain economically meaningful information for investors in futures-based commodity ETPs.

2.1.2 Term Structure of Futures Prices

At any given point in time, several futures contracts with different maturities are traded simultaneously, forming the term structure of futures prices. The shape of this futures curve reflects the interaction between storage costs, convenience yield, and risk premia. When inventories are abundant and carrying costs dominate, futures prices tend to exceed spot prices, producing an upward-sloping curve commonly referred to as contango (Working, 1949).

Brennan (1958) formalizes the role of convenience yield in this setting. When inventories are scarce, the implicit benefit of holding the physical commodity may exceed observable storage costs, making near-term contracts relatively more valuable than longer-dated contracts. Under such conditions, the futures curve may slope downward, a situation commonly described as backwardation. The shape of the futures curve can therefore be interpreted as an indicator of underlying market tightness and inventory conditions.

The term structure may also reflect time-varying risk premia rather than physical market conditions alone. Bessembinder and Lemmon (2002) show that forward prices can depend on hedging pressure and demand uncertainty, while Hamilton and Wu (2014) provide evidence that crude oil futures premia vary over time with broader macroeconomic conditions. These findings suggest that the futures curve should not be interpreted purely as a storage-based relationship, but rather as the outcome of both physical fundamentals and financial market forces.

For this reason, the term structure of commodity futures is economically informative beyond the pricing of individual contracts. It summarizes market conditions that are relevant for investors maintaining futures exposure over time. This makes the futures curve especially important in the context of futures-based commodity ETPs, whose returns depend not only on movements in the underlying commodity but also on the maturity structure of the contracts used to maintain exposure.

2.1.3 Futures Returns and the Economic Meaning of Roll Yield

The economic importance of the futures curve becomes particularly clear when examining the returns of investors who maintain continuous futures exposure over time. Unlike an investor who invests in the physical commodity, a futures investor does not simply earn the return of the underlying commodity. Instead, realized returns depend on both price movements in the underlying market and the process through which expiring contracts are replaced with new ones.

Fama and French (1987) show that commodity futures returns reflect expected spot-price changes as well as risk premia. This implies that futures returns cannot be interpreted as a direct equivalent of spot returns. Erb and Harvey (2006) further argue that the roll component is a central part of long-run commodity futures performance. When investors maintain exposure by selling a near-term contract and buying a longer-dated contract, the return consequences depend on the shape of the futures curve at the time of the roll.

In contango, rolling exposure forward typically involves replacing a cheaper near-term contract with a more expensive longer-dated contract, which is associated with a negative roll yield. In backwardation, the opposite holds: the investor replaces a more expensive near-term contract with a cheaper longer-dated contract, which is associated with a positive roll yield. For this reason, the slope of the futures curve can materially influence realized returns even when the underlying commodity price itself does not move strongly.

This mechanism is especially important for futures-based commodity ETPs, because such products maintain exposure through repeated contract rolls rather than through direct ownership of the underlying commodity. As a result, the return implications of contango and backwardation are not merely theoretical, but directly relevant for product performance. At the same time, it is important to distinguish between the general return implications of futures curve shape and the exact realized roll outcome of an individual product. Actual product-level outcomes may also depend on benchmark design, contract selection, and roll timing. Nevertheless, the concept of roll yield provides the key economic link between futures curve structure and the returns of futures-based commodity investment products.

2.2 Characteristics of Commodity Markets

2.2.1 Physical Market Fundamentals

Commodity markets differ from most financial asset markets because the underlying assets are physical goods that must be produced, transported, stored, and consumed. As a result, commodity prices are closely linked to real economic conditions such as production capacity, supply disruptions, transportation constraints, and end-user demand. These features make commodity markets particularly sensitive to changes in physical availability and inventory conditions.

Inventories play a central role in this setting because they help balance temporary mismatches between supply and demand. When inventories are abundant, markets are generally better able to absorb short-run shocks, and price fluctuations may remain relatively contained. When inventories are low, even small disruptions in supply or demand can lead to sharper price responses. Gorton and Rouwenhorst (2006) emphasize that this physical-market dimension is one of the key characteristics distinguishing commodities from traditional financial assets. In a later study, Gorton et al. (2013) show that inventory conditions are closely related to expected commodity futures returns.

Storage capacity is another important feature of commodity markets. Because commodities require physical storage infrastructure, excess supply may create market stress when inventories accumulate faster than they can be stored efficiently. This issue becomes especially relevant in energy markets, where storage constraints can materially affect both spot and futures prices. Hamilton (2009) highlights the importance of such mechanisms in oil markets, where imbalances between production and consumption may amplify price volatility when storage capacity becomes binding.

Taken together, these features imply that commodity price dynamics are shaped not only by financial expectations but also by the underlying physical structure of the market. This is important for the present study because the shape of the futures curve and the returns of futures-based commodity ETPs are ultimately linked to market conditions that originate in the physical commodity sector.

2.2.2 Commodities as an Investment Class

Over time, commodities have become more than instruments for commercial hedging and risk transfer. They have also developed into an investment asset class that attracts institutional and retail investors seeking diversification and alternative sources of return. This shift has increased the importance of understanding commodity futures not only as tools for producers and consumers, but also as investment vehicles in their own right.

Gorton and Rouwenhorst (2006) show that commodity futures have historically generated returns comparable to equities while displaying relatively low correlation with traditional financial assets. These characteristics have contributed to the view that commodities can play a useful role in diversified portfolios. In this sense, the appeal of commodity investment does not arise solely from exposure to individual raw materials, but also from the broader portfolio properties of the asset class.

At the same time, investing in commodities differs from investing in conventional financial securities. For most investors, exposure is obtained through futures markets or products linked to futures rather than through direct ownership of physical commodities. This means that realized investment performance depends not only on movements in the underlying commodity price, but also on the structure of futures markets and the way exposure is maintained over time.

This distinction is particularly important for the present study. If commodity investment is accessed mainly through futures-based instruments, then the slope of the futures curve becomes relevant not only for futures pricing but also for the returns ultimately experienced by investors. The investment perspective therefore provides an important link between commodity market theory and the performance of futures-based single-commodity ETPs.

2.2.3 Financialization and Market Dynamics

The growing role of commodities as an investment asset class has been accompanied by the broader financialization of commodity markets. In this context, financialization refers to the increasing participation of institutional and other non-commercial investors in commodity futures trading. As commodity exposure has become more accessible through index products and exchange-traded instruments, commodity futures markets have become more closely connected to developments in the wider financial system.

Tang and Xiong (2012) show that the expansion of commodity index investment increased the co-movement between commodity futures and broader financial markets. Cheng and Xiong (2014) similarly document the growing presence of financial investors in commodity futures markets and argue that this development has strengthened the link between commodity prices and portfolio allocation decisions. These studies suggest that commodity futures prices reflect not only physical-market conditions but also the behaviour of financial investors.

Financialization may affect market dynamics through several channels. Basak and Pavlova (2016) show theoretically that investment flows can influence equilibrium futures prices when arbitrage is limited, while Mou (2011) provides evidence that predictable index-roll trading can create temporary price pressure in commodity futures markets. Together, these findings imply that futures prices may be shaped not only by storage conditions and hedging needs, but also by systematic trading behaviour and portfolio rebalancing.

For the present study, this is relevant because futures-based commodity ETPs operate at the intersection of commodity markets and financial markets. Their returns depend on futures-market conditions, but those conditions may themselves be influenced by financial participation and trading flows. This does not remove the importance of physical fundamentals, but it does imply that the relationship between term structure and ETP returns may vary across market environments and may become especially relevant during periods of market stress.

2.3 Structure and Performance of Futures-Based Commodity ETPs

2.3.1 Structure of Single-Commodity Exchange-Traded Products

Single-commodity exchange-traded products (ETPs) are listed financial instruments that provide investors with exposure to a single underlying commodity through securities traded on a stock exchange. In the context of this study, we use the term ETP because it captures both exchange-traded funds (ETFs) and exchange-traded commodities (ETCs). Although these instruments differ in legal structure, they serve a similar economic function when they provide non-leveraged long exposure to a single commodity (Marszk, 2017; Orlando, 2025).

In the literature regarding exchange-traded products, ETFs are usually treated as the most established segment. ETFs are generally structured as pooled investment vehicles and the shares are traded continuously on an exchange. ETFs typically aim to track the performance of a benchmark index or asset class. (Ben-David et al., 2017; Orlando, 2025)

A key distinction in this study concerns ETCs. As Marszk (2017) emphasizes, ETCs should not simply be treated as commodity ETFs under another name. The reason for this is that Marszk notes ETCs being commonly structured as debt instruments rather than investment funds. Besides this difference, what is important in the context of our study is that both ETFs and ETCs are exchange-traded instruments and designed to deliver returns linked to the underlying commodity market. Therefore, the thesis distinguishes between the legal structures of ETFs and ETCs while emphasizing the economic similarity relevant for the analysis. In other words, products may differ formally, while still offering investors comparable access to commodity price movements. (Marszk, 2017; Orlando, 2025)

Single-commodity ETPs also differ from broad commodity index products in an important way. Rather than offering diversified exposure across several commodities, they are tied to one underlying market, such as crude oil, natural gas, copper, or wheat. This makes their performance more directly linked to commodity-specific market conditions, including the structure of the relevant futures curve. It also makes them particularly suitable for examining how differences in futures-market conditions are reflected in product returns.

For the present thesis, the key point is that the products in the sample are comparable at the economic level even if their legal wrappers are not identical. All selected products are non-leveraged, long-only, and linked to a single underlying commodity. This makes them an appropriate group for examining whether the shape of the underlying commodity futures curve helps explain differences in realized ETP returns.

