Universidad del País Vasco
Euskal Herriko Unibertsitatea
BILBOKO INGENIARITZA ESKOLA
ESCUELA DE INGENIERÍA DE BILBAO

MÁSTER UNIVERSITARIO EN INGENIERÍA INDUSTRIAL

TRABAJO FIN DE MASTER

HEDGING PRESSURE DYNAMICS IN LME ZINC FUTURES MARKETS

Estudiante: Gallardo Cabrera, Ania

Director/Directora: Herrero Villalibre, Saioa

Curso: 2023-2024

Fecha: Bilbao, junio de 2024

ABSTRACT

This study explores the hedging pressure dynamics within the London Metal Exchange (LME) zinc futures market, emphasizing the impact of hedging activities on futures prices. Zinc, an essential industrial metal, is subject to price volatility influenced by various market participants, including producers, consumers, and speculators. The research investigates the relationship between net hedging positions and zinc futures prices, particularly in the context of the LME's shift to the Commitment of Traders Report (COTR) format. By analyzing weekly data from January 2018 to May 2024, the study employs econometric models to assess how changes in hedging pressure influence futures prices. Results indicate that while hedging pressure does exhibit some predictive power, its impact on returns is limited and influenced by other market dynamics and structural changes. This research contributes to the understanding of market behavior and risk management strategies in the zinc futures market, offering insights for market participants on the role of hedging in price formation.

RESUMEN

Este estudio explora la dinámica de la presión de cobertura dentro del mercado de futuros de zinc en la Bolsa de Metales de Londres (LME), destacando el impacto de las actividades de cobertura en los precios de los futuros. El zinc, un metal industrial esencial, está sujeto a la volatilidad de precios, influenciada por diversos participantes del mercado, incluidos productores, consumidores y especuladores. La investigación analiza la relación entre las posiciones netas de cobertura y los precios de los futuros de zinc, particularmente en el contexto de la transición de la LME al formato del Informe de Compromiso de los Operadores (COTR, por sus siglas en inglés). Al analizar datos semanales desde enero de 2018 hasta mayo de 2024, el estudio emplea modelos econométricos para evaluar cómo los cambios en la presión de cobertura influyen en los precios de los futuros. Los resultados indican que, si bien la presión de cobertura exhibe cierto poder predictivo, su impacto en los rendimientos es limitado y está influenciado por otras dinámicas de mercado y cambios estructurales. Esta investigación contribuye a la comprensión del comportamiento del mercado y las estrategias de gestión de riesgos en el mercado de futuros de zinc, ofreciendo ideas para los participantes del mercado sobre el papel de la cobertura en la formación de precios.

LABURPENA

Azterlan honek Londresko Metalen Burtsan (LME) zinkaren gerokoen merkatuaren barruan estalduraren presioaren dinamika aztertzen du, eta estaldura-jarduerek gerokoen prezioetan duten eragina nabarmentzen du. Zinka, funtsezko metal industriala, prezioen hegazkortasunaren mende dago, eta merkatuko hainbat parte-hartzailek eragiten dute, ekoizleak, kontsumitzaileak eta espekulatzaileak barne. Ikerketak estaldura-posizio garbien eta zink-gerokoen prezioen arteko erlazioa aztertzen du, batez ere LMEtik Operadoreen Konpromiso Txostenaren formaturako trantsizioaren testuinguruan (COTR, ingelesezko sigletan). 2018ko urtarriletik 2024ko maiatzera arteko asteroko datuak aztertzean, ikerketak eredu ekonometrikoak erabiltzen ditu estaldura-presioaren aldaketek gerokoen prezioetan nola eragiten duten ebaluatzeko. Emaitzek adierazten dutenez, estaldura-presioak nolabaiteko ahalmen iragarlea badu ere, etekinetan duen eragina mugatua da, eta beste merkatu-dinamika eta egitura-aldaketa batzuen eragina du. Merkatuaren portaera eta zink-etorkizunen merkatuko arriskuak kudeatzeko estrategiak ulertzen laguntzen du ikerketa honek, merkatuko parte-hartzaileei ideiak eskainiz estaldurak prezioen eraketan duen eginkizunari buruz.

INDEX

FIGURE INDEX

1. INTRODUCTION

BACKGROUND AND CONTEXT OF THE STUDY

The dynamics of commodity markets, particularly within the metal sector, have witnessed significant fluctuations and volatility over the years. Zinc, as a vital industrial metal, plays a crucial role in various sectors, including manufacturing, construction, and infrastructure development. The management of risk associated with zinc price movements becomes necessary for market participants, including producers, consumers, and investors.

The mid-2000s saw a notable surge in commodity prices, leading to a growing discourse on the role of speculators in influencing price levels. This period coincided with substantial changes in the commodity futures and options markets, commonly referred to as the "financialization" of these markets. As these markets experienced substantial growth in both held positions and traded volumes around 2004, questions arose regarding the impact of hedging activities on zinc futures prices (*Cheng and Xiong, 2014*).

A significant number of studies and papers (*Sanders et al. (2010), Irwin and Sanders (2012*) have been released investigating the relationship between these factors across a wide range of commodities. However, this writing will focus on the hedging activities' impact on zinc 3-month future prices on the London Metal Exchange (LME). The exclusivity of this focus is motivated by the recent shift in the LME's information-sharing policy, transitioning to the Commitment of Trader Report format (which has been used for years in the exchanges in the United States). This document, which will be further described in a following section, intricately details the positions of distinct hedging groups in the zinc futures market. This shift in reporting methodology presents an opportune moment to conduct a causal analysis, delving into the potential influence of net hedger positions on commodity futures prices.

PROPOSED RESEARCH QUESTIONS

The purpose of this research is to examine data from the COTR (Commitment of Traders Reports) to test the relationship between LME futures prices and the trading positions of various types of investors. In light of the ongoing discussions regarding hedging and speculative influences within zinc futures markets, this research aims to explore three fundamental questions: Do market participants' trading activities, specifically net hedging pressure, drive price movements? Are the trading positions of different market participant groups influenced by price changes? Have the dynamics of these relationships shifted because of structural changes, such as Covid-19, in futures markets?

Through a thorough and systematic examination, this study seeks to analyze the lead-lag relationships between traders' positions and prices in zinc futures market.

STRUCTURE AND METHODOLOGY OF THE PROJECT

This research aims to analyze the hedging pressure dynamics in the LME Zinc futures markets, focusing on how changes in hedging pressure influence futures prices. The methodology involves collecting relevant market data, applying econometric models, and conducting a series of robustness checks to ensure the validity of the results.

Firstly, data collection has been conducted using Bloomberg and London Metal Exchange webpage as main sources. This data collects the COTR and a detailed description of the values of each market participant, such as closing prices and trading volume or open interest of each considered group. This research will cover the data from the 26th of January 2018 to the 29th of May 2024, given by the COTR with a weekly frequency.

The data has been downloaded and preprocessed before analyzing it on the programming language Python. The data has been carefully cleaned by dropping any nonexistent value among all the data collection and it has been transformed in an excel sheet for a more comfortable way of working with it. Additionally, many measures and new variables have been added to this excel sheet in order to make additional calculations.

All the data has been thoroughly examined in Section 5 of this project. This section delves into market sentiment and explores the potential reasons behind the variations observed among all market participants. Descriptive statistics have been conducted on the time series for a better understanding of the positions acquired by the different participants on the zinc futures market.

In Section 6, a regression analysis is conducted to examine whether the relationships hypothesized between variables hold true. Specifically, the analysis assesses whether various net hedging pressures and other considered variables can predict returns. This involves using statistical techniques like regression to quantify how changes in one variable (such as net hedging pressure) relate to changes in another variable (returns). By analyzing these relationships, the study aims to determine if and to what extent these variables influence or predict returns in the context being studied. The analysis was conducted using the programming language Pyhton and Microsoft Excel. One of the biggest limitations of this methodology is the potential for omitted variable bias due to unobserved factors affecting futures prices. Additionally, the results are specific to the time period studied and may not generalize to other periods.

To sum up, the methodology outlined provides a comprehensive approach to analyzing the hedging pressure dynamics in the LME Zinc futures markets, ensuring robust and reliable results.

PREVIOUS STUDIES

The role of hedgers and speculators in commodity markets has been an issue of considerable interest and controversy (Brunetti, 2015). However, consensus on the impact of the different groups of investors on commodity futures markets remains elusive.

(Park J. , 2018) found that the evidence of volatility transmission between oil and base metals was somewhat strong. He argued that considering this result, the behavior of volatility in oil and LME futures prices applied to hedge decisions across the commodity market was useful. (Park J. a., 2018) reported that the LME market was not efficient due to the false premise that the financialization of commodities had been growing. This result implied that the LME futures market could generate somewhat possible excess returns, which could attract speculators such as hedge funds. (Park J. , 2019) argued that the canceled warrants (CWs) in the LME market were key indicators to explain the financialization effect in the market price directly. He pointed out that the CWs variable was crucial because it encompassed two factors—fundamentals and non-fundamentals. He found that the rise in CWs increased the LME metal prices, including aluminum, zinc, tin, and nickel. He also found that the positive impact of the CWs on metal returns was transitory.

(Arouri, 2011) demonstrated the speculative efficiency of the aluminum contract traded on the LME through cointegration methodologies. Their study revealed that futures aluminum prices exhibited cointegration with spot prices and served as biased estimators of future spot prices. On a somewhat divergent note, (Figuerola-Ferretti, 2008) analyzed the volatility characteristics of aluminum and copper on the LME. They discovered that both aluminum and copper volatilities displayed statistically long memory processes, attributed to speculative trading activities.

2. Zinc Futures Market

OVERVIEW OF THE LONDON METAL EXCHANGE (LME)

The London Metal Exchange (LME) is a futures and forwards exchange based in London, United Kingdom. Formed in 1877, the LME is one of the world's largest markets for trading non-ferrous metals, which include aluminum, copper, lead, nickel, tin, and zinc. This market allows participants to trade in different types of contracts of the metals, with the option to settle them both physically and financially. LME contracts are standardized with respect to expiration dates and size. They are always traded in lots, which vary in size from 1 to 50 metric tons depending on contract type and the underlying metal and are priced in US dollars. However, the LME also publishes official exchange rates to enable settlement in pound sterling, Japanese yen, and euros.

The LME has three transparent and regulated platforms for trading metal contracts: LMEselect and LMEbullion (electronic), "The Ring" (outcry open system) and a 24-hour telephone market [4]. The trend is moving in the direction of electronic trading, distancing from the conventional open outcry trading method, wherein traders convene either face-to-face or in trading pits. The LME outcry open system, "The Ring" was the only European exchange where open-outcry trading took place. It was temporarily closed in March 2020 due to Covid-19, and restarted in September 2021. The distinctive feature of "The Ring" is that participants have the ability to hedge metals prices for specific days into the future, instead of using a standardized one- or three-month contract that are offered on most futures exchanges.

ZINC

Zinc is a versatile and essential metal with a wide range of applications. It is extracted from ore through mining, primarily from underground or open-pit mines. The ore undergoes crushing, grinding, and flotation to produce zinc concentrate, which is then roasted to produce zinc oxide. The oxide is reduced using carbon or treated with sulfuric acid in electrolytic processes to produce pure zinc.

Zinc is essential for galvanizing steel to prevent rust and corrosion. This process is widely used in construction, automotive, and infrastructure projects. Zinc is also used in the production of alloys such as brass (zinc and copper) and bronze (zinc, copper, and tin). These alloys have various applications in machinery, electrical components, and decorative items. Finally, one of the most important uses of zinc is the production of batteries (such as zinc-carbon and zinc-air batteries) and in various chemical processes.

ZINC SPOT PRICE

The London Metal Exchange (LME) zinc spot price refers to the current price at which zinc can be bought or sold for immediate delivery on the London Metal Exchange. The LME zinc spot price is influenced by various factors, including supply and demand dynamics, production levels, and global economic conditions. When demand outpaces supply, prices tend to rise, and vice versa. Major zinc-producing countries include China, Australia, Peru, and the United States. Changes in production levels due to new mines, closures, or production halts can significantly impact prices.

RELATIONSHIP BETWEEN ZINC SPOT PRICE AND ZINC 3-MONTH FUTURES PRICE

The relationship between the spot price and the 3-month future price of zinc is influenced by several key factors. In Figure 1, we observe the trend of the 3-month futures price of zinc from 1992 to present day. This price reflects the market's expectations for zinc prices three months into the future, incorporating various factors such as anticipated supply and demand, production costs, and macroeconomic indicators. As the futures price fluctuates, it provides insights into how traders and market participants view the potential changes in the zinc market.

Figure 1: Zinc 3 month futures settle price

Periods of rising futures prices might indicate expectations of increased demand or potential supply constraints, while declining futures prices could suggest anticipated reductions in demand or an increase in supply. Analyzing the 3-month futures price trend helps market participants make informed decisions about trading, hedging, and planning for future price movements. As a matter of fact, there is a relationship between the futures settle price and the volume. Typically, higher trading volumes in zinc 3-month futures can indicate increased market activity and liquidity. This liquidity often supports more stable and reflective price discovery in futures markets. Also, most times, high trading volume is a signal. A significant increase in trading volume can sometimes precede changes in the settlement price. For example, a surge in volume might signal a shift in sentiment or expectations regarding future zinc prices. In Figure 2, it is clear how in times of high volume trading, the settlement price has also been driven up, such as 2016 or 2024. However, this relationship does not always hold true. Around the 2007 financial crisis, the settlement price highly increased, but the traded volume remained low.

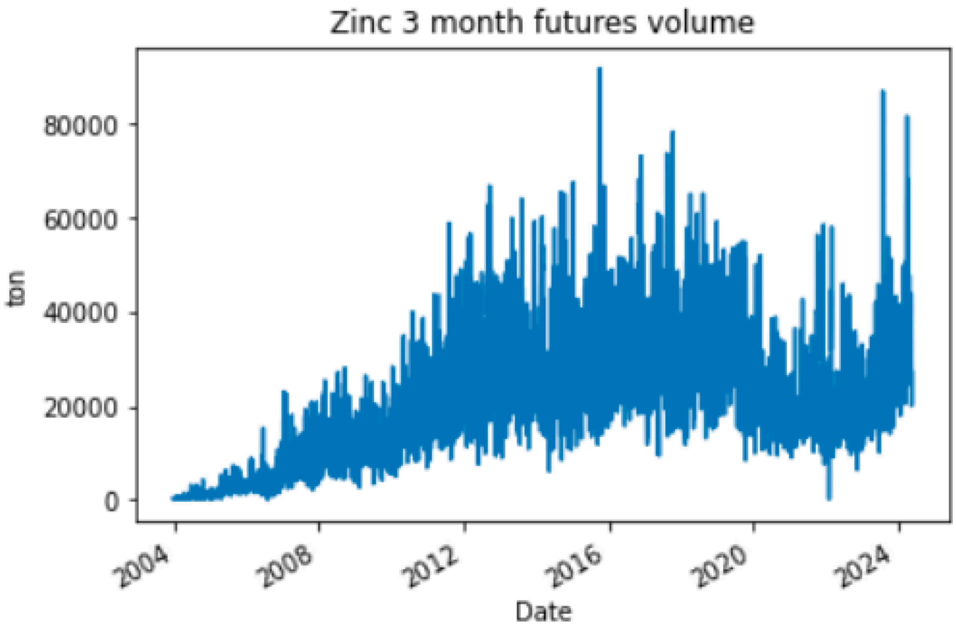

Figure 2: Zinc 3 month futures volume

The main factors affecting zinc futures prices are described. Interest rates, or the cost of carry, play a crucial role as higher interest rates increase the cost of holding the physical commodity, leading to higher future prices relative to the spot price. Supply and demand expectations also affect this relationship, giving place to two common conditions named contango and backwardation.

- Contango: When the future price is higher than the spot price, it suggests that the market expects prices to rise over time. This is often due to costs of carry, such as storage and interest rates, or expectations of higher future demand or lower future supply.
- Backwardation: When the future price is lower than the spot price, it indicates that the market expects prices to fall. This can occur when there are current supply constraints expected to ease or high current demand expected to decline.

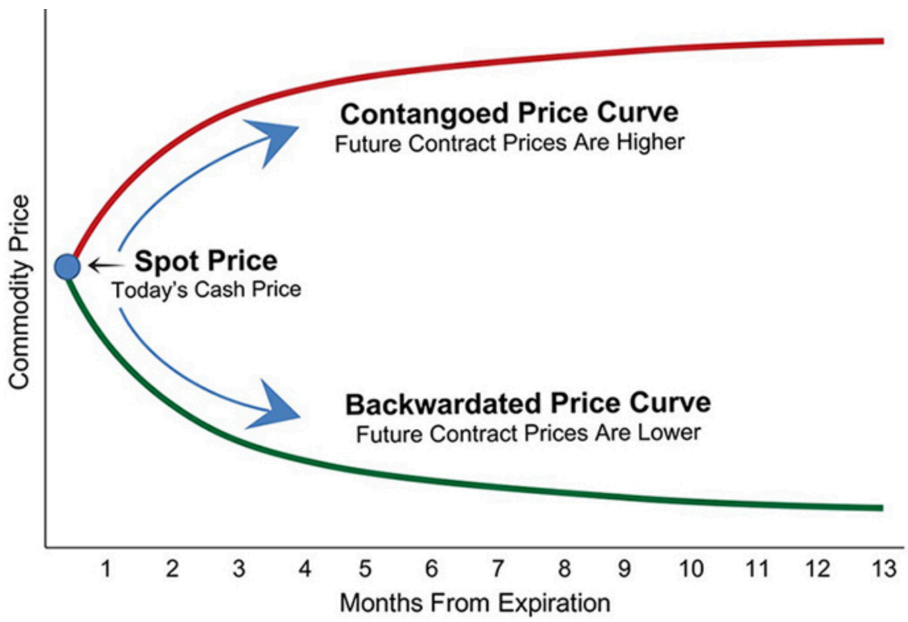

Figure 3: Illustrative explanation of contango and backwardation

Inventory levels are another important factor, with high current inventories potentially lowering the spot price and expectations of decreasing inventories pushing future prices higher. In Figure 3, we see the closing stock of zinc in LME from 1992 to present day.

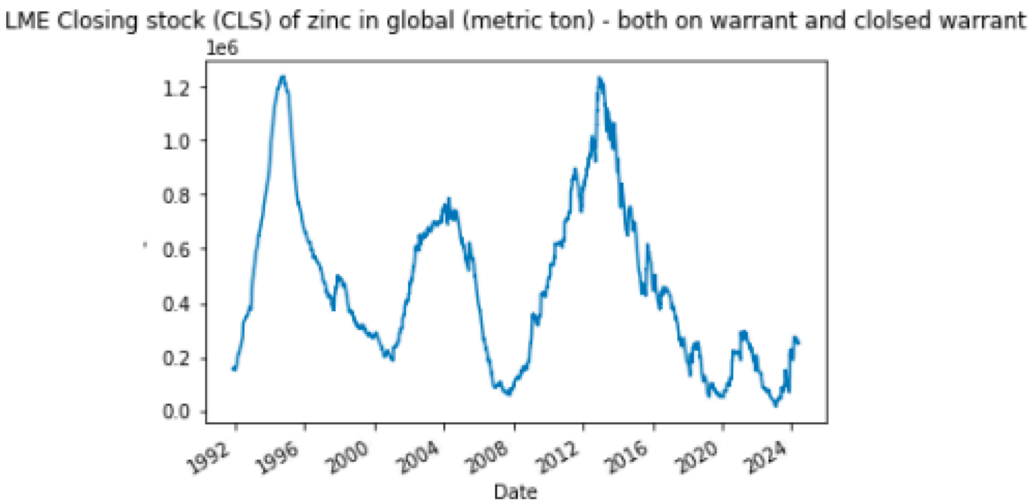

Figure 4: LME Closing Stock of Zinc

This graph shows the closing stock (CLS) of zinc on the London Metal Exchange (LME) in metric tons, both on warrant and closed warrant. The warrant refers to a certificate that gives the holder the right to ownership of a certain amount of a commodity, in this case, zinc. On-warrant metal is metal that is being held in LME warehouses and is readily available for delivery, while closed warrant metal is metal that is being held outside of LME warehouses but is still registered with the LME. The graph shows that the CLS of zinc has fluctuated over the past three decades, but there is no clear upward or downward trend. To demonstrate this, Figure 5 calculates the weekly difference. The weekly difference of CLS exhibits notable fluctuations over time, displaying both positive and negative values. This variability suggests significant changes in zinc stock levels on the LME from week to week.

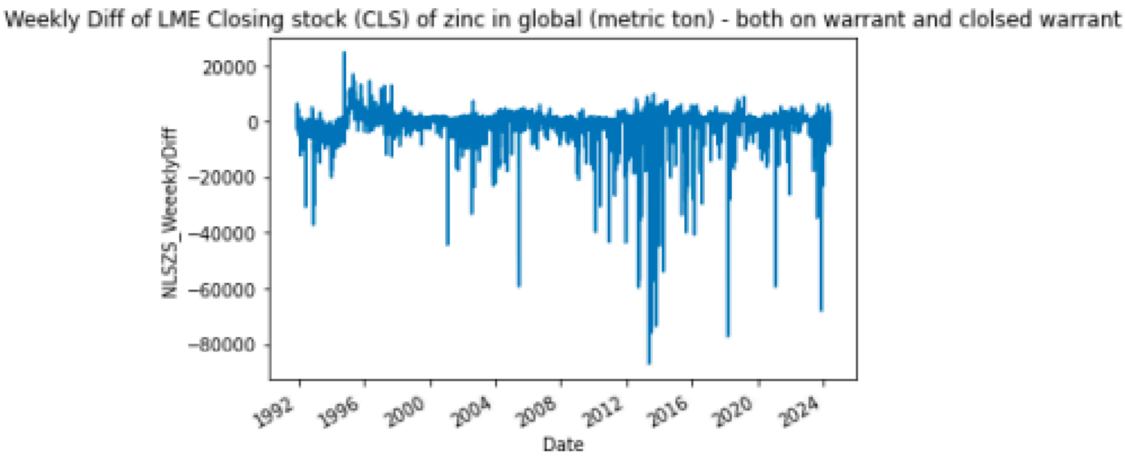

Figure 5: Weekly difference of Closing Stock of Zinc

Arbitrage opportunities help maintain balance, as traders exploit price discrepancies between spot and futures markets, bringing prices closer together. Geopolitical and economic factors, such as trade policies, macroeconomic conditions, and regulatory changes, can cause sudden shifts in supply and demand expectations, impacting both spot and future prices.

ZINC FUTURES MARKET ON THE LME

The London Metal Exchange (LME) offers a robust market for zinc futures, which is vital for those involved in the metal supply chain and the investment community. LME zinc futures contracts allow market participants to hedge against price fluctuations or take on price risks. These contracts specify the purchase or sale of zinc at a future date at an agreed price, providing an effective tool for price risk management. All the metals traded on the LME must conform to strict specifications regarding quality, lot size and shape and their contracts are subject to LME's rules.

The most important contract specifications of the zinc futures:

- Contract size: 25 metric tons.

- Quality: Special high-grade zinc of 99.995% purity (minimum) that must conform to a specific chemical composition.
- Delivery locations: LME approved warehouses around the world.
- Settlement type: Physical.
- Expiry dates: Monthly contracts expiring on the third Wednesday of each month, with contracts traded up to 12 months forward.
- Price quotes: Prices are quoted in US dollars per ton.

Zinc futures are traded at “The Ring”, electronically on the LME's electronic trading platform or by telephone. Many market participants engage in zinc futures trading, including producers, consumers, speculators, and institutional investors. This diversity of participants contributes to market liquidity of zinc futures.

COMMITMENT OF TRADERS REPORT (COTR)

A Commitment of Traders Report (COTR) is a weekly publication that provides a breakdown of the open interest positions held by different types of traders in futures markets, in this case, the focus will be on the zinc market. The COTR aims to provide transparency in the futures market by showing the distribution of positions among various trader categories which helps market participants understand market dynamics and the potential impact of different trader behaviors on prices.

From December 2017, the London Metal Exchange publishes a new Commitment of Traders Report (COTR) in accordance with MiFID II (formally known as Markets in Financial Instruments Directive II, is a regulatory framework established by the European Union which aims to increase transparency, investor protection, and fairness in the trading of financial instruments across the EU).

The new design of COTR aims to characterize the primary business activities of LME members and their clients and is published on a weekly basis. The COTR compiles data from the daily open positions reported in the Commodity Position Reports (CPR), submitted by members through the LME universal data gateway (UDG). For each metal, a dedicated COTR is generated and published on the LME website every Tuesday, reflecting positions held at the close of business on the preceding Friday.

The classification of Members and their Clients is the following:

a) Investment firms or credit institutions
b) Investment funds
c) Other financial institutions
d) Commercial undertakings

Each member must exercise judgement as to which classification they should be categorized in, as well as that of their clients. However, LME reserves the right to re-classify any entity where it deems necessary to do so.

Concerning the report, it must provide the following information as shown on Figure 1:

- NUMBER OF LONG POSITIONS (L): the aggregate quantity of long positions per classification held on Friday at the end of the trading day.
- NUMBER OF SHORT POSITIONS (S): the aggregate quantity of short positions per classification held on Friday at the end of the trading day.
- NOTATION OF POSITION QUANTITY: the units used to report the number of positions.
- CHANGES SINCE THE PREVIOUS REPORT (L): increase or decrease in the long position for each classification with respect to the previous Friday.
- CHANGES SINCE THE PREVIOUS REPORT (S): increase or decrease in the short position for each classification with respect to the previous Friday.
- % OF TOTAL LONG OPEN INTEREST (L): % of total open interest represented by the long positions.
- % OF TOTAL SHORT OPEN INTEREST (S): % of total open interest represented by the short positions.
- NUMBER OF PERSONS HOLDING A POSITION IN EACH CATEGORY (L AND S COMBINED): total number of Members and Clients holding a position in each classification.

Similarly, activities can be categorized into two types: risk-reducing activities, primarily involving hedging, and other activities that serve different purposes or objectives, such as speculation. This distinction is essential for understanding the motives behind participants' actions in the market.

Figure 6: Commitment of Traders Report of Zinc Futures

	Risk Reducing			Other Activities			Total		
	Position	Chg	% OI	Position	Chg	% OI	Position	Chg	% OI
Investment Firms or Credit Institutions									
Long	19369.57	529.08	5.84	170029.1	1247.93	51.23	189398.6	1776.99	57.07
Short	11900.55	-866.25	3.59	172889.1	4358.29	52.21	184789.7	3492.06	55.8
Net	7469.02	1395.33		-2860.05	-3110.33		4608.95	-1715.02	
Investment Funds									
Long	570	4	0.17	66908.94	1757.77	20.16	67478.94	1761.77	20.33
Short	112	4	0.03	27521.13	-1035.91	8.31	27633.13	-1031.91	8.34
Net	458	0		39387.8	2793.67		39845.8	2793.67	
Other Financial Institutions									
Long	3087	-149	0.93	20366.25	12.78	6.14	23453.25	-136.22	7.07
Short	13429	181	4.06	4683	130	1.41	18112	311	5.47
Net	-10342	-330		15683.25	-117.22		5341.25	-447.22	
Commercial Undertakings									
Long	32674.76	-5721.6	9.85	18868.24	-1563.15	5.69	51543	-7284.75	15.54
Short	77513.56	-3827.47	23.41	23120.14	-2783.2	6.98	100633.7	-6610.68	30.39
Net	-44838.8	-1894.13		-4251.9	1220.05		-49090.7	-674.08	
Directive Compliance Obligations									
Long	0	#N/A N/A	0	0	0	0	0	#N/A N/A	0
Short	0	#N/A N/A	0	0	#N/A N/A	0	0	#N/A N/A	0
Net	0	#N/A N/A		0	#N/A N/A		0	#N/A N/A	

3. Hedging Pressure, Hedgers and Indexes

SIGNIFICANCE OF HEDGING IN COMMODITY MARKETS

Hedging is a financial strategy that consists in taking and offsetting position in an asset or investment that reduces the risk of adverse price movements of the existing portfolio. That is, it is a strategic mechanism to manage and mitigate risks associated with price volatility. Derivatives can be effective hedges against their underlying assets because the relationship between the two is usually defined. They include options, swaps, futures, and forward contracts. The underlying assets can be stocks, bonds, commodities, currencies, indexes, or interest rates. The focus of this writing will be set on hedging in commodity markets, more specifically in Zinc futures market on LME.

The significance of hedging can be summarized in two main key points. The first and most important one is risk mitigation. Commodity prices are inherently volatile due to various factors, such as geopolitical events, supply and demand imbalances, and weather conditions. Hedging allows market participants, including producers, consumers, and investors, to safeguard against adverse price movements. However, there is a risk-reward tradeoff inherent in hedging; while it reduces potential risk, it may also take away potential gains. Nevertheless, most individuals or firms would opt for accepting an expected and limited loss instead of facing a sudden unforeseeable bigger loss. The second reason of why hedging is important in the financial world is the stability it provides to producers and consumers. Hedging provides a mechanism to stabilize revenues and costs, in which producers can lock in favorable prices for their goods, while consumers can secure predictable input costs.

Net hedging pressure (HP) is constructed using COTR data on commercial positions in futures contracts traded on futures exchanges. The CFTC defines a ‘commercial’ position as one belonging to a participant that uses the futures market to hedge exposures that arise as part of their usual operations. As such, it includes positions of producers and consumers, but also of swap dealers who act as intermediaries in these markets.

The net hedging pressure is defined as the aggregate difference between the volume of long (buy) and short (sell) positions in futures contracts held by hedgers.

$$HP_t = Long_t - Short_t$$

The percent net long position is referred to as hedging pressure variable:

$$HP_t(\%) = \frac{Long_t - Short_t}{Long_t + Short_t}$$

The net hedging pressure variable (HP) is bound to be between -1 and 1, and to be interpreted as follows. For example, a HP of -0.3 means that 30% of commercials are net short.

However, there is not one single correct way to calculate the net hedging pressure. Some other researchers define the net hedging pressure as the aggregate difference between the volume of short (sell) and long (buy) positions in futures contracts held by hedgers.

$$HP_t = Short_t - Long_t$$

And therefore, the percent net hedging position will be defined as follows:

$$HP_t(\%) = \frac{Short_t - Long_t}{Long_t + Short_t}$$

In the context of this study, the latter one will be used for all calculations. Therefore, positive values will indicate a higher volume of short positions compared to long positions.

All definitions, however, have one thing in common. It considers the positions held by hedgers. Thus, the next question naturally arises: “Who exactly are considered hedgers?”

Traditionally, hedgers in the financial markets have been defined as those participants who use derivatives and other financial instruments to mitigate the risk of adverse price movements in the assets they deal with. These participants primarily consisted of commercial undertakings, such as manufacturers and producers, who hedge to manage the price risks associated with the commodities they use or produce. For instance, a manufacturing company might hedge against the price volatility of metals required for production, while an agricultural producer might hedge against fluctuations in crop prices. Producers are naturally long in their own commodities due to their production activities. They may use short positions in futures or options markets to hedge against price risks and lock in favorable prices for future production. For example, a wheat farmer is naturally long in wheat because they grow and sell it. They benefit from higher prices because it directly affects their revenue. Manufacturers may take a long position in commodities when they anticipate needing a particular raw material or commodity in the future for production. By going long, they can lock in a favorable price now to hedge against potential price increases in the future.

Regardless, the hedging group of commercial undertakings has historically been short because they often hedge their price risk by taking positions opposite to their physical exposure in the commodity market. This allows them to protect their revenues or costs from adverse price movements.

Most studies in the field of commodities and futures markets have focused on examining exclusively the impact of commercial undertakings' short positions on the overall dynamics and revenues of the futures market, assuming they play the central role as the primary hedgers. While it is true that commercial undertakings have traditionally been the primary participants using these markets for hedging purposes, in recent years, with the financialization of markets and structural changes, it has become necessary to consider the impact of hedging activities by other investor groups, their motivations, and their influence on revenues in the zinc futures market. The detailed classification of trading members reflected in the Commitments of Traders Report (COTR) will be crucial to address these dynamics that will be examined.