2.3.2 Futures Curve Exposure and ETP Returns

For futures-based single-commodity ETPs, the structure of the underlying futures market is directly relevant for realized returns. Prior research supports this link. Stewart et al. (2023) show that ETP tracking performance varies across agricultural and energy markets, while Cortazar et al. (2024) show that expected returns on commodity ETFs are closely related to the characteristics of their underlying futures exposure. Chincarini and Moneta (2021) similarly emphasize that in oil markets the roll component can form an economically important part of investor returns. Together, these studies indicate that futures-based commodity ETP returns inherit important features from the futures positions through which exposure is maintained.

At the same time, the relationship between futures curve shape and ETP returns is not mechanically one-to-one. Actual product performance also depends on benchmark design, contract selection, roll timing, and other implementation details. This distinction is important for the present thesis. The empirical analysis does not attempt to measure the exact realized roll yield of each individual product. Instead, it uses a market-level proxy for the slope of the underlying commodity's futures curve. The purpose is therefore not to reproduce each product's exact rolling outcome, but to examine whether ETP returns are systematically related to the broader term-structure environment of the underlying commodity market.

This distinction does not weaken the economic relevance of the analysis. On the contrary, it makes it possible to study whether a common futures-curve measure helps explain return differences across a set of comparable single-commodity ETPs. In this sense, futures curve exposure provides the key conceptual bridge between commodity futures pricing and the realized returns of futures-based commodity ETPs.

2.3.3 Market Conditions and ETP Performance

The return performance of futures-based commodity ETPs may vary not only with the shape of the futures curve, but also with the broader market environment in which that exposure is maintained. Periods of elevated uncertainty may affect commodity futures prices through changes in volatility, liquidity, hedging pressure, and risk premia. This means that the relationship between term structure and ETP returns may differ over time.

Prior research supports this view. Hamilton and Wu (2014) show that risk premia in crude oil futures vary over time with broader macroeconomic conditions, while Chen et al. (2023) show that financial stress is associated with higher commodity price volatility, especially in energy markets. Bakas and Triantafyllou (2020) similarly show that pandemic-related uncertainty increased commodity market volatility. Together, these studies suggest that the pricing environment faced by futures-based commodity products may become materially different during episodes of market stress.

For futures-based ETPs, this matters because stressed market environments may alter both the level and the return implications of futures curve conditions. If uncertainty, disrupted demand, or changing hedging needs affect the slope of the futures curve, they may also affect how strongly that slope is reflected in realized product returns. The economic importance of term structure may therefore depend not only on the commodity itself, but also on the surrounding market regime.

This perspective is especially relevant for the present thesis because it motivates the possibility that the relationship between term structure and ETP returns changed during the COVID-19 period. The next section therefore turns to commodity market conditions during the pandemic and explains why that period provides a useful setting for examining whether the role of term structure differed across market phases.

2.4 Commodity Markets During the COVID-19 Crisis

The COVID-19 period provides an especially relevant setting for examining commodity futures markets because the pandemic disrupted both physical commodity flows and financial market conditions at the same time. Restrictions on mobility, changes in industrial activity, supply-chain disruptions, and shifts in investor uncertainty affected commodity markets through several channels simultaneously. As a result, the pandemic created unusually sharp changes in demand conditions, inventories, and futures pricing.

These effects were particularly visible in energy markets. Borgards et al. (2021) show that commodity futures markets exhibited strong price reactions during the pandemic, with crude oil displaying especially unusual behaviour. This was consistent with the collapse in transport demand and the rapid deterioration in storage conditions during the early phase of the crisis. In such an environment, the slope of the futures curve became especially informative because it reflected not only expected market conditions but also severe short-run dislocations in physical and financial market balance.

The pandemic period was also characterized by heightened uncertainty. Szczygielski et al. (2022) show that uncertainty played an important role in the energy sector during COVID-19, while Bakas and Triantafyllou (2020) show that pandemic-related uncertainty increased commodity price volatility more broadly. These findings suggest that the relationship between futures curve conditions and realized returns may have been different under pandemic market conditions than during more normal periods.

For the purposes of this thesis, the key implication is not that COVID-19 created one single uniform regime across all commodities. Rather, the pandemic provides a useful setting for examining whether the relationship between term structure and ETP returns changed across distinct market phases and whether such changes were similar across commodity groups. This is particularly relevant because the economic mechanisms discussed earlier in this section—inventory pressure, storage constraints, time-varying risk premia, and financial stress—were all likely to be unusually important during this period.

2.5 Research Gap and Hypotheses

2.5.1 Research Gap

The existing literature provides a strong foundation for understanding commodity futures pricing, the role of inventories, and the economic significance of futures curve structure. Working (1949) shows that the shape of the futures curve reflects storage-related market conditions, while Brennan (1958) formalizes the role of convenience yield in explaining backwardation when inventories are scarce. Later empirical studies further show that inventory conditions and futures curve shape are closely related to expected returns in commodity futures markets. In particular, Gorton et al. (2013) show that commodities with tighter inventory conditions and stronger backwardation tend to earn higher subsequent futures returns.

Previous research also demonstrates that returns on futures-based investment strategies differ from movements in the underlying spot commodity alone. Fama and French (1987) show that commodity futures returns reflect both expected spot-price movements and risk premia, while Erb and Harvey (2006) emphasize that the roll component is an economically important part of long-run commodity futures returns. These insights are directly relevant for futures-based commodity ETPs, because such products maintain exposure through rolling futures contracts rather than through physical ownership of the underlying commodity.

At the same time, a growing literature has examined the performance and tracking characteristics of commodity ETPs. Stewart et al. (2023) show that ETP tracking performance varies across agricultural and energy markets, while Cortazar et al. (2024) show that expected returns on commodity ETFs are closely linked to the properties of their underlying futures exposure. Together, these studies indicate that the relationship between futures curve conditions and ETP returns is economically plausible, but not mechanically one-to-one, because realized product performance also depends on benchmark construction, contract selection, roll methodology, and other product-level frictions.

Accordingly, the contribution of this thesis is not to study a completely unexplored topic. Rather, it is to examine this relationship in a more unified empirical setting. More specifically, this thesis studies whether a common market-level measure of futures term structure helps explain the returns of multiple non-leveraged single-commodity ETPs within one panel framework and whether that relationship differs across COVID-19 subperiods and commodity groups. In this sense, the study extends the existing literature by combining insights from commodity futures pricing and commodity ETP performance in one comparative framework.

The COVID-19 period provides a particularly useful setting for this analysis. Commodity markets experienced exceptional demand disruptions, inventory pressures, and uncertainty during the pandemic, especially in energy markets. These conditions make it possible to examine not only whether term structure matters for ETP returns on average, but also whether its importance changed across distinct market environments.

2.5.2 Hypotheses

The theoretical and empirical literature reviewed above suggests that the shape of the futures curve should matter for the returns of futures-based single-commodity ETPs. Because these products maintain exposure through rolling futures contracts, contango is expected to be associated with less favourable return outcomes, while backwardation is expected to be associated with more favourable return outcomes. This leads to the first hypothesis:

- H1: Contango is negatively associated with the returns of futures-based single-commodity ETPs

The literature also suggests that crisis conditions may alter this relationship. During the COVID-19 period, commodity markets were affected by heightened uncertainty, large demand disruptions, and sharp changes in futures curve conditions, particularly in energy markets. If these conditions strengthen the return implications of futures curve shape, the negative association between contango and ETP returns would be expected to become stronger during the pandemic. This leads to the second hypothesis:

- H2: The negative relationship between contango and futures-based single-commodity ETP returns is stronger during the COVID-19 period.

Together, these hypotheses translate the main theoretical arguments of this chapter into an empirical setting in which futures curve shape, ETP returns, and pandemic-period market conditions can be examined simultaneously.

3 Methodology

3.1 Research Design

To address the research question, this study follows a structured empirical approach. First, weekly data on commodity ETPs, underlying commodity prices, commodity futures settlement prices and the VIX index are collected and organized into a panel dataset. Second, the main variables used in the analysis are constructed, including ETP returns, a commodity-specific term-structure measure based on front-month and second-month futures prices, and relevant control variables. In addition, the analysis considers COVID-period specifications to assess whether the relationship differs across market environments and commodity groups. Finally, robustness checks are conducted to evaluate the stability of the findings.

This section is structured as follows: Section 3.2 describes the data and sample selection. Section 3.3 defines the variables used in the empirical analysis. Section 3.4 presents the empirical model, while Section 3.5 discusses the subsample analyses. Finally, Section 3.6 explains the estimation method and robustness considerations.

3.2 Data and sample selection

The empirical analysis is based on financial market data obtained from LSEG Workspace. The sample period runs from 1 January 2015 to 25 December 2025. The dataset combines information on single-commodity exchange-traded products, commodity futures settlement prices, underlying commodity spot prices, and the VIX index. All series are aligned into a weekly panel format to ensure consistent timing across variables used in the empirical analysis.

Weekly observations are defined using Thursday as the reference day throughout the sample. When a Thursday observation is not available because of market holiday or missing trading data, the observation is taken from the previously available trading day. The same observation rule is applied consistently across ETP prices, futures prices, underlying commodity prices, and the VIX index. This approach ensures that all weekly variables are measured at a common point in time and avoids mismatches caused by differences in trading calendars across markets. Compared with daily observations, weekly data helps to reduce short-term fluctuations and reduce sensitivity to high-frequency market noise. (Goodhart & O'Hara, 1997)

The empirical sample consists of 16 non-leveraged long-only single-commodity exchange-traded products, listed in Appendix 1 (Table 7). The products are grouped into three commodity sectors:

energy, metals, and agriculture. This classification is used in the subsample analysis, as the relationship between futures term structure and product returns may differ across commodity groups and market environments.