COTR CLASSIFICATION OF MEMBERS

As previously explained, the formal classification of Members and Clients is the following:

a) **Investment firms or credit institutions:** Investment firms, also known as asset management companies or investment management firms, specialize in managing investments on behalf of their clients, which include individuals, corporations, and institutions. Their purpose is to create and manage investment portfolios tailored to the goals, risk tolerance and time horizon of their clients, but they also offer financial advice, and execute trades on behalf of their clients, which include individuals, corporations, and institutions. Credit institutions, commonly referred to as banks or financial institutions, primarily focus on providing credit and other financial services to individuals, businesses, and governments. Investment firms and credit institutions are integral to the financial system, each offering distinct services that support individuals, businesses, and the broader economy.

b) **Investment funds:** Investment funds in the LME COTR include entities that manage pooled funds from multiple investors, which are then used to trade metals contracts on the exchange. These funds allow individual investors to access a diversified portfolio of assets, which might be difficult to achieve on their own due to limited capital or expertise. These funds are usually strategically diversified to reduce the risk associated to individual securities, they provide a professional management to make informed decisions and offer high liquidity and accessibility. The most common types of investment funds are: Mutual funds (both open- and closed-end funds), Exchange-Traded Funds (ETFs), Hedge funds, Index funds, Money Market funds, Bond funds and Real Estate Investment Trusts (REITs).

While both investment funds and investment firms participate in trading activities on the LME, they differ in their structures, roles, and classifications within the LME COTR report based on whether they are managing pooled investments (funds) or providing investment services (firms). The main differences are the following 3:

1. Nature: Investment funds are pooled investment vehicles managed collectively to achieve specific investment objectives, while investment firms are professional entities that provide investment management and advisory services to clients.

2. Function: Investment funds primarily manage pooled assets and directly trade in the market, whereas investment firms manage client portfolios and execute trades on behalf of clients.

3. Regulation: Both investment funds and investment firms are subject to regulatory oversight, but investment funds are often regulated based on the type of fund structure (e.g., mutual funds regulated by SEC in the US), while investment firms are regulated as financial service providers.

c) **Other financial institutions:** In the context of the LME (London Metal Exchange) Commitments of Traders Report (COTR), "Other financial institutions" refers to a specific category of market participants that typically includes financial entities other than commercial traders and non-commercial traders. This category may include investment banks, pension funds, insurance companies or sovereign wealth funds. These institutions play a crucial role in the commodities market by providing liquidity, conducting arbitrage, managing risk, and seeking profit opportunities through trading activities on the LME.

d) **Commercial undertakings:** This refers to the entities that engage in trading a ctivities on the exchange primarily for commercial purposes related to their business operations. These entities are classified based on their involvement in the production, processing, or consumption of metals (zinc in this case), as opposed to trading for speculative purposes or managing investments on behalf of clients. There are 3 types of entities in this group: producers (companies involved in extracting a processing zinc), consumers (manufacturers and industrial companies that use raw zinc in their production processes) and merchants (entities that are involved in buying, selling and distributing the zinc in the physical market). Commercial undertakings use futures contracts on the LME to hedge against adverse price movements that could impact their profitability or operational costs. For example, a zinc producer might hedge against falling zinc prices by taking a long position in zinc futures.

FACTORS INFLUENCING HEDGING PRESSURE AND FUTURE PRICES ON THE LME

Discussion on the factors influencing hedging pressure and future prices on the London Metal Exchange (LME) encompasses various elements that drive market dynamics. Some key factors to consider are:

1. Supply and Demand Fundamentals: The fundamental factors of supply and demand for metals play a significant role in shaping hedging pressure and future prices on the LME. Changes in global economic conditions, industrial production, infrastructure investment, and technological advancements can impact demand for metals. Similarly, factors such as supply disruptions, geopolitical tensions, and regulatory changes can influence supply dynamics. Fluctuations in supply-demand fundamentals can lead to shifts in hedging strategies and price expectations among market participants.

2. Macroeconomic Indicators: Macroeconomic variables such as interest rates, inflation, exchange rates, and overall economic growth have implications for metal prices and hedging pressure on the LME. For instance, changes in interest rates can affect the cost of carrying inventories and financing hedging positions, influencing market participants' hedging decisions. Exchange rate movements can also impact metal prices by affecting the competitiveness of exports and imports.

3. Market Sentiment and Speculation: Market sentiment and speculative trading activities can exert significant influence on hedging pressure and future prices on the LME. Speculators,

including hedge funds, commodity trading advisors, and proprietary trading firms, often engage in futures trading to capitalize on short-term price movements or hedge against risk. Changes in investor sentiment, risk appetite, and market expectations can lead to fluctuations in hedging pressure and contribute to price volatility.

4. Inventory Levels and Storage Costs: Inventory levels in LME-approved warehouses and associated storage costs can affect hedging pressure and future prices. High inventory levels may exert downward pressure on prices as market participants anticipate oversupply, leading to increased hedging activity to protect against price declines. Conversely, low inventory levels or supply disruptions may trigger hedging to guard against potential shortages and price spikes (Cootner, 1967).

5. Regulatory Environment: Regulatory policies and market regulations can influence hedging pressure and future prices on the LME. Changes in position limits, margin requirements, and reporting obligations may impact the behavior of market participants, affecting hedging strategies and market dynamics. Regulatory interventions aimed at addressing market manipulation, fraud, or systemic risks can also influence investor confidence and market efficiency.

6. Technological Advances and Market Structure: Advances in technology, such as electronic trading platforms and algorithmic trading strategies, have transformed the structure of commodity markets, including the LME. High-frequency trading, algorithmic execution, and order-routing algorithms have increased market liquidity and efficiency but also introduced new sources of volatility. Understanding the implications of technological innovations and market structure changes is crucial for analyzing hedging pressure and future prices on the LME.

RELATIONSHIP BETWEEN HEDGING PRESSURE AND HISTORICAL RETURNS

The relationship between three-month future hedging pressure and historical returns can be complex and multifaceted. As previously explained, hedging pressure in futures markets refers to the imbalance between hedgers' short and long positions. This is often quantified as the net position of hedgers (short minus long positions) relative to the total open interest in the futures market. Historical returns are the past performance of an asset, zinc in this case, typically measured as percentage changes in price over specific periods.

According to traditional theory, when there is a high demand for hedging (more short positions by hedgers), it suggests that hedgers are willing to pay a premium to transfer risk. This risk premium is then captured by speculators who take the long positions. Consequently, futures prices might be lower than expected future spot prices, leading to higher future returns.

Many studies have shown that there is often a negative relationship between hedging pressure and future returns. When hedging pressure is high (more short positions), future returns tend to be higher as the market adjusts for the increased risk premium. Conversely, low hedging pressure (more long positions) might indicate lower future returns.

Hedgers' risk aversion can drive the demand for futures contracts. In other words, high hedging pressure indicates greater demand for hedging risk, which speculators accommodate, expecting higher returns as compensation. Studies in other various futures markets (commodities, financial futures) generally support the notion that hedging pressure is a predictor of future returns. For example, the Commodity Futures Trading Commission (CFTC) data often reveals this relationship in commodity futures. If true, this will have a very important practical implication because understanding the relationship between hedging pressure and historical returns can help in devising trading strategies. For example, traders might use hedging pressure as a signal for taking long or short positions in futures markets. On the other hand, firms and investors might adjust their risk management practices based on anticipated returns associated with varying levels of hedging pressure.

POSSIBLE PREDICTIVE INDEXES

The following indexes are traditionally known to be predictive due to their ability to capture underlying economic conditions, market sentiment, and risk perceptions, which directly influence hedging behaviors. They often act as leading indicators, providing early signals about future economic and market conditions, but they also reflect investor behavior and expectations and help in assessing risk levels in the market, which are essential for predicting the risk premium and future returns. The selected indexes for this research are the following:

a. **VIX (Volatility Index):** Whaley (Whaley, 2009) brought up the definition of VIX. VIX was seen as a forward-looking record of the normal return unpredictability of the US S&P 500 Index over the course of the following 30 days. Today, the VIX index is also known as the fear index, and it is represented in Figure 2. The VIX index tends to spike sharply during periods of high uncertainty, such as the 2007 financial crisis or the 2020 COVID-19 pandemic. The VIX index is a real-time measure of the market's expectations for volatility over the next 30 days. A high VIX indicates that investors expect significant price fluctuations, reflecting heightened uncertainty and risk aversion. Empirical studies often find that higher VIX levels are associated with higher future returns on assets, as the increased demand for hedging during volatile periods results in a risk premium.

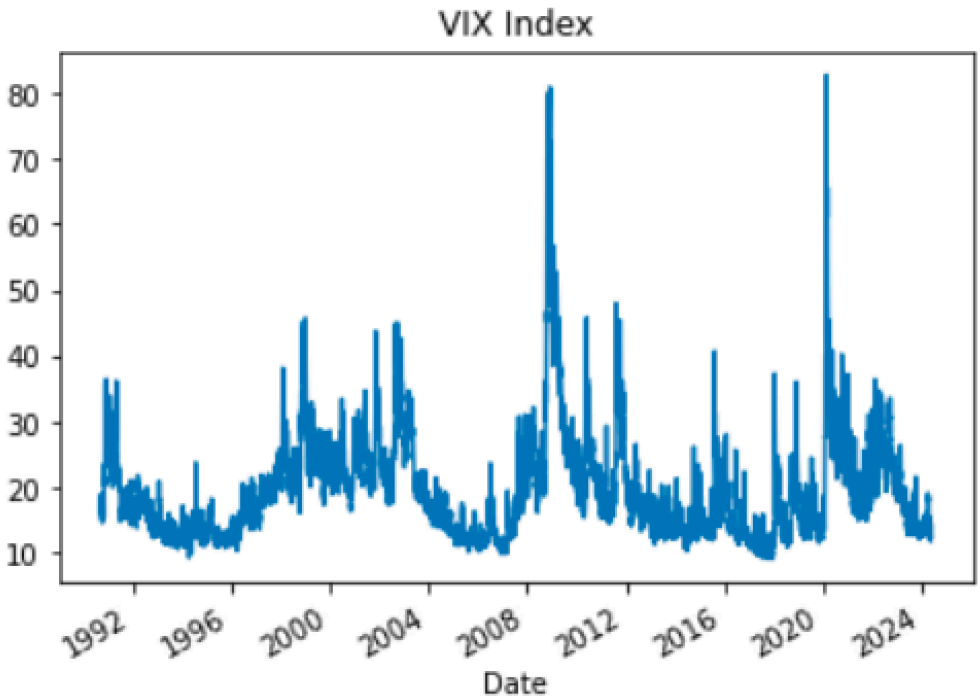

Figure 7: VIX Index

b. **PMI (Purchasing Managers' Index):** The PMI is an indicator of the economic health of the manufacturing and service sectors, with values above 50 indicating expansion and below 50 indicating contraction. PMI is based on a monthly survey of supply chain managers across 19 industries, covering both upstream and downstream activity. A higher PMI reflects positive business sentiment and economic growth. During such periods, companies might hedge less aggressively, reducing hedging pressure.

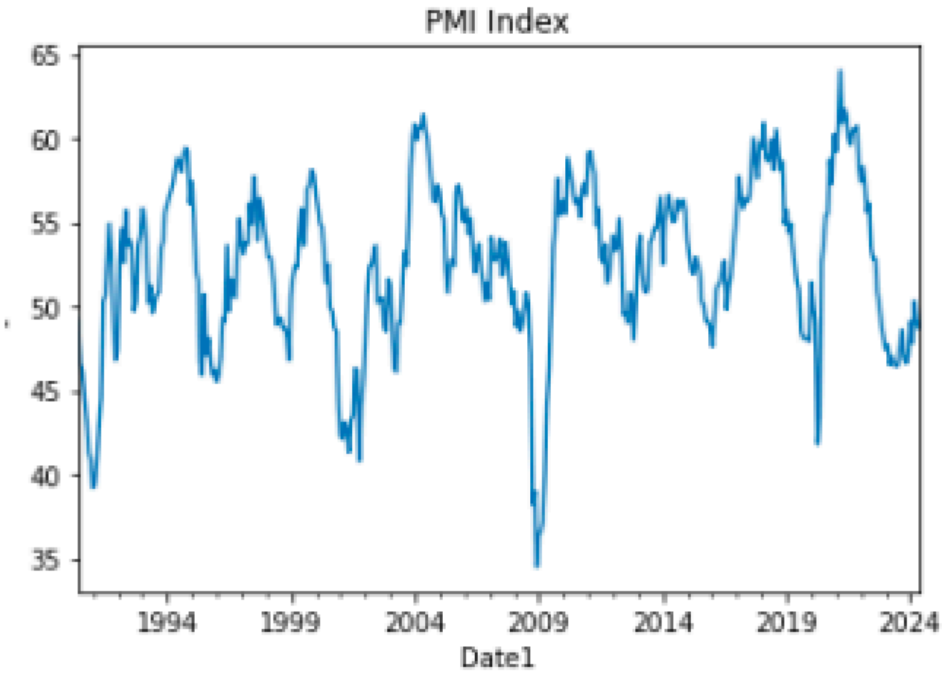

Figure 8: PMI Index

c. **BDI (Baltic Dry Index):** The BDI monitors rates for ships carrying dry bulk commodities such as coal, iron ore, and grain, and is an indicator of global economic activity. A higher BDI indicates robust global trade and economic activity, which might reduce the need for hedging by commodity producers and traders due to favorable market conditions, leading to lower hedging pressure.

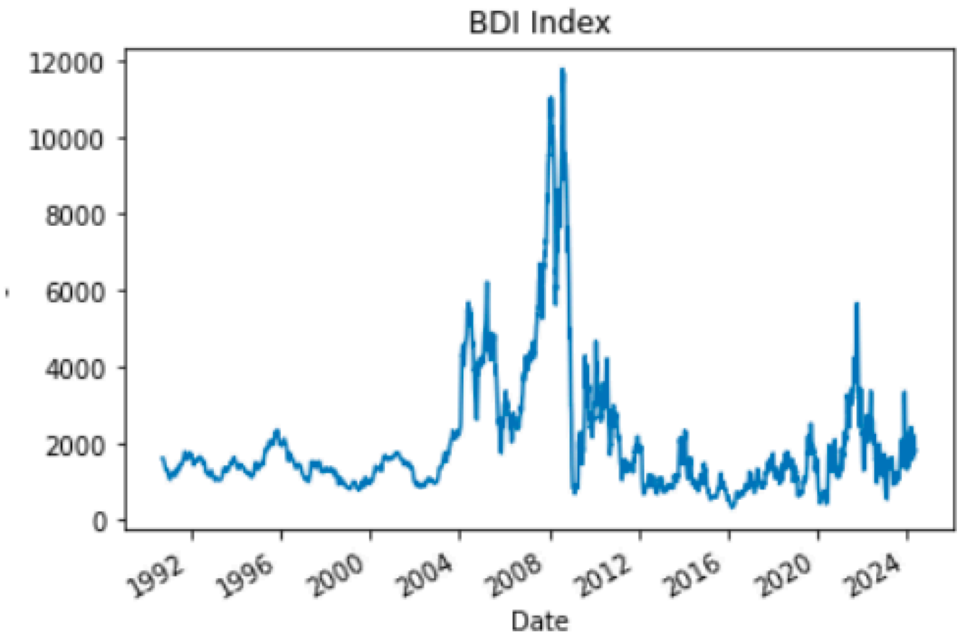

Figure 9: BDI Index

d. **SPX (S&P500 Index):** The SPX is a stock market index that measures the performance of 500 large companies listed on stock exchanges in the United States. The SPX often reflects broader market trends and investor sentiment. During bull markets, hedging pressure might decrease as confidence in rising stock prices increases, potentially leading to lower future returns. In bear markets, hedging pressure typically increases as market participants seek to protect against downside risk, leading to higher future returns due to the increased risk premium.

Figure 10: SPX Index

4. Data Collection and Processing

COMMITMENT OF TRADERS REPORT DATA RETRIEVAL

London Metal Exchange (LME) publishes the Commitment of Traders (COT) report, which provides a weekly snapshot of the positions held by traders in the zinc market. This report is a crucial resource for market participants, as it reveals the levels of long and short positions in zinc futures contracts. The COT data is widely used by traders, analysts, and researchers to gauge market sentiment and to make informed trading decisions.

It categorizes traders into several groups, such as commercial undertakings or investment funds. Each category provides insight into the behavior of different market participants. To effectively utilize COT data, it is essential to systematically collect and process the data. The data has been downloaded from Bloomberg and it is as recent as May 2023. The data has been preprocessed, cleaned and checked to use it in the main program of this project. The program has been developed in the programming language Python, in a Spider environment.

As this research is focused on the hedging process, the data for the "Risk Redcution" column of Figure 1 will be downloaded. For each participant, both long and short positions history will be downloaded. Additionally, the weekly difference of the values will be computed, to see the changes in each variable from week to week. Finally, the hedging pressure of the variable will be calculated as the total short position minus total long position, and it will be depicted.

5. DATA ANALYSIS AND RESULTS

ZINC COTR RISK REDUCTION OPEN INTERESTS

Commercial Undertakings Open Interest Risk Reduction

As previously explained in detail, commercial undertakings, such as manufacturers and producers, are those who hedge to manage the price risks associated with the commodities they use or produce. Open interest is the number of active contracts that are still open and have not been offset by an opposing transaction. Therefore, the following figures, show the quantity of futures contracts that are both long (participants that anticipate a price increase) or short (those who anticipate a price decrease).

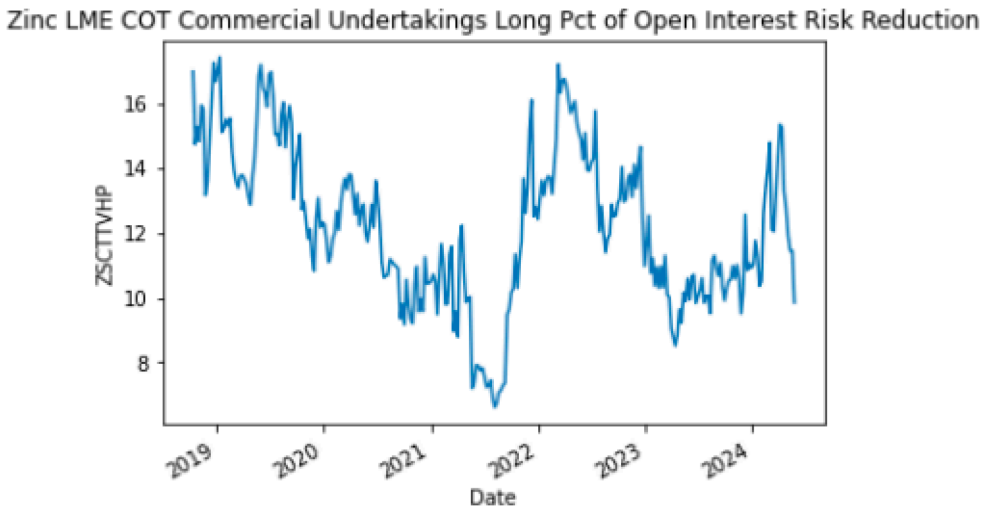

Figure 11: Commercial Undertakings Long Percentage Open Interest Risk Reduction

Additionally, descriptive statistics have been conducted on the time series. In a subsequent section, a regression analysis will be performed, which requires ensuring that all time series used are stationary.

The results produced by the commercial undertakings long percentage of open interest are the following:

ADF Statistic	-2.325237
p-value	0.163958
Critical Values:	
1%	-3.453
5%	-2.872
10%	-2.572

Figure 12: ADF Commercial Undertakings Long Percentage

The performed Augmented Dickey-Fuller Test is used to test for the presence of a unit root in a time series sample. This unit root indicates that the time series is non-stationary, meaning it has some form of trend or seasonality.

In this case, the ADF statistic value is -2.325, which is greater that the critical values at any significance level. Therefore, it is not possible to reject the null hypothesis. Additionally, the p-value is 0.163958, which is greater than 0.05, indicating that the evidence against the null hypothesis is not strong.

The conclusion is that this time series is likely non-stationary, as there is not enough evidence to reject the null hypothesis of a unit root. For making the time series stationary, the weekly difference variable is used. We can confirm from the partial autocorrelation graph shown in Figure 13 that this differencing will probably be enough to make it stationary.

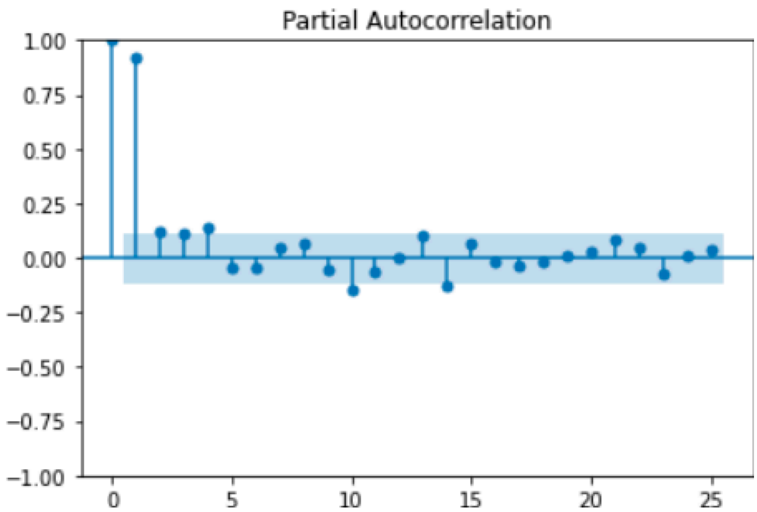

Figure 13: PACF of Commercial Undertakings Long Percentage

In Figure 14, the weekly difference of the previous variable is shown. Visually, this graph appears much more stable with less apparent trend. However, the same statistical examination will be conducted.

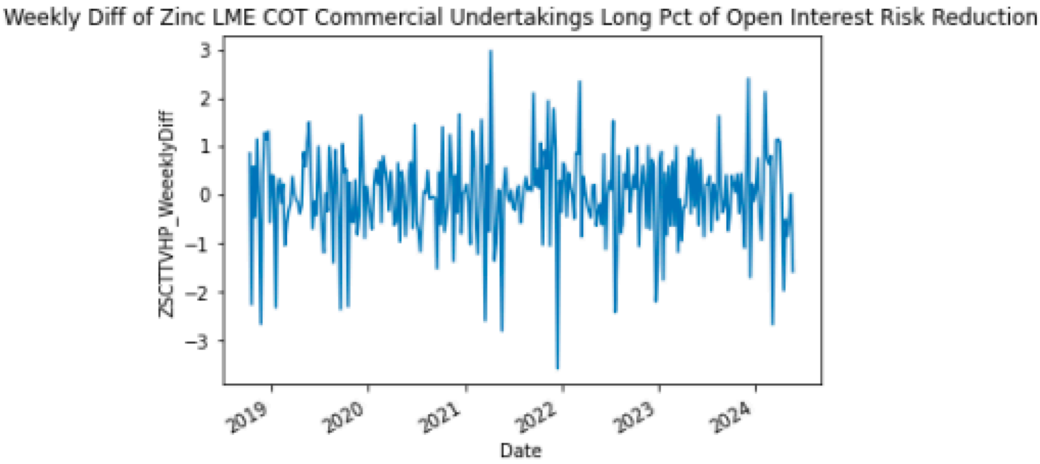

Figure 14: Weekly difference of Commercial Undertakings Long Percentage Open Interest Risk Reduction

The results of the Augmented Dickey-Fuller Test (ADF) are shown in the following table:

ADF Statistic	-12.821685
p-value	0
Critical Values:	
1%	-3.453
5%	-2.872
10%	-2.572

Figure 15: ADF test results for Commercial Undertakings Long percentage weekly difference

A ADF value of -12.821, which is less than any of the values at any significance level, indicates that the null hypothesis can be rejected. The p-value also indicates strong evidence against the null hypothesis. This suggests that after differencing, the series no longer has trends, seasonality, or other non-stationary behavior, making it suitable for further time series analysis or modeling.

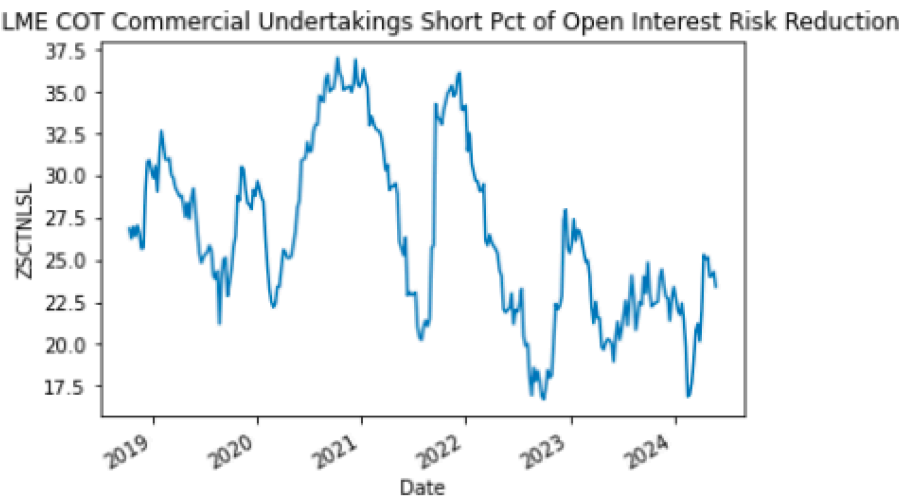

Figure 16: Commercial Undertakings Short Percentage Open Interest Risk Reduction

The same procedure is done for the Commercial Undertakings Short percentage open interest that is represented in Figure 16. Clearly, the time series is not stationary, but the Augmented Dickey-Fuller test is conducted.

Very similarly to the long percentage open interest, the ADF statistic (-2.022) has a value greater than any of the critical values, meaning that the null hypothesis will not be rejected. Additionally, the p-value is greater than 0.05, indicating that the unit root hypothesis should not be rejected.

ADF Statistic	-2.021848
p-value	0.277076
Critical Values:	
1%	-3.453
5%	-2.871
10%	-2.572

Figure 17: ADF Commercial Undertakings Short Percentage

Therefore, the time series is differenced to achieve a stationary time series. Figure 18 shows the Weekly Difference of Commercial Undertakings short percentage of open interest, which looks more stationary than the previous one. The Augmented Dickey-Fuller test is conducted on this time series, giving the results on Figure 19.

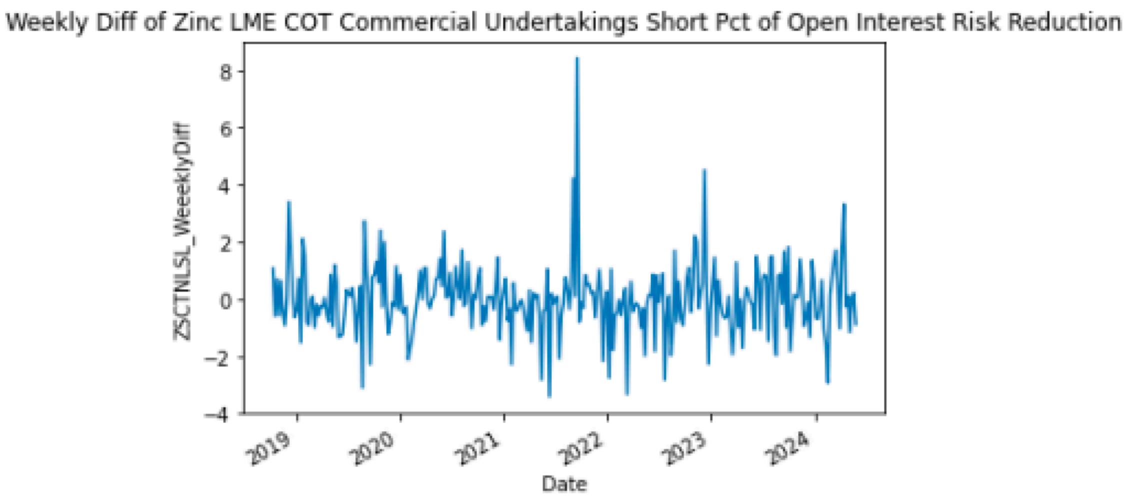

Figure 18: Weekly difference of Commercial Undertakings Short Percentage Open Interest Risk Reduction

ADF Statistic	-12.821685
p-value	0
Critical Values:	
1%	-3.453
5%	-2.871
10%	-2.572

Figure 19: ADF test results for Commercial Undertakings Short percentage weekly difference

The ADF shows a value of -12.821, which is less than any of the values at any significance level, indicates that the null hypothesis can be rejected. The p-value is 0, indicating strong evidence against the null hypothesis and meaning that the time series is highly probable stationary.

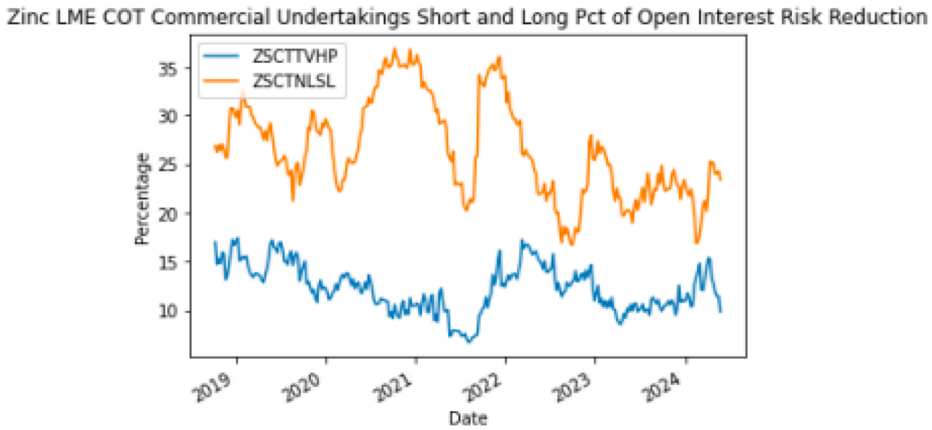

Figure 20: Comparison between short and long positions for Commercial Undertakings

The comparison between both long and short positions is depicted in Figure 20. In this graph, it is clear that the short position percentage of open interest is higher than that of the long position. This might be because commercial entities have a significant amount of the commodity or asset in inventory or expected future production, so they may take short positions to protect against price declines.

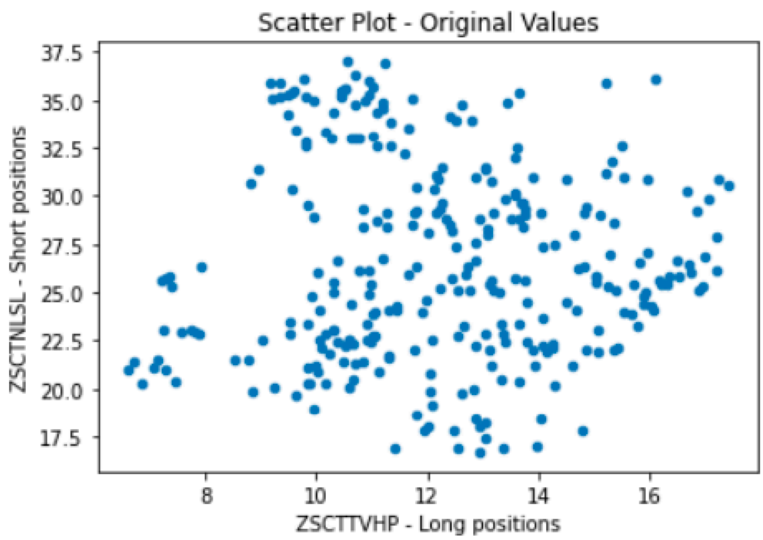

Figure 21: Scatter Plot for Long and Short Positions of Commercial Undertakings

The scatter plot is not very informative in this case. There is no clear trend (dots appear randomly scattered), which might suggest a lack of a consistent hedging strategy. The correlation coefficient for these two variables is 0.007138, very close to 0, which suggests that there is almost no linear relationship between the two variables. A value this close to zero indicates that

changes in one variable are not linearly associated with changes in the other variable. This lack of a strong linear relationship suggests that the movements in long and short positions are not predictable based on each other, at least in a linear sense.
Entities could be taking opportunistic positions based on market conditions. However, when we make a scatter plot of the weekly difference values, there is a clearer pattern. The data points are somewhat clustered in the center around a zero difference on both axes, indicating that the weekly differences between the long positions and short positions tend to be relatively small. There is a slight positive correlation between the long and short positions. This means that when the weekly difference for the long positions is positive (meaning the long position increased), the weekly difference for the short positions is also likely to be positive (meaning the short position increased as well). This could be a sign that commercial hedgers are increasing both their long and short positions proportionately. Additionally, the correlation coefficient has a value of 0.269265 is closer to 0 than to 1, suggesting a weak but positive linear relationship between the two variables. The weak positive correlation implies that there is a slight tendency for the weekly changes in the two variables to move together. However, this relationship is not strong enough to be relied upon heavily for predictive purposes.

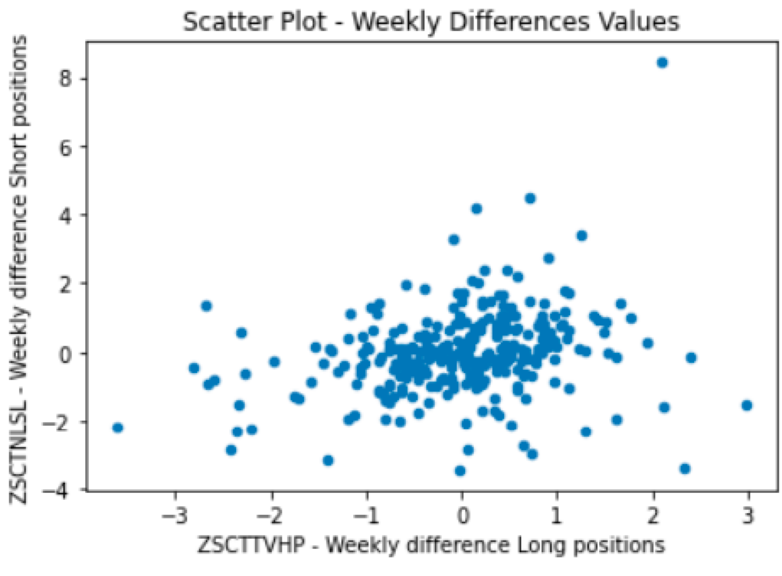

Figure 22: Scatter Plot for the Weekly Differences of Long and Short Positions of Commercial Undertakings

Overall, the scatter plot suggests that commercial hedgers are using a variety of strategies to manage their risk in the zinc market. Some hedgers may be more focused on locking in profits by selling short contracts while simultaneously buying long contracts. Others may be more focused on minimizing risk by entering into spread contracts. The data also suggests that the hedging strategies of commercial hedgers may vary from week to week.

The Net Hedging Pressure is calculated by subtracting long positions from short positions for each day. The resultant graph is depicted in Figure 23. Hedging pressure is clearly very positive, indicating that short positions significantly exceed long positions. As previously mentioned, this phenomenon can be explained by commercial entities holding a substantial amount of the commodity or asset in inventory or anticipating future production, leading them to take short positions to protect against price declines.

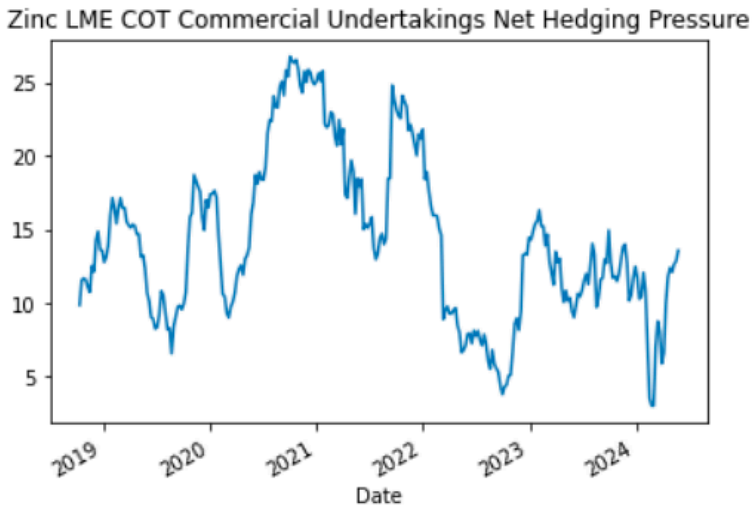

Figure 23: Commercial Undertakings Net Hedging Pressure

Finally, the descriptive statistics of both the long and short position variables and their weekly differences. The mean value for short positions (26.50) is significantly higher than the mean for long positions (12.23), indicating a general bias towards holding short positions. The standard deviation for the short percentage of open interest is 5.10 which suggests that the data points for short positions are more spread out or varied compared to long positions, which have a standard deviation of 2.43.

	ZSCTTVHP	ZSCTTVHP_WeeklyDiff
Count	294	294
Mean	12.223163	-0.021395
Std	2.429201	0.914951
Min	6.6	-3.61
Max	17.42	2.97
Kurtosis	-0.52080	1.553717
Skewness	0.03386	-0.459612

	ZSCTNLSL	ZSCTNLSL_WeeklyDiff
Count	294	294
Mean	26.508231	-0.008061
Std	5.102458	1.22844
Min	16.69	-3.43
Max	37	8.44
Kurtosis	-0.86457	8.050290
Skewness	0.27635	1.205619

Figure 24: Descriptive Statistics of long and short positions and their weekly differences for Commercial Undertakings

Investment Firms or Credit Institutions Open Interest Risk Reduction

Investment firms are entities that manage investments on behalf of their clients or for their own portfolios. They include a variety of institutions such as hedge funds or mutual funds. Credit institutions are financial entities primarily involved in lending and other credit-related activities but also engage in trading financial instruments like zinc futures. These might include commercial or investment banks.

As depicted in Figure 25, the long position percentage held by investment firms and credit institutions is lower compared to that of commercial undertakings. This discrepancy can be attributed to the differing primary objectives of these market participants. Unlike commercial undertakings, which primarily engage in hedging to manage risk related to their physical holdings of zinc, investment firms and credit institutions often prioritize speculative activities and market making.

Investment firms, including hedge funds, mutual funds, and asset management companies, seek to profit from price movements in zinc futures. They use sophisticated trading strategies, including leveraging and short-selling, to capitalize on market trends. Their speculative nature means they are more likely to take short positions when they anticipate a decline in zinc prices, contributing to lower long position percentages.

Credit institutions, such as commercial and investment banks, also participate in the zinc futures market primarily for purposes other than hedging. They provide liquidity through market making, offering buy and sell quotes to facilitate trades for other market participants. This role is crucial for ensuring an efficient and liquid market but does not necessarily involve taking significant long positions.

Furthermore, these financial entities engage in arbitrage opportunities, exploiting price discrepancies between different markets or instruments. This activity can lead to more balanced positions rather than a heavy tilt towards long positions. Additionally, their involvement in providing client services, including trade execution and risk management, means they often hold positions that mirror the needs of their clients, which can vary widely.

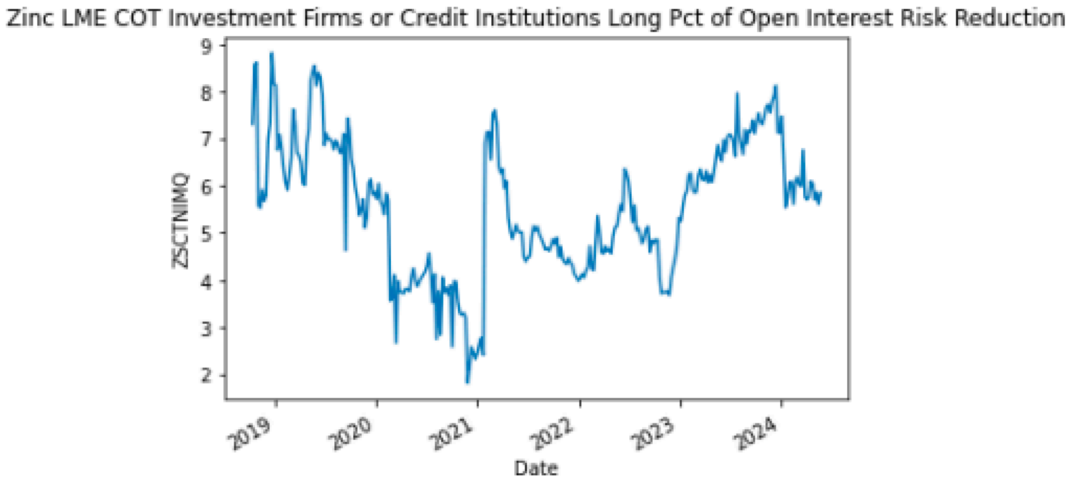

Figure 25: Investment Firms or Credit Institutions Long Percentage of Open Interest Risk Reduction

The Augmented Dickey-Fuller test has also been conducted in this time series in order to check its stationarity. The results can be seen in Figure 26.

ADF Statistic	-3.296754
p-value	0.015024
Critical Values:	
1%	-3.453
5%	-2.871
10%	-2.572

Figure 26: ADF Investment Firms or Credit Institutions Long Percentage

In this case, because the ADF statistic (-3.297) is more negative than all the critical values (-3.453, -2.871, -2.572), we can reject the null hypothesis at all significance levels (1%, 5%, and 10%). This means that the time series is likely stationary. The low p-value also suggest that this is very likely true.

Figure 27: Weekly difference of Investment Firms or Credit Institutions Long Percentage Open Interest Risk Reduction

The weekly difference of the long percentage open interest is graphed in Figure 27. Most values are around the zero mark, meaning there is little variability on the weekly difference.

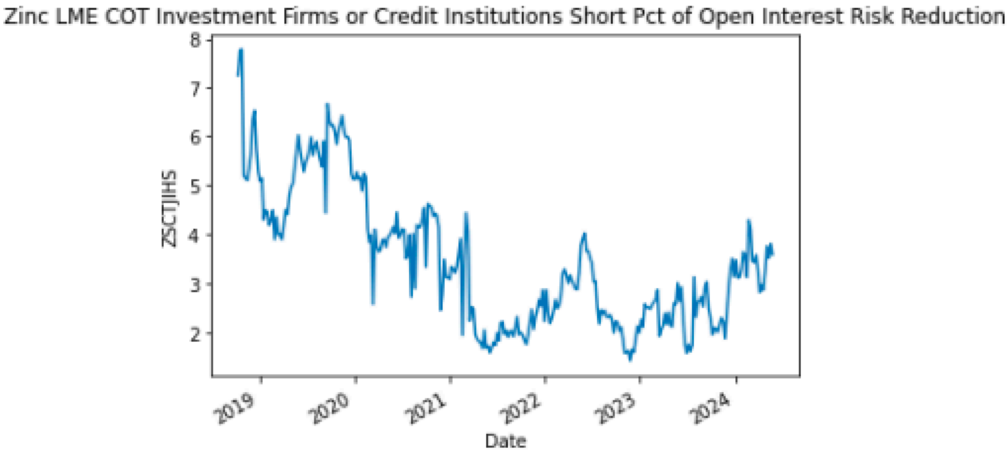

Figure 28: Investment Firms or Credit Institutions Short Percentage of Open Interest Risk Reduction

For the short percentage of open interest, there has been a clear decline over the past 4 years. This decline could be due to traders who were previously betting on zinc prices to fall (holding short positions) might be changing their minds. This could be due to positive news about future zinc demand, production disruptions, or other factors that could lead to a price increase.

Investment firms might reduce their short positions on zinc futures due to the consequences of COVID-19, which could have decreased supply. The pandemic has disrupted mining operations, supply chains, and production processes globally, leading to lower availability of zinc. Anticipating higher future prices due to these supply constraints, firms may reduce their short positions to mitigate potential losses.

ADF Statistic	-3.342576
p-value	0.013079
Critical Values:	
1%	-3.453
5%	-2.872
10%	-2.572

Figure 29: ADF Investment Firms or Credit Institutions Short Percentage

The p-value is the probability of observing a test statistic as extreme as the one calculated, assuming the null hypothesis (that the time series is not stationary) is true. In this case, the p-value is 0.013, which is less than all the standard significance levels (1%, 5%, and 10%). This means that we reject the null hypothesis. In other words, there is enough evidence to conclude that the time series is likely stationary.

We take the weekly difference, resulting in a very spiky graph, as seen on Figure 30, indicating great variability due to the decrease on the short positions over the last years. The Augmented Dickey-Fuller Tests shows the absolute stationarity of the differenced time series. Figure 31 shows an ADF statistic of -15.498, which means that the series is very likely stationary.

Figure 30: Weekly difference of Investment Firms or Credit Institutions Short Percentage Open Interest Risk Reduction

ADF Statistic	-15.498494
p-value	0
Critical Values:	
1%	-3.453
5%	-2.871
10%	-2.572

Figure 31: ADF test results for Investment Firms or Credit Institutions Short percentage weekly difference

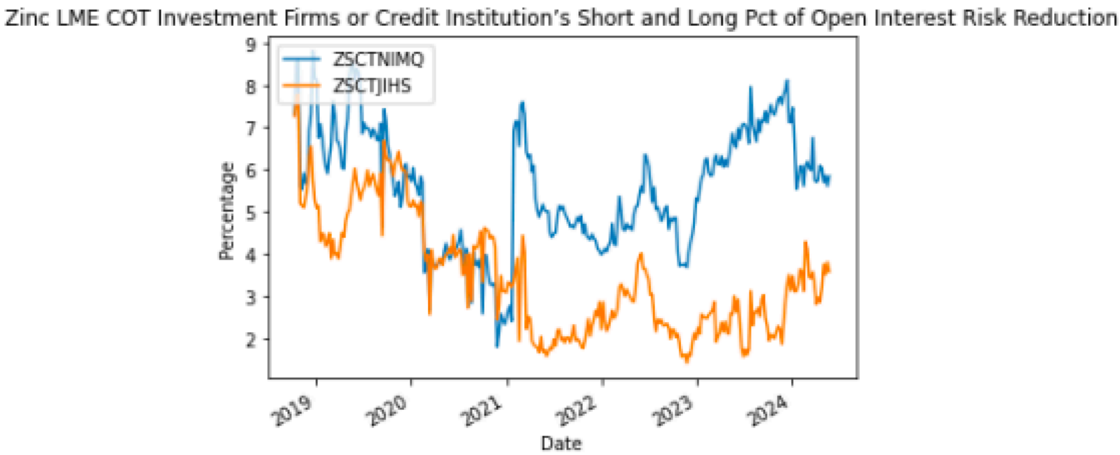

Figure 32: Comparison between short and long positions for Investment Firms or Credit Institutions

In Figure 32, there is a comparison between long and short positions. It is clear that until 2021, both long and short positions moved nearly identically, with both percentages fluctuating between 3% and 8%. From 2021 on, a divergence between long and short positions becomes apparent. The percentage of long positions (blue line) increased significantly, reaching peaks above 8% at various points. Conversely, the percentage of short positions (orange line) generally decreased and stabilized around 4%, with slight fluctuations.

The graph indicates a strategic shift among investment firms or credit institutions towards holding more long positions relative to short positions from 2021 onwards, suggesting a more bullish outlook on Zinc prices in the LME during this period.

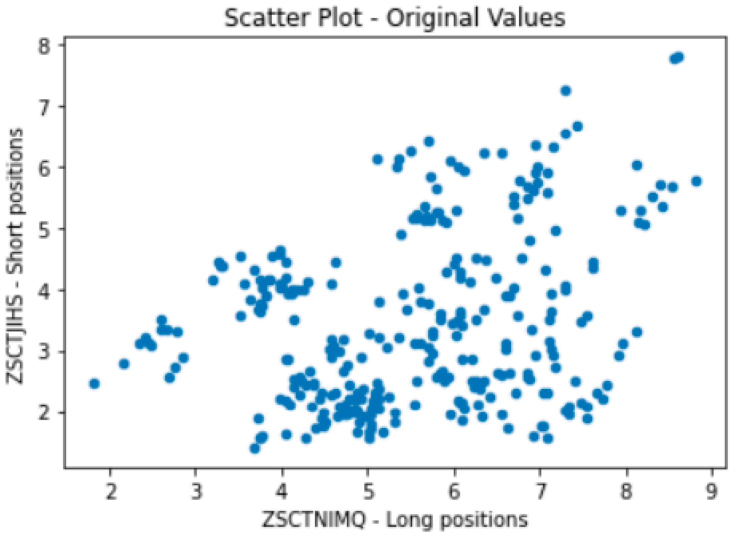

Figure 33: Scatter Plot for Long and Short Positions of Investment Firms or Credit Institutions

The scatter plot shown in Figure 33, shows that there is a noticeable positive correlation between long and short positions. As the number of long positions increases, the number of short positions tends to increase as well, although the relationship is not perfectly linear. The correlation coefficient between the long and short positions has a value of 0.300917 is closer to 1 than to 0, suggesting a moderate positive linear relationship between the two variables. This indicates that as one variable increases, there is a moderate tendency for the other variable to also increase. The relationship is not strong but is noticeable. The plot shows a dense cluster of points around the 4% to 5% range for both long and short positions, suggesting that this range is common for both types of positions among investment firms or credit institutions.

Overall, the scatter plot reinforces the observation from the previous line graph that there is a relationship between long and short positions, but it also highlights the variability and occasional divergence between the two types of positions.

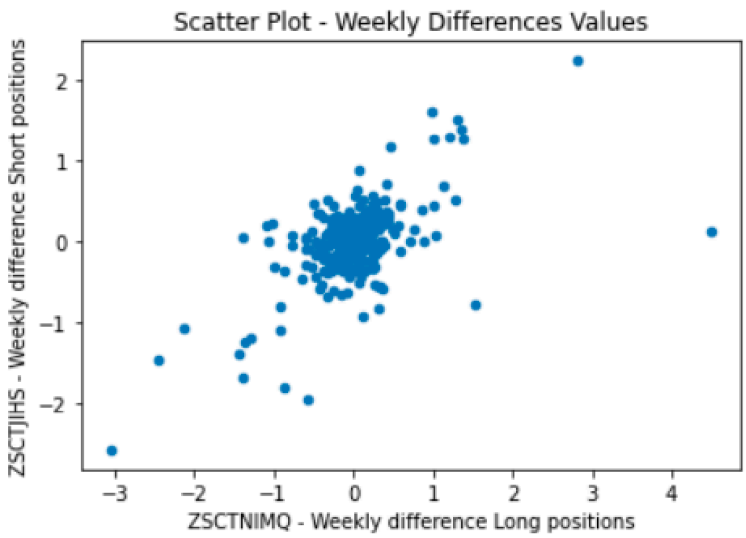

Figure 34: Scatter Plot for the Weekly Differences of Long and Short Positions of Investment Firms or Credit Institutions

When making the weekly difference of the time series, most data points are concentrated around the origin (0, 0), indicating that many weeks see little to no change in either long or short positions. This plot highlights that while there is a general tendency for long and short positions to change together on a weekly basis, there are periods of significant variability and outliers that deviate from this trend, especially, and as previously seen, from 2021 on. The correlation coefficient of the weekly difference is of 0.599776, which is closer to 1 than to 0, indicating a moderate to strong positive linear relationship between the weekly differences of the two variables. Traders could consider using changes in one variable's weekly difference to predict or inform decisions about changes in the other variable's weekly difference.

Figure 35: Investment Firms or Credit Institutions Net Hedging Pressure

In Figure 35, we can see the net hedging pressure graph. The hedging pressure is calculated as the difference between the number of short positions and long positions that investment firms or credit institutions have held over the past year. Therefore, a negative hedging pressure means that there are more long positions than short positions (net bullish pressure). This is supported by the descriptive statistics seen on Figure 36. Long positions have a higher mean than short ones, suggesting a general tendency for holding more long positions than short. ones. Additionally, both long and short positions exhibit relatively similar standard deviations, indicating a comparable level of volatility.

As previously explained, the bullish sentiment started in 2021, when the long positions of the investors started to rise. One of the most logic explanations is that the pandemic caused disruptions in mining operations and global supply chains. Reduced supply, combined with recovering demand, created a favorable environment for higher zinc prices, prompting financial firms to take longer positions. Also, positive economic indicators from China, such as higher industrial production or stronger GDP growth, could signal increased demand for zinc, as China is a major consumer of industrial metals. The market's reaction to such news is a testament to how dynamic and sensitive the positioning of institutional traders can be, driven by both macroeconomic indicators and specific market events.

	ZSCTNIMQ	ZSCTNIMQ_WeeklyDiff
Count	294	294
Mean	5.535782	-0.006701
Std	1.432183	0.589688
Min	1.81	-3.04
Max	8.82	4.48
Kurtosis	-0.54789	15.852487
Skewness	-0.06420	1.028531

	ZSCTJIHS	ZSCTJIHS_WeeklyDiff
Count	294	294
Mean	3.452041	-0.010918
Std	1.384861	0.477458
Min	1.41	-2.59
Max	7.79	2.24
Kurtosis	-0.28119	6.961386
Skewness	0.71172	-0.560579

Figure 36: Descriptive Statistics of long and short positions and their weekly differences for Investment Firms or Credit Institutions

Investment Funds Open Interest Risk Reduction

Investment funds in the context of LME (London Metal Exchange) zinc futures are entities that pool capital from multiple investors to invest in zinc futures contracts. These funds can include hedge funds, mutual funds, pension funds, and other types of investment vehicles. These entities have several key motivators for trading zinc on the LME futures market, such as, risk management and portfolio protection, speculation, diversification, taking advantage of arbitrage opportunities or liquidity provision.

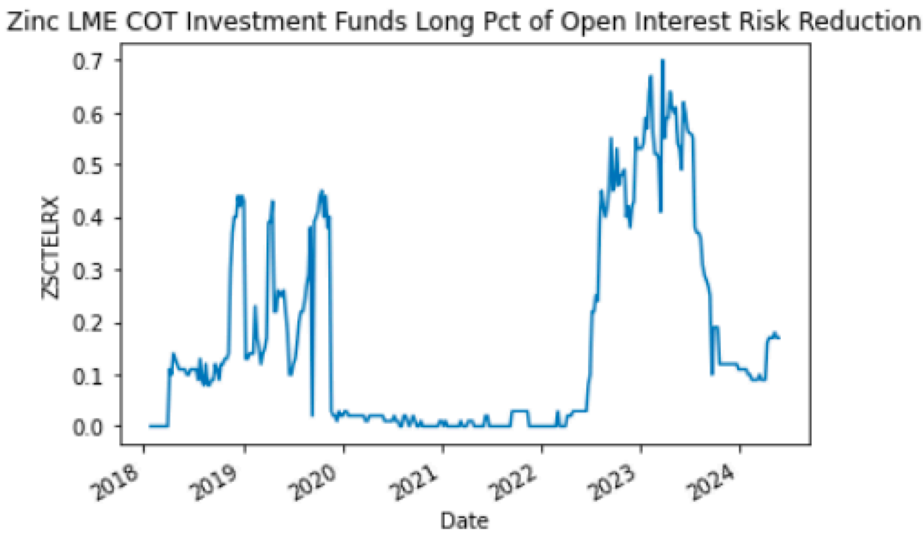

Figure 37: Investment Funds Long Percentage of Open Interest Risk Reduction

In Figure 37, we see the long positions taken by Investment Funds from 2018 until 2024. It is remarkable that from 2020 to mid 2022, the long percentage of open interest declined almost to zero. The explanation to this could be Covid-19. The pandemic has created economic uncertainty, affecting investor sentiment in various ways. While some investor groups view the pandemic as an opportunity to take a bullish position due to reduced zinc supply and anticipated recovery, others perceive it as a time of heightened risk and downturns. This divergence in perception leads some investors to reduce long positions to mitigate risk and preserve capital. Moreover, increased volatility in zinc prices further contributes to hesitancy among investors to maintain long positions, anticipating greater risks associated with price fluctuations. Overall, the pandemic underscored the complexity of investor decision-making in commodity markets like zinc, where perceptions of supply dynamics, economic recovery prospects, and price volatility all play significant roles in shaping investment strategies.

ADF Statistic	-2.205669
p-value	0.204195
Critical Values:	
1%	-3.45
5%	-2.87
10%	-2.571

Figure 38: ADF Investment Funds Long Percentage

Based on the p-value of 0.204, we can say that the test statistic is not statistically significant at the 5% level. This means that we cannot reject the null hypothesis, which states that the time series data has a unit root, meaning it's non-stationary. We perform the difference of the weekly values to make a stationary time series.

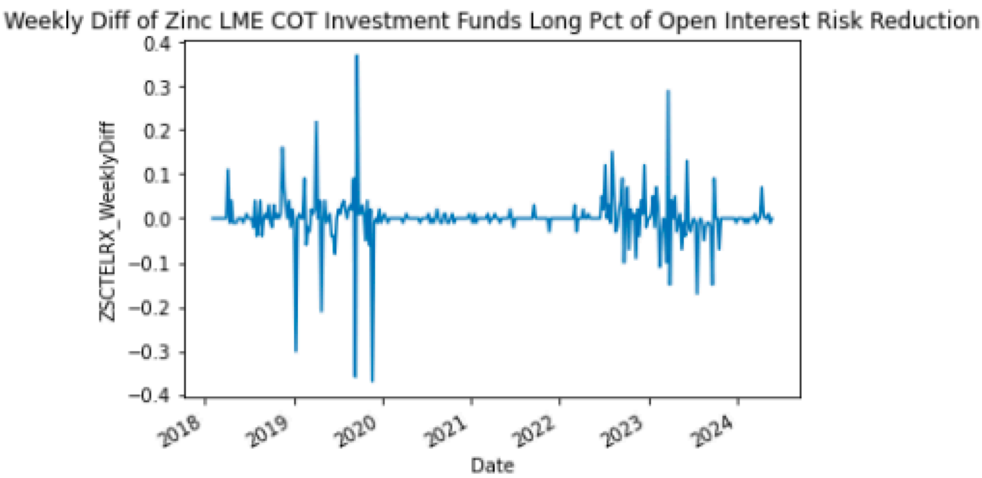

Figure 39: Weekly difference of Investment Funds Long Percentage Open Interest Risk Reduction

ADF Statistic	-23.246524
p-value	0
Critical Values:	
1%	-3.45
5%	-2.87
10%	-2.571

Figure 40: ADF test results for Investment Funds Long percentage weekly difference

The Augmented Dickey-Fuller test results indicate that the differenced time series is now stationary, with an ADF statistic significantly lower than any critical value. Figure 39 visually confirms this stationarity, despite some spikes present in the graph.

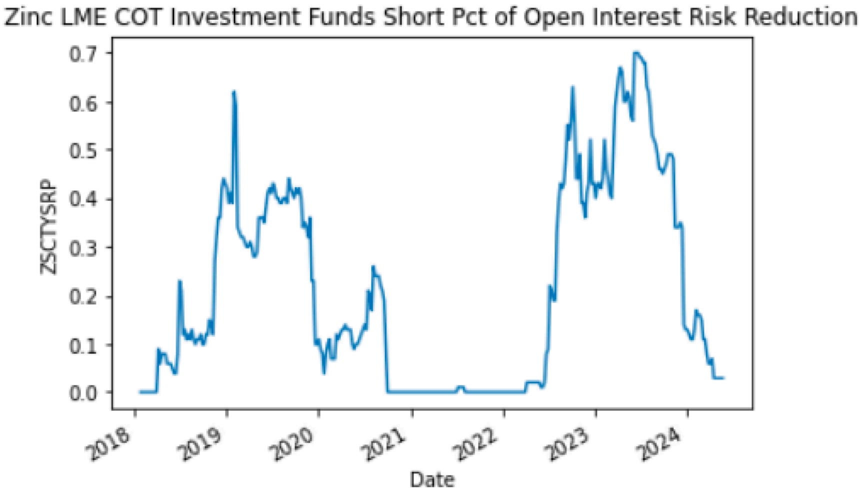

Figure 41: Investment Funds Short Percentage of Open Interest Risk Reduction

Just like long positions held by investment funds, short positions also decrease during certain periods. The reasons for this decline could mirror those affecting long positions, potentially linked to uncertainties surrounding zinc prices. Investment funds may manage risk exposure by reducing short positions, especially if they foresee reduced downside potential or increased volatility.

ADF Statistic	-1.737379
p-value	0.411964
Critical Values:	
1%	-3.45
5%	-2.87
10%	-2.571

Figure 42: ADF Investment Funds Short Percentage

The ADF Statistic (-1.737379) is higher (less negative) than all the critical values at the 1%, 5%, and 10% significance levels. This means the test does not reject the null hypothesis of a unit root. The p-value (0.411964) is significantly higher than common significance levels (0.01, 0.05, 0.10), further indicating that there is insufficient evidence to reject the null hypothesis. Therefore, the time series data likely has a unit root and is non-stationary. This implies that its properties change over time, and it may require to do the weekly differencing represented in Figure 43.

The weekly differencing shows the spikes between 2018 and 2021, and 2023 and 2024, and the little movement of short positions between 2021 and 2023. This differenced time series has a ADF statistic value of -13.468109, which implies stationarity.

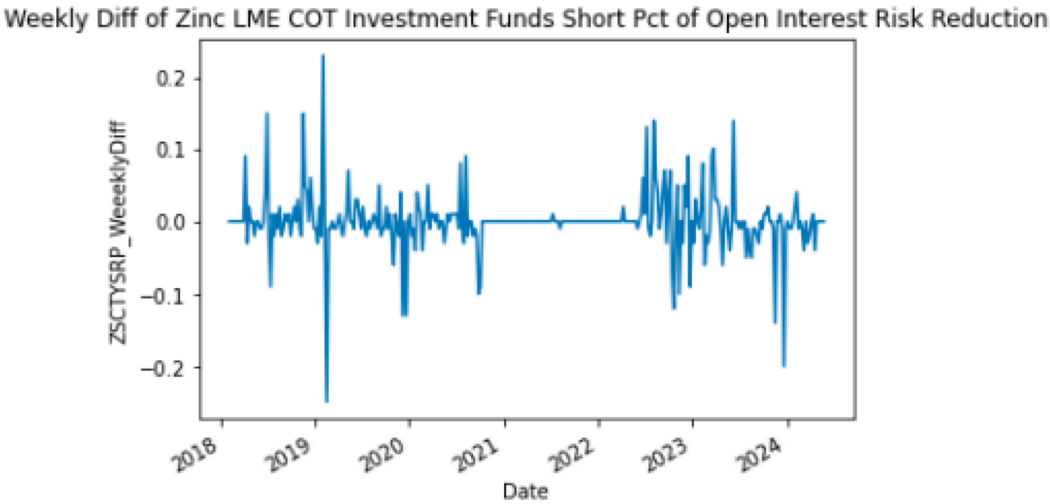

Figure 43: Weekly difference of Investment Funds Short Percentage Open Interest Risk Reduction

The graph in Figure 44, compares the percentage of open interest for investment funds both in short and long positions. As previously mentioned, the percentage for both positions declines drastically in 2020-2021, with an important rise in both positions, peaking in mid-2022, suggesting a period of high volatility or significant market events leading funds to adjust their positions substantially.

The spikes and trends in the graph suggest periods of high market volatility or significant events affecting zinc prices, prompting investment funds to adjust their risk exposure. The more pronounced fluctuations in the short positions compared to the long positions might indicate that funds were more reactive or speculative in their short positions during certain periods.

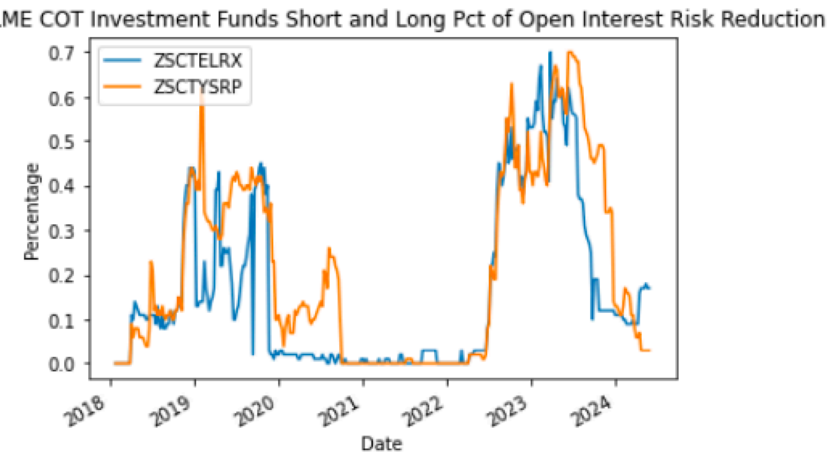

Figure 44: Comparison between short and long positions for Investment Funds

The scatter plot in Figure 45, shows the relationship between long and short positions. There appears to be a positive correlation between the long positions and short positions. As the percentage of long positions increases, the percentage of short positions also tends to increase. There are clusters of points at different ranges, particularly a dense cluster around low values for both long and short positions (close to (0,0)). The dense cluster at low values might indicate periods of low market activity or low volatility, where funds are not heavily invested in either direction (period of 2020-2022).