As mentioned in Chapter 2, the sample includes both exchange-traded funds (ETFs) and exchange-traded commodities (ETCs). These instruments are analysed jointly because, despite differences in legal structure, they serve the same economic purpose in the context of this study: they provide non-leveraged long exposure to a single underlying commodity. Leveraged and inverse products are excluded from the sample because their return dynamics are mechanically affected by leverage, compounding, and asymmetric exposure, which would make them less comparable to the non-leveraged long-only products.

3.3 Variable construction

In this section, we define the variables used in the empirical analysis. The dependent variable is the weekly return of each product, and the main explanatory variable is a commodity-specific term-structure measure. Control variables are the underlying spot price of the commodity and broader market uncertainty. In addition, the analysis includes COVID-period variables to examine whether the relationship between term structure and product returns changed during the pandemic.

3.3.1 Dependent variable: product returns

The dependent variable is the weekly return of product *i* in week *t*. Returns are calculated from the total return index series of each product using consecutive weekly observations. In the empirical analysis, product returns are measured as weekly logarithmic returns.

3.3.2 Main explanatory variable: term structure

The main explanatory variable is a commodity-specific term-structure measure that is constructed from the settlement prices of the underlying commodity's front-month and second-month futures contracts. The measure is defined as:

$$TS_{j,t} = \frac{F2_{j,t} - F1_{j,t}}{F1_{j,t}}$$

where $F1_{j,t}$ represents the settlement price of the front-month futures contract and $F2_{j,t}$ the settlement price of the second-month futures contract for commodity *j* in week *t*. Positive values of the measure indicate contango and negative values indicate backwardation. The choice is consistent with the commodity futures literature, which treats the shape of the futures curve as informative about storage conditions, convenience yield and expected futures returns. (Fama & French, 1987)

The futures prices are obtained from LSEG Workspace using generic front-month and second-month futures series based on settlement prices. Contract identification and contract switching are not imposed in this study. Instead, the term-structure measure relies on LSEG's generic continuous futures series. According to LSEG documentation, rolling continuous RICs roll automatically, and Refinitiv's continuation rolls on the last trading day (Aramyan, 2023). Accordingly, the front-month and second-month contracts used in this study follow the provider's continuation rule rather than a researcher-imposed roll schedule.

This has two implications for interpretation. First, roll dates may introduce mechanical changes into the generic futures series when the underlying contract switches from one maturity to the next. The use of weekly observations helps reduce the visibility of such effects, although it does not eliminate them completely. Second, futures contracts are not perfectly identical across commodities in terms of listing conventions and available contract months. As a result, cross-commodity comparability is practical but not exact.

The same term-structure series is assigned to all ETPs tracking the same underlying commodity. This is appropriate because the purpose of the variable is to capture the market-level slope of the underlying commodity futures curve rather than product-specific implementation details. Importantly, the variable does not measure the exact roll yield of each product. Differences in benchmark construction, contract selection, and roll methodology may cause the realized roll component of an individual ETP to differ from the simple front-to-second-month slope used here. Therefore, the term-structure measure should be interpreted as a proxy for the underlying commodity's futures curve rather than an exact product-level roll measure.

3.3.3 Control variables

The first control variable is the weekly spot return of the underlying commodity. This variable is included to separate the effect of changes in the futures curve from the effect of movements in the underlying commodity price itself. Each product is matched with the spot return series of the commodity it uses. While the spot series may not always match the exact benchmark followed by each product, it provides a close proxy for the underlying commodity price movement and captures the same economic information relevant for the analysis. This variable is important because spot prices and futures prices are closely related, but the futures curve shape may provide additional information beyond spot price movements (Fama & French, 1987). In the empirical analysis, underlying commodity returns are measured as weekly logarithmic returns.

The second control variable is the VIX index, which is included as a proxy for general market uncertainty. In the empirical analysis, the VIX is measured in logarithmic form. Weekly observations are again taken using Thursday as the reference day, with the previously available trading day used when necessary. Using the logarithm of the VIX helps reduce skewness in the series and allows the coefficient to be interpreted as the association between changes in general market uncertainty and commodity ETP returns in a more stable functional form. Using the VIX as a proxy for general market uncertainty is consistent with Bekaert & Hoerova (2014), who show that the VIX reflects both expected market volatility and a variance premium component. However, while VIX captures an important dimension of aggregate market uncertainty, it does not fully account for all common shocks affecting commodity ETP returns simultaneously. For this reason, it should be interpreted as a control for general market uncertainty rather than as a substitute for time fixed effects.

3.3.4 COVID-period variables

To assess whether the relationship between term structure and product returns changed during the pandemic, the empirical analysis distinguishes between three COVID-19 subperiods:

- Initial shock: 1.3.2020 – 30.6.2020
- Post-shock: 1.7.2020 – 31.1.2021
- Recovery: 1.2.2021 – 31.12.2021

These periods are defined as analytical windows intended to separate the acute disruption phase of the pandemic from the subsequent adjustment and early recovery phases. The initial shock period begins in March 2020, when the World Health Organization characterized the COVID-19 as a pandemic, and covers the months of the most severe initial market disruption. This period is particularly relevant for commodity markets because the pandemic and associated containment measures triggered a collapse in demand, especially in energy markets, while commodity prices fell sharply during the first half of 2020 (Cucinotta & Vanelli, 2020; World Bank Group, 2020a, 2020b).

The post-shock period, from July 2020 to January 2021, is intended to capture the transition from the initial collapse to a more stable but still fragile market environment. By the second half of 2020, some commodity prices had begun to recover from their earlier lows, yet the global outlook remained highly uncertain, and the broader recovery was still weak and uneven. This makes the post-shock period conceptually distinct from both the acute disruption phase and the more visible recovery dynamics that emerged later in 2021 (World Bank Group, 2020b, 2021).

The recovery period is defined as beginning in February 2021. This choice does not imply that recovery started everywhere at exactly the same time. Rather, it marks the beginning of an early recovery phase in which vaccination campaigns, reopening expectations, and improving macroeconomic and commodity-demand prospects became more clearly visible. In early 2021, major international institutions increasingly linked the global recovery outlook to vaccine rollout, and commodity and energy market reports described 2021 as a year of rebounding demand supported by improving growth expectations, policy stimulus, and the acceleration of vaccinations in major economies (IEA, 2021; International Monetary Fund, 2021; World Bank Group, 2021).

3.4 Empirical Model

This section presents the regression framework used to examine whether the term structure of commodity futures helps explain the returns of futures-based single-commodity ETPs and whether this relationship changed during the COVID-19 period. The empirical analysis is conducted using a fixed-effects panel model estimated at the product-week level. This framework is appropriate because the main interest lies in within-product variation over time rather than in permanent differences across products. Fixed effects are used to absorb time-invariant product-specific heterogeneity that could otherwise bias coefficient estimates. (Arellano, 1987; Stock & Watson, 2008)

3.4.1 Main regression

The baseline regression model is specified as follows:

$$R_{i,t} = \alpha + \beta_1 TS_{j,t} + \beta_2 Shock_t + \beta_3 PostShock_t + \beta_4 Recovery_t + \beta_5 (TS_{j,t} \times Shock_t) + \beta_6 (TS_{j,t} \times PostShock_t) + \beta_7 (TS_{j,t} \times Recovery_t) + \beta_8 (SpotR_{i,t}) + \beta_9 \log(VIX_t) + \mu_i + \varepsilon_{i,t},$$

where $R_{i,t}$ represents the weekly logarithmic return of product *i* in week *t*; $TS_{j,t}$ is the term-structure measure of the underlying commodity *j* in week *t*; $Shock_t$, $PostShock_t$ and $Recovery_t$ are indicator variables for the three COVID-19 subperiods; $(SpotR_{i,t})$ denotes the weekly logarithmic return of underlying commodity's price; $\log(VIX_t)$ is the logarithm of the weekly VIX observation; $\mu_i$captures product fixed effects; and $\varepsilon_{i,t}$ is the error term.

The model includes the three previously introduced COVID-windows using period-specific dummy variables and interaction terms with the term-structure measure. The reference category consists of all observations outside these three windows and is interpreted as normal market conditions. Therefore, the reference period includes both the pre-pandemic period and the later post-recovery period from 2022 onward.

3.4.2 Interpretation of coefficients

The coefficient of primary interest is β_1, which captures the association between term structure and product returns during normal market conditions. A negative estimate implies that a steeper contango is associated with lower ETP returns under normal market conditions, whereas a positive estimate would indicate the opposite.

The coefficients on the interaction terms show whether this relationship changes during the COVID-19 subperiods. Because the model includes interactions, the effect of term structure in each market period must be interpreted as a total effect rather than through the interaction coefficient alone. Specifically, the effect of term structure is given by β_1 in the reference period, $\beta_1 + \beta_5$ during the initial shock period, $\beta_1 + \beta_6$ during the post-shock period, and $\beta_1 + \beta_7$ during the recovery period. These linear combinations are reported separately in the results section to make the economic interpretation of the estimates more transparent.

The coefficients β_2, β_3 and β_4 capture changes in average return during those stress periods relative to the reference-period benchmark. This structure makes it possible for us to distinguish general pandemic-related changes in returns and the specific effect of futures curve shape on product performance. (Borgards et al., 2021)

3.4.3 Model rationale

The product fixed-effects specification is used to account for time-invariant heterogeneity across ETPs. This is important because the products in the sample may differ in benchmark design, replication structure, fee levels, and other institutional characteristics that could affect their return dynamics. To the extent that such differences are constant over time, product fixed effects help account for these differences and thereby reduce omitted-variable bias in the estimated relationship between term structure and returns.