The spread of points at higher values, indicates more variability and potential periods of higher activity or volatility in the market. At higher percentages of long positions (above 0.3), the short positions exhibit more spread. This indicates that there might be different strategies or market conditions influencing the funds' decisions, leading to varying short positions for similar long positions. Overall, the graph indicates a generally positive correlation with notable clustering and variability.

The calculation of the correlation coefficient shows that this linear relationship is highly probable. The value of 0.841587 indicates a strong positive linear relationship between short and long positions for investment funds. The strong positive correlation between short and long positions suggests that investment funds tend to take positions that move together. This knowledge can be used to inform investment strategies. Understanding this correlation is crucial for managing risk exposures in investment portfolios. Changes in one type of position could indicate potential changes in the other type, influencing risk management decisions. While there is a strong correlation, it’s important to remember that correlation does not imply causation. Additional analysis would be needed to determine if changes in one type of position directly cause changes in the other.

When making the weekly difference time series shown in Figure 46, the data points are clustered around the center (0,0), indicating that, in most cases, the weekly changes in both long and short positions are relatively small. The correlation coefficient between the weekly differences of the long and short positions is 0.241068. This weaker correlation between the weekly differences suggests that while there is some relationship between short-term changes in the variables, this relationship is less reliable compared to their overall movement.

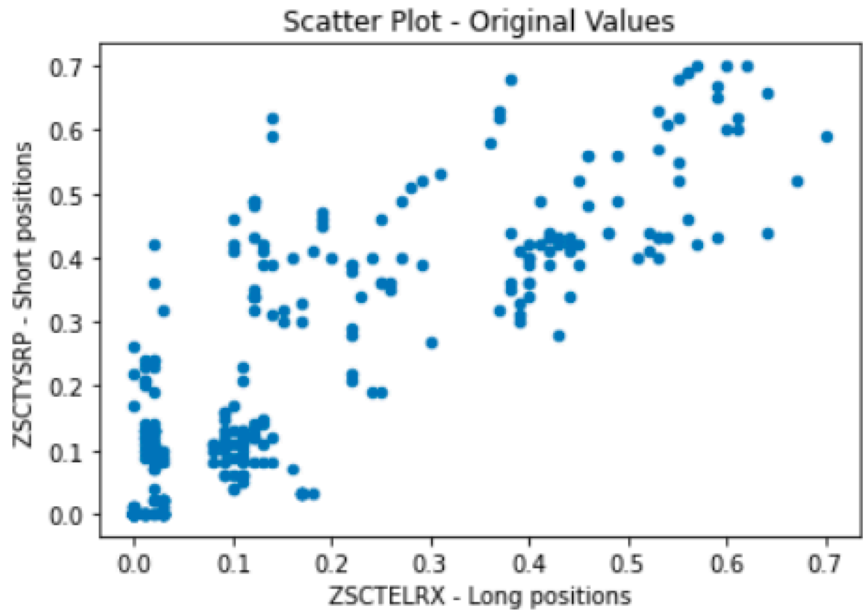

Figure 45: Scatter Plot for Long and Short Positions of Investment Funds

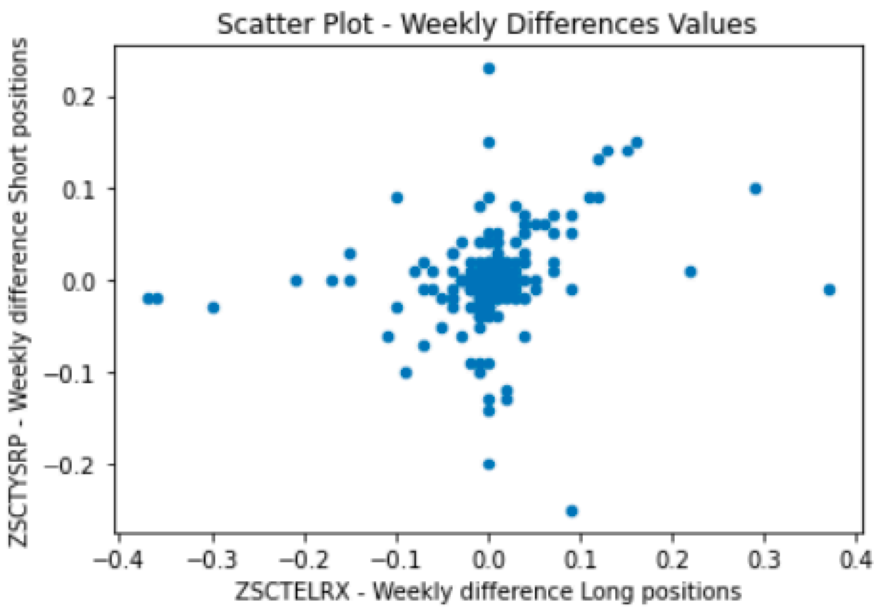

Figure 46: Scatter Plot for the Weekly Differences of Long and Short Positions of Investment Funds

The net hedging pressure shown in Figure 47, exhibits significant volatility throughout the period, with frequent and sharp swings both upwards and downwards. There doesn't seem to be a consistent upward or downward trend over the entire period. The index fluctuates around the zero line, suggesting periods of both net long and net short positions. The descriptive statistics presented in Figure 48 show the mean values for both long and short positions time series, with the mean and standard deviation for the short positions being slightly higher.

However, the net hedging pressure of Investment Funds is low compared to that of other market participants, such as Investment Firms or Credit Institutions.

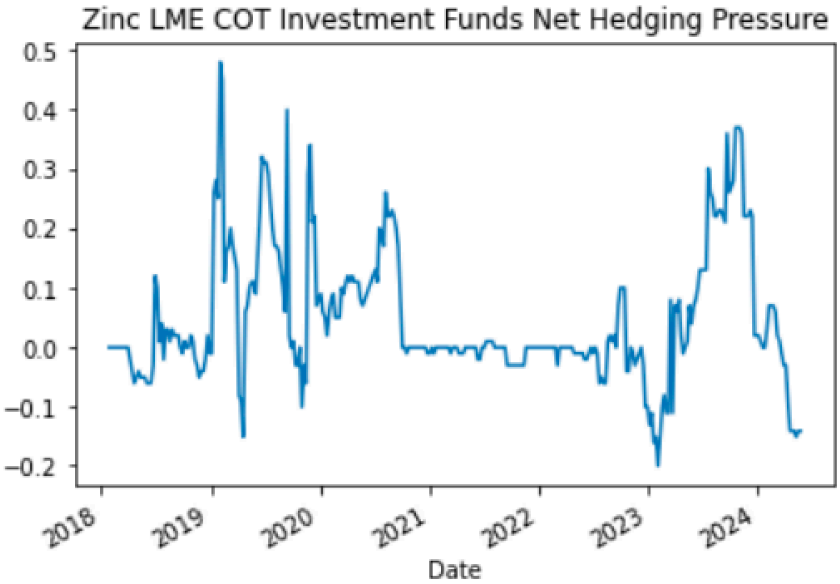

Figure 47: Investment Funds Net Hedging Pressure

	ZSCTELRX	ZSCTELRX_ WeeklyDiff
Count	330	330
Mean	0.16303	0.000515
Std	0.190564	0.056424
Min	0	-0.37
Max	0.7	0.37
Kurtosis	-0.14736	1.079954
Skewness	20.67600	-0.808650

	ZSCTYSRP	ZSCTYSRP_ WeeklyDiff
Count	330	330
Mean	0.211364	0.000091
Std	0.208081	0.040622
Min	0	-0.25
Max	0.7	0.23
Kurtosis	-0.88687	11.167665
Skewness	0.62828	-0.192145

Figure 48: Descriptive Statistics of long and short positions and their weekly differences for Investment Funds

Other Financial Institutions Open Interest Risk Reduction

Other financial institutions typically refer to a category of market participants that are distinct from traditional commercial traders, producers, consumers, or individual speculators. They can include a variety of financial entities that participate in the futures markets for purposes other than hedging physical zinc production or consumption. Some examples can be hedge funds, insurance companies or investment banks.

The long percentage of open interest of this group of investors exhibits significant fluctuations throughout the period, indicating dynamic changes.

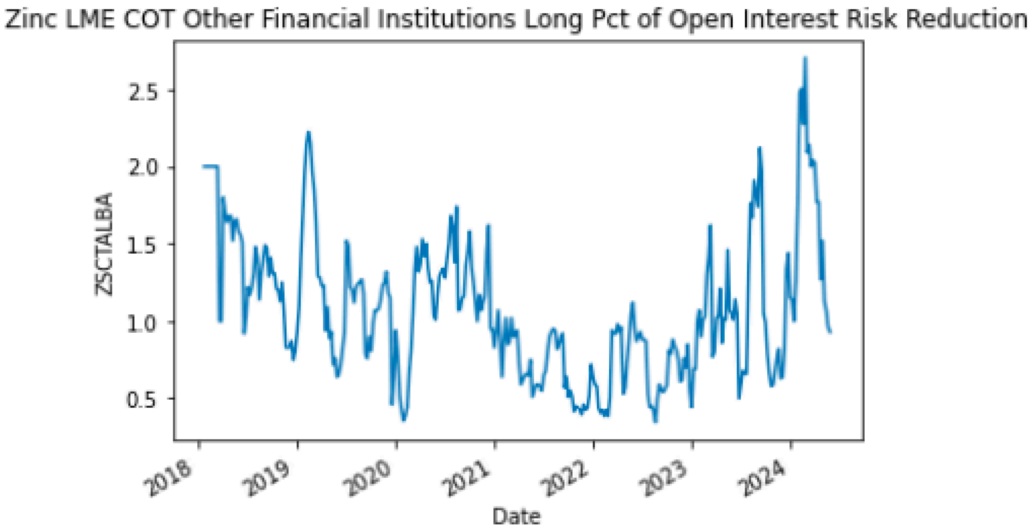

Figure 49: Other Financial Institutions Long Percentage of Open Interest Risk Reduction

ADF Statistic	-3.011795
p-value	0.033807
Critical Values:	
1%	-3.451
5%	-2.871
10%	-2.572

Figure 50: ADF Other Financial Institutions Long Percentage

Despite the significant fluctuations in the time series, the results of the Augmented Dickey-Fuller test (Figure 50) indicate a strong probability of rejecting the null hypothesis, suggesting that the time series is likely stationary. Additionally, a very low p-value of 0.0338 which is less than the conventional significance levels of 0.05 and 0.10, indicates that we can reject the null hypothesis.

The graph in Figure 51 illustrates the weekly change in the percentage of open interest held by Other Financial Institutions in long positions. Although it may appear highly volatile, it's important to consider the scale, with peaks reaching 0.75 or 1.00. In comparison, other groups like Commercial Undertakings have peaks around 3.00, indicating that this time series is relatively centered around 0.

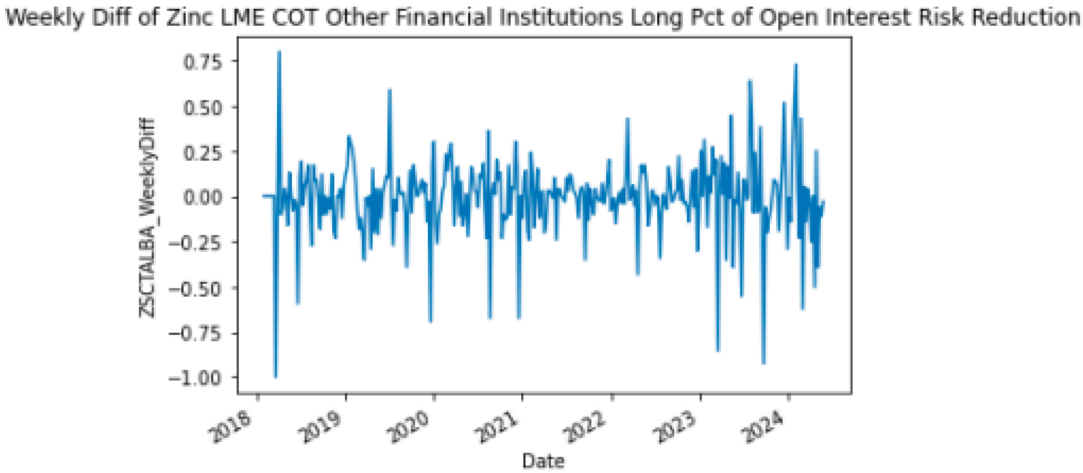

Figure 51: Weekly difference of Other Financial Institutions Long Percentage Open Interest Risk Reduction

The percentage of short positions shown in Figure 52 seems to be substantially higher for Other Financial Institutions, peaking several times at 5.5. Financial institutions might short zinc futures if they anticipate a decline in zinc prices. By selling futures contracts now and buying them back at a lower price later, they can profit from the price difference. The substantial rise in short positions from 2020 onward suggests that the pandemic prompted financial institutions to short zinc futures in order to profit from the anticipated decline in prices during this period.

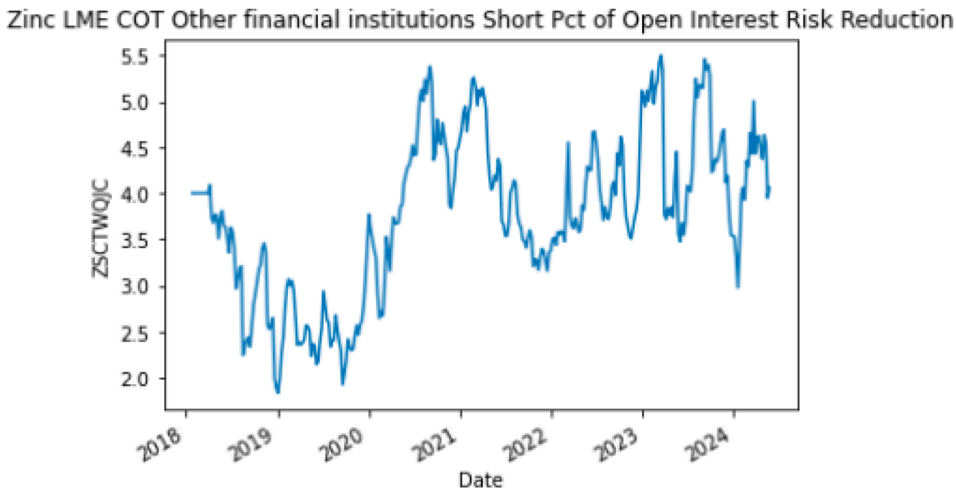

Figure 52: Other Financial Institutions Short Percentage of Open Interest Risk Reduction

ADF Statistic	-2.950349
p-value	0.039799
Critical Values:	
1%	-3.45
5%	-2.87
10%	-2.571

Figure 53: ADF Other Financial Institutions Short Percentage

Based on the ADF test results, we can reject the null hypothesis of a unit root at the 5% significance level. This suggests that the time series is likely stationary. However, the weekly difference values are calculated. In this case, the ADF statistic is -13.450, which indicates that this time series is very likely stationary.

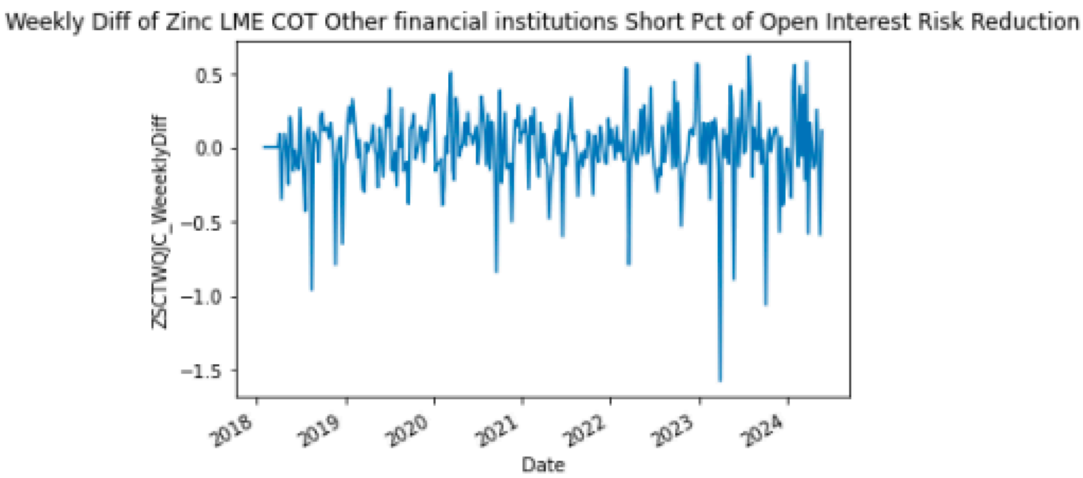

Figure 54: Weekly difference of Other Financial Institutions Short Percentage Open Interest Risk Reduction

The percentage of open interest held by Other Financial Institutions in both short and long Zinc LME contracts fluctuates significantly as seen in the comparison graph in Figure 55, indicating rapid adjustments to their market positions. This volatility suggests that they are actively responding to changing market conditions, potentially influenced by economic indicators, supply-demand dynamics, and other geopolitical events previously mentioned.

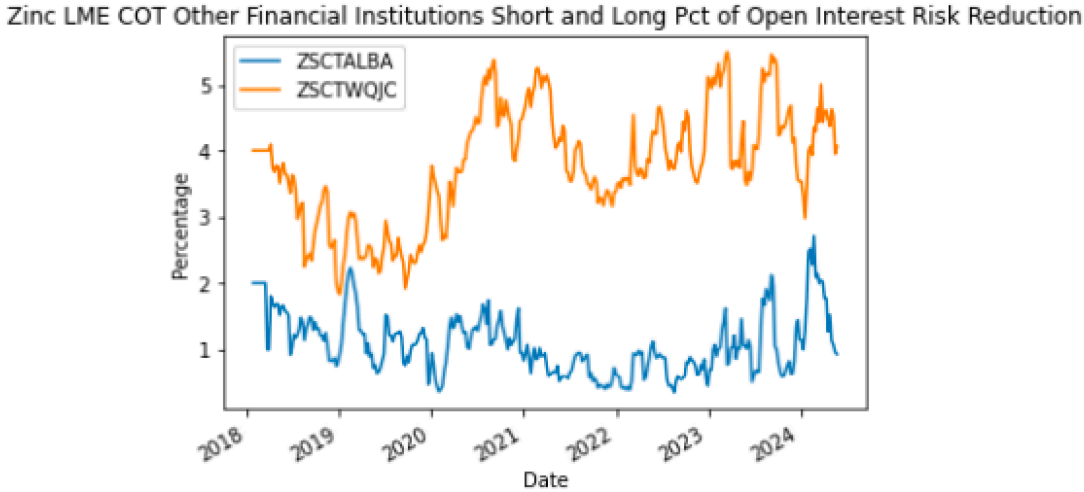

Figure 55: Comparison between short and long positions for Other Financial Institutions

The scatter plot in Figure 56 illustrates a relatively weak positive relationship between long and short positions in Zinc LME contracts. While there's a general tendency for both positions to increase or decrease simultaneously, the data points are widely dispersed, indicating that other factors significantly influence these positions. The correlation coefficient of 0.085634 indicates a very weak positive linear relationship between the long and short positions in financial institutions. The very weak correlation in original variables (0.085634) suggests that decisions based solely on the levels of long or short positions in financial institutions may not be reliable predictors of each other's movements. Investors may need to consider other factors or indicators when making decisions.

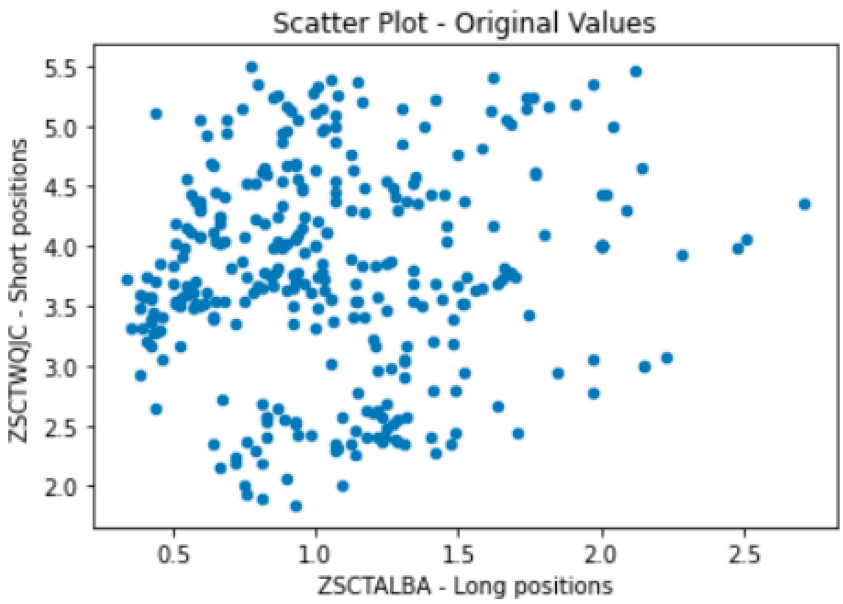

Figure 56: Scatter Plot for Long and Short Positions of Other Financial Institutions

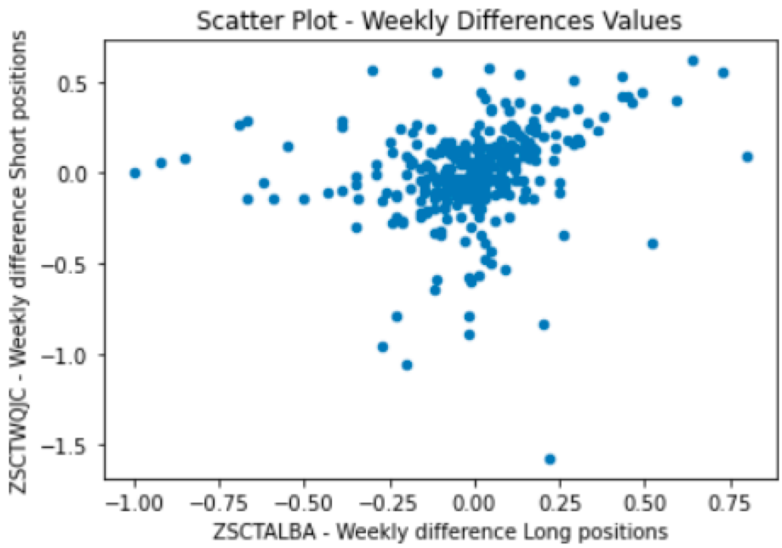

Figure 57: Scatter Plot for the Weekly Differences of Long and Short Positions of Other Financial Institutions

The scatter plot reveals a weak positive correlation between weekly changes in long and short positions held by Other Financial Institutions in Zinc LME contracts, suggesting a limited tendency for these positions to move in tandem. This indicates that while there might be some coordinated behavior among Other Financial Institutions, other factors significantly influence their trading decisions. Indeed, the weak correlation coefficient in weekly differences (0.245190) implies that short-term trading strategies based on changes in weekly long or short positions across financial institutions should be approached cautiously. The relationship is not strong enough to rely on for predictive purposes without additional analysis.

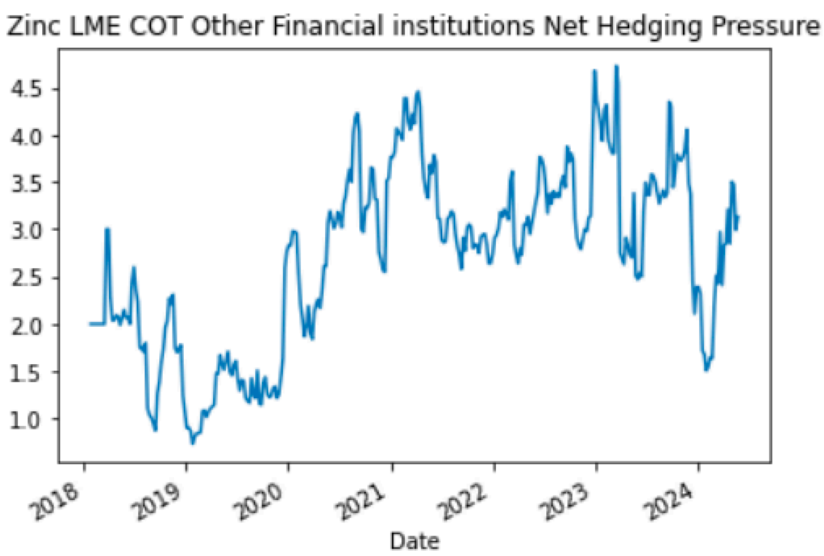

Figure 58: Other Financial Institutions Net Hedging Pressure

Figure 58 shows the net hedging pressure of Other Financial Institutions. The metric exhibits significant volatility, with periods of both high and low pressure. There's no clear trend, suggesting the sentiment towards Zinc has been highly variable. The frequent spikes indicate periods of intense directional bets, potentially influencing Zinc prices.

	ZSCTALBA	ZSCTALBA_ WeeklyDiff
Count	330	330
Mean	1.072273	-0.003242
Std	0.458847	0.209768
Min	0.34	-1
Max	2.71	0.8
Kurtosis	0.31209	4.776282
Skewness	0.77032	-0.709189

	ZSCTWQJC	ZSCTWQJC_ WeeklyDiff
Count	330	330
Mean	3.774303	0.000182
Std	0.86972	0.254819
Min	1.83	-1.58
Max	5.5	0.62
Kurtosis	-0.68621	6.171757
Skewness	-0.12354	-1.400576

Figure 59: Descriptive Statistics of long and short positions and their weekly differences for Other Financial Institutions

All Financials Open Interest Risk Reduction

The sum of all financials' net hedging pressures is represented in Figure 60. The plot shows fluctuations over the years with both positive and negative values. The series exhibits considerable volatility. After mid-2021, it shows a period of relative stability, but still with noticeable fluctuations. The sharp increase and subsequent decrease around 2021 might correspond to investment firms or credit institutions and investment funds sudden decrease on their long positions.

This group includes all financial institutions, which might engage in frequent trading, leading to higher volatility. Understanding the power that financials have in the hedging pressure and therefore in the Zinc market is essential. Financial institutions often hold large positions in the market and their hedging activities can significantly impact market prices and volatility. Understanding their behavior helps in anticipating and responding to these impacts. To begin with, changes in financials' hedging pressure can be predictive of future price movements. For instance, a significant increase in hedging pressure might signal expectations of higher future volatility or price changes. By monitoring hedging activities, other market participants can adjust their strategies accordingly. Finally, it is important to notice that hedging pressure can serve as an economic indicator. For example, increased hedging pressure in commodities might reflect broader economic trends such as inflation expectations, changes in industrial demand, or shifts in economic policies.

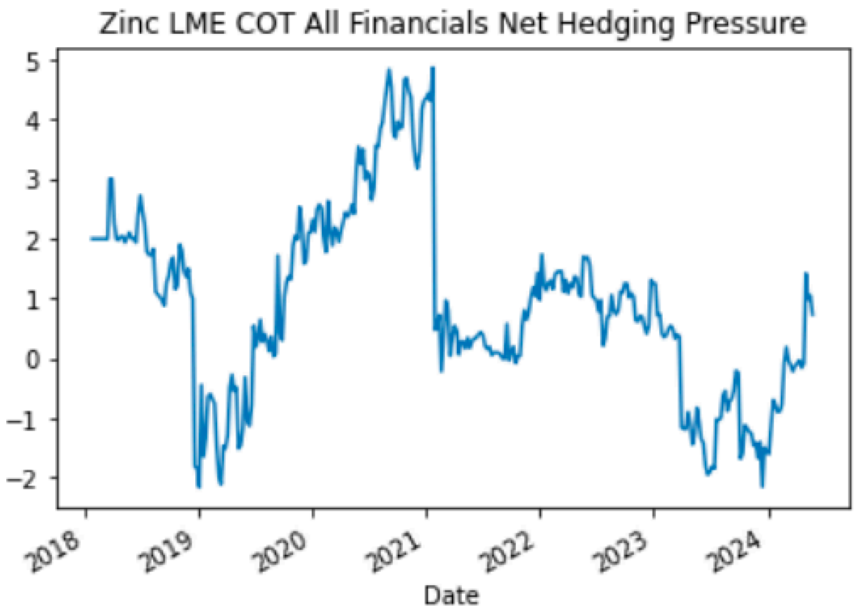

Figure 60: All Financials Net Hedging Pressure

Total Open Interest Risk Reduction

Total net hedging pressure is the sum of all the groups' hedging pressures trading zinc futures in the LME Zinc market. Comparing the total net hedging pressure with the sum of all financials hedging pressure we see that there is an increased participation by financial players, such as hedge funds, which has led to higher hedging pressure. Speculation on price movements in the context of recovering demand (post pandemic) and supply constraints might have led to increased hedging activities. Additionally, rising concerns about inflation in 2021 due to economic recovery and stimulus measures could have prompted investors to hedge against potential price increases in commodities.

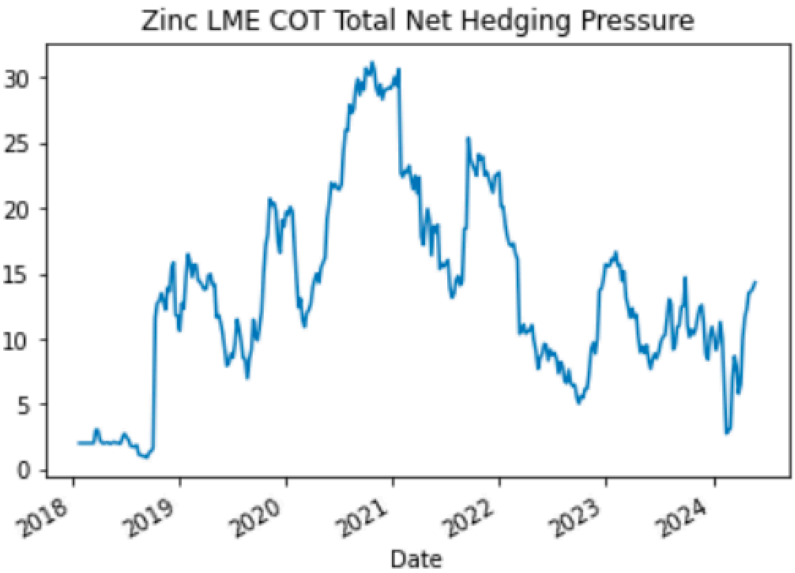

Figure 61: Total Net Hedging Pressure

6. Regression Results

In this section of the research, a variety of regressions will be tested to evaluate the predictive power of different variables, such as various investors' hedging pressures and control variables like VIX. Specifically, three different types of models will be tested.

Predictive regression 1: Univariate model

The first predictive regression model is a univariate model, where the daily returns variable is presented as a dependent variable of the net hedging pressure, as described in the following formula.

$$y_i = a_0 + a_1 x_k$$

In this formula, the dependent varible y_i refers to the i-day return on Friday, calculated as the natural logarithm of the settle price of th.e i-day over the settle price on Friday. For instance, the 2-day log return of Friday 17th of May 2024 is calculated as the natural logarithm of the settle price of Tuesday 21st of May 2024 over the settle price on Friday 17.

Natural logarithms are used instead of normal returns because they provide more normally distributed data, additive properties over time, relative change measurement and alignment with continuously compounded returns.

In the formula, a_0 is a constant. This constant is added to the regression model to account for the intercept term, which represents the expected value of the dependent variable when all independent variables are equal to zero, ensuring the model can fit data more accurately by capturing the baseline level of the dependent variable. a_1 represents the slope coefficient of the independent variable x_k. Specifically, a_1 indicates the change in the dependent variable a_1 for a one-unit change in x_k, assuming all other variables remain constant.

Finally, x_k represents the independent variable or predictor variable, which is set as the different COTR groups' net hedging pressure. The calculation of the net hedging pressure is explained in a previous section. This variable is used to explain or predict changes in the dependent variable y_i, as it will be shown in the results.

Table 1: Total net hedging pressure univariate model results

	Dependent variable				
	1-day log return	2-day log return	3-day log return	4-day log return	5-day log return
$\boldsymbol{a_0}$ coefficient	-0.0033	-0.0044	-0.0056	-0.0071	-0.0062
$\boldsymbol{a_1}$ coefficient	0.0002	0.0002	0.0003	0.0004	0.0004
Adj. R-squared	0.006	0.001	0.003	0.0005	0.003
F-statistic	3.002	1.167	2.011	2.656	2.057
Probability (F-stat)	0.0841	0.281	0.157	0.104	0.152

Table 2: Commercial undertakings net hedging pressure univariate model results

	Dependent variable				
	1-day log return	2-day log return	3-day log return	4-day log return	5-day log return
$\boldsymbol{a_0}$ coefficient	-00027	-0.042	-0.0043	-0.0044	-0.0049
$\boldsymbol{a_1}$ coefficient	0.0002	0.0002	0.0002	0.003	0.0004
Adj. R-squared	0.001	0.002	-0.001	-0.002	-0.001
F-statistic	1.142	0.5616	0.6848	0.5658	0.7947
Probability (F-stat)	0.286	0.454	0.409	0.453	0.373

Table 3: Investment Firms and Credit Institutions net hedging pressure univariate model results

	Dependent variable				
	1-day log return	2-day log return	3-day log return	4-day log return	5-day log return
$\boldsymbol{a_0}$ coefficient	0.0006	-0.0012	0.0009	0.0010	0.0017
$\boldsymbol{a_1}$ coefficient	0.0003	0.0002	0.0009	0.0008	0.0007
Adj. R-squared	0.001	-0.003	0.000	-0.002	-0.003
F-statistic	0.3236	0.1039	0.8660	0.5057	0.2703
Probability (F-stat)	0.570	0.747	0.353	0.478	0.604

Table 4: Investment Funds net hedging pressure univariate model results

	Dependent variable				
	1-day log return	2-day log return	3-day log return	4-day log return	5-day log return
$\boldsymbol{a_0}$ coefficient	0.0002	-0.0015	-0.0013	-0.0009	0.00093
$\boldsymbol{a_1}$ coefficient	-0.0052	-0.0031	0.0061	0.0033	0.0025
Adj. R-squared	-0.002	-0.003	-0.003	-0.003	-0.003
F-statistic	0.4197	0.08049	0.2099	0.04241	0.01679
Probability (F-stat)	0.518	0.777	0.647	0.837	0.897

Table 5: Other Financial Institutions net hedging pressure univariate model results

	Dependent variable				
	1-day log return	2-day log return	3-day log return	4-day log return	5-day log return
a_0 coefficient	-0.0024	-0.0043	-0.0016	-0.0037	-0.0031
a_1 coefficient	0.0008	0.0009	0.0002	0.0011	0.0012
Adj. R-squared	0.003	-0.002	-0.003	-0.003	-0.003
F-statistic	0.7112	0.4842	0.01743	0.2842	0.2522
Probability (F-stat)	0.400	0.487	0.895	0.594	0.616

Table 6: All Financials net hedging pressure univariate model results

	Dependent variable				
	1-day log return	2-day log return	3-day log return	4-day log return	5-day log return
a_0 coefficient	-0.0006	-0.0022	-0.0018	-0.0017	-0.0007
a_1 coefficient	0.0006	0.0006	0.0011	0.0013	0.0012
Adj. R-squared	0.000	-0.002	0.001	0.001	-0.001
F-statistic	1.101	0.5335	1.178	1.160	0.7268
Probability (F-stat)	0.295	0.466	0.279	0.282	0.395

The results indicate that the a_1 coefficients for net hedging pressure are small and the adjusted R-squared values are low, suggesting limited predictive power of net hedging pressure on log returns over different periods (1 to 5 days). The F-statistic probabilities are also not significant, reinforcing the conclusion that the univariate model is not a good predictor of daily returns in the zinc futures market.

Predictive regression 2: Bivariate model

The second predictive regression model is a bivariate model, where the daily returns variable is presented as a dependent variable of both the net hedging pressure and historical returns, as described in the following formula:

$$y_i = a_0 + a_1 x_k + a_2 w_j$$

As previously explained, y_i is the dependent variable which refers to the i-day return on Friday. a_0 is the constant added to the regression model to account for the intercept term. Finally, x_k represents the first independent variable or predictor variable, which is set as the different COTR groups' net hedging pressure, and w_j represents the second independent variable or predictor variable, which is set as the historical returns. The calculation of the net hedging pressure is explained in a previous section.

Table 7: Total net hedging pressure bivariate model results

	Dependent variable				
	1-day log return	2-day log return	3-day log return	4-day log return	5-day log return
a_0 coefficient	-0.0026	-0.0042	-0.0047	-0.0047	-0.0046
a_1 coefficient	0.0002	0.0002	0.0003	0.0003	0.0003
a_2 coefficient	0.0884	-0.0021	-0.0455	0.0112	0.0263
Adj. R-squared	0.007	-0.005	-0.001	-0.004	-0.003
F-statistic	1.928	0.3406	0.8426	0.4344	0.5666
Probability (F-stat)	0.147	0.712	0.432	0.648	0.568

Table 8: Commercial undertakings net hedging pressure bivariate model results

	Dependent variable				
	1-day log return	2-day log return	3-day log return	4-day log return	5-day log return
a_0 coefficient	-0.0025	-0.0042	-0.0046	-0.0043	-0.0046
a_1 coefficient	0.0002	0.0002	0.0003	0.0002	0.0003
a_2 coefficient	0.0875	-0.0026	-0.0460	0.0113	0.0264
Adj. R-squared	0.005	-0.005	-0.002	-0.005	-0.004
F-statistic	1.773	0.2810	0.7100	0.3012	0.4866
Probability (F-stat)	0.172	0.755	0.493	0.740	0.615

Table 9: Investment Firms and Credit Institutions net hedging pressure bivariate model results

	Dependent variable				
	1-day log return	2-day log return	3-day log return	4-day log return	5-day log return
a_0 coefficient	0.0005	-0.0012	0.0009	0.0010	0.0017
a_1 coefficient	0.0003	0.0002	0.0009	0.0008	0.0007
a_2 coefficient	0.0926	0.0008	-0.0406	0.0161	0.0331
Adj. R-squared	0.004	-0.007	-0.002	-0.005	-0.005
F-statistic	1.511	0.05191	0.7215	0.2913	0.2786
Probability (F-stat)	0.223	0.949	0.487	0.748	0.757

Table 10: Investment Funds net hedging pressure bivariate model results

	Dependent variable				
	1-day log return	2-day log return	3-day log return	4-day log return	5-day log return
a_0 coefficient	0.0001	-0.0015	-0.0009	0.0010	0.00065
a_1 coefficient	-0.0054	-0.0031	0.0032	0.0008	0.0024
a_2 coefficient	0.0923	0.0005	0.0155	0.0161	0.0336
Adj. R-squared	0.004	-0.007	-0.007	-0.005	-0.006
F-statistic	1.550	0.04015	0.05748	0.2913	0.1568
Probability (F-stat)	0.214	0.961	0.944	0.748	0.855

Table 11: Other Financial Institutions net hedging pressure bivariate model results

	Dependent variable				
	1-day log return	2-day log return	3-day log return	4-day log return	5-day log return
$\boldsymbol{a_0}$ coefficient	-0.0028	-0.0043	-0.0015	-0.0037	-0.0031
$\boldsymbol{a_1}$ coefficient	0.0009	-0.0009	0.0002	0.0011	0.0012
$\boldsymbol{a_2}$ coefficient	0.0945	0.0010	-0.0415	0.0157	0.0329
Adj. R-squared	0.005	-0.005	-0.005	-0.006	-0.005
F-statistic	1.760	0.2414	0.3104	0.1792	0.2683
Probability (F-stat)	0.174	0.786	0.733	0.836	0.765

Table 12: All Financials net hedging pressure bivariate model results

	Dependent variable				
	1-day log return	2-day log return	3-day log return	4-day log return	5-day log return
$\boldsymbol{a_0}$ coefficient	-0.0007	-0.0022	-0.0017	-0.0017	-0.0007
$\boldsymbol{a_1}$ coefficient	0.0007	0.0006	0.0011	0.0013	0.0012
$\boldsymbol{a_2}$ coefficient	0.0954	0.0016	-0.0400	0.0162	0.0318
Adj. R-squared	0.007	-0.005	-0.001	-0.003	-0.004
F-statistic	1.986	0.2663	0.8688	0.6181	0.4958
Probability (F-stat)	0.139	0.766	0.421	0.540	0.610

The results from the bivariate model show that adding historical returns as a control variable does not significantly enhance the model's fit. The adjusted R-squared values remain low or even negative in many cases, indicating poor model performance. The F-statistic values are also not significant, suggesting that neither net hedging pressure nor historical returns are strong predictors of daily log returns. The a_2 coefficients for historical returns vary widely, indicating inconsistent effects across different return periods. The adjusted R-squared values are generally low, demonstrating that the bivariate model does not capture much of the variance in the dependent variable. Finally, the F-statistic values are insignificant, further indicating that the model does not provide a meaningful prediction of returns.

This analysis implies that historical returns, when combined with net hedging pressure, do not significantly contribute to predicting future returns in the zinc futures market. The relationship between past performance and future returns is not strong enough to rely upon for predictive modeling in this context.

Predictive regression 3: Multivariate model

The third predictive regression model is a multivariate model, where the daily returns variable is presented as a dependent variable of net hedging pressure, historical returns, and additional market and economic indicators, as described in the following formula:

$$y_i = a_0 + a_1 x_k + a_2 w_j + a_3 z_n$$

The terms $y_i, a_{0,}, a_1, x_k, a_2$ and w_j have previously been explained. In this model their meaning remains equal. Finally, z_n represents the independent variable or predictor variables, which include additional market and economic indicators such as the VIX index, BDI (Baltic Dry Index), PMI (Purchasing Managers' Index), and SPX (S&P 500 Index). The coefficient a_3 corresponds to these predictor variables. The inclusion of these indicators aims to capture broader market and economic influences on zinc futures returns, potentially enhancing the model's explanatory power.

In this model, only 1-day returns have been considered, as they have been proved to give the best results in the programmed Python algorithm. The results are summarized in the tables below for all the groups of the COT report.

Table 13: Total net hedging pressure multivariate model results

	Dependent variable			
	1-day log return	1-day log return	1-day log return	1-day log return
$\boldsymbol{a_0}$ coefficient	0.0004	-0.0028	-0.0031	-0.0056
$\boldsymbol{a_1}$ coefficient	0.0002	0.0002	0.0010	0.0002
$\boldsymbol{a_2}$ coefficient	0.0799	0.0876	0.104	0.0870
$\boldsymbol{a_3}$ coefficient (VIX)	-0.0002			
$\boldsymbol{a_3}$ coefficient (BDI)		0.000		
$\boldsymbol{a_3}$ coefficient (PMI)			-0.001	
$\boldsymbol{a_3}$ coefficient (SPX)				0.000
Adj. R-squared	0.010	0.003	0.020	0.004
F-statistic	1.956	1.284	1.579	1.387
Probability (F-stat)	0.121	0.280	0.201	0.247

Table 14: Commercial Undertakings net hedging pressure multivariate model results

	Dependent variable			
	1-day log return	1-day log return	1-day log return	1-day log return
$\boldsymbol{a_0}$ coefficient	0.0005	-0.0026	0.0027	-0.0050
$\boldsymbol{a_1}$ coefficient	0.0002	0.0002	0.0006	0.0002
$\boldsymbol{a_2}$ coefficient	0.0003	0.0869	0.1478	0.0864
$\boldsymbol{a_3}$ coefficient (VIX)	-0.0001			
$\boldsymbol{a_3}$ coefficient (BDI)		0.000		
$\boldsymbol{a_3}$ coefficient (PMI)			-0.002	
$\boldsymbol{a_3}$ coefficient (SPX)				0.000
Adj. R-squared	0.008	0.002	-0.002	0.003
F-statistic	1.718	1.180	0.9327	1.257
Probability (F-stat)	0.163	0.318	0.429	0.289

Table 15: Investment Firms and Credit Institutions net hedging pressure multivariate model results

	Dependent variable			
	1-day log return	1-day log return	1-day log return	1-day log return
$\boldsymbol{a_0}$ coefficient	0.0056	0.0000	0.0147	-0.0039
$\boldsymbol{a_1}$ coefficient	0.0007	0.0004	0.0028	0.0007
$\boldsymbol{a_2}$ coefficient	0.0839	0.0902	0.1459	0.0914
$\boldsymbol{a_3}$ coefficient (VIX)	-0.0002			
$\boldsymbol{a_3}$ coefficient (BDI)		0.000		
$\boldsymbol{a_3}$ coefficient (PMI)			0.000	
$\boldsymbol{a_3}$ coefficient (SPX)				0.000
Adj. R-squared	0.009	0.000	0.020	0.003
F-statistic	1.879	1.038	1.571	1.261
Probability (F-stat)	0.133	0.376	0.203	0.288

Table 16: Investment Funds net hedging pressure multivariate model results

	Dependent variable			
	1-day log return	1-day log return	1-day log return	1-day log return
$\boldsymbol{a_0}$ coefficient	0.0034	0.0000	0.0078	-0.0011
$\boldsymbol{a_1}$ coefficient	-0.0066	-0.0053	-0.0156	-0.0048
$\boldsymbol{a_2}$ coefficient	0.0851	0.0918	0.1551	0.0917
$\boldsymbol{a_3}$ coefficient (VIX)	-0.0002			
$\boldsymbol{a_3}$ coefficient (BDI)		0.000		
$\boldsymbol{a_3}$ coefficient (PMI)			-0.0001	
$\boldsymbol{a_3}$ coefficient (SPX)				0.000
Adj. R-squared	0.006	0.000	0.006	0.001
F-statistic	1.617	1.031	1.177	1.048
Probability (F-stat)	0.186	0.379	0.324	0.372

Table 17: Other Financial Institutions net hedging pressure multivariate model results

	Dependent variable			
	1-day log return	1-day log return	1-day log return	1-day log return
$\boldsymbol{a_0}$ coefficient	0.000	-0.0029	-0.0014	-0.0025
$\boldsymbol{a_1}$ coefficient	0.0011	0.0009	0.0025	0.0010
$\boldsymbol{a_2}$ coefficient	0.0874	0.0939	0.1544	0.0948
$\boldsymbol{a_3}$ coefficient (VIX)	-0.0002			
$\boldsymbol{a_3}$ coefficient (BDI)		0.000		
$\boldsymbol{a_3}$ coefficient (PMI)			-0.0001	
$\boldsymbol{a_3}$ coefficient (SPX)				0.000
Adj. R-squared	0.008	0.002	0.000	0.002
F-statistic	1.803	1.170	0.9989	1.171
Probability (F-stat)	0.147	0.321	0.398	0.321

Table 18: All Financials net hedging pressure multivariate model results

	Dependent variable			
	1-day log return	1-day log return	1-day log return	1-day log return
$\boldsymbol{a_0}$ coefficient	0.0045	-0.0013	0.0011	-0.0047
$\boldsymbol{a_1}$ coefficient	0.0014	0.0007	0.0049	0.0008
$\boldsymbol{a_2}$ coefficient	0.0862	0.0931	0.1680	0.0946
$\boldsymbol{a_3}$ coefficient (VIX)	-0.0003			
$\boldsymbol{a_3}$ coefficient (BDI)		0.000		
$\boldsymbol{a_3}$ coefficient (PMI)			0.000	
$\boldsymbol{a_3}$ coefficient (SPX)				0.000
Adj. R-squared	0.018	0.004	0.063	0.005
F-statistic	2.763	1.354	2.847	1.511
Probability (F-stat)	0.0424	0.257	0.0427	0.212

The adjusted R-squared values in the multivariate model are slightly higher than those in the univariate and bivariate models. This indicates that the multivariate model explains a bit more of the variance in the dependent variable (y_i) compared to simpler models. However, the increase in adjusted R-squared is modest, implying that the additional variables contribute only marginally to the overall explanatory power of the model. Despite the inclusion of multiple predictor variables, most of the coefficients for these variables are not statistically significant. This means that the null hypothesis, which states that the coefficient is equal to zero (indicating no effect), cannot be rejected for most of the additional variables. As a result, the additional market and economic indicators do not provide substantial predictive power for zinc futures returns.

While variables such as the VIX index, BDI, PMI, and SPX are included to capture broader market and economic influences, their lack of significance suggests that they do not have a strong or consistent impact on daily returns in the zinc futures market. This could be due to the specific nature of the zinc market, where other unobserved factors or market-specific dynamics play a more crucial role.

The inclusion of multiple variables in the multivariate model increases the risk of overfitting, where the model captures noise rather than the underlying pattern. This is particularly a concern when the additional variables do not show significant coefficients, as it indicates that they do not consistently contribute to predicting the dependent variable.

The results from the multivariate model analysis indicate that while including additional market and economic indicators may seem to enhance the model's explanatory power slightly, the improvement is not substantial enough to warrant their inclusion. Future research might focus on identifying other relevant factors or using different modeling techniques to better capture the dynamics of the zinc futures market.

7. CONCLUSIONS

The analysis of the hedging pressure dynamics in the LME zinc futures markets aimed to uncover the relationship between hedging activities of various market participants and zinc futures prices. This study specifically investigated the impact of net hedging pressure on zinc futures prices, utilizing data from the Commitment of Traders Reports (COTR) and incorporating additional market and economic indicators. The conclusions drawn from this research provide insights into the predictive power of hedging pressure and other factors on zinc futures returns.

The analysis of univariate, bivariate, and multivariate regression models revealed that net hedging pressure from various market participants has limited predictive power over zinc futures returns. The adjusted R-squared values were generally low, indicating that the models did not capture a significant portion of the variance in returns. However, nearly all coefficients are positive, aligning with the hypotheis that hedging pressure has predictive information context.This suggests that while hedging pressure may influence zinc prices to some extent, it is not a dominant factor in predicting daily returns.

On the other hand, the study considered different groups of market participants, including commercial undertakings, investment firms, credit institutions, investment funds, and other financial institutions. The regression results indicated that the net hedging pressures from these groups individually and collectively did not significantly enhance the models' predictive capability. This implies that the trading activities and hedging behaviors of these groups are not strong predictors of zinc futures price movements. The inclusion of historical returns in the bivariate models also did not significantly improve the models' explanatory power. The coefficients for historical returns were often insignificant, and the adjusted R-squared values remained low. This indicates that past performance of zinc futures does not provide substantial information for predicting future returns in the context of this study.

The multivariate models incorporated additional variables such as the VIX index, BDI (Baltic Dry Index), PMI (Purchasing Managers' Index), and SPX (S&P 500 Index). While these variables slightly increased the adjusted R-squared values, the improvements were modest, and most coefficients were not significant. This suggests that broader market and economic conditions, while relevant, do not strongly predict zinc futures returns when considered alongside net hedging pressure.

Despite the comprehensive data analysis and regression models, the study faced limitations such as potential omitted variable bias and the specific time period covered, which may not generalize to other periods.

As for the implications for the market participants, the findings suggest that while hedging pressure is an important market dynamic, it is not a strong standalone predictor of zinc futures returns. Market participants should consider a broader range of factors, including supply and demand fundamentals, macroeconomic indicators, and geopolitical events in their trading strategies.

The results of this study highlight the complexity of the zinc futures market and the need for further research to identify additional factors that may influence price movements. Future research with a slightly different approach might document statistical evidence for the stated hypotheses. Some future research could include analyzing data over longer time periods or different market cycles as they could provide more robust insights into the relationships between hedging pressure and zinc futures prices. Also, incorporating other relevant variables, such as inventory levels, production costs, and technological advancements in zinc extraction and processing, could enhance the predictive models. Lastly, utilizing advanced econometric and machine learning techniques could uncover nonlinear relationships and interactions between variables that traditional regression models may not capture.

In conclusion, the hedging pressure dynamics in the LME zinc futures market exhibit limited predictive power over daily returns. While net hedging pressure and additional market indicators provide some insights, their overall impact on zinc futures prices is modest. Market participants should adopt a comprehensive approach to trading and risk management, considering a wide range of factors and utilizing diverse strategies. Continued research in this field is essential to deepen our understanding of commodity markets and enhance predictive capabilities.

BIBLIOGRAPHY

Arouri, M. E. (2011). The Speculative Efficiency of the Aluminum Market: A Nonlinear Investigation. *International Economics* .

Brunetti, C. B. (2015). Speculators, Prices and Market Volatility. *Journal of Financial and Quantitative Analysis 51*.

Cootner, P. H. (1967). *Speculation and Hedging.* Massachusetts Institute of Technology, Cambridge 02139.

Figuerola-Ferretti, I. a. (2008). Commonality in the LME aluminum and copper volatility processes through a FIGARCH lens. *Journal of Futures Markets 28*.

Park, J. (2018). *Volatility Transmission between Oil and LME Futures.* (Vol. Applied Economics and Finance 5).

Park, J. (2019). The Role of Cancelled Warrants in the LME Market. *International Journal of Financial Studies 7*.

Park, J. (2019). The Role of Cancelled Warrants in the LME Market. *International Journal of Financial Studies 7*, 10.

Park, J. a. (2018). *Testing Efficiency of the London Metal Exchange: New Evidence.* International Journal of Financial Studies 6.