The baseline specification does not include time fixed effects. The purpose of the main model is to estimate how variation in the underlying commodity's term structure is associated with variation in product returns after controlling for product-specific fixed effects, spot-price movements, and general market uncertainty. However, common shocks affecting all products simultaneously remain an important econometric concern in weekly commodity panel data. For this reason, the role of common time shocks is addressed later through robustness analysis rather than treated as resolved by the baseline specification alone.

The interaction structure is included because the COVID-19 crisis is unlikely to represent a single uniform regime. Instead, the pandemic period involved an acute disruption phase, a subsequent adjustment period, and an early recovery phase. Allowing the slope coefficient on term structure to vary across these subperiods makes it possible to test not only whether term structure matters for ETP returns on average, but also whether its importance changed across distinct market environments.

3.5 Subsample analyses

In addition to the full-sample estimation, the baseline model is re-estimated separately for three commodity groups: energy, metals and agriculture. The purpose of these subsample analyses is to examine whether the association between futures term structure and ETP returns differs across commodity sectors and whether the full-sample results reflect a broad-based pattern or are driven mainly by one segment of the commodity market.

The subsample analysis is well motivated by existing literature on commodities. Prior research shows that determining factors for returns vary across commodities and over time (G. B. Gorton et al., 2013). An additional reason for examining commodities by group is the COVID-19 stress period. Borgards et al. (2021) document substantial disruptions in commodity futures markets, while the intensity of these effects varied across different groups.

Methodologically, the subsample analyses use the same regression specification introduced in Section 3.4. Applying the same model separately to each commodity group improves the comparability of coefficient estimates across sectors and allows the full-sample findings to be interpreted more carefully. In this sense, the subsample analysis is used to assess whether the estimated relationship between term structure and returns is relatively similar across commodity groups or whether it appears to be concentrated in specific segments of the sample.

The subsample results are interpreted as complementary to the full-sample analysis rather than as equally precise evidence. This is because the number of products in each sector is substantially smaller than in the full sample, which weakens statistical precision and makes cluster-robust inference less stable. For this reason, the full-sample specification remains the main empirical model, while the commodity-group estimations are used primarily to identify heterogeneity in the relationship across sectors and to support the interpretation of the baseline findings.

3.6 Estimation method and robustness considerations

The regression models are estimated using ordinary least squares (OLS) within a fixed-effects framework. Because the data form a weekly product-level panel, statistical inference requires particular care. In this setting, regression residuals may exhibit heteroskedasticity, serial correlation within products, and exposure to common shocks affecting multiple products simultaneously. These concerns are especially relevant in commodity markets during periods of elevated market stress.

The baseline specification is estimated with product fixed effects and standard errors clustered at the product level. Clustering at the product level allows the error terms to be correlated within products over time and therefore provides a more appropriate inference framework than conventional homoscedastic standard errors. This is used as the main inference method throughout the analysis. At the same time, the number of clusters in the full sample remains modest, as the sample contains 16 products. For this reason, inference based on product-clustered standard errors should be interpreted with some caution, and this concern becomes even more important in the commodity-group subsamples.

To assess whether heteroskedasticity is present in the residuals, the analysis applies the Breusch-Pagan test. To examine whether the residuals exhibit within-product serial correlation, the analysis applies Wooldridge’s test for serial correlation in fixed-effects panel models. The diagnostic tests are used to evaluate whether the error structure of the weekly panel data is consistent with the use of robust inference methods. Their role is diagnostic rather than determinative, meaning that the estimation strategy is not based mechanically on a single test result.

To evaluate the sensitivity of the main findings to the treatment of the error structure, the analysis also considers alternative covariance estimators. Specifically, Newey-West standard errors are used as a robustness check against heteroskedasticity and serial correlation, while Driscoll-Kraay standard errors are used as an additional robustness check that allows for cross-sectional dependence in the residuals. The purpose of these alternative covariance estimators is to assess whether the main conclusions remain stable under different assumptions about the dependence structure of the errors.

A separate econometric concern arises from common time shocks. In weekly commodity panel data, broad market developments such as macroeconomic news, changes in global demand conditions, or commodity-wide stress episodes may affect several products at the same time. Although the baseline model controls for general market uncertainty through the VIX index, this does not fully eliminate the possibility of omitted common time effects. For this reason, an additional robustness check is conducted by re-estimating the baseline model with week fixed effects. In this specification, time-specific variables that vary only across weeks, such as the COVID-period dummies and the VIX measure, are absorbed by the week fixed effects. The week fixed-effects model therefore provides a stricter test of whether the estimated association between term structure and product returns is sensitive to shocks common to all products in each week.

Overall, the estimation strategy is designed to balance interpretability and econometric caution. The baseline fixed-effects model provides the main framework for analysis, while diagnostic tests, alternative covariance estimators, and week fixed effects are used to evaluate the robustness of the findings and to clarify the main limitations of the empirical design.

4 Results

This section presents the empirical results for the weekly panel of 16 exchange-traded commodity products over the period from 8 January 2015 to 25 December 2025. The sample is constructed from 9,168 observations including products from agriculture, energy and metals. Of total observations, 2 292 belong to the agriculture group, 4 011 to the energy group and 2 865 to the metals group.

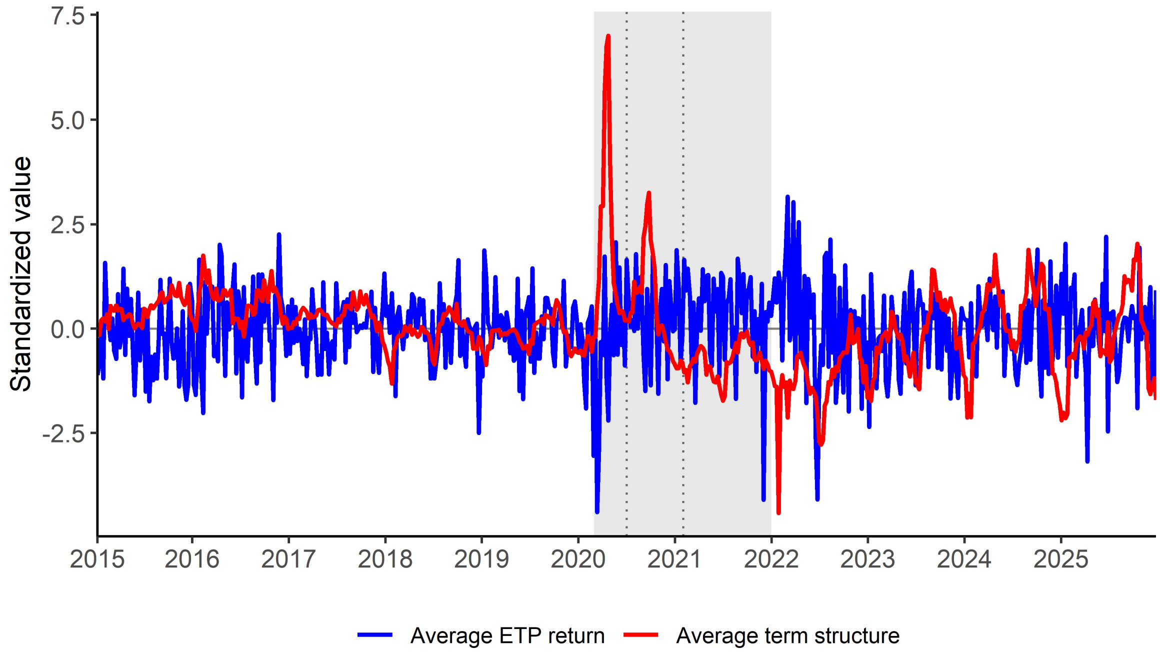

Figure 1: Average standardized weekly ETP returns and term structure, 2015–2025

The figure shows the weekly cross-sectional averages of ETP returns and the term-structure measure for the 16 single-commodity ETPs in the sample. Both series are standardized to mean zero and unit variance. The shaded area marks the COVID-19 period, and the dotted vertical lines indicate the boundaries between the initial shock, post-shock, and recovery subperiods.

Figure 1 provides descriptive evidence on the evolution of average product returns and the average term structure over time. The figure shows substantial variation over time in both variables. The variation is particularly pronounced during the COVID-19 stress period, which is shown in the figure as the shaded area. More generally, the figure suggests that the relationship between product returns and the term structure varies over time, which further motivates the regression-based analysis reported below.

4.1 Baseline regression results

4.1.1 Statistical interpretation

Table 1 reports the baseline fixed-effects regression results for the full sample. The coefficient on the term-structure measure is negative and statistically significant at the one-tenth percent level. This indicates that, during the baseline period, a steeper contango is associated with lower ETP returns. The result is consistent with existing literature and the view that futures curve shape is relevant for explaining the returns of futures-based commodity ETPs and that contango is associated with lower return outcomes than backwardation.

Table 1: Baseline fixed-effects regression results for the full sample

This table reports fixed-effects panel regression results for the full sample of 9,168 weekly observations on 16 non-leveraged single-commodity ETPs over the period from 8 January 2015 to 25 December 2025. The dependent variable is the weekly logarithmic return of product *i* in week *t*. Product fixed effects are included, and standard errors are clustered at the product level. The term-structure variable is constructed from the front-month and second-month futures settlement prices of the underlying commodity. The control variables are the weekly logarithmic return of the underlying commodity and the logarithm of the VIX index. Significance markers: * p<0.05, **p<0.01, ***p<0.001. Standard errors (SEs) are in parenthesis.

Variable	
Term Structure	-0.0849***
	(0.0167)
Shock	0.0024
	(0.0036)
Post Shock	0.0066***
	(0.0013)
Recovery	0.0067***
	(0.0009)
Spot Return	0.4016***
	(0.0748)
VIX	-0.0081**
	(0.0020)
Term Structure: Shock	-0.0253
	(0.1503)
Term Structure: Post Shock	0.0731***
	(0.0147)
Term Structure: Recovery	0.2276**
	(0.0763)

Further examination of Table 1 shows that the estimated coefficients on the COVID-period dummies suggest that average returns were higher during the post-shock and recovery periods compared to the baseline period. However, the coefficient on the initial shock dummy is not statistically significant. These coefficients should not, however, be interpreted as evidence about the role of term structure by themselves, because the model also includes interaction terms between term structure and the COVID-period indicators.

The interaction estimates suggest that the baseline negative relationship between term structure and returns did not remain constant across pandemic subperiods. The interaction for the initial shock period is statistically insignificant, which indicates no clear evidence that the baseline slope changed during the earliest phase of the pandemic. The interaction for the post-shock period is positive and statistically significant in the baseline clustered specification, implying that the negative baseline relationship became weaker after the initial shock. The interaction for the recovery period is also positive and statistically significant, indicating that the slope became more positive relative to the baseline period. However, as shown more clearly by the total-effects reported later, the recovery-period interpretation should remain cautious in the full sample.

The control variables behave as expected. The coefficient on the underlying commodity return is positive and highly statistically significant, indicating that ETP returns move strongly with changes in the underlying commodity price. The coefficient on the VIX is negative and statistically significant, suggesting that higher general market uncertainty is associated with lower commodity ETP returns. These findings are consistent with the idea that both commodity-specific price movements and broader financial market conditions are relevant for explaining ETP performance.

Taken together, the coefficient estimates indicate that futures term structure helps explain the returns of futures-based commodity ETPs under normal market conditions. At the same time, the interaction terms suggest that this relationship did not remain constant across pandemic subperiods. Because the model includes interactions, the economic meaning of these coefficients is best evaluated through the implied total effects, which are discussed in the following subsection.

4.1.2 Economic interpretation

Because the specification includes interaction terms, the economic meaning of the estimates is more transparent when the implied total effects of term structure are considered. The results for the total effects are presented in Table 2. In the reference period, the estimated total effect of term structure is -0.0849. This implies that a 0.1 increase in the term-structure measure is associated with approximately a 0.0085 decrease in weekly log returns, or about a 0.85 percentage point lower weekly return. This is economically meaningful at the weekly frequency and indicates that changes in the slope of the futures curve are not only statistically significant but also relevant in magnitude.

Table 2: Total effects of term structure across market periods

This table reports the implied total effects of the term-structure measure across the reference, shock, post-shock, and recovery periods. Total effects are calculated as linear combinations of the baseline term-structure coefficient and the relevant interaction coefficient from the full-sample fixed-effects model. Standard errors are based on the same product-clustered covariance matrix as in the baseline regression. The reference period includes observations outside the three COVID-19 subperiods. Significance markers: * p<0.05, **p<0.01, ***p<0.001. Standard errors (SEs) are in parenthesis.

Baseline	Shock	Post Shock	Recovery
-0.0849***	-0.1102	-0.0118	0.1427
(0.0167)	(0.1509)	(0.0127)	(0.0740)

During the initial shock period, the estimated total effect remains negative at -0.1102, but the estimate is statistically insignificant. This means that, although the point estimate is somewhat more negative than in the reference period, the data do not provide reliable evidence that the effect of contango on ETP returns became stronger during the earliest phase of the pandemic. During the post-shock period, the estimated total effect is close to zero at -0.0118 and statistically insignificant. Accordingly, the results do not provide clear evidence of a negative term-structure effect during that phase.

The recovery-period estimate is economically notable but statistically weaker than the baseline effect. The estimated total effect is 0.1427, which implies that a 0.1 increase in the term-structure measure is associated with approximately a 0.0143 increase in weekly log returns, or about a 1.43 percentage point higher weekly return. However, this estimate is not statistically significant. For this reason, it is best interpreted as suggesting a possible reversal in the relationship rather than establishing one conclusively. This distinction is important considering the robustness analysis, which shows that the negative baseline relationship is stable across alternative inference methods, whereas the post-shock effect is more sensitive to specification choice.

Overall, the full-sample results suggest that futures curve shape is economically relevant for explaining commodity ETP returns. The clearest and most robust result is the negative baseline association between contango and returns. By contrast, the COVID-period evidence is more nuanced. The full-sample estimates do not support a broad claim that the negative contango effect became stronger during the pandemic. Instead, they suggest that the relationship changed across pandemic subperiods, weakening after the initial shock and becoming more positive during the recovery period. However, the latter finding should be interpreted cautiously and alongside the commodity-group and robustness results discussed later.

4.2 Commodity group results

4.2.1 Statistical interpretation

To examine whether the full-sample results are broad-based across commodity sectors or driven mainly by a specific segment, the baseline model is re-estimated separately for the energy, metals, and agriculture subsamples. This allows a more careful assessment of whether the relationship between futures term structure and ETP returns differs across commodity groups. The results are reported in Table 3.

Table 3: Fixed-effects regression results by commodity group

This table reports fixed-effects panel regression results separately for the energy, metals, and agriculture subsamples. The dependent variable is the weekly logarithmic return of product *i* in week *t*. Product fixed effects are included in all specifications, and standard errors are clustered at the product level. The energy subsample contains 4,011 observations across 7 products, the metals subsample 2,865 observations across 5 products, and the agriculture subsample 2,292 observations across 4 products. Significance markers: * p<0.05, **p<0.01, ***p<0.001. Standard errors (SEs) are in parenthesis.

Variable	Energy	Metals	Agriculture
Term Structure	-0.0926*	-0.1061	-0.0498
	(0.025)	(0.064)	(0.0243)
Shock	0.0040	-0.0009	-0.0002
	(0.0096)	(0.0018)	(0.0020)
Post Shock	0.0092*	-0.0001	0.0018
	(0.0027)	(0.0014)	(0.0008)
Recovery	0.0099**	0.0019*	0.0048*
	(0.0019)	(0.0006)	(0.0010)
Spot Return	0.3140**	0.8332***	0.7198***
	(0.0687)	(0.074)	(0.039)
VIX	-0.0133*	-0.0018	-0.0024*
	(0.0041)	(0.0015)	(0.0006)
Term Structure: Shock	-0.0347	0.2909	0.0453
	(0.1866)	(0.2849)	(0.1975)
Term Structure: Post Shock	0.0825**	-0.0141	0.0381
	(0.0179)	(0.1512)	(0.0167)
Term Structure: Recovery	0.5304*	0.2278	0.0852
	(0.1514)	(0.1183)	(0.0389)

The strongest evidence is found in the energy subsample. In the reference period, the coefficient on term structure is negative and statistically significant, indicating that steeper contango is associated with lower ETP returns in energy markets. In addition, the interaction terms for the post-shock and recovery periods are positive and statistically significant. This suggests that, relative to the baseline period, the slope of the relationship became less negative after the initial shock and more positive during the recovery period. The interaction for the initial shock period is not statistically significant, which implies that there is no clear evidence of a distinct slope change during the earliest phase of the pandemic.

By contrast, the evidence is much weaker in the metals and agriculture subsamples. In both cases, the coefficient on term structure remains negative, but is not statistically significant. Likewise, none of the interaction terms provide strong statistical evidence that the term-structure effect changed systematically across the COVID-19 subperiods. These results do not support the view that the time-varying pattern observed in the full sample applies uniformly across the commodity groups. Instead, they suggest that the full-sample time variation is driven primarily by the energy segment.

The control variables are broadly consistent with the full-sample results, although their precision differs across groups. The coefficient on the underlying commodity return is positive and statistically significant in all three subsamples, indicating that changes in the underlying commodity price remain an important driver of ETP returns across sectors. The coefficient on the VIX is negative in all three groups, but the statistical strength of this result is less stable, especially in the smaller subsamples.

These subsample results should be interpreted with caution. The energy subsample contains 7 products, the metals subsample 5 products, and the agriculture subsample only 4 products. With such a small number of clusters, cluster-robust inference becomes less stable, and coefficient estimates are more sensitive to individual products. For this reason, the energy results provide suggestive and economically interpretable evidence of time variation, whereas the metals and agriculture results are better interpreted as showing limited and imprecise evidence rather than evidence of no effect.

Taken together, the subsample regressions refine the full-sample interpretation. They indicate that the baseline negative relationship between contango and ETP returns is most clearly identifiable in the energy segment, while the stronger pandemic-period time variation also appears mainly in energy markets. The results for metals and agriculture do not provide comparable statistical support for the same pattern. Accordingly, the commodity-group analysis does not support a broad commodity-wide amplification story during COVID-19 but rather points to a more sector-specific pattern concentrated in energy markets.

4.2.2 Economic interpretation

The economic interpretation of the commodity-group results is most clearly seen through the total effects of term structure across market periods. The total effects are reported in Table 4. These estimates help assess whether the negative baseline relationship between contango and ETP returns is economically similar across groups and whether that relationship changed during the COVID-19 subperiods.

Table 4: Total effects of term structure by commodity group

This table reports the implied total effects of the term-structure measure across market periods for the energy, metals, and agriculture subsamples. Total effects are calculated as linear combinations of the baseline term-structure coefficient and the relevant interaction coefficient from each commodity-group regression. Standard errors are based on product-clustered covariance matrices within each subsample. Significance markers: * p<0.05, **p<0.01, ***p<0.001. Standard errors (SEs) are in parenthesis.