Whaley, R. E. (2009). *Understanding the vix. Journal of Portfolio Management.*

ANNEX: PYTHON CODE

```
#!/usr/bin/env python3
# -*- coding: utf-8 -*-
"""
Created on Thu Jun 13 12:06:51 2024

@author: aniagallardo
"""

# Import libraries
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
import statsmodels.api as sm
from scipy.stats import skew, kurtosis
from statsmodels.graphics.tsaplots import plot_pacf


#Augmented dickey fuller test for predictors
from statsmodels.tsa.stattools import adfuller

def test_stationarity(ts):
    # Perform the ADF test
    adf_test = adfuller(ts, autolag='AIC')
    # Output ADF test results
    print('ADF Statistic: %f' % adf_test[0])
    print('p-value: %f' % adf_test[1])
    print('Critical Values:')
    for key, value in adf_test[4].items():
        print('\t%s: %.3f' % (key, value))


#Read excel file LME calculations -> VIX
df_VIX=pd.read_excel('Return calculations.xlsx', sheet_name='3MO
FUTURES return', \
                      usecols=[1,14],skiprows=6)
# Convert 'Date' column to datetime format
df_VIX['Date'] = pd.to_datetime(df_VIX['Date'], errors='coerce')

 # Drop rows where 'Date' is NaT
```

```
df_VIX = df_VIX.dropna(subset=['Date'])
df_VIX= df_VIX.sort_values(by='Date', ascending=True)
df_VIX.set_index('Date', inplace=True)

# Display the columns to verify
print(df_VIX.columns)

# Double-check the output
print(df_VIX.head())
print(df_VIX.tail())
df_VIX['VIX'].plot();plt.ylabel('-')
plt.title('VIX Index')
plt.show()

#Read excel file LME calculations -> BDI
df_BDI=pd.read_excel('Return calculations.xlsx', sheet_name='3MO
FUTURES return', \
                      usecols=[1,15],skiprows=6)
# Convert 'Date' column to datetime format
df_BDI['Date'] = pd.to_datetime(df_BDI['Date'], errors='coerce')

 # Drop rows where 'Date' is NaT
df_BDI = df_BDI.dropna(subset=['Date'])
df_BDI= df_BDI.sort_values(by='Date', ascending=True)
df_BDI.set_index('Date', inplace=True)

# Display the columns to verify
print(df_BDI.columns)

# Double-check the output
print(df_BDI.head())
print(df_BDI.tail())
df_BDI['BDI'].plot();plt.ylabel('-')
plt.title('BDI Index')
plt.show()



#Read excel file LME calculations -> PMI
df_PMI=pd.read_excel('Return calculations.xlsx', sheet_name='3MO
FUTURES return', \
                      usecols=[19,20],skiprows=6)
print(df_PMI.columns)

# Convert 'Date' column to datetime format
df_PMI['Date1'] = pd.to_datetime(df_PMI['Date1'], errors='coerce')
# Drop rows where 'Date' is NaT
df_PMI = df_PMI.dropna(subset=['Date1'])
df_PMI= df_PMI.sort_values(by='Date1', ascending=True)
df_PMI.set_index('Date1', inplace=True)

# Display the columns to verify
print(df_PMI.columns)

# Double-check the output
print(df_PMI.head())
print(df_PMI.tail())
```

```
df_PMI['PMI'].plot();plt.ylabel('-')
plt.title('PMI Index')
plt.show()

#Read excel file LME calculations -> SPX
df_SPX=pd.read_excel('Return calculations.xlsx', sheet_name='3MO
FUTURES return', \
                      usecols=[1,17],skiprows=6)
 # Convert 'Date' column to datetime format
df_SPX['Date'] = pd.to_datetime(df_SPX['Date'], errors='coerce')

# Drop rows where 'Date' is NaT
df_SPX = df_SPX.dropna(subset=['Date'])
df_SPX= df_SPX.sort_values(by='Date', ascending=True)
df_SPX.set_index('Date', inplace=True)

# Display the columns to verify
print(df_SPX.columns)

# Double-check the output
print(df_SPX.head())
print(df_SPX.tail())
df_SPX['SPX'].plot();plt.ylabel('-')
plt.title('SPX Index')
plt.show()

#Read excel file LME calculations -> LMZSDY Zinc spot price
df_LMZSDY=pd.read_excel('Return calculations.xlsx', sheet_name='LME
spot', \
                      usecols=[0,1,2],skiprows=6)
 # Convert 'Date' column to datetime format
df_LMZSDY['Date'] = pd.to_datetime(df_LMZSDY['Date'], errors='coerce')

# Drop rows where 'Date' is NaT
df_LMZSDY = df_LMZSDY.dropna(subset=['Date'])
df_LMZSDY= df_LMZSDY.sort_values(by='Date', ascending=True)
df_LMZSDY.set_index('Date', inplace=True)


# Display the columns to verify
print(df_LMZSDY.columns)

# Double-check the output
print(df_LMZSDY.head())
print(df_LMZSDY.tail())

df_LMZSDY['PX_SETTLE'].plot();plt.ylabel('$/ton')
plt.title('Zinc spot price')
plt.show()

df_LMZSDY['PX_VOLUME'].plot();plt.ylabel('ton')
plt.title('Zinc spot volume')
plt.show()

#Read excel file LME calculations -> LMZSDS03 Zinc 3 month futures
settle price
```

```
df_LMZSDS03=pd.read_excel('Return calculations.xlsx', sheet_name='3MO 
FUTURES return', \
                        usecols=[1,3,18],skiprows=6)
 # Convert 'Date' column to datetime format
df_LMZSDS03['Date'] = pd.to_datetime(df_LMZSDS03['Date'], 
errors='coerce')

# Drop rows where 'Date' is NaT
df_LMZSDS03 = df_LMZSDS03.dropna(subset=['Date'])
df_LMZSDS03= df_LMZSDS03.sort_values(by='Date', ascending=True)
df_LMZSDS03.set_index('Date', inplace=True)


# Display the columns to verify
print(df_LMZSDS03.columns)

# Double-check the output
print(df_LMZSDS03.head())
print(df_LMZSDS03.tail())

df_LMZSDS03['PX_SETTLE'].plot();plt.ylabel('$/ton')
plt.title('Zinc 3 month futures settle price')
plt.show()

df_LMZSDS03['PX_VOLUME'].plot();plt.ylabel('ton')
plt.title('Zinc 3 month futures volume')
plt.show()

#Read excel file LME calculations -> LME Closing stock (CLS) of zinc 
in global (metric ton) - both on warrant and clolsed warrant
df_NLSZS=pd.read_excel('Return calculations.xlsx', sheet_name='LME 
Closing stock', \
                        usecols=range(2),skiprows=6)
df_NLSZS= df_NLSZS.sort_values(by='Date', ascending=True)
df_NLSZS.set_index('Date', inplace=True)

df_NLSZS.rename(columns={'PX_LAST': 'NLSZS'}, inplace=True)
NLSZS_WeeklyDiff=df_NLSZS.shift(+1)-df_NLSZS
df_NLSZS['NLSZS_WeeklyDiff']=NLSZS_WeeklyDiff

# Display the columns to verify
print(df_NLSZS.columns)

# Double-check the output and plot
print(df_NLSZS.head())
print(df_NLSZS.tail())
df_NLSZS['NLSZS'].plot();plt.ylabel('-')
plt.title('LME Closing stock (CLS) of zinc in global (metric ton) - 
both on warrant and clolsed warrant')
plt.show()

df_NLSZS['NLSZS_WeeklyDiff'].plot();plt.ylabel('NLSZS_WeeeklyDiff')
plt.title('Weekly Diff of LME Closing stock (CLS) of zinc in global 
(metric ton) - both on warrant and clolsed warrant')
plt.show()
```

```
##########################################################################
#Risk reduction Open Interest for different categories of the COT report
##########################################################################

# ZSCTTVHP Index: Zinc LME COT Commercial Undertakings Long Pct of Open Interest Risk Reduction
#   (Bloomberg Ticker=ZSCTTVHP Index)

df_ZSCTTVHP = pd.read_excel('Return calculations.xlsx', sheet_name='Commercial undertakings', \
                    usecols=range(3),skiprows=5)
df_ZSCTTVHP = df_ZSCTTVHP.sort_values(by='Date', ascending=True)
df_ZSCTTVHP.set_index('Date', inplace=True)
df_ZSCTTVHP.rename(columns={'PX_LAST': 'ZSCTTVHP'}, inplace=True)
df_ZSCTTVHP.rename(columns={'CHG_NET_1D': 'ZSCTTVHP_WeeklyDiff'}, inplace=True)

# Double-check the output
print(df_ZSCTTVHP.head())
print(df_ZSCTTVHP.tail())
df_ZSCTTVHP['ZSCTTVHP'].plot();plt.ylabel('ZSCTTVHP')
plt.title('Zinc LME COT Commercial Undertakings Long Pct of Open Interest Risk Reduction')
plt.show()
df_ZSCTTVHP['ZSCTTVHP_WeeklyDiff'].plot();plt.ylabel('ZSCTTVHP_Weeekly Diff')
plt.title('Weekly Diff of Zinc LME COT Commercial Undertakings Long Pct of Open Interest Risk Reduction')
plt.show()



# ZSCTNLSL Index: Zinc LME COT Commercial Undertakings Short Pct of Open Interest Risk Reduction
#   (Bloomberg Ticker=ZSCTNLSL Index)

df_ZSCTNLSL = pd.read_excel('Return calculations.xlsx', sheet_name='Commercial undertakings', \
                    usecols=[0,5,6],skiprows=5)
print(df_ZSCTNLSL.columns)
df_ZSCTNLSL = df_ZSCTNLSL.sort_values(by='Date', ascending=True)
df_ZSCTNLSL.set_index('Date', inplace=True)
df_ZSCTNLSL.rename(columns={'PX_LAST.1': 'ZSCTNLSL'}, inplace=True)
df_ZSCTNLSL.rename(columns={'CHG_NET_1D.1': 'ZSCTNLSL_WeeklyDiff'}, inplace=True)

# Double-check the output
print(df_ZSCTNLSL.head())
print(df_ZSCTNLSL.tail())
df_ZSCTNLSL['ZSCTNLSL'].plot();plt.ylabel('ZSCTNLSL')
plt.title('Zinc LME COT Commercial Undertakings Short Pct of Open Interest Risk Reduction')
```

```
plt.show()
df_ZSCTNLSL['ZSCTNLSL_WeeklyDiff'].plot();plt.ylabel('ZSCTNLSL_Weeekly
Diff')
plt.title('Weekly Diff of Zinc LME COT Commercial Undertakings Short
Pct of Open Interest Risk Reduction')
plt.show()

#Hedging pressure

df_ComUnd_HP = pd.read_excel('Return calculations.xlsx',
sheet_name='Commercial undertakings', \
                   usecols=[0,9],skiprows=5)
print(df_ComUnd_HP.columns)
df_ComUnd_HP = df_ComUnd_HP.sort_values(by='Date', ascending=True)
df_ComUnd_HP.set_index('Date', inplace=True)
print(df_ComUnd_HP.head())
print(df_ComUnd_HP.tail())
df_ComUnd_HP['Commercial Undertakings Net Hedging
Pressure'].plot();plt.ylabel('')
plt.title('Zinc LME COT Commercial Undertakings Net Hedging Pressure')
plt.show()

#Comparison graphs
df_ZSCTTVHP['ZSCTTVHP'].plot();
df_ZSCTNLSL['ZSCTNLSL'].plot();
plt.ylabel('Percentage')
plt.legend(loc='upper left')
plt.title('Zinc LME COT Commercial Undertakings Short and Long Pct of
Open Interest Risk Reduction')
plt.show()

pd.merge(df_ZSCTTVHP,df_ZSCTNLSL,on='Date',how='inner').plot.scatter(x
='ZSCTTVHP', y='ZSCTNLSL');plt.ylabel('ZSCTNLSL - Short
positions');plt.xlabel('ZSCTTVHP - Long positions')
plt.title('Scatter Plot - Original Values')
pd.merge(df_ZSCTTVHP,df_ZSCTNLSL,on='Date',how='inner').plot.scatter(x
='ZSCTTVHP_WeeklyDiff', \

y='ZSCTNLSL_WeeklyDiff',);plt.ylabel('ZSCTNLSL - Weekly difference
Short positions');plt.xlabel('ZSCTTVHP - Weekly difference Long
positions')
plt.title('Scatter Plot - Weekly Differences Values')

######stats
#long

series_ZSCTTVHP = df_ZSCTTVHP['ZSCTTVHP']
test_stationarity(series_ZSCTTVHP)


series_diff_ZSCTTVHP = df_ZSCTTVHP['ZSCTTVHP_WeeklyDiff']
test_stationarity(series_diff_ZSCTTVHP)

plot_pacf(series_ZSCTTVHP)
plot_pacf(series_diff_ZSCTTVHP)

#short
```

```
series_ZSCTNLSL = df_ZSCTNLSL['ZSCTNLSL']
test_stationarity(series_ZSCTNLSL)

series_diff_ZSCTNLSL = df_ZSCTNLSL['ZSCTNLSL_WeeklyDiff']
test_stationarity(series_diff_ZSCTNLSL)

plot_pacf(series_ZSCTNLSL)
plot_pacf(series_diff_ZSCTNLSL)

## CORRELATION COEFFICIENTS
# OG - Create a DataFrame from the variables
data = {
    'ZSCTTVHP': series_ZSCTTVHP,
    'ZSCTNLSL': series_ZSCTNLSL
}
df = pd.DataFrame(data)

# Calculate the correlation coefficient
correlation_coefficient = df['ZSCTTVHP'].corr(df['ZSCTNLSL'])

# Display the correlation coefficient
print(correlation_coefficient)

# WEEKLY DIFF - Create a DataFrame from the variables
data = {
    'ZSCTTVHP_WeeklyDiff': series_diff_ZSCTTVHP,
    'ZSCTNLSL_WeeklyDiff': series_diff_ZSCTNLSL
}
df = pd.DataFrame(data)

# Calculate the correlation coefficient
correlation_coefficient =
df['ZSCTTVHP_WeeklyDiff'].corr(df['ZSCTNLSL_WeeklyDiff'])

# Display the correlation coefficient
print(correlation_coefficient)

# Calculate descriptive statistics
stats_ZSCTTVHP = df_ZSCTTVHP.describe()
stats_ZSCTNLSL = df_ZSCTNLSL.describe()

# Calculate descriptive statistics
stats_ZSCTTVHP = df_ZSCTTVHP.describe()
stats_ZSCTNLSL = df_ZSCTNLSL.describe()


# Calculate kurtosis and skewness
kurt_ZSCTTVHP = kurtosis(df_ZSCTTVHP['ZSCTTVHP'])
skew_ZSCTTVHP = skew(df_ZSCTTVHP['ZSCTTVHP'])
kurt_ZSCTTVHP_WeeklyDiff =
kurtosis(df_ZSCTTVHP['ZSCTTVHP_WeeklyDiff'])
skew_ZSCTTVHP_WeeklyDiff = skew(df_ZSCTTVHP['ZSCTTVHP_WeeklyDiff'])

kurt_ZSCTNLSL = kurtosis(df_ZSCTNLSL['ZSCTNLSL'])
skew_ZSCTNLSL = skew(df_ZSCTNLSL['ZSCTNLSL'])
```

```
kurt_ZSCTNLSL_WeeklyDiff =
kurtosis(df_ZSCTNLSL['ZSCTNLSL_WeeklyDiff'])
skew_ZSCTNLSL_WeeklyDiff = skew(df_ZSCTNLSL['ZSCTNLSL_WeeklyDiff'])

# Print descriptive statistics, kurtosis, and skewness
print("Descriptive Statistics for ZSCTTVHP:")
print(stats_ZSCTTVHP)
print(f"Kurtosis for ZSCTTVHP: {kurt_ZSCTTVHP}")
print(f"Skewness for ZSCTTVHP: {skew_ZSCTTVHP}")
print(f"Kurtosis for ZSCTTVHP_WeeklyDiff: {kurt_ZSCTTVHP_WeeklyDiff}")
print(f"Skewness for ZSCTTVHP_WeeklyDiff: {skew_ZSCTTVHP_WeeklyDiff}")
print()

print("Descriptive Statistics for ZSCTNLSL:")
print(stats_ZSCTNLSL)
print(f"Kurtosis for ZSCTNLSL: {kurt_ZSCTNLSL}")
print(f"Skewness for ZSCTNLSL: {skew_ZSCTNLSL}")
print(f"Kurtosis for ZSCTNLSL_WeeklyDiff: {kurt_ZSCTNLSL_WeeklyDiff}")
print(f"Skewness for ZSCTNLSL_WeeklyDiff: {skew_ZSCTNLSL_WeeklyDiff}")
print()
######################################################################
#########


# ZSCTNIMQ Index: Zinc LME COT Investment Firms or Credit Institutions
Long Pct of Open Interest Risk Reduction
#   (Bloomberg Ticker=ZSCTNIMQ Index)

df_ZSCTNIMQ = pd.read_excel('Return calculations.xlsx',
sheet_name='Investment Firms or Credit Inst', \
                   usecols=range(3),skiprows=5)
df_ZSCTNIMQ = df_ZSCTNIMQ.sort_values(by='Date', ascending=True)
df_ZSCTNIMQ.set_index('Date', inplace=True)
df_ZSCTNIMQ.rename(columns={'PX_LAST': 'ZSCTNIMQ'}, inplace=True)
df_ZSCTNIMQ.rename(columns={'CHG_NET_1D': 'ZSCTNIMQ_WeeklyDiff'},
inplace=True)

# Double-check the output
print(df_ZSCTNIMQ.head())
print(df_ZSCTNIMQ.tail())
df_ZSCTNIMQ['ZSCTNIMQ'].plot();plt.ylabel('ZSCTNIMQ')
plt.title('Zinc LME COT Investment Firms or Credit Institutions Long
Pct of Open Interest Risk Reduction')
plt.show()
df_ZSCTNIMQ['ZSCTNIMQ_WeeklyDiff'].plot();plt.ylabel('ZSCTNIMQ_WeeklyD
iff')
plt.title('Weekly Diff of Zinc LME COT Investment Firms or Credit
Institutions Long Pct of Open Interest Risk Reduction')
plt.show()


# ZSCTJIHS Index.xlsx: Zinc LME COT Investment Firms or Credit
Institutions Short Pct of Open Interest Risk Reduction
#   (Bloomberg Ticker=ZSCTJIHS Index)

df_ZSCTJIHS = pd.read_excel('Return calculations.xlsx',
sheet_name='Investment Firms or Credit Inst', \
```

```
                    usecols=[0,5,6],skiprows=5)
print(df_ZSCTJIHS.columns)
df_ZSCTJIHS = df_ZSCTJIHS.sort_values(by='Date', ascending=True)
df_ZSCTJIHS.set_index('Date', inplace=True)
df_ZSCTJIHS.rename(columns={'PX_LAST.1': 'ZSCTJIHS'}, inplace=True)
df_ZSCTJIHS.rename(columns={'CHG_NET_1D.1': 'ZSCTJIHS_WeeklyDiff'},
inplace=True)

# Double-check the output
print(df_ZSCTJIHS.head())
print(df_ZSCTJIHS.tail())
df_ZSCTJIHS['ZSCTJIHS'].plot();plt.ylabel('ZSCTJIHS')
plt.title('Zinc LME COT Investment Firms or Credit Institutions Short
Pct of Open Interest Risk Reduction')
plt.show()
df_ZSCTJIHS['ZSCTJIHS_WeeklyDiff'].plot();plt.ylabel('ZSCTJIHS_Weeekly
Diff')
plt.title('Weekly Diff of Zinc LME COT Investment Firms or Credit
Institutions Short Pct of Open Interest Risk Reduction')
plt.show()

#Hedging pressure

df_InvFirm_HP = pd.read_excel('Return calculations.xlsx',
sheet_name='Investment Firms or Credit Inst', \
                    usecols=[0,9],skiprows=5)

df_InvFirm_HP = df_InvFirm_HP.sort_values(by='Date', ascending=True)
df_InvFirm_HP.set_index('Date', inplace=True)

print(df_InvFirm_HP.head())
print(df_InvFirm_HP.tail())
df_InvFirm_HP['Investment Firms or Credit Institution's net hedging
pressure'].plot();plt.ylabel('')
plt.title('Zinc LME COT Investment Firms or Credit Institutions Net
Hedging Pressure')
plt.show()

#Comparison graphs
df_ZSCTNIMQ['ZSCTNIMQ'].plot();
df_ZSCTJIHS['ZSCTJIHS'].plot();
plt.ylabel('Percentage')
plt.legend(loc='upper left')
plt.title('Zinc LME COT Investment Firms or Credit Institution's Short
and Long Pct of Open Interest Risk Reduction')
plt.show()

pd.merge(df_ZSCTNIMQ,df_ZSCTJIHS,on='Date',how='inner').plot.scatter(x
='ZSCTNIMQ', y='ZSCTJIHS');plt.ylabel('ZSCTJIHS - Short
positions');plt.xlabel('ZSCTNIMQ - Long positions')
plt.title('Scatter Plot - Original Values')
pd.merge(df_ZSCTNIMQ,df_ZSCTJIHS,on='Date',how='inner').plot.scatter(x
='ZSCTNIMQ_WeeklyDiff', \

y='ZSCTJIHS_WeeklyDiff',);plt.ylabel('ZSCTJIHS - Weekly difference
Short positions');plt.xlabel('ZSCTNIMQ - Weekly difference Long
positions')
```

```
plt.title('Scatter Plot - Weekly Differences Values')


#####stats
#long

series_ZSCTNIMQ = df_ZSCTNIMQ['ZSCTNIMQ']
test_stationarity(series_ZSCTNIMQ)

series_diff_ZSCTNIMQ = df_ZSCTNIMQ['ZSCTNIMQ_WeeklyDiff']
test_stationarity(series_diff_ZSCTNIMQ)

plot_pacf(series_ZSCTNIMQ)
plot_pacf(series_diff_ZSCTNIMQ)

#short

series_ZSCTJIHS = df_ZSCTJIHS['ZSCTJIHS']
test_stationarity(series_ZSCTJIHS)

series_diff_ZSCTJIHS = df_ZSCTJIHS['ZSCTJIHS_WeeklyDiff']
test_stationarity(series_diff_ZSCTJIHS)

plot_pacf(series_ZSCTJIHS)
plot_pacf(series_diff_ZSCTJIHS)

## CORRELATION COEFFICIENTS
# OG - Create a DataFrame from the variables
data = {
    'ZSCTNIMQ': series_ZSCTNIMQ,
    'ZSCTJIHS': series_ZSCTJIHS
}
df = pd.DataFrame(data)

# Calculate the correlation coefficient
correlation_coefficient = df['ZSCTNIMQ'].corr(df['ZSCTJIHS'])

# Display the correlation coefficient
print(correlation_coefficient)

# WEEKLY DIFF - Create a DataFrame from the variables
data = {
    'ZSCTNIMQ_WeeklyDiff': series_diff_ZSCTNIMQ,
    'ZSCTJIHS_WeeklyDiff': series_diff_ZSCTJIHS
}
df = pd.DataFrame(data)

# Calculate the correlation coefficient
correlation_coefficient =
df['ZSCTNIMQ_WeeklyDiff'].corr(df['ZSCTJIHS_WeeklyDiff'])

# Display the correlation coefficient
print(correlation_coefficient)

# Calculate descriptive statistics
stats_ZSCTNIMQ = df_ZSCTNIMQ.describe()
stats_ZSCTJIHS = df_ZSCTJIHS.describe()
```

```
# Calculate kurtosis and skewness
kurt_ZSCTNIMQ = kurtosis(df_ZSCTNIMQ['ZSCTNIMQ'])
skew_ZSCTNIMQ = skew(df_ZSCTNIMQ['ZSCTNIMQ'])
kurt_ZSCTNIMQ_WeeklyDiff =
kurtosis(df_ZSCTNIMQ['ZSCTNIMQ_WeeklyDiff'])
skew_ZSCTNIMQ_WeeklyDiff = skew(df_ZSCTNIMQ['ZSCTNIMQ_WeeklyDiff'])

kurt_ZSCTJIHS = kurtosis(df_ZSCTJIHS['ZSCTJIHS'])
skew_ZSCTJIHS = skew(df_ZSCTJIHS['ZSCTJIHS'])
kurt_ZSCTJIHS_WeeklyDiff =
kurtosis(df_ZSCTJIHS['ZSCTJIHS_WeeklyDiff'])
skew_ZSCTJIHS_WeeklyDiff = skew(df_ZSCTJIHS['ZSCTJIHS_WeeklyDiff'])

# Print descriptive statistics, kurtosis, and skewness
print("Descriptive Statistics for ZSCTNIMQ:")
print(stats_ZSCTNIMQ)
print(f"Kurtosis for ZSCTNIMQ: {kurt_ZSCTNIMQ}")
print(f"Skewness for ZSCTNIMQ: {skew_ZSCTNIMQ}")
print(f"Kurtosis for ZSCTNIMQ_WeeklyDiff: {kurt_ZSCTNIMQ_WeeklyDiff}")
print(f"Skewness for ZZSCTNIMQ_WeeklyDiff:
{skew_ZSCTNIMQ_WeeklyDiff}")
print()

print("Descriptive Statistics for ZSCTJIHS:")
print(stats_ZSCTJIHS)
print(f"Kurtosis for ZSCTJIHS: {kurt_ZSCTJIHS}")
print(f"Skewness for ZSCTJIHS: {skew_ZSCTJIHS}")
print(f"Kurtosis for ZSCTJIHS_WeeklyDiff: {kurt_ZSCTJIHS_WeeklyDiff}")
print(f"Skewness for ZSCTJIHS_WeeklyDiff: {skew_ZSCTJIHS_WeeklyDiff}")
print()

######################################################################
#########

# ZSCTELRX Index: Zinc LME COT Investment Funds Long Pct of Open
Interest Risk Reduction
#  (Bloomberg Ticker=ZSCTELRX Index)

df_ZSCTELRX = pd.read_excel('Return calculations.xlsx',
sheet_name='Investment funds', \
                    usecols=range(3),skiprows=5)
print(df_ZSCTELRX.columns)
df_ZSCTELRX = df_ZSCTELRX.sort_values(by='Date', ascending=True)
df_ZSCTELRX.set_index('Date', inplace=True)
df_ZSCTELRX.rename(columns={'PX_LAST': 'ZSCTELRX'}, inplace=True)
df_ZSCTELRX.rename(columns={'CHG_NET_1D': 'ZSCTELRX_WeeklyDiff'},
inplace=True)

# Double-check the output
print(df_ZSCTELRX.head())
print(df_ZSCTELRX.tail())
df_ZSCTELRX['ZSCTELRX'].plot();plt.ylabel('ZSCTELRX')
plt.title('Zinc LME COT Investment Funds Long Pct of Open Interest
Risk Reduction')
plt.show()
```

```
df_ZSCTELRX['ZSCTELRX_WeeklyDiff'].plot();plt.ylabel('ZSCTELRX_WeeklyD
iff')
plt.title('Weekly Diff of Zinc LME COT Investment Funds Long Pct of
Open Interest Risk Reduction')
plt.show()


# ZSCTYSRP Index.xlsx: Zinc LME COT Investment Funds Short Pct of Open
Interest Risk Reduction
#  (Bloomberg Ticker=ZSCTYSRP Index)

df_ZSCTYSRP = pd.read_excel('Return calculations.xlsx',
sheet_name='Investment funds', \
                   usecols=[0,5,6],skiprows=5)
print(df_ZSCTYSRP.columns)
df_ZSCTYSRP = df_ZSCTYSRP.sort_values(by='Date', ascending=True)
df_ZSCTYSRP.set_index('Date', inplace=True)
df_ZSCTYSRP.rename(columns={'PX_LAST.1': 'ZSCTYSRP'}, inplace=True)
df_ZSCTYSRP.rename(columns={'CHG_NET_1D.1': 'ZSCTYSRP_WeeklyDiff'},
inplace=True)

# Double-check the output
print(df_ZSCTYSRP.head())
print(df_ZSCTYSRP.tail())
df_ZSCTYSRP['ZSCTYSRP'].plot();plt.ylabel('ZSCTYSRP')
plt.title('Zinc LME COT Investment Funds Short Pct of Open Interest
Risk Reduction')
plt.show()
df_ZSCTYSRP['ZSCTYSRP_WeeklyDiff'].plot();plt.ylabel('ZSCTYSRP_Weeekly
Diff')
plt.title('Weekly Diff of Zinc LME COT Investment Funds Short Pct of
Open Interest Risk Reduction')
plt.show()

#Hedging pressure
df_InvFund_HP = pd.read_excel('Return calculations.xlsx',
sheet_name='Investment funds', \
                   usecols=[0,9],skiprows=5)
df_InvFund_HP = df_InvFund_HP.sort_values(by='Date', ascending=True)
df_InvFund_HP.set_index('Date', inplace=True)
print(df_InvFund_HP.head())
print(df_InvFund_HP.tail())
df_InvFund_HP['Investment funds net hedging
pressure'].plot();plt.ylabel('')
plt.title('Zinc LME COT Investment Funds Net Hedging Pressure')
plt.show()

#Comparison graphs
df_ZSCTELRX['ZSCTELRX'].plot();
df_ZSCTYSRP['ZSCTYSRP'].plot();
plt.ylabel('Percentage')
plt.legend(loc='upper left')
plt.title('Zinc LME COT Investment Funds Short and Long Pct of Open
Interest Risk Reduction')
plt.show()
```

```
pd.merge(df_ZSCTELRX,df_ZSCTYSRP,on='Date',how='inner').plot.scatter(x
='ZSCTELRX', y='ZSCTYSRP');plt.ylabel('ZSCTYSRP - Short
positions');plt.xlabel('ZSCTELRX - Long positions')
plt.title('Scatter Plot - Original Values')
pd.merge(df_ZSCTELRX,df_ZSCTYSRP,on='Date',how='inner').plot.scatter(x
='ZSCTELRX_WeeklyDiff', \

y='ZSCTYSRP_WeeklyDiff',);plt.ylabel('ZSCTYSRP - Weekly difference
Short positions');plt.xlabel('ZSCTELRX - Weekly difference Long
positions')
plt.title('Scatter Plot - Weekly Differences Values')

#####stats
#long

series_ZSCTELRX = df_ZSCTELRX['ZSCTELRX']
test_stationarity(series_ZSCTELRX)

series_diff_ZSCTELRX = df_ZSCTELRX['ZSCTELRX_WeeklyDiff']
test_stationarity(series_diff_ZSCTELRX)

plot_pacf(series_ZSCTELRX)
plot_pacf(series_diff_ZSCTELRX)

#short

series_ZSCTYSRP = df_ZSCTYSRP['ZSCTYSRP']
test_stationarity(series_ZSCTYSRP)

series_diff_ZSCTYSRP= df_ZSCTYSRP['ZSCTYSRP_WeeklyDiff']
test_stationarity(series_diff_ZSCTYSRP)

plot_pacf(series_ZSCTYSRP)
plot_pacf(series_diff_ZSCTYSRP)

## CORRELATION COEFFICIENTS
# OG - Create a DataFrame from the variables
data = {
    'ZSCTELRX': series_ZSCTELRX,
    'ZSCTYSRP': series_ZSCTYSRP
}
df = pd.DataFrame(data)

# Calculate the correlation coefficient
correlation_coefficient = df['ZSCTELRX'].corr(df['ZSCTYSRP'])

# Display the correlation coefficient
print(correlation_coefficient)

# WEEKLY DIFF - Create a DataFrame from the variables
data = {
    'ZSCTELRX_WeeklyDiff': series_diff_ZSCTELRX,
    'ZSCTYSRP_WeeklyDiff': series_diff_ZSCTYSRP
}
df = pd.DataFrame(data)

# Calculate the correlation coefficient
```

```
correlation_coefficient =
df['ZSCTELRX_WeeklyDiff'].corr(df['ZSCTYSRP_WeeklyDiff'])

# Display the correlation coefficient
print(correlation_coefficient)

# Calculate descriptive statistics
stats_ZSCTELRX = df_ZSCTELRX.describe()
stats_ZSCTYSRP = df_ZSCTYSRP.describe()

# Calculate kurtosis and skewness
kurt_ZSCTELRX = kurtosis(df_ZSCTELRX['ZSCTELRX'])
skew_ZSCTELRX = skew(df_ZSCTELRX['ZSCTELRX'])
kurt_ZSCTELRX_WeeklyDiff =
kurtosis(df_ZSCTELRX['ZSCTELRX_WeeklyDiff'])
skew_ZSCTELRX_WeeklyDiff = skew(df_ZSCTELRX['ZSCTELRX_WeeklyDiff'])

kurt_ZSCTYSRP = kurtosis(df_ZSCTYSRP['ZSCTYSRP'])
skew_ZSCTYSRP = skew(df_ZSCTYSRP['ZSCTYSRP'])
kurt_ZSCTYSRP_WeeklyDiff =
kurtosis(df_ZSCTYSRP['ZSCTYSRP_WeeklyDiff'])
skew_ZSCTYSRP_WeeklyDiff = skew(df_ZSCTYSRP['ZSCTYSRP_WeeklyDiff'])

# Print descriptive statistics, kurtosis, and skewness
print("Descriptive Statistics for ZSCTELRX:")
print(stats_ZSCTELRX)
print(f"Kurtosis for ZSCTELRX: {kurt_ZSCTELRX}")
print(f"Skewness for ZSCTELRX: {skew_ZSCTELRX}")
print(f"Kurtosis for ZSCTELRX_WeeklyDiff: {kurt_ZSCTELRX_WeeklyDiff}")
print(f"Skewness for ZSCTELRX_WeeklyDiff: {skew_ZSCTELRX_WeeklyDiff}")
print()

print("Descriptive Statistics for ZSCTYSRP:")
print(stats_ZSCTYSRP)
print(f"Kurtosis for ZSCTYSRP: {kurt_ZSCTYSRP}")
print(f"Skewness for ZSCTYSRP: {skew_ZSCTYSRP}")
print(f"Kurtosis for ZSCTYSRP_WeeklyDiff: {kurt_ZSCTYSRP_WeeklyDiff}")
print(f"Skewness for ZSCTYSRP_WeeklyDiff: {skew_ZSCTYSRP_WeeklyDiff}")
print()


######################################################################
#########

# ZSCTALBA Index: Zinc LME COT Other Financial Institutions Long Pct
of Open Interest Risk Reduction
#  (Bloomberg Ticker=ZSCTALBA Index)

df_ZSCTALBA = pd.read_excel('Return calculations.xlsx',
sheet_name='Other financial institutions', \
                   usecols=range(3),skiprows=5)
print(df_ZSCTALBA.columns)
df_ZSCTALBA = df_ZSCTALBA.sort_values(by='Date', ascending=True)
df_ZSCTALBA.set_index('Date', inplace=True)
df_ZSCTALBA.rename(columns={'PX_LAST': 'ZSCTALBA'}, inplace=True)
df_ZSCTALBA.rename(columns={'CHG_NET_1D': 'ZSCTALBA_WeeklyDiff'},
inplace=True)
```

```
# Double-check the output
print(df_ZSCTALBA.head())
print(df_ZSCTALBA.tail())
df_ZSCTALBA['ZSCTALBA'].plot();plt.ylabel('ZSCTALBA')
plt.title('Zinc LME COT Other Financial Institutions Long Pct of Open
Interest Risk Reduction')
plt.show()
df_ZSCTALBA['ZSCTALBA_WeeklyDiff'].plot();plt.ylabel('ZSCTALBA_WeeklyD
iff')
plt.title('Weekly Diff of Zinc LME COT Other Financial Institutions
Long Pct of Open Interest Risk Reduction')
plt.show()


# ZSCTWQJC Index.xlsx: Zinc LME COT Other Financial Institutions Short
Pct of Open Interest Risk Reduction
#  (Bloomberg Ticker=ZSCTWQJC Index)

df_ZSCTWQJC = pd.read_excel('Return calculations.xlsx',
sheet_name='Other financial institutions', \
                    usecols=[0,5,6],skiprows=5)
print(df_ZSCTWQJC.columns)
df_ZSCTWQJC = df_ZSCTWQJC.sort_values(by='Date', ascending=True)
df_ZSCTWQJC.set_index('Date', inplace=True)
df_ZSCTWQJC.rename(columns={'PX_LAST.1': 'ZSCTWQJC'}, inplace=True)
df_ZSCTWQJC.rename(columns={'CHG_NET_1D.1': 'ZSCTWQJC_WeeklyDiff'},
inplace=True)

# Double-check the output
print(df_ZSCTWQJC.head())
print(df_ZSCTWQJC.tail())
df_ZSCTWQJC['ZSCTWQJC'].plot();plt.ylabel('ZSCTWQJC')
plt.title('Zinc LME COT Other financial institutions Short Pct of Open
Interest Risk Reduction')
plt.show()
df_ZSCTWQJC['ZSCTWQJC_WeeklyDiff'].plot();plt.ylabel('ZSCTWQJC_Weeekly
Diff')
plt.title('Weekly Diff of Zinc LME COT Other financial institutions
Short Pct of Open Interest Risk Reduction')
plt.show()

#Hedging pressure
df_OtherFin_HP = pd.read_excel('Return calculations.xlsx',
sheet_name='Other financial institutions', \
                    usecols=[0,9],skiprows=5)

df_OtherFin_HP = df_OtherFin_HP.sort_values(by='Date', ascending=True)
df_OtherFin_HP.set_index('Date', inplace=True)

print(df_OtherFin_HP .head())
print(df_OtherFin_HP .tail())
df_OtherFin_HP ['Other Financial institutions net hedging
pressure'].plot();plt.ylabel('')
plt.title('Zinc LME COT Other Financial institutions Net Hedging
Pressure')
plt.show()
```

```
#Comparison graphs
df_ZSCTALBA['ZSCTALBA'].plot();
df_ZSCTWQJC['ZSCTWQJC'].plot();
plt.ylabel('Percentage')
plt.legend(loc='upper left')
plt.title('Zinc LME COT Other Financial Institutions Short and Long
Pct of Open Interest Risk Reduction')
plt.show()

pd.merge(df_ZSCTALBA,df_ZSCTWQJC,on='Date',how='inner').plot.scatter(x
='ZSCTALBA', y='ZSCTWQJC');plt.ylabel('ZSCTWQJC - Short
positions');plt.xlabel('ZSCTALBA - Long positions')
plt.title('Scatter Plot - Original Values')
pd.merge(df_ZSCTALBA,df_ZSCTWQJC,on='Date',how='inner').plot.scatter(x
='ZSCTALBA_WeeklyDiff', \

y='ZSCTWQJC_WeeklyDiff',);plt.ylabel('ZSCTWQJC - Weekly difference
Short positions');plt.xlabel('ZSCTALBA - Weekly difference Long
positions')
plt.title('Scatter Plot - Weekly Differences Values')

######stats
#long

series_ZSCTALBA = df_ZSCTALBA['ZSCTALBA']
test_stationarity(series_ZSCTALBA)

series_diff_ZSCTALBA = df_ZSCTALBA['ZSCTALBA_WeeklyDiff']
test_stationarity(series_diff_ZSCTALBA)

plot_pacf(series_ZSCTALBA)
plot_pacf(series_diff_ZSCTALBA)

#short

series_ZSCTWQJC = df_ZSCTWQJC['ZSCTWQJC']
test_stationarity(series_ZSCTWQJC)

series_diff_ZSCTWQJC= df_ZSCTWQJC['ZSCTWQJC_WeeklyDiff']
test_stationarity(series_diff_ZSCTWQJC)

plot_pacf(series_ZSCTWQJC)
plot_pacf(series_diff_ZSCTWQJC)

## CORRELATION COEFFICIENTS
# OG - Create a DataFrame from the variables
data = {