	Energy	Metals	Agriculture
Baseline	-0.0926*	-0.1061	-0.0498
	(0.025)	(0.0642)	(0.0243)
Shock	-0.1273	0.1847	-0.0045
	(0.1851)	(0.3118)	(0.2170)
Post Shock	-0.0100	-0.1203	-0.0116
	(0.0181)	(0.1261)	(0.0089)
Recovery	0.4378*	0.1217	0.0355
	(0.1610)	(0.1372)	(0.024)

The clearest economic pattern is found in the energy subsample. In the reference period, the total effect of term structure is negative and statistically significant at -0.0926. This implies that a 0.10 increase in the term-structure measure is associated with approximately a 0.0093 decrease in weekly log returns, or about a 0.93 percentage point lower weekly return. This is close in magnitude to the full-sample baseline estimate and indicates that the negative contango effect is economically meaningful in energy markets. During the shock and post-shock periods, the estimated total effects remain statistically insignificant. By contrast, the recovery-period total effect becomes positive and statistically significant at 0.4378, implying that a 0.10 increase in the term-structure measure is associated with approximately a 0.0438 increase in weekly log returns, or about a 4.38 percentage point higher weekly return. This is a large effect in economic terms and provides the strongest evidence in the thesis that the return implications of term structure changed materially during the pandemic.

For metals and agriculture, the economic interpretation is more limited. In both groups, the baseline total effect is negative, which is directionally consistent with the full-sample results, but the estimates are not statistically significant in the clustered subsample inference. Additionally, the total effects for the shock, post-shock, and recovery periods are statistically insignificant in both groups. As a result, the data do not provide reliable evidence that the overall effect of term structure changed systematically across the COVID-19 subperiods in these two sectors. In economic terms, this means that the results are too imprecise to support strong conclusions about time variation outside the energy segment.

Taken together, the total-effects estimates indicate that the economically meaningful pandemic-related time variation in the term-structure effect is concentrated mainly in energy markets. This is important for interpreting the full-sample results. The aggregate pattern does not appear to reflect a broad commodity-wide shift across all sectors. Rather, it appears to be driven primarily by the energy subsample, where the recovery-period estimate is both economically large and statistically more convincing than in the overall sample.

These results also help clarify the interpretation of the second research question. The commodity-group evidence does not support a general conclusion that the negative contango effect became stronger during COVID-19. Instead, the results suggest a more nuanced pattern: the baseline negative relationship is most clearly visible in energy markets, and the most notable pandemic-related change is a later recovery-period reversal that is concentrated in that sector. Given the small number of products within each commodity group, especially in metals and agriculture, these subsample estimates should nevertheless be interpreted as supportive rather than definitive evidence.

4.3 Robustness checks

4.3.1 Diagnostic tests and alternative covariance estimators

Because the empirical analysis is based on weekly panel data, the reliability of statistical inference depends on how the residual dependence structure is handled. As discussed in Section 3.6, heteroskedasticity and serial correlation are relevant concerns in this setting. To assess whether these issues are present in the data, the analysis applies a Breusch-Pagan test for heteroskedasticity and Wooldridge's test for serial correlation in fixed-effects panels. Diagnostic tests indicate that the error terms do not satisfy the classical homoskedasticity and no-serial-correlation assumptions. The studentized Breusch–Pagan test rejects the null of homoskedasticity ($BP = 606.43$, $df = 9$, $p < 0.001$), and Wooldridge's test for serial correlation in fixed-effects panels rejects the null of no first-order serial correlation ($F(1,9150) = 10.54$, $p = 0.001$). Therefore, the baseline specifications are reported with standard errors clustered at the product level, and robustness is further assessed using alternative covariance estimators.

To examine whether the full-sample results are sensitive to the treatment of the error structure, the baseline fixed-effects model is re-estimated using three alternative covariance estimators: product-clustered standard errors, Newey-West standard errors, and Driscoll-Kraay standard errors. The results for the alternative covariance estimators are presented in Appendix 2 (Table 8). Across all three specifications, the main baseline findings remain stable. Most importantly, the coefficient on the term structure measure remains negative and statistically significant in each case. This indicates that the negative baseline relationship between contango and ETP returns is not driven by the choice of covariance estimator.

The broader COVID-period pattern is also qualitatively similar across the alternative inference methods, although some differences in statistical precision emerge. The interaction between term structure and the recovery period remains positive and statistically significant across all three specifications. By contrast, the interaction between term structure and the post-shock period is statistically significant only under product-clustered standard errors and becomes statistically insignificant when Newey-West and Driscoll-Kraay standard errors are applied. This implies that the evidence for a distinct post-shock slope change is more sensitive to the treatment of serial correlation than the baseline negative association and the recovery-period interaction and should therefore be interpreted with more caution.

The control variables remain broadly consistent across the alternative covariance estimators. The coefficient on the underlying commodity return remains positive and highly significant throughout, while the coefficient on the VIX remains negative, although its level of statistical significance varies across specifications. Taken together, the alternative covariance estimators strengthen the main interpretation of the full-sample analysis: the negative baseline association between term structure and product returns is robust, while the evidence for a distinct post-shock slope change should be interpreted more cautiously.

Importantly, however, alternative covariance estimators do not by themselves resolve the omitted common time shock issue created by the absence of time fixed effects. They improve inference under different assumptions about the error structure, but they do not absorb shocks that affect all products simultaneously in each week. For this reason, a separate robustness check using week fixed effects is required.

4.3.2 Week fixed effects

A separate robustness concern relates to omitted common time shocks. In weekly commodity panel data, macroeconomic news, global demand developments, and market-wide stress events may affect several products simultaneously. This issue is particularly relevant in commodity markets, where broad market conditions may shift the returns of many products at the same time. Although the baseline model controls for general market uncertainty through the VIX index, this does not fully eliminate the possibility of omitted common time effects.

To address this concern, the baseline model is re-estimated with week fixed effects. The results are reported in Table 5. The purpose of this specification is not to replace the baseline model, but to provide a stricter robustness test of whether the estimated relationship between term structure and product returns survives after controlling for shocks common to all products in each week.

Table 5: Week fixed-effects robustness results

This table reports robustness results from a fixed-effects specification that includes both product and week fixed effects. The dependent variable is the weekly logarithmic return of product *i* in week *t*. Standard errors are clustered at the product level. Variables that vary only over time, including the COVID-period dummies and the logarithm of the VIX index, are absorbed by the week fixed effects and therefore are not separately identified in this specification. The table is used to assess whether the estimated relationship between term structure and ETP returns is robust to shocks common to all products in a given week. Significance markers: * p<0.05, **p<0.01, ***p<0.001. Standard errors (SEs) are in parenthesis.

Term Structure	Spot Return	Term Structure: Shock	Term Structure: Post Shock	Term Structure: Recovery
-0.0720*** (0.012)	0.3408*** (0.067)	-0.1309 (0.1576)	0.051** (0.014)	0.2016** (0.062)

The week fixed-effects results support the main full-sample interpretation. The baseline coefficient on term structure remains negative and statistically significant. In addition, the interaction terms for the post-shock and recovery periods remain positive and statistically significant, while the interaction for the initial shock period remains insignificant. This suggests that the main full-sample findings are not driven solely by omitted common week-level shocks. At the same time, the week fixed-effects specification should be interpreted as a robustness test rather than as the preferred baseline model, because it absorbs a substantial amount of common time variation and prevents separate estimation of purely time-varying controls.

Overall, the week fixed-effects results strengthen the credibility of the baseline findings. They show that the negative association between contango and ETP returns remains visible even after controlling for common week-level shocks, and that the more positive slope estimates in the later COVID subperiods are not purely an artifact of omitted time effects.

4.3.3 Excluding energy products

As an additional robustness check, we re-estimated the baseline model after excluding all energy products from the sample. The results are presented in Table 6. This test is particularly relevant because the commodity-group analysis suggested that the stronger time variation in the full sample may be driven mainly by the energy segment. The purpose of the non-energy specification is therefore to assess whether the baseline negative relationship between term structure and product returns remains present outside energy markets.

Table 6: Fixed-effects regression results for the non-energy sample

This table reports fixed-effects panel regression results for the non-energy sample, which excludes all energy products. The dependent variable is the weekly logarithmic return of product *i* in week *t*. Product fixed effects are included, and standard errors are clustered at the product level. The purpose of this specification is to assess whether the baseline relationship between term structure and ETP returns remains present outside the energy sector. Significance markers: * p<0.05, **p<0.01, ***p<0.001. Standard errors (SEs) are in parenthesis.

Variable	
Term Structure	-0.0518*
	(0.021)
Shock	0.0001
	(0.0013)
Post Shock	0.0008
	(0.0009)
Recovery	0.0033**
	(0.0007)
Spot Return	0.7801***
	(0.0413)
VIX	-0.0024**
	(0.0007)
Term Structure: Shock	0.07289
	(0.1784)
Term Structure: Post Shock	0.0420*
	(0.0160)
Term Structure: Recovery	0.0766
	(0.0359)

The non-energy results show that the coefficient on term structure remains negative and statistically significant in the clustered baseline specification. This indicates that the negative relationship between the term-structure measure and product returns is not driven solely by energy products.

The control variables also remain economically meaningful: the coefficient on the underlying commodity return remains strongly positive, and the coefficient on the VIX remains negative. The corresponding results using alternative covariance estimators are reported in Appendix 2 (Table 9).

At the same time, the time-varying COVID-period effects become clearly weaker once the energy segment is removed. The post-shock and recovery interaction terms remain positive, but their statistical strength is more limited and more sensitive to the choice of covariance estimator than in the full sample. More importantly, the total-effects estimates for the non-energy sample do not provide evidence that the overall effect of term structure differs systematically across COVID subperiods. The baseline total effect remains negative and statistically significant, whereas the total effects for the shock, post-shock, and recovery periods are all statistically insignificant. The total-effect results are presented in Appendix 2 (Table 10).

This contrast with the full-sample and energy results sharpens the substantive conclusions of the thesis. The negative baseline association between contango and futures-based ETP returns appears to extend beyond the energy sector. However, the stronger pandemic-related time variation is much less evident once energy products are excluded. In this sense, the robustness analysis supports a more nuanced conclusion: the baseline term-structure effect is more broad-based, but the time-varying COVID-pattern is driven mainly by energy products.

Because the non-energy sample contains only nine products, these results should still be interpreted with caution. The clustered covariance matrix required a numerical correction, which is consistent with the broader difficulty of conducting cluster-robust inference when the number of clusters is small. For this reason, the non-energy specification is best viewed as supportive evidence rather than a standalone basis for strong inference.

5 Conclusions

This thesis examined whether the term structure of commodity futures helps explain the returns of futures-based single-commodity exchange-traded products and whether this relationship changed during the COVID-19 period. Using weekly panel data for 16 non-leveraged single-commodity ETPs over 2015–2025, the analysis showed that futures curve conditions are relevant for understanding product returns. In the full-sample baseline specification, a steeper contango was associated with lower ETP returns, even after controlling for changes in the underlying commodity price and general market uncertainty. This indicates that the shape of the futures curve contains economically meaningful information beyond contemporaneous commodity price movements alone.

The results provide support for the main hypothesis. In the baseline full-sample setting, the relationship between contango and ETP returns is negative and statistically significant, which is consistent with the view that adverse roll conditions are reflected in the performance of futures-based commodity products. The magnitude of the estimate also suggests that the effect is not only statistically detectable but economically relevant at the weekly frequency. In this sense, the first research question receives a clear positive answer: futures term structure helps explain the returns of futures-based single-commodity ETPs.

The second hypothesis is not supported in its general form. The results do not show that the negative contango effect became uniformly stronger during the COVID-19 period. During the initial shock phase, the estimated total effect remained negative, but there is no statistically reliable evidence that the relationship became more strongly negative than in the baseline period. During the post-shock phase, the negative baseline association weakened substantially, and during the recovery period the full-sample estimate became positive. However, the recovery-period result is only weakly significant in the full sample and should therefore be interpreted cautiously. The evidence is more consistent with a changing relationship across pandemic subperiods than with a simple amplification of the negative contango effect during COVID-19.

The commodity-group analysis helps clarify this pattern. The strongest and clearest time variation is found in the energy subsample. In energy markets, the baseline relationship between contango and returns remains negative and statistically significant, while the recovery-period estimate turns positive and is statistically stronger than in the aggregate sample. By contrast, the metals and agriculture subsamples do not provide comparable statistical support for similar time variation.

Their coefficient estimates are directionally similar, but the evidence is substantially less precise. This means that the pandemic-related time variation observed in the full sample should not be interpreted as a broad commodity-market result. Rather, it appears to be driven mainly by energy products.

The robustness checks sharpen this conclusion further. Diagnostic tests indicate that heteroskedasticity and serial correlation are relevant concerns in the weekly panel, which supports the use of robust inference methods. The negative baseline relationship remains present under alternative covariance estimators, and it also survives the week fixed-effects specification. This strengthens confidence in the main baseline result. At the same time, the post-shock estimate proves more sensitive to specification choice than the baseline term-structure effect, and the non-energy specification shows that the broader negative baseline relationship remains visible even after excluding energy products, whereas the stronger COVID-period time variation largely does not. Taken together, the robustness evidence supports a nuanced interpretation: the baseline negative association is relatively broad-based, but the pandemic-period slope changes are mainly an energy-market phenomenon.

These findings contribute to the literature in two ways. First, the thesis brings together commodity futures term structure and single-commodity ETP returns within one unified panel framework. Second, it shows that the relationship is state-dependent, but not in the simple form initially hypothesized. For investors, the results imply that attention should be paid not only to expected commodity price movements, but also to the slope of the futures curve, especially in futures-based energy products where term-structure conditions appear to matter most.

The study also has limitations. The sample size is modest, and the number of products in the commodity-group subsamples is small, which weakens statistical precision and calls for caution in interpreting sector-specific results. In addition, the term-structure variable is designed to capture the market-level futures curve environment rather than each product's exact realized roll yield. Although the robustness checks address several inferential concerns, the baseline specification also remains exposed to the broader challenge of common market-wide shocks in weekly commodity data. Future research could extend the analysis by using larger product samples, more product-specific roll measures, alternative definitions of market regimes, and additional approaches to modelling common time effects.

Overall, the results suggest that futures curve dynamics matter for the returns of futures-based single-commodity ETPs, but their effects are neither constant over time nor uniform across commodity sectors. The clearest conclusion is that contango is associated with lower returns under normal market conditions. The clearest qualification is that the pandemic-related evolution of this relationship was driven mainly by energy markets rather than by a general strengthening of the negative contango effect across all commodity ETPs.

References

Aramyan, H. (2023, January 24). *Reconstructing RICs for expired futures contracts*. LSEG Developer Portal.

Arellano, M. (1987). Computing Robust Standard Errors for Within-groups Estimators. *Oxford Bulletin of Economics and Statistics*, *49*(4), 431–434. https://doi.org/10.1111/j.1468-0084.1987.mp49004006.x

Bakas, D., & Triantafyllou, A. (2020). Commodity price volatility and the economic uncertainty of pandemics. *Economics Letters*, *193*, 109283. https://doi.org/10.1016/j.econlet.2020.109283

Basak, S., & Pavlova, A. (2016). A Model of Financialization of Commodities. *Journal of Finance*, *71*(4), 1511–1556. https://doi.org/10.1111/jofi.12408

Bekaert, G., & Hoerova, M. (2014). The VIX, the variance premium and stock market volatility. *Journal of Econometrics*, *183*(2), 181–192. https://doi.org/10.1016/j.jeconom.2014.05.008

Ben-David, I., Franzoni, F., & Moussawi, R. (2017). Exchange-Traded Funds. *Annual Review of Financial Economics*, *9*(1), 169–189. https://doi.org/10.1146/annurev-financial-110716-032538

Bessembinder, H., & Lemmon, M. L. (2002). Equilibrium pricing and optimal hedging in electricity forward markets. *Journal of Finance*, *57*(3), 1347–1382. https://doi.org/10.1111/1540-6261.00463

Borgards, O., Czudaj, R. L., & Hoang, T. H. Van. (2021). Price overreactions in the commodity futures market: An intraday analysis of the Covid-19 pandemic impact. *Resources Policy*, *71*, 101966. https://doi.org/10.1016/j.resourpol.2020.101966

Brennan, M. J. (1958). The Supply of Storage. *The American Economic Review*, *48*(1), 50–72.

Chen, L., Verousis, T., Wang, K., & Zhou, Z. (2023). Financial stress and commodity price volatility. *Energy Economics*, *125*, 106874. https://doi.org/10.1016/j.eneco.2023.106874

Cheng, I.-H., & Xiong, W. (2014). Financialization of Commodity Markets. *Annual Review of Financial Economics*, *6*(1), 419–441. https://doi.org/10.1146/annurev-financial-110613-034432

Chincarini, L. B., & Moneta, F. (2021). The challenges of oil investing: Contango and the financialization of commodities. *Energy Economics*, *102*(3), 105443. https://doi.org/10.1016/j.eneco.2021.105443

Cortazar, G., Ortega, H., Santa Maria, J., & Schwartz, E. S. (2024). Expected returns on commodity ETFs and their underlying assets. *Journal of Commodity Markets*, *36*, 100439. https://doi.org/10.1016/j.jcomm.2024.100439

Cucinotta, D., & Vanelli, M. (2020). WHO declares COVID-19 a pandemic. *Acta Biomedica,* 91(1), 157–160. https://doi.org/10.23750/abm.v91i1.9397

Dusak, K. (1973). Futures Trading and Investor Returns: An Investigation of Commodity Market Risk Premiums. *The Journal of Political Economy*, *81*(6), 1387–1406. https://doi.org/10.1086/260133

Erb, C. B., & Harvey, C. R. (2006). The strategic and tactical value of commodity futures. *Financial Analysts Journal,* 62(2), 69–97. https://doi.org/10.2469/faj.v62.n2.4084

Fama, E. F., & French, K. R. (1987). Commodity Futures Prices: Some Evidence on Forecast Power, Premiums, and the Theory of Storage. *The Journal of Business*, *60*(1), 55–73. https://doi.org/10.1086/296385

Goodhart, C. A. E., & O'Hara, M. (1997). High frequency data in financial markets: Issues and applications. *Journal of Empirical Finance*, *4*(2), 73–114. https://doi.org/10.1016/S0927-5398(97)00003-0

Gorton, G. B., Hayashi, F., & Rouwenhorst, K. G. (2013). The Fundamentals of Commodity Futures Returns. Review of Finance, *17*(1), 35–105. https://doi.org/10.1093/rof/rfs019

Gorton, G., & Geert Rouwenhorst, K. (2006). Facts and fantasies about commodity futures. *Financial Analysts Journal*, *62*(2), 47–68. https://doi.org/10.2469/faj.v62.n2.4083

Hamilton, J. D. (2009). Causes and Consequences of the Oil Shock of 2007-08. NBER Working Paper No. 15002. https://doi.org/10.3386/w15002

Hamilton, J. D., & Wu, J. C. (2014). Risk premia in crude oil futures prices. *Journal of International Money and Finance*, *42*, 9–37. https://doi.org/10.1016/j.jimonfin.2013.08.003

International Energy Agency. (2021). Global Energy Review *2021*. IEA.

International Monetary Fund. (2021). World Economic Outlook, April 2021: Managing Divergent Recoveries. IMF.