    'ZSCTALBA': series_ZSCTALBA,
    'ZSCTWQJC': series_ZSCTWQJC
}
df = pd.DataFrame(data)

# Calculate the correlation coefficient
correlation_coefficient = df['ZSCTALBA'].corr(df['ZSCTWQJC'])

# Display the correlation coefficient
```

```
print(correlation_coefficient)

# WEEKLY DIFF - Create a DataFrame from the variables
data = {
    'ZSCTALBA_WeeklyDiff': series_diff_ZSCTALBA,
    'ZSCTWQJC_WeeklyDiff': series_diff_ZSCTWQJC
}
df = pd.DataFrame(data)

# Calculate the correlation coefficient
correlation_coefficient =
df['ZSCTALBA_WeeklyDiff'].corr(df['ZSCTWQJC_WeeklyDiff'])

# Display the correlation coefficient
print(correlation_coefficient)

# Calculate descriptive statistics
stats_ZSCTALBA = df_ZSCTALBA.describe()
stats_ZSCTWQJC = df_ZSCTWQJC.describe()

# Calculate kurtosis and skewness
kurt_ZSCTALBA = kurtosis(df_ZSCTALBA['ZSCTALBA'])
skew_ZSCTALBA = skew(df_ZSCTALBA['ZSCTALBA'])
kurt_ZSCTALBA_WeeklyDiff =
kurtosis(df_ZSCTALBA['ZSCTALBA_WeeklyDiff'])
skew_ZSCTALBA_WeeklyDiff = skew(df_ZSCTALBA['ZSCTALBA_WeeklyDiff'])

kurt_ZSCTWQJC = kurtosis(df_ZSCTWQJC['ZSCTWQJC'])
skew_ZSCTWQJC = skew(df_ZSCTWQJC['ZSCTWQJC'])
kurt_ZSCTWQJC_WeeklyDiff =
kurtosis(df_ZSCTWQJC['ZSCTWQJC_WeeklyDiff'])
skew_ZSCTWQJC_WeeklyDiff = skew(df_ZSCTWQJC['ZSCTWQJC_WeeklyDiff'])

# Print descriptive statistics, kurtosis, and skewness
print("Descriptive Statistics for ZSCTALBA:")
print(stats_ZSCTALBA)
print(f"Kurtosis for ZSCTALBA: {kurt_ZSCTALBA}")
print(f"Skewness for ZSCTALBA: {skew_ZSCTALBA}")
print(f"Kurtosis for ZSCTALBA_WeeklyDiff: {kurt_ZSCTALBA_WeeklyDiff}")
print(f"Skewness for ZSCTALBA_WeeklyDiff: {skew_ZSCTALBA_WeeklyDiff}")
print()

print("Descriptive Statistics for ZSCTWQJC:")
print(stats_ZSCTWQJC)
print(f"Kurtosis for ZSCTWQJC: {kurt_ZSCTWQJC}")
print(f"Skewness for ZSCTWQJC: {skew_ZSCTWQJC}")
print(f"Kurtosis for ZSCTWQJC_WeeklyDiff: {kurt_ZSCTWQJC_WeeklyDiff}")
print(f"Skewness for ZSCTWQJC_WeeklyDiff: {skew_ZSCTWQJC_WeeklyDiff}")
print()

######################################################################
#########

# ZSCTWZPM Index: Zinc LME COT Directive compliance obligations Long
Pct of Open Interest Risk Reduction
#  (Bloomberg Ticker=ZSCTWZPM Index)
```

```
df_ZSCTWZPM = pd.read_excel('Return calculations.xlsx',
sheet_name='Directive compliance obl', \
                   usecols=range(3),skiprows=5)
print(df_ZSCTWZPM.columns)
df_ZSCTWZPM = df_ZSCTWZPM.sort_values(by='Date', ascending=True)
df_ZSCTWZPM.set_index('Date', inplace=True)
df_ZSCTWZPM.rename(columns={'PX_LAST': 'ZSCTWZPM'}, inplace=True)
df_ZSCTWZPM.rename(columns={'CHG_NET_1D': 'ZSCTWZPM_WeeklyDiff'},
inplace=True)

# Double-check the output
print(df_ZSCTWZPM.head())
print(df_ZSCTWZPM.tail())
df_ZSCTWZPM['ZSCTWZPM'].plot();plt.ylabel('ZSCTWZPM')
plt.title('Zinc LME COT Directive compliance obligations Long Pct of
Open Interest Risk Reduction')
plt.show()
df_ZSCTWZPM['ZSCTWZPM_WeeklyDiff'].plot();plt.ylabel('ZSCTWZPM_WeeklyD
iff')
plt.title('Weekly Diff of Zinc LME COT Directive compliance
obligations Long Pct of Open Interest Risk Reduction')
plt.show()


# ZSCTXAMA Index.xlsx: Zinc LME COT Directive compliance obligations
Short Pct of Open Interest Risk Reduction
#  (Bloomberg Ticker=ZSCTXAMA Index)

df_ZSCTXAMA = pd.read_excel('Return calculations.xlsx',
sheet_name='Directive compliance obl', \
                   usecols=[0,5,6],skiprows=5)
print(df_ZSCTXAMA.columns)
df_ZSCTXAMA = df_ZSCTXAMA.sort_values(by='Date', ascending=True)
df_ZSCTXAMA.set_index('Date', inplace=True)
df_ZSCTXAMA.rename(columns={'PX_LAST.1': 'ZSCTXAMA'}, inplace=True)
df_ZSCTXAMA.rename(columns={'CHG_NET_1D.1': 'ZSCTXAMA_WeeklyDiff'},
inplace=True)

# Double-check the output
print(df_ZSCTXAMA.head())
print(df_ZSCTXAMA.tail())
df_ZSCTXAMA['ZSCTXAMA'].plot();plt.ylabel('ZSCTXAMA')
plt.title('Zinc LME COT Directive compliance obligations Short Pct of
Open Interest Risk Reduction')
plt.show()
df_ZSCTXAMA['ZSCTXAMA_WeeklyDiff'].plot();plt.ylabel('ZSCTXAMA_Weeekly
Diff')
plt.title('Weekly Diff of Zinc LME COT Directive compliance
obligations Short Pct of Open Interest Risk Reduction')
plt.show()

df_DirComp_HP = pd.read_excel('Return calculations.xlsx',
sheet_name='Directive compliance obl', \
                   usecols=[0,9],skiprows=5)

print(df_DirComp_HP .head())
print(df_DirComp_HP .tail())
```

```
df_DirComp_HP ['Directive compliance obligations net hedging
pressure'].plot();plt.ylabel('Directive compliance obligations net
hedging pressure')
plt.title('Zinc LME COT Directive compliance obligations Net Hedging
Pressure')
plt.show()

######################################################################
#########

#All financials net hedging pressure (calculated as Short-Long of
Investment Firms or Credit Inst+Investment funds+Other financial
institutions


df_AllFin_HP = pd.read_excel('Return calculations.xlsx',
sheet_name='All financials', \
                   usecols=[0,1],skiprows=1)
df_AllFin_HP = df_AllFin_HP.sort_values(by='Date', ascending=True)
df_AllFin_HP.set_index('Date', inplace=True)

print(df_AllFin_HP .head())
print(df_AllFin_HP .tail())
df_AllFin_HP ['All financials net hedging
pressure'].plot();plt.ylabel('')
plt.title('Zinc LME COT All Financials Net Hedging Pressure')
plt.show()

######################################################################
#########

#Total net hedging pressure net hedging pressure (calculated as Short-
Long of Investment Firms or Credit Inst+Investment funds+Other
financial institutions


df_Total_HP = pd.read_excel('Return calculations.xlsx',
sheet_name='Total HP', \
                   usecols=[0,1],skiprows=1)

df_Total_HP = df_Total_HP.sort_values(by='Date', ascending=True)
df_Total_HP.set_index('Date', inplace=True)
print(df_Total_HP .head())
print(df_Total_HP .tail())
df_Total_HP ['Total net hedging pressure'].plot();plt.ylabel('')
plt.title('Zinc LME COT Total Net Hedging Pressure')
plt.show()

######################################################################
#########
########################################
#Create the main variables and dataframe
########################################

#Create spot price, futures, price, and basis
LMZSDY_PX_SETTLE=df_LMZSDY['PX_SETTLE'] #LME Zinc Spot price
```

```
LMZSDS03_PX_SETTLE=df_LMZSDS03['PX_SETTLE'] #LME Zinc 3-month Futures
price
LMZSD_basis=LMZSDS03_PX_SETTLE/LMZSDY_PX_SETTLE-1 #LME Zinc 3-month-
spot basis

Oneday_ret= pd.read_excel('Return calculations.xlsx', sheet_name='3MO
FUTURES return', \
                   usecols=[1,4],skiprows=6)
print(Oneday_ret)
Twoday_ret= pd.read_excel('Return calculations.xlsx', sheet_name='3MO
FUTURES return', \
                       usecols=[5],skiprows=6)
Threeday_ret= pd.read_excel('Return calculations.xlsx',
sheet_name='3MO FUTURES return', \
                           usecols=[6],skiprows=6)
Fourday_ret= pd.read_excel('Return calculations.xlsx', sheet_name='3MO
FUTURES return', \
                       usecols=[7],skiprows=6)
Fiveday_ret= pd.read_excel('Return calculations.xlsx', sheet_name='3MO
FUTURES return', \
                           usecols=[8],skiprows=6)

Oneday_hist_ret= pd.read_excel('Return calculations.xlsx',
sheet_name='3MO FUTURES return', \
                   usecols=[9],skiprows=6)
Twoday_hist_ret= pd.read_excel('Return calculations.xlsx',
sheet_name='3MO FUTURES return', \
                       usecols=[10],skiprows=6)
Threeday_hist_ret= pd.read_excel('Return calculations.xlsx',
sheet_name='3MO FUTURES return', \
                           usecols=[11],skiprows=6)
Fourday_hist_ret= pd.read_excel('Return calculations.xlsx',
sheet_name='3MO FUTURES return', \
                       usecols=[12],skiprows=6)
Fiveday_hist_ret= pd.read_excel('Return calculations.xlsx',
sheet_name='3MO FUTURES return', \
                           usecols=[13],skiprows=6)

    ########MODEL 1##########
#REGRESSIONS total HP
Oneday_ret= pd.read_excel('Return calculations.xlsx', sheet_name='3MO
FUTURES return', \
                   usecols=[1,4],skiprows=6)
Oneday_ret = Oneday_ret.sort_values(by='Date', ascending=False)
Oneday_ret.set_index('Date', inplace=True)

start_date = pd.to_datetime('2018-02-02')
Oneday_ret_filtered = Oneday_ret.loc[Oneday_ret.index >= start_date]
Oneday_ret_friday=
Oneday_ret_filtered[Oneday_ret_filtered.index.dayofweek == 4]  # 0
corresponds to Monday
print(Oneday_ret_friday)


df_oneday_HP = pd.read_excel('Regression.xlsx', sheet_name='TOTALHP',
usecols=[0, 1, 2], skiprows=1)
df_oneday_HP = df_oneday_HP.sort_values(by='Date', ascending=True)
```

```
df_oneday_HP.set_index('Date', inplace=True)
df_oneday_HP.index = pd.to_datetime(df_oneday_HP.index)
print(df_oneday_HP)

# Extract variables for regression
X = df_oneday_HP['Total net hedging pressure']
y = df_oneday_HP['1 day log return']

X = X.reindex(y.index)


# Add a constant to the independent variable (Total HP)
X = sm.add_constant(X)

# Fit the regression model
model = sm.OLS(y, X)
results = model.fit()

# Print regression summary
print(results.summary())


#TWO DAYS
Twoday_ret= pd.read_excel('Return calculations.xlsx', sheet_name='3MO
FUTURES return', \
                   usecols=[1,5],skiprows=6)
Twoday_ret = Twoday_ret.sort_values(by='Date', ascending=False)
Twoday_ret.set_index('Date', inplace=True)

start_date = pd.to_datetime('2018-02-02')
Twoday_ret_filtered = Twoday_ret.loc[Twoday_ret.index >= start_date]
Twoday_ret_friday =
Twoday_ret_filtered[Twoday_ret_filtered.index.dayofweek == 4]  # 0
corresponds to Monday
print(Twoday_ret_friday)


df_twoday_HP = pd.read_excel('Regression.xlsx', sheet_name='TOTALHP',
usecols=[0, 1, 3], skiprows=1)
df_twoday_HP = df_twoday_HP.sort_values(by='Date', ascending=True)
df_twoday_HP.set_index('Date', inplace=True)
df_twoday_HP.index = pd.to_datetime(df_twoday_HP.index)
print(df_twoday_HP)

# Extract variables for regression
X = df_twoday_HP['Total net hedging pressure']
y = df_twoday_HP['2 day log return']

X = X.reindex(y.index)


# Add a constant to the independent variable (Total HP)
X = sm.add_constant(X)

# Fit the regression model
model = sm.OLS(y, X)
results = model.fit()
```

```
# Print regression summary
print(results.summary())

#THREE DAYS
Threeday_ret= pd.read_excel('Return calculations.xlsx',
sheet_name='3MO FUTURES return', \
                   usecols=[1,6],skiprows=6)
Threeday_ret = Threeday_ret.sort_values(by='Date', ascending=False)
Threeday_ret.set_index('Date', inplace=True)

start_date = pd.to_datetime('2018-02-02')
Threeday_ret_filtered = Threeday_ret.loc[Twoday_ret.index >=
start_date]
Threeday_ret_friday =
Threeday_ret_filtered[Threeday_ret_filtered.index.dayofweek == 4]  # 0
corresponds to Monday
print(Threeday_ret_friday)


df_threeday_HP = pd.read_excel('Regression.xlsx',
sheet_name='TOTALHP', usecols=[0, 1, 4], skiprows=1)
df_threeday_HP = df_threeday_HP.sort_values(by='Date', ascending=True)
df_threeday_HP.set_index('Date', inplace=True)
df_threeday_HP.index = pd.to_datetime(df_threeday_HP.index)
print(df_threeday_HP)

# Extract variables for regression
X = df_threeday_HP['Total net hedging pressure']
y = df_threeday_HP['3 day log return']

X = X.reindex(y.index)


# Add a constant to the independent variable (Total HP)
X = sm.add_constant(X)

# Fit the regression model
model = sm.OLS(y, X)
results = model.fit()

# Print regression summary
print(results.summary())

#FOUR DAYS
Fourday_ret= pd.read_excel('Return calculations.xlsx', sheet_name='3MO
FUTURES return', \
                   usecols=[1,7],skiprows=6)
Fourday_ret = Fourday_ret.sort_values(by='Date', ascending=False)
Fourday_ret.set_index('Date', inplace=True)

start_date = pd.to_datetime('2018-02-02')
Fourday_ret_filtered = Fourday_ret.loc[Twoday_ret.index >= start_date]
Fourday_ret_friday =
Fourday_ret_filtered[Fourday_ret_filtered.index.dayofweek == 4]  # 0
corresponds to Monday
print(Fourday_ret_friday)
```

```
df_fourday_HP = pd.read_excel('Regression.xlsx', sheet_name='TOTALHP',
usecols=[0, 1, 5], skiprows=1)
df_fourday_HP = df_fourday_HP.sort_values(by='Date', ascending=True)
df_fourday_HP.set_index('Date', inplace=True)
df_fourday_HP.index = pd.to_datetime(df_fourday_HP.index)
print(df_fourday_HP)

# Extract variables for regression
X = df_fourday_HP['Total net hedging pressure']
y = df_fourday_HP['4 day log return']

X = X.reindex(y.index)


# Add a constant to the independent variable (Total HP)
X = sm.add_constant(X)

# Fit the regression model
model = sm.OLS(y, X)
results = model.fit()

# Print regression summary
print(results.summary())

#FIVE DAYS
Fiveday_ret= pd.read_excel('Return calculations.xlsx', sheet_name='3MO
FUTURES return', \
                         usecols=[1,8],skiprows=6)
Fiveday_ret = Fiveday_ret.sort_values(by='Date', ascending=False)
Fiveday_ret.set_index('Date', inplace=True)

start_date = pd.to_datetime('2018-02-02')
Fiveday_ret_filtered = Fiveday_ret.loc[Twoday_ret.index >= start_date]
Fiveday_ret_friday =
Fiveday_ret_filtered[Fiveday_ret_filtered.index.dayofweek == 4]   # 0
corresponds to Monday
print(Fiveday_ret_friday)


df_fiveday_HP = pd.read_excel('Regression.xlsx', sheet_name='TOTALHP',
usecols=[0, 1, 6], skiprows=1)
df_fiveday_HP = df_fiveday_HP.sort_values(by='Date', ascending=True)
df_fiveday_HP.set_index('Date', inplace=True)
df_fiveday_HP.index = pd.to_datetime(df_fiveday_HP.index)
print(df_fiveday_HP)

# Extract variables for regression
X = df_fiveday_HP['Total net hedging pressure']
y = df_fiveday_HP['5 day log return']

X = X.reindex(y.index)


# Add a constant to the independent variable (Total HP)
X = sm.add_constant(X)
```

```
# Fit the regression model
model = sm.OLS(y, X)
results = model.fit()

# Print regression summary
print(results.summary())


##COMMERCIAL UNDERTAKINGS
#ONE DAY
df_oneday_COMHP = pd.read_excel('Regression.xlsx',
sheet_name='COMMUND', usecols=[0, 1, 2], skiprows=1)
df_oneday_COMHP = df_oneday_COMHP.sort_values(by='Date',
ascending=True)
df_oneday_COMHP.set_index('Date', inplace=True)
df_oneday_COMHP.index = pd.to_datetime(df_oneday_COMHP.index)
print(df_oneday_COMHP)

# Extract variables for regression
X = df_oneday_COMHP['Commercial Undertakings Net Hedging Pressure']
y = df_oneday_COMHP['1 day log return']

X = X.reindex(y.index)


# Add a constant to the independent variable (Total HP)
X = sm.add_constant(X)

# Fit the regression model
model = sm.OLS(y, X)
results = model.fit()

# Print regression summary
print(results.summary())

#TWO DAYS
df_twoday_COMHP = pd.read_excel('Regression.xlsx',
sheet_name='COMMUND', usecols=[0, 1, 3], skiprows=1)
df_twoday_COMHP = df_twoday_COMHP.sort_values(by='Date',
ascending=True)
df_twoday_COMHP.set_index('Date', inplace=True)
df_twoday_COMHP.index = pd.to_datetime(df_twoday_COMHP.index)
print(df_twoday_COMHP)

# Extract variables for regression
X = df_twoday_COMHP['Commercial Undertakings Net Hedging Pressure']
y = df_twoday_COMHP['2 day log return']

X = X.reindex(y.index)


# Add a constant to the independent variable (Total HP)
X = sm.add_constant(X)

# Fit the regression model
model = sm.OLS(y, X)
```

```
results = model.fit()

# Print regression summary
print(results.summary())

#THREE DAYS
df_threeday_COMHP = pd.read_excel('Regression.xlsx',
sheet_name='COMMUND', usecols=[0, 1, 4], skiprows=1)
df_threeday_COMHP = df_threeday_COMHP.sort_values(by='Date',
ascending=True)
df_threeday_COMHP.set_index('Date', inplace=True)
df_threeday_COMHP.index = pd.to_datetime(df_threeday_COMHP.index)
print(df_threeday_COMHP)

# Extract variables for regression
X = df_threeday_COMHP['Commercial Undertakings Net Hedging Pressure']
y = df_threeday_COMHP['3 day log return']

X = X.reindex(y.index)


# Add a constant to the independent variable (Total HP)
X = sm.add_constant(X)

# Fit the regression model
model = sm.OLS(y, X)
results = model.fit()

# Print regression summary
print(results.summary())

#FOUR DAYS
df_fourday_COMHP = pd.read_excel('Regression.xlsx',
sheet_name='COMMUND', usecols=[0, 1, 5], skiprows=1)
df_fourday_COMHP = df_fourday_COMHP.sort_values(by='Date',
ascending=True)
df_fourday_COMHP.set_index('Date', inplace=True)
df_fourday_COMHP.index = pd.to_datetime(df_fourday_COMHP.index)
print(df_fourday_COMHP)

# Extract variables for regression
X = df_fourday_COMHP['Commercial Undertakings Net Hedging Pressure']
y = df_fourday_COMHP['4 day log return']

X = X.reindex(y.index)


# Add a constant to the independent variable (Total HP)
X = sm.add_constant(X)

# Fit the regression model
model = sm.OLS(y, X)
results = model.fit()

# Print regression summary
print(results.summary())
```

```
#FIVE DAYS
df_fiveday_COMHP = pd.read_excel('Regression.xlsx',
sheet_name='COMMUND', usecols=[0, 1, 6], skiprows=1)
df_fiveday_COMHP = df_fiveday_COMHP.sort_values(by='Date',
ascending=True)
df_fiveday_COMHP.set_index('Date', inplace=True)
df_fiveday_COMHP.index = pd.to_datetime(df_fiveday_COMHP.index)
print(df_fiveday_COMHP)

# Extract variables for regression
X = df_fiveday_COMHP['Commercial Undertakings Net Hedging Pressure']
y = df_fiveday_COMHP['5 day log return']

X = X.reindex(y.index)


# Add a constant to the independent variable (Total HP)
X = sm.add_constant(X)

# Fit the regression model
model = sm.OLS(y, X)
results = model.fit()

# Print regression summary
print(results.summary())


##INVESTMENT FIRMS
#ONE DAY
df_oneday_INVFIR = pd.read_excel('Regression.xlsx',
sheet_name='InvFirms', usecols=[0, 1, 2], skiprows=1)
df_oneday_INVFIR = df_oneday_INVFIR.sort_values(by='Date',
ascending=True)
df_oneday_INVFIR.set_index('Date', inplace=True)
df_oneday_INVFIR.index = pd.to_datetime(df_oneday_INVFIR.index)
print(df_oneday_INVFIR)

# Extract variables for regression
X = df_oneday_INVFIR['Investment Firms or Credit Institution’s net
hedging pressure']
y = df_oneday_INVFIR['1 day log return']

X = X.reindex(y.index)


# Add a constant to the independent variable (Total HP)
X = sm.add_constant(X)

# Fit the regression model
model = sm.OLS(y, X)
results = model.fit()

# Print regression summary
print(results.summary())

#TWO DAY
```

```
df_twoday_INVFIR = pd.read_excel('Regression.xlsx',
sheet_name='InvFirms', usecols=[0, 1, 3], skiprows=1)
df_twoday_INVFIR = df_twoday_INVFIR.sort_values(by='Date',
ascending=True)
df_twoday_INVFIR.set_index('Date', inplace=True)
df_twoday_INVFIR.index = pd.to_datetime(df_twoday_INVFIR.index)
print(df_twoday_INVFIR)

# Extract variables for regression
X = df_twoday_INVFIR['Investment Firms or Credit Institution's net
hedging pressure']
y = df_twoday_INVFIR['2 day log return']

X = X.reindex(y.index)


# Add a constant to the independent variable (Total HP)
X = sm.add_constant(X)

# Fit the regression model
model = sm.OLS(y, X)
results = model.fit()

# Print regression summary
print(results.summary())


#THREE DAY
df_threeday_INVFIR = pd.read_excel('Regression.xlsx',
sheet_name='InvFirms', usecols=[0, 1, 4], skiprows=1)
df_threeday_INVFIR = df_threeday_INVFIR.sort_values(by='Date',
ascending=True)
df_threeday_INVFIR.set_index('Date', inplace=True)
df_threeday_INVFIR.index = pd.to_datetime(df_threeday_INVFIR.index)
print(df_threeday_INVFIR)

# Extract variables for regression
X = df_threeday_INVFIR['Investment Firms or Credit Institution's net
hedging pressure']
y = df_threeday_INVFIR['3 day log return']

X = X.reindex(y.index)


# Add a constant to the independent variable (Total HP)
X = sm.add_constant(X)

# Fit the regression model
model = sm.OLS(y, X)
results = model.fit()

# Print regression summary
print(results.summary())


#FOUR DAY
```

```
df_fourday_INVFIR = pd.read_excel('Regression.xlsx',
sheet_name='InvFirms', usecols=[0, 1, 5], skiprows=1)
df_fourday_INVFIR = df_fourday_INVFIR.sort_values(by='Date',
ascending=True)
df_fourday_INVFIR.set_index('Date', inplace=True)
df_fourday_INVFIR.index = pd.to_datetime(df_fourday_INVFIR.index)
print(df_fourday_INVFIR)

# Extract variables for regression
X = df_fourday_INVFIR['Investment Firms or Credit Institution’s net
hedging pressure']
y = df_fourday_INVFIR['4 day log return']

X = X.reindex(y.index)


# Add a constant to the independent variable (Total HP)
X = sm.add_constant(X)

# Fit the regression model
model = sm.OLS(y, X)
results = model.fit()

# Print regression summary
print(results.summary())



#FIVE DAY
df_fiveday_INVFIR = pd.read_excel('Regression.xlsx',
sheet_name='InvFirms', usecols=[0, 1, 6], skiprows=1)
df_fiveday_INVFIR = df_fiveday_INVFIR.sort_values(by='Date',
ascending=True)
df_fiveday_INVFIR.set_index('Date', inplace=True)
df_fiveday_INVFIR.index = pd.to_datetime(df_fiveday_INVFIR.index)
print(df_fiveday_INVFIR)

# Extract variables for regression
X = df_fiveday_INVFIR['Investment Firms or Credit Institution’s net
hedging pressure']
y = df_fiveday_INVFIR['5 day log return']

X = X.reindex(y.index)


# Add a constant to the independent variable (Total HP)
X = sm.add_constant(X)

# Fit the regression model
model = sm.OLS(y, X)
results = model.fit()

# Print regression summary
print(results.summary())

##INVESTMENT FUNDS
#ONE DAY
```

```
df_oneday_INVFUN = pd.read_excel('Regression.xlsx',
sheet_name='InvFunds', usecols=[0, 1, 2], skiprows=1)
df_oneday_INVFUN = df_oneday_INVFUN.sort_values(by='Date',
ascending=True)
df_oneday_INVFUN.set_index('Date', inplace=True)
df_oneday_INVFUN.index = pd.to_datetime(df_oneday_INVFUN.index)
print(df_oneday_INVFUN)

# Extract variables for regression
X = df_oneday_INVFUN['Investment funds net hedging pressure']
y = df_oneday_INVFUN['1 day log return']

X = X.reindex(y.index)


# Add a constant to the independent variable (Total HP)
X = sm.add_constant(X)

# Fit the regression model
model = sm.OLS(y, X)
results = model.fit()

# Print regression summary
print(results.summary())

#TWO DAY
df_twoday_INVFUN = pd.read_excel('Regression.xlsx',
sheet_name='InvFunds', usecols=[0, 1, 3], skiprows=1)
df_twoday_INVFUN = df_twoday_INVFUN.sort_values(by='Date',
ascending=True)
df_twoday_INVFUN.set_index('Date', inplace=True)
df_twoday_INVFUN.index = pd.to_datetime(df_twoday_INVFUN.index)
print(df_twoday_INVFUN)

# Extract variables for regression
X = df_twoday_INVFUN['Investment funds net hedging pressure']
y = df_twoday_INVFUN['2 day log return']

X = X.reindex(y.index)


# Add a constant to the independent variable (Total HP)
X = sm.add_constant(X)

# Fit the regression model
model = sm.OLS(y, X)
results = model.fit()

# Print regression summary
print(results.summary())


#THREE DAY
df_threeday_INVFUN = pd.read_excel('Regression.xlsx',
sheet_name='InvFunds', usecols=[0, 1, 4], skiprows=1)
df_threeday_INVFUN = df_threeday_INVFUN.sort_values(by='Date',
ascending=True)
```

```python
df_threeday_INVFUN.set_index('Date', inplace=True)
df_threeday_INVFUN.index = pd.to_datetime(df_threeday_INVFUN.index)
print(df_threeday_INVFUN)

# Extract variables for regression
X = df_threeday_INVFUN['Investment funds net hedging pressure']
y = df_threeday_INVFUN['3 day log return']

X = X.reindex(y.index)


# Add a constant to the independent variable (Total HP)
X = sm.add_constant(X)

# Fit the regression model
model = sm.OLS(y, X)
results = model.fit()

# Print regression summary
print(results.summary())

#FOUR DAY
df_fourday_INVFUN = pd.read_excel('Regression.xlsx',
sheet_name='InvFunds', usecols=[0, 1, 5], skiprows=1)
df_fourday_INVFUN = df_fourday_INVFUN.sort_values(by='Date',
ascending=True)
df_fourday_INVFUN.set_index('Date', inplace=True)
df_fourday_INVFUN.index = pd.to_datetime(df_fourday_INVFUN.index)
print(df_fourday_INVFUN)

# Extract variables for regression
X = df_fourday_INVFUN['Investment funds net hedging pressure']
y = df_fourday_INVFUN['4 day log return']

X = X.reindex(y.index)


# Add a constant to the independent variable (Total HP)
X = sm.add_constant(X)

# Fit the regression model
model = sm.OLS(y, X)
results = model.fit()

# Print regression summary
print(results.summary())


#FIVE DAY
df_five_INVFUN = pd.read_excel('Regression.xlsx',
sheet_name='InvFunds', usecols=[0, 1, 6], skiprows=1)
df_five_INVFUN = df_five_INVFUN.sort_values(by='Date', ascending=True)
df_five_INVFUN.set_index('Date', inplace=True)
df_five_INVFUN.index = pd.to_datetime(df_five_INVFUN.index)
print(df_five_INVFUN)

# Extract variables for regression
```

```
X = df_five_INVFUN['Investment funds net hedging pressure']
y = df_five_INVFUN['5 day log return']

X = X.reindex(y.index)


# Add a constant to the independent variable (Total HP)
X = sm.add_constant(X)

# Fit the regression model
model = sm.OLS(y, X)
results = model.fit()

# Print regression summary
print(results.summary())


##OTHER FINANCIAL INSTITUTIONS
#ONE DAY
df_oneday_OTHFIN = pd.read_excel('Regression.xlsx',
sheet_name='OtherFin', usecols=[0, 1, 2], skiprows=1)
df_oneday_OTHFIN = df_oneday_OTHFIN.sort_values(by='Date',
ascending=True)
df_oneday_OTHFIN.set_index('Date', inplace=True)
df_oneday_OTHFIN.index = pd.to_datetime(df_oneday_OTHFIN.index)
print(df_oneday_OTHFIN)

# Extract variables for regression
X = df_oneday_OTHFIN['Other Financial institutions net hedging
pressure']
y = df_oneday_OTHFIN['1 day log return']

X = X.reindex(y.index)


# Add a constant to the independent variable (Total HP)
X = sm.add_constant(X)

# Fit the regression model
model = sm.OLS(y, X)
results = model.fit()

# Print regression summary
print(results.summary())

#TWO DAY
df_twoday_OTHFIN = pd.read_excel('Regression.xlsx',
sheet_name='OtherFin', usecols=[0, 1, 3], skiprows=1)
df_twoday_OTHFIN = df_twoday_OTHFIN.sort_values(by='Date',
ascending=True)
df_twoday_OTHFIN.set_index('Date', inplace=True)
df_twoday_OTHFIN.index = pd.to_datetime(df_twoday_OTHFIN.index)
print(df_twoday_OTHFIN)

# Extract variables for regression
X = df_twoday_OTHFIN['Other Financial institutions net hedging
pressure']
```

```
y = df_twoday_OTHFIN['2 day log return']

X = X.reindex(y.index)


# Add a constant to the independent variable (Total HP)
X = sm.add_constant(X)

# Fit the regression model
model = sm.OLS(y, X)
results = model.fit()

# Print regression summary
print(results.summary())


#THREE DAY
df_threeday_OTHFIN = pd.read_excel('Regression.xlsx',
sheet_name='OtherFin', usecols=[0, 1, 4], skiprows=1)
df_threeday_OTHFIN = df_threeday_OTHFIN.sort_values(by='Date',
ascending=True)
df_threeday_OTHFIN.set_index('Date', inplace=True)
df_threeday_OTHFIN.index = pd.to_datetime(df_threeday_OTHFIN.index)
print(df_threeday_OTHFIN)

# Extract variables for regression
X = df_threeday_OTHFIN['Other Financial institutions net hedging
pressure']
y = df_threeday_OTHFIN['3 day log return']

X = X.reindex(y.index)


# Add a constant to the independent variable (Total HP)
X = sm.add_constant(X)

# Fit the regression model
model = sm.OLS(y, X)
results = model.fit()

# Print regression summary
print(results.summary())

#FOUR DAY
df_fourday_OTHFIN = pd.read_excel('Regression.xlsx',
sheet_name='OtherFin', usecols=[0, 1, 5], skiprows=1)
df_fourday_OTHFIN = df_fourday_OTHFIN.sort_values(by='Date',
ascending=True)
df_fourday_OTHFIN.set_index('Date', inplace=True)
df_fourday_OTHFIN.index = pd.to_datetime(df_fourday_OTHFIN.index)
print(df_fourday_OTHFIN)

# Extract variables for regression
X = df_fourday_OTHFIN['Other Financial institutions net hedging
pressure']
y = df_fourday_OTHFIN['4 day log return']
```

```
X = X.reindex(y.index)


# Add a constant to the independent variable (Total HP)
X = sm.add_constant(X)

# Fit the regression model
model = sm.OLS(y, X)
results = model.fit()

# Print regression summary
print(results.summary())

#FIVE DAY
df_fiveday_OTHFIN = pd.read_excel('Regression.xlsx',
sheet_name='OtherFin', usecols=[0, 1, 6], skiprows=1)
df_fiveday_OTHFIN = df_fiveday_OTHFIN.sort_values(by='Date',
ascending=True)
df_fiveday_OTHFIN.set_index('Date', inplace=True)
df_fiveday_OTHFIN.index = pd.to_datetime(df_fiveday_OTHFIN.index)
print(df_fiveday_OTHFIN)

# Extract variables for regression
X = df_fiveday_OTHFIN['Other Financial institutions net hedging
pressure']
y = df_fiveday_OTHFIN['5 day log return']

X = X.reindex(y.index)


# Add a constant to the independent variable (Total HP)
X = sm.add_constant(X)

# Fit the regression model
model = sm.OLS(y, X)
results = model.fit()

# Print regression summary
print(results.summary())


##ALL FINANCIAL INSTITUTIONS
#ONE DAY
df_oneday_ALLFIN = pd.read_excel('Regression.xlsx',
sheet_name='AllFin', usecols=[0, 1, 2], skiprows=1)
df_oneday_ALLFIN = df_oneday_ALLFIN.sort_values(by='Date',