Marszk, A. (2017). Exchange traded commodities as a category of innovative products on European financial markets. *E-Finanse*, *13*(2), 14–21. https://doi.org/10.1515/fiqf-2016-0019

Mou, Y. (2011). Limits to Arbitrage and Commodity Index Investment: Front-Running the Goldman Roll. SSRN Working Paper. https://doi.org/10.2139/ssrn.1716841

Orlando, G. (2025). Exchange traded products: Taxonomy, risk and mitigations. *International Review of Financial Analysis*, *101*, 103969. https://doi.org/10.1016/j.irfa.2025.103969

Stewart, S. L., Massa, O. I., Hassman, C., & Leon, M. de. (2023). ETP tracking of U.S. agricultural and energy markets. *Journal of Commodity Markets*, *31*, 100344. https://doi.org/10.1016/j.jcomm.2023.100344

Stock, J. H., & Watson, M. W. (2008). Heteroskedasticity-Robust Standard Errors for Fixed Effects Panel Data Regression. *Econometrica*, *76*(1), 155–174. https://doi.org/10.1111/j.0012-9682.2008.00821.x

Szczygielski, J. J., Brzeszczyński, J., Charteris, A., & Bwanya, P. R. (2022). The COVID-19 storm and the energy sector: The impact and role of uncertainty. *Energy Economics*, *109*, 105258. https://doi.org/10.1016/j.eneco.2021.105258

Tang, K., & Xiong, W. (2012). Index investment and the financialization of commodities. *Financial Analysts Journal*, *68*(6), 54–74. https://doi.org/10.2469/faj.v68.n6.5

Working, H. (1949). The Theory of Price of Storage. *The American Economic Review,* 39(6), 1254–1262.

World Bank Group. (2020a). *Commodity Markets Outlook, April 2020*. https://doi.org/10.1596/33624

World Bank Group. (2020b). *Commodity Markets Outlook, October 2020*. https://doi.org/10.1596/34621

World Bank Group. (2021). *Commodity Markets Outlook, April 2021: Causes and Consequences of Metal Price Shocks*. https://doi.org/10.1596/35458

Appendices

Appendix 1: Sample of single-commodity ETPs included in the analysis

Table 7: Sample of single-commodity ETPs included in the analysis

This table lists the single-commodity ETPs included in the empirical sample. The products are classified into the energy, metals, and agriculture sectors and include both ETFs and ETCs. All selected products provide non-leveraged long exposure to one underlying commodity.

Product name	Product Type	Category	Commodity
United States Oil Fund, LP (USO)	ETF	Energy	WTI
Invesco DB Oil Fund (DBO)	ETF	Energy	WTI
United States 12 Month Oil Fund, LP (USL)	ETF	Energy	WTI
United States Brent Oil Fund, LP (BNO)	ETF	Energy	Brent Crude
United States Natural Gas Fund, LP (UNG)	ETF	Energy	Natural Gas
United States 12 Month Natural Gas Fund, LP (UNL)	ETF	Energy	Natural Gas
WisdomTree Natural Gas (OD7L)	ETC	Energy	Natural Gas
United States Copper Index Fund, LP (CPER)	ETF	Metals	Copper
WisdomTree Aluminium (OD7A)	ETC	Metals	Aluminium
WisdomTree Copper (OD7C)	ETC	Metals	Copper
WisdomTree Nickel (OD7M)	ETC	Metals	Nickel
WisdomTree Zinc (OD7T)	ETC	Metals	Zinc
Teucrium Corn Fund (CORN)	ETF	Agriculture	Corn
Teucrium Soybean Fund (SOYB)	ETF	Agriculture	Soybean
Teucrium Wheat Fund (WEAT)	ETF	Agriculture	Wheat
Teucrium Sugar Fund (CANE)	ETF	Agriculture	Sugar

Appendix 2: Robustness checks

Table 8: Full-sample fixed-effects results with alternative covariance estimators

This table reports full-sample fixed-effects regression results using three alternative inference methods: product-clustered standard errors, Newey–West standard errors, and Driscoll–Kraay standard errors. The dependent variable is the weekly logarithmic return of product *i* in week *t*. Product fixed effects are included in all specifications. The table is used to assess whether the baseline findings are robust to alternative treatments of heteroskedasticity, serial correlation, and cross-sectional dependence. Significance markers: * $p<0.05$, ** $p<0.01$, *** $p<0.001$. Standard errors (SEs) are in parenthesis.

Variable	Clustered by Product	Newey-West (L=4)	Driscoll-Kraay (L=4)
Term Structure	-0.0849***	-0.0849***	-0.0849**
	(0.0167)	(0.0216)	(0.0305)
Shock	0.0024	0.0024	0.0024
	(0.0036)	(0.0044)	(0.0049)
Post Shock	0.0066***	0.0066***	0.0066*
	(0.0013)	(0.0016)	(0.0026)
Recovery	0.0067***	0.0067***	0.0067**
	(0.0009)	(0.0013)	(0.0024)
Spot Return	0.4016***	0.4016***	0.4016***
	(0.0748)	(0.0338)	(0.047)
VIX	-0.0081**	-0.0081***	-0.0081*
	(0.0020)	(0.0016)	(0.0034)
Term Structure: Shock	-0.0253	-0.0253	-0.0253
	(0.1503)	(0.1465)	(0.1296)
Term Structure: Post Shock	0.0731***	0.0731	0.073
	(0.0147)	(0.054)	(0.070)
Term Structure: Recovery	0.2276**	0.2276***	0.2276**
	(0.0763)	(0.067)	(0.078)

Table 9: Non-energy fixed-effects results with alternative covariance estimators

This table reports fixed-effects regression results for the non-energy sample using three alternative inference methods: product-clustered standard errors, Newey–West standard errors, and Driscoll–Kraay standard errors. The dependent variable is the weekly logarithmic return of product *i* in week *t*. Product fixed effects are included in all specifications. The table is used to assess whether the non-energy results are sensitive to the treatment of the error structure. Significance markers: * p<0.05, **p<0.01, ***p<0.001. Standard errors (SEs) are in parenthesis.

Variable	Clustered by Product	Newey-West (L=4)	Driscoll-Kraay (L=4)
Term Structure	-0.0518* (0.0206)	-0.0518** (0.0192)	-0.0518* (0.0209)
Shock	0.0001 (0.0013)	0.0001 (0.0014)	0.0001 (0.0018)
Post Shock	0.0008 (0.0009)	0.0008 (0.0009)	0.0008 (0.0010)
Recovery	0.0033** (0.0007)	0.0033*** (0.0007)	0.0033*** (0.0007)
Spot Return	0.7801*** (0.0413)	0.7801*** (0.0368)	0.7801*** (0.0323)
VIX	-0.0024** (0.0007)	-0.0024* (0.0010)	-0.0024 (0.0013)
Term Structure: Shock	0.0729 (0.1784)	0.0729 (0.1620)	0.0729 (0.1045)
Term Structure: Post Shock	0.0420* (0.0160)	0.0420 (0.0247)	0.0420 (0.0252)
Term Structure: Recovery	0.0766 (0.0359)	0.0766* (0.0312)	0.0766* (0.0332)

Table 10: Total effects of term structure for the non-energy sample

This table reports the implied total effects of the term-structure measure across the reference, shock, post-shock, and recovery periods for the non-energy sample. Total effects are calculated as linear combinations of the baseline term-structure coefficient and the relevant interaction coefficient from the non-energy fixed-effects model. Standard errors are based on the same product-clustered covariance matrix as in the corresponding regression specification. Significance markers: * p<0.05, **p<0.01, ***p<0.001. Standard errors (SEs) are in parenthesis.

Baseline	Shock	Post Shock	Recovery
-0.0518* (0.0206)	0.0211 (0.1911)	-0.0098 (0.0095)	0.0248 (0.0218)

Appendix 3 Explanation of the use of AI (Selvitys tekoälyn käytöstä)

In the preparation of this thesis, I used generative artificial intelligence as a support tool at different stages of the research and writing process. The tool used, the purpose of its use, and the measures taken to verify the AI-assisted output are described below. I affirm that I used AI with appropriate care, disclosed its use in accordance with the applicable guidelines, and accept full responsibility for the entire content of this thesis.

1. **Tool Used: OpenAI's ChatGPT (GPT-5.4 Thinking)**

- **Phase of use:** topic ideation, research planning, and source search support
    - **Purpose of use:** I used AI tools as a support in the early stages of the thesis process to refine the thesis topic and to clarify the research focus. It was also used to narrow the scope of the study and to identify relevant literature related to commodity futures term structure, contango, backwardation, and roll yield.
    - **Verification:** All suggestions provided by AI were treated as preliminary support. All the actual literature searches were conducted independently in academic databases, and all sources included in the thesis were selected, read, and evaluated independently.
- **Phase of use:** Methodology and data analysis (R code)
    - **Purpose of use:** I used AI tools as a support tool in working with R code for the empirical analysis. It was used to help identify coding errors, improve code structure, troubleshoot regression and plotting issues, and suggest technical adjustments related to data handling, model implementation, and figure formatting. The tool was used only for technical coding support and not for making methodological decisions or generating the interpretation of empirical results.
    - **Verification:** All code suggestions were tested independently in R before use. I understood the entire working script and verified several times that all key parts of the code worked as intended. All methodological choices, model specifications, and interpretations of results were made independently.
- **Phase of use:** composition, editing and revision

- **Purpose of use**: I used AI to improve the language and readability of text. This included proofreading, correcting grammar and spelling, improving sentence clarity, and rephrasing individual sentences or short passages in a more academic style without changing the intended meaning.
- **Verification**: I reviewed all language suggestions before use and only made changes after verifying that the intended meaning and academic accuracy did not change. The final wording, structure, arguments, and conclusions remained fully under my control and responsibility. AI was not used for ghost-writing or to generate any key arguments of the thesis.