ascending=True)
df_oneday_ALLFIN.set_index('Date', inplace=True)
df_oneday_ALLFIN.index = pd.to_datetime(df_oneday_ALLFIN.index)
print(df_oneday_ALLFIN)

# Extract variables for regression
X = df_oneday_ALLFIN['All financials net hedging pressure']
y = df_oneday_ALLFIN['1 day log return']

X = X.reindex(y.index)
```

```
# Add a constant to the independent variable (Total HP)
X = sm.add_constant(X)

# Fit the regression model
model = sm.OLS(y, X)
results = model.fit()

# Print regression summary
print(results.summary())


#TWO DAY
df_twoday_ALLFIN = pd.read_excel('Regression.xlsx',
sheet_name='AllFin', usecols=[0, 1, 3], skiprows=1)
df_twoday_ALLFIN = df_twoday_ALLFIN.sort_values(by='Date',
ascending=True)
df_twoday_ALLFIN.set_index('Date', inplace=True)
df_twoday_ALLFIN.index = pd.to_datetime(df_twoday_ALLFIN.index)
print(df_twoday_ALLFIN)

# Extract variables for regression
X = df_twoday_ALLFIN['All financials net hedging pressure']
y = df_twoday_ALLFIN['2 day log return']

X = X.reindex(y.index)


# Add a constant to the independent variable (Total HP)
X = sm.add_constant(X)

# Fit the regression model
model = sm.OLS(y, X)
results = model.fit()

# Print regression summary
print(results.summary())

#THREE DAY
df_threeday_ALLFIN = pd.read_excel('Regression.xlsx',
sheet_name='AllFin', usecols=[0, 1, 4], skiprows=1)
df_threeday_ALLFIN = df_threeday_ALLFIN.sort_values(by='Date',
ascending=True)
df_threeday_ALLFIN.set_index('Date', inplace=True)
df_threeday_ALLFIN.index = pd.to_datetime(df_threeday_ALLFIN.index)
print(df_threeday_ALLFIN)

# Extract variables for regression
X = df_threeday_ALLFIN['All financials net hedging pressure']
y = df_threeday_ALLFIN['3 day log return']

X = X.reindex(y.index)


# Add a constant to the independent variable (Total HP)
X = sm.add_constant(X)
```

```
# Fit the regression model
model = sm.OLS(y, X)
results = model.fit()

# Print regression summary
print(results.summary())


#FOUR DAY
df_fourday_ALLFIN = pd.read_excel('Regression.xlsx',
sheet_name='AllFin', usecols=[0, 1, 5], skiprows=1)
df_fourday_ALLFIN = df_fourday_ALLFIN.sort_values(by='Date',
ascending=True)
df_fourday_ALLFIN.set_index('Date', inplace=True)
df_fourday_ALLFIN.index = pd.to_datetime(df_fourday_ALLFIN.index)
print(df_fourday_ALLFIN)

# Extract variables for regression
X = df_fourday_ALLFIN['All financials net hedging pressure']
y = df_fourday_ALLFIN['4 day log return']

X = X.reindex(y.index)


# Add a constant to the independent variable (Total HP)
X = sm.add_constant(X)

# Fit the regression model
model = sm.OLS(y, X)
results = model.fit()

# Print regression summary
print(results.summary())


#FIVE DAY
df_fiveday_ALLFIN = pd.read_excel('Regression.xlsx',
sheet_name='AllFin', usecols=[0, 1, 6], skiprows=1)
df_fiveday_ALLFIN = df_fiveday_ALLFIN.sort_values(by='Date',
ascending=True)
df_fiveday_ALLFIN.set_index('Date', inplace=True)
df_fiveday_ALLFIN.index = pd.to_datetime(df_fiveday_ALLFIN.index)
print(df_fiveday_ALLFIN)

# Extract variables for regression
X = df_fiveday_ALLFIN['All financials net hedging pressure']
y = df_fiveday_ALLFIN['5 day log return']

X = X.reindex(y.index)


# Add a constant to the independent variable (Total HP)
X = sm.add_constant(X)

# Fit the regression model
model = sm.OLS(y, X)
results = model.fit()
```

```
# Print regression summary
print(results.summary())


#HISTORICAL RETURN  only fridays
Oneday_histret= pd.read_excel('Return calculations.xlsx',
sheet_name='3MO FUTURES return', \
                   usecols=[1,9],skiprows=6)
Oneday_histret = Oneday_histret.sort_values(by='Date',
ascending=False)
Oneday_histret.set_index('Date', inplace=True)

start_date = pd.to_datetime('2018-10-12')
Oneday_histret_filtered = Oneday_histret.loc[Oneday_ret.index >=
start_date]
Oneday_histret_friday=
Oneday_histret_filtered[Oneday_histret_filtered.index.dayofweek == 4]
# 0 corresponds to Monday
print(Oneday_histret_friday)

Twoday_histret= pd.read_excel('Return calculations.xlsx',
sheet_name='3MO FUTURES return', \
                   usecols=[1,10],skiprows=6)
Twoday_histret = Twoday_histret.sort_values(by='Date',
ascending=False)
Twoday_histret.set_index('Date', inplace=True)

start_date = pd.to_datetime('2018-10-12')
Twoday_histret_filtered = Twoday_histret.loc[Twoday_ret.index >=
start_date]
Twoday_histret_friday=
Twoday_histret_filtered[Twoday_histret_filtered.index.dayofweek == 4]
# 0 corresponds to Monday
print(Twoday_histret_friday)

Threeday_histret= pd.read_excel('Return calculations.xlsx',
sheet_name='3MO FUTURES return', \
                   usecols=[1,11],skiprows=6)
Threeday_histret = Threeday_histret.sort_values(by='Date',
ascending=False)
Threeday_histret.set_index('Date', inplace=True)

start_date = pd.to_datetime('2018-10-12')
Threeday_histret_filtered = Threeday_histret.loc[Threeday_ret.index >=
start_date]
Threeday_histret_friday=
Threeday_histret_filtered[Threeday_histret_filtered.index.dayofweek ==
4]  # 0 corresponds to Monday
print(Threeday_histret_friday)


Fourday_histret= pd.read_excel('Return calculations.xlsx',
sheet_name='3MO FUTURES return', \
                   usecols=[1,12],skiprows=6)
Fourday_histret = Fourday_histret.sort_values(by='Date',
ascending=False)
```

```
Fourday_histret.set_index('Date', inplace=True)

start_date = pd.to_datetime('2018-10-12')
Fourday_histret_filtered = Fourday_histret.loc[Fourday_ret.index >=
start_date]
Fourday_histret_friday=
Fourday_histret_filtered[Fourday_histret_filtered.index.dayofweek ==
4]  # 0 corresponds to Monday
print(Fourday_histret_friday)


Fiveday_histret= pd.read_excel('Return calculations.xlsx',
sheet_name='3MO FUTURES return', \
                   usecols=[1,13],skiprows=6)
Fiveday_histret = Fiveday_histret.sort_values(by='Date',
ascending=False)
Fiveday_histret.set_index('Date', inplace=True)

start_date = pd.to_datetime('2018-10-12')
Fiveday_histret_filtered = Fiveday_histret.loc[Fiveday_ret.index >=
start_date]
Fiveday_histret_friday=
Fiveday_histret_filtered[Fiveday_histret_filtered.index.dayofweek ==
4]  # 0 corresponds to Monday
print(Fiveday_histret_friday)


VIX= pd.read_excel('Return calculations.xlsx', sheet_name='3MO FUTURES
return', \
                   usecols=[1,14],skiprows=6)
VIX = VIX.sort_values(by='Date', ascending=False)
VIX.set_index('Date', inplace=True)

start_date = pd.to_datetime('2018-10-12')
VIX_filtered = VIX.loc[VIX.index >= start_date]
VIX_friday= VIX_filtered[VIX_filtered.index.dayofweek == 4]  # 0
corresponds to Monday
print(VIX_friday)


BDI = pd.read_excel('Return calculations.xlsx', sheet_name='3MO
FUTURES return', \
                   usecols=[1,15],skiprows=6)
BDI = BDI.sort_values(by='Date', ascending=False)
BDI.set_index('Date', inplace=True)

start_date = pd.to_datetime('2018-10-12')
BDI_filtered = BDI.loc[BDI.index >= start_date]
BDI_friday= BDI_filtered[BDI_filtered.index.dayofweek == 4]  # 0
corresponds to Monday
print(BDI_friday)

PMI = pd.read_excel('Return calculations.xlsx', sheet_name='3MO
FUTURES return', \
                   usecols=[1,20],skiprows=6)
PMI = PMI.sort_values(by='Date', ascending=False)
PMI.set_index('Date', inplace=True)
```

```
start_date = pd.to_datetime('2018-10-12')
PMI_filtered = PMI.loc[BDI.index >= start_date]
PMI_friday= PMI_filtered[PMI_filtered.index.dayofweek == 4]   # 0
corresponds to Monday
print(PMI_friday)

SPX = pd.read_excel('Return calculations.xlsx', sheet_name='3MO
FUTURES return', \
                   usecols=[1,17],skiprows=6)
SPX = SPX.sort_values(by='Date', ascending=False)
SPX.set_index('Date', inplace=True)

start_date = pd.to_datetime('2018-10-12')
SPX_filtered = SPX.loc[BDI.index >= start_date]
SPX_friday= SPX_filtered[SPX_filtered.index.dayofweek == 4]   # 0
corresponds to Monday
print(SPX_friday)


###########MODEL 2##############
##TOTAL HP
#ONE DAY
df_oneday2_HP = pd.read_excel('Regression.xlsx', sheet_name='TOTALHP',
usecols=[0, 1, 2, 7], skiprows=1)
df_oneday2_HP = df_oneday2_HP.sort_values(by='Date', ascending=True)
df_oneday2_HP.set_index('Date', inplace=True)
df_oneday2_HP.index = pd.to_datetime(df_oneday2_HP.index)
# Drop rows with any NaN values
df_oneday2_HP.dropna(inplace=True)
print(df_oneday2_HP)


# Extract variables for regression
X = df_oneday2_HP[['Total net hedging pressure','1 day hist log
return']]
y = df_oneday2_HP['1 day log return']

X = X.reindex(y.index)


# Add a constant to the independent variable (Total HP)
X = sm.add_constant(X)

# Fit the regression model
model = sm.OLS(y, X)
results = model.fit()

# Print regression summary
print(results.summary())

#TWO DAY
df_twoday2_HP = pd.read_excel('Regression.xlsx', sheet_name='TOTALHP',
usecols=[0, 1, 3, 8], skiprows=1)
df_twoday2_HP = df_twoday2_HP.sort_values(by='Date', ascending=True)
df_twoday2_HP.set_index('Date', inplace=True)
df_twoday2_HP.index = pd.to_datetime(df_twoday2_HP.index)
```

```
# Drop rows with any NaN values
df_twoday2_HP.dropna(inplace=True)
print(df_twoday2_HP)


# Extract variables for regression
X = df_twoday2_HP[['Total net hedging pressure','2 day hist log 
return']]
y = df_twoday2_HP['2 day log return']

X = X.reindex(y.index)


# Add a constant to the independent variable (Total HP)
X = sm.add_constant(X)

# Fit the regression model
model = sm.OLS(y, X)
results = model.fit()

# Print regression summary
print(results.summary())

#THREE DAY
df_threeday2_HP = pd.read_excel('Regression.xlsx', 
sheet_name='TOTALHP', usecols=[0, 1, 4, 9], skiprows=1)
df_threeday2_HP = df_threeday2_HP.sort_values(by='Date', 
ascending=True)
df_threeday2_HP.set_index('Date', inplace=True)
df_threeday2_HP.index = pd.to_datetime(df_threeday2_HP.index)
# Drop rows with any NaN values
df_threeday2_HP.dropna(inplace=True)
print(df_threeday2_HP)


# Extract variables for regression
X = df_threeday2_HP[['Total net hedging pressure','3 day hist log 
return']]
y = df_threeday2_HP['3 day log return']

X = X.reindex(y.index)


# Add a constant to the independent variable (Total HP)
X = sm.add_constant(X)

# Fit the regression model
model = sm.OLS(y, X)
results = model.fit()

# Print regression summary
print(results.summary())


#FOUR DAY
df_fourday2_HP = pd.read_excel('Regression.xlsx', 
sheet_name='TOTALHP', usecols=[0, 1, 5, 10], skiprows=1)
```

```python
df_fourday2_HP = df_fourday2_HP.sort_values(by='Date', ascending=True)
df_fourday2_HP.set_index('Date', inplace=True)
df_fourday2_HP.index = pd.to_datetime(df_fourday2_HP.index)
# Drop rows with any NaN values
df_fourday2_HP.dropna(inplace=True)
print(df_fourday2_HP)


# Extract variables for regression
X = df_fourday2_HP[['Total net hedging pressure','4 day hist log
return']]
y = df_fourday2_HP['4 day log return']

X = X.reindex(y.index)


# Add a constant to the independent variable (Total HP)
X = sm.add_constant(X)

# Fit the regression model
model = sm.OLS(y, X)
results = model.fit()

# Print regression summary
print(results.summary())

#FIVE DAY
df_fiveday2_HP = pd.read_excel('Regression.xlsx',
sheet_name='TOTALHP', usecols=[0, 1, 6, 11], skiprows=1)
df_fiveday2_HP = df_fiveday2_HP.sort_values(by='Date', ascending=True)
df_fiveday2_HP.set_index('Date', inplace=True)
df_fiveday2_HP.index = pd.to_datetime(df_fiveday2_HP.index)
# Drop rows with any NaN values
df_fiveday2_HP.dropna(inplace=True)
print(df_fiveday2_HP)


# Extract variables for regression
X = df_fiveday2_HP[['Total net hedging pressure','5 day hist log
return']]
y = df_fiveday2_HP['5 day log return']

X = X.reindex(y.index)


# Add a constant to the independent variable (Total HP)
X = sm.add_constant(X)

# Fit the regression model
model = sm.OLS(y, X)
results = model.fit()

# Print regression summary
print(results.summary())


##COMMERCIAL UNDERTAKINGS
```

```
#ONE DAY
df_oneday2_COMHP = pd.read_excel('Regression.xlsx',
sheet_name='COMMUND', usecols=[0, 1, 2, 7], skiprows=1)
df_oneday2_COMHP = df_oneday2_COMHP.sort_values(by='Date',
ascending=True)
df_oneday2_COMHP.set_index('Date', inplace=True)
df_oneday2_COMHP.index = pd.to_datetime(df_oneday2_COMHP.index)
print(df_oneday2_COMHP)

# Extract variables for regression
X = df_oneday2_COMHP[['Commercial Undertakings Net Hedging
Pressure','1 day hist log return']]
y = df_oneday2_COMHP['1 day log return']

X = X.reindex(y.index)


# Add a constant to the independent variable (Total HP)
X = sm.add_constant(X)

# Fit the regression model
model = sm.OLS(y, X)
results = model.fit()

# Print regression summary
print(results.summary())


#TWO DAY
df_twoday2_COMHP = pd.read_excel('Regression.xlsx',
sheet_name='COMMUND', usecols=[0, 1, 3, 8], skiprows=1)
df_twoday2_COMHP = df_twoday2_COMHP.sort_values(by='Date',
ascending=True)
df_twoday2_COMHP.set_index('Date', inplace=True)
df_twoday2_COMHP.index = pd.to_datetime(df_twoday2_COMHP.index)
print(df_twoday2_COMHP)

# Extract variables for regression
X = df_twoday2_COMHP[['Commercial Undertakings Net Hedging
Pressure','2 day hist log return']]
y = df_twoday2_COMHP['2 day log return']

X = X.reindex(y.index)


# Add a constant to the independent variable (Total HP)
X = sm.add_constant(X)

# Fit the regression model
model = sm.OLS(y, X)
results = model.fit()

# Print regression summary
print(results.summary())

#THREE DAY
```

```
df_threeday2_COMHP = pd.read_excel('Regression.xlsx',
sheet_name='COMMUND', usecols=[0, 1, 4, 9], skiprows=1)
df_threeday2_COMHP = df_threeday2_COMHP.sort_values(by='Date',
ascending=True)
df_threeday2_COMHP.set_index('Date', inplace=True)
df_threeday2_COMHP.index = pd.to_datetime(df_threeday2_COMHP.index)
print(df_threeday2_COMHP)

# Extract variables for regression
X = df_threeday2_COMHP[['Commercial Undertakings Net Hedging
Pressure','3 day hist log return']]
y = df_threeday2_COMHP['3 day log return']

X = X.reindex(y.index)


# Add a constant to the independent variable (Total HP)
X = sm.add_constant(X)

# Fit the regression model
model = sm.OLS(y, X)
results = model.fit()

# Print regression summary
print(results.summary())

#FOUR DAY
df_fourday2_COMHP = pd.read_excel('Regression.xlsx',
sheet_name='COMMUND', usecols=[0, 1, 5, 10], skiprows=1)
df_fourday2_COMHP = df_fourday2_COMHP.sort_values(by='Date',
ascending=True)
df_fourday2_COMHP.set_index('Date', inplace=True)
df_fourday2_COMHP.index = pd.to_datetime(df_fourday2_COMHP.index)
print(df_fourday2_COMHP)

# Extract variables for regression
X = df_fourday2_COMHP[['Commercial Undertakings Net Hedging
Pressure','4 day hist log return']]
y = df_fourday2_COMHP['4 day log return']

X = X.reindex(y.index)


# Add a constant to the independent variable (Total HP)
X = sm.add_constant(X)

# Fit the regression model
model = sm.OLS(y, X)
results = model.fit()

# Print regression summary
print(results.summary())

#FIVE DAY
df_fiveday2_COMHP = pd.read_excel('Regression.xlsx',
sheet_name='COMMUND', usecols=[0, 1, 6, 11], skiprows=1)
```

```
df_fiveday2_COMHP = df_fiveday2_COMHP.sort_values(by='Date',
ascending=True)
df_fiveday2_COMHP.set_index('Date', inplace=True)
df_fiveday2_COMHP.index = pd.to_datetime(df_fiveday2_COMHP.index)
print(df_fiveday2_COMHP)

# Extract variables for regression
X = df_fiveday2_COMHP[['Commercial Undertakings Net Hedging
Pressure','5 day hist log return']]
y = df_fiveday2_COMHP['5 day log return']

X = X.reindex(y.index)


# Add a constant to the independent variable (Total HP)
X = sm.add_constant(X)

# Fit the regression model
model = sm.OLS(y, X)
results = model.fit()

# Print regression summary
print(results.summary())


##INVESTMENT FIRMS
#ONE DAY
df_oneday2_INVFIR = pd.read_excel('Regression.xlsx',
sheet_name='InvFirms', usecols=[0, 1, 2, 7], skiprows=1)
df_oneday2_INVFIR = df_oneday2_INVFIR.sort_values(by='Date',
ascending=True)
df_oneday2_INVFIR.set_index('Date', inplace=True)
df_oneday2_INVFIR.index = pd.to_datetime(df_oneday2_INVFIR.index)
print(df_oneday2_INVFIR)

# Extract variables for regression
X = df_oneday2_INVFIR[['Investment Firms or Credit Institution’s net
hedging pressure','1 day hist log return']]
y = df_oneday2_INVFIR['1 day log return']

X = X.reindex(y.index)


# Add a constant to the independent variable (Total HP)
X = sm.add_constant(X)

# Fit the regression model
model = sm.OLS(y, X)
results = model.fit()

# Print regression summary
print(results.summary())

#TWO DAY
df_twoday2_INVFIR = pd.read_excel('Regression.xlsx',
sheet_name='InvFirms', usecols=[0, 1, 3, 8], skiprows=1)
```

```
df_twoday2_INVFIR = df_twoday2_INVFIR.sort_values(by='Date',
ascending=True)
df_twoday2_INVFIR.set_index('Date', inplace=True)
df_twoday2_INVFIR.index = pd.to_datetime(df_twoday2_INVFIR.index)
print(df_twoday2_INVFIR)

# Extract variables for regression
X = df_twoday2_INVFIR[['Investment Firms or Credit Institution's net
hedging pressure','2 day hist log return']]
y = df_twoday2_INVFIR['2 day log return']

X = X.reindex(y.index)


# Add a constant to the independent variable (Total HP)
X = sm.add_constant(X)

# Fit the regression model
model = sm.OLS(y, X)
results = model.fit()

# Print regression summary
print(results.summary())

#THREE DAY
df_threeday2_INVFIR = pd.read_excel('Regression.xlsx',
sheet_name='InvFirms', usecols=[0, 1, 4, 9], skiprows=1)
df_threeday2_INVFIR = df_threeday2_INVFIR.sort_values(by='Date',
ascending=True)
df_threeday2_INVFIR.set_index('Date', inplace=True)
df_threeday2_INVFIR.index = pd.to_datetime(df_threeday2_INVFIR.index)
print(df_threeday2_INVFIR)

# Extract variables for regression
X = df_threeday2_INVFIR[['Investment Firms or Credit Institution's net
hedging pressure','3 day hist log return']]
y = df_threeday2_INVFIR['3 day log return']

X = X.reindex(y.index)


# Add a constant to the independent variable (Total HP)
X = sm.add_constant(X)

# Fit the regression model
model = sm.OLS(y, X)
results = model.fit()

# Print regression summary
print(results.summary())

#FOUR DAY
df_fourday2_INVFIR = pd.read_excel('Regression.xlsx',
sheet_name='InvFirms', usecols=[0, 1, 5, 10], skiprows=1)
df_fourday2_INVFIR = df_fourday2_INVFIR.sort_values(by='Date',
ascending=True)
df_fourday2_INVFIR.set_index('Date', inplace=True)
```

```
df_fourday2_INVFIR.index = pd.to_datetime(df_fourday2_INVFIR.index)
print(df_fourday2_INVFIR)

# Extract variables for regression
X = df_fourday2_INVFIR[['Investment Firms or Credit Institution’s net
hedging pressure','4 day hist log return']]
y = df_fourday2_INVFIR['4 day log return']

X = X.reindex(y.index)


# Add a constant to the independent variable (Total HP)
X = sm.add_constant(X)

# Fit the regression model
model = sm.OLS(y, X)
results = model.fit()

# Print regression summary
print(results.summary())

#FIVE DAY
df_fiveday2_INVFIR = pd.read_excel('Regression.xlsx',
sheet_name='InvFirms', usecols=[0, 1, 6, 11], skiprows=1)
df_fiveday2_INVFIR = df_fiveday2_INVFIR.sort_values(by='Date',
ascending=True)
df_fiveday2_INVFIR.set_index('Date', inplace=True)
df_fiveday2_INVFIR.index = pd.to_datetime(df_fiveday2_INVFIR.index)
print(df_fiveday2_INVFIR)

# Extract variables for regression
X = df_fiveday2_INVFIR[['Investment Firms or Credit Institution’s net
hedging pressure','5 day hist log return']]
y = df_fiveday2_INVFIR['5 day log return']

X = X.reindex(y.index)


# Add a constant to the independent variable (Total HP)
X = sm.add_constant(X)

# Fit the regression model
model = sm.OLS(y, X)
results = model.fit()

# Print regression summary
print(results.summary())


##INVESTMENT FUNDS
#ONE DAY
df_oneday2_INVFUN = pd.read_excel('Regression.xlsx',
sheet_name='InvFunds', usecols=[0, 1, 2, 7], skiprows=1)
df_oneday2_INVFUN = df_oneday2_INVFUN.sort_values(by='Date',
ascending=True)
df_oneday2_INVFUN.set_index('Date', inplace=True)
df_oneday2_INVFUN.index = pd.to_datetime(df_oneday2_INVFUN.index)
```

```
print(df_oneday2_INVFUN)

# Extract variables for regression
X = df_oneday2_INVFUN[['Investment funds net hedging pressure','1 day
hist log return']]
y = df_oneday2_INVFUN['1 day log return']

X = X.reindex(y.index)


# Add a constant to the independent variable
X = sm.add_constant(X)

# Fit the regression model
model = sm.OLS(y, X)
results = model.fit()

# Print regression summary
print(results.summary())


#TWO DAY
df_twoday2_INVFUN = pd.read_excel('Regression.xlsx',
sheet_name='InvFunds', usecols=[0, 1, 3, 8], skiprows=1)
df_twoday2_INVFUN = df_twoday2_INVFUN.sort_values(by='Date',
ascending=True)
df_twoday2_INVFUN.set_index('Date', inplace=True)
df_twoday2_INVFUN.index = pd.to_datetime(df_twoday2_INVFUN.index)
print(df_twoday2_INVFUN)

# Extract variables for regression
X = df_twoday2_INVFUN[['Investment funds net hedging pressure','2 day
hist log return']]
y = df_twoday2_INVFUN['2 day log return']

X = X.reindex(y.index)

# Add a constant to the independent variable
X = sm.add_constant(X)

# Fit the regression model
model = sm.OLS(y, X)
results = model.fit()

# Print regression summary
print(results.summary())

#THREE DAY
df_threeday2_INVFUN = pd.read_excel('Regression.xlsx',
sheet_name='InvFunds', usecols=[0, 1, 4, 9], skiprows=1)
df_threeday2_INVFUN = df_threeday2_INVFUN.sort_values(by='Date',
ascending=True)
df_threeday2_INVFUN.set_index('Date', inplace=True)
df_threeday2_INVFUN.index = pd.to_datetime(df_threeday2_INVFUN.index)
print(df_threeday2_INVFUN)

# Extract variables for regression
```

```
X = df_threeday2_INVFUN[['Investment funds net hedging pressure','3 
day hist log return']]
y = df_threeday2_INVFUN['3 day log return']

X = X.reindex(y.index)

# Add a constant to the independent variable
X = sm.add_constant(X)

# Fit the regression model
model = sm.OLS(y, X)
results = model.fit()

# Print regression summary
print(results.summary())


#FOUR DAY
df_fourday2_INVFUN = pd.read_excel('Regression.xlsx', 
sheet_name='InvFunds', usecols=[0, 1, 5, 10], skiprows=1)
df_fourday2_INVFUN = df_fourday2_INVFUN.sort_values(by='Date', 
ascending=True)
df_fourday2_INVFUN.set_index('Date', inplace=True)
df_fourday2_INVFUN.index = pd.to_datetime(df_fourday2_INVFUN.index)
print(df_fourday2_INVFUN)

# Extract variables for regression
X = df_fourday2_INVFUN[['Investment funds net hedging pressure','4 day 
hist log return']]
y = df_fourday2_INVFUN['4 day log return']

X = X.reindex(y.index)

# Add a constant to the independent variable
X = sm.add_constant(X)

# Fit the regression model
model = sm.OLS(y, X)
results = model.fit()

# Print regression summary
print(results.summary())


#FIVE DAY
df_fiveday2_INVFUN = pd.read_excel('Regression.xlsx', 
sheet_name='InvFunds', usecols=[0, 1, 6, 11], skiprows=1)
df_fiveday2_INVFUN = df_fiveday2_INVFUN.sort_values(by='Date', 
ascending=True)
df_fiveday2_INVFUN.set_index('Date', inplace=True)
df_fiveday2_INVFUN.index = pd.to_datetime(df_fiveday2_INVFUN.index)
print(df_fiveday2_INVFUN)

# Extract variables for regression
X = df_fiveday2_INVFUN[['Investment funds net hedging pressure','5 day 
hist log return']]
y = df_fiveday2_INVFUN['5 day log return']
```

```
X = X.reindex(y.index)

# Add a constant to the independent variable
X = sm.add_constant(X)

# Fit the regression model
model = sm.OLS(y, X)
results = model.fit()

# Print regression summary
print(results.summary())

##OTHER FINANCIAL INSTITUTIONS
#ONE DAY
df_oneday2_OTHFIN = pd.read_excel('Regression.xlsx',
sheet_name='OtherFin', usecols=[0, 1, 2, 7], skiprows=1)
df_oneday2_OTHFIN = df_oneday2_OTHFIN.sort_values(by='Date',
ascending=True)
df_oneday2_OTHFIN.set_index('Date', inplace=True)
df_oneday2_OTHFIN.index = pd.to_datetime(df_oneday2_OTHFIN.index)
print(df_oneday2_OTHFIN)

# Extract variables for regression
X = df_oneday2_OTHFIN[['Other Financial institutions net hedging
pressure','1 day hist log return']]
y = df_oneday2_OTHFIN['1 day log return']

X = X.reindex(y.index)


# Add a constant to the independent variable (Total HP)
X = sm.add_constant(X)

# Fit the regression model
model = sm.OLS(y, X)
results = model.fit()

# Print regression summary
print(results.summary())

#TWO DAY
df_twoday2_OTHFIN = pd.read_excel('Regression.xlsx',
sheet_name='OtherFin', usecols=[0, 1, 3, 8], skiprows=1)
df_twoday2_OTHFIN = df_twoday2_OTHFIN.sort_values(by='Date',
ascending=True)
df_twoday2_OTHFIN.set_index('Date', inplace=True)
df_twoday2_OTHFIN.index = pd.to_datetime(df_twoday2_OTHFIN.index)
print(df_twoday2_OTHFIN)

# Extract variables for regression
X = df_twoday2_OTHFIN[['Other Financial institutions net hedging
pressure','2 day hist log return']]
y = df_twoday2_OTHFIN['2 day log return']

X = X.reindex(y.index)
```

```
# Add a constant to the independent variable (Total HP)
X = sm.add_constant(X)

# Fit the regression model
model = sm.OLS(y, X)
results = model.fit()

# Print regression summary
print(results.summary())

#THREE DAY
df_threeday2_OTHFIN = pd.read_excel('Regression.xlsx',
sheet_name='OtherFin', usecols=[0, 1, 4, 9], skiprows=1)
df_threeday2_OTHFIN = df_threeday2_OTHFIN.sort_values(by='Date',
ascending=True)
df_threeday2_OTHFIN.set_index('Date', inplace=True)
df_threeday2_OTHFIN.index = pd.to_datetime(df_threeday2_OTHFIN.index)
print(df_threeday2_OTHFIN)

# Extract variables for regression
X = df_threeday2_OTHFIN[['Other Financial institutions net hedging
pressure','3 day hist log return']]
y = df_threeday2_OTHFIN['3 day log return']

X = X.reindex(y.index)


# Add a constant to the independent variable (Total HP)
X = sm.add_constant(X)

# Fit the regression model
model = sm.OLS(y, X)
results = model.fit()

# Print regression summary
print(results.summary())


#FOUR DAY
df_fourday2_OTHFIN = pd.read_excel('Regression.xlsx',
sheet_name='OtherFin', usecols=[0, 1, 5, 10], skiprows=1)
df_fourday2_OTHFIN = df_fourday2_OTHFIN.sort_values(by='Date',
ascending=True)
df_fourday2_OTHFIN.set_index('Date', inplace=True)
df_fourday2_OTHFIN.index = pd.to_datetime(df_fourday2_OTHFIN.index)
print(df_fourday2_OTHFIN)

# Extract variables for regression
X = df_fourday2_OTHFIN[['Other Financial institutions net hedging
pressure','4 day hist log return']]
y = df_fourday2_OTHFIN['4 day log return']

X = X.reindex(y.index)


# Add a constant to the independent variable (Total HP)
```

```
X = sm.add_constant(X)

# Fit the regression model
model = sm.OLS(y, X)
results = model.fit()

# Print regression summary
print(results.summary())

#FIVE DAY
df_fiveday2_OTHFIN = pd.read_excel('Regression.xlsx',
sheet_name='OtherFin', usecols=[0, 1, 6, 11], skiprows=1)
df_fiveday2_OTHFIN = df_fiveday2_OTHFIN.sort_values(by='Date',
ascending=True)
df_fiveday2_OTHFIN.set_index('Date', inplace=True)
df_fiveday2_OTHFIN.index = pd.to_datetime(df_fiveday2_OTHFIN.index)
print(df_fiveday2_OTHFIN)

# Extract variables for regression
X = df_fiveday2_OTHFIN[['Other Financial institutions net hedging
pressure','5 day hist log return']]
y = df_fiveday2_OTHFIN['5 day log return']

X = X.reindex(y.index)


# Add a constant to the independent variable (Total HP)
X = sm.add_constant(X)

# Fit the regression model
model = sm.OLS(y, X)
results = model.fit()

# Print regression summary
print(results.summary())

##ALL FINANCIAL INSTITUTIONS
#ONE DAY
df_oneday2_ALLFIN = pd.read_excel('Regression.xlsx',
sheet_name='AllFin', usecols=[0, 1, 2, 7], skiprows=1)
df_oneday2_ALLFIN = df_oneday2_ALLFIN.sort_values(by='Date',
ascending=True)
df_oneday2_ALLFIN.set_index('Date', inplace=True)
df_oneday2_ALLFIN.index = pd.to_datetime(df_oneday2_ALLFIN.index)
print(df_oneday2_ALLFIN)

# Extract variables for regression
X = df_oneday2_ALLFIN[['All financials net hedging pressure','1 day
hist log return']]
y = df_oneday2_ALLFIN['1 day log return']

X = X.reindex(y.index)


# Add a constant to the independent variable (Total HP)
X = sm.add_constant(X)
```

```
# Fit the regression model
model = sm.OLS(y, X)
results = model.fit()

# Print regression summary
print(results.summary())

#TWO DAY
df_twoday2_ALLFIN = pd.read_excel('Regression.xlsx',
sheet_name='AllFin', usecols=[0, 1, 3, 8], skiprows=1)
df_twoday2_ALLFIN = df_twoday2_ALLFIN.sort_values(by='Date',
ascending=True)
df_twoday2_ALLFIN.set_index('Date', inplace=True)
df_twoday2_ALLFIN.index = pd.to_datetime(df_twoday2_ALLFIN.index)
print(df_twoday2_ALLFIN)

# Extract variables for regression
X = df_twoday2_ALLFIN[['All financials net hedging pressure','2 day
hist log return']]
y = df_twoday2_ALLFIN['2 day log return']

X = X.reindex(y.index)


# Add a constant to the independent variable (Total HP)
X = sm.add_constant(X)

# Fit the regression model
model = sm.OLS(y, X)
results = model.fit()

# Print regression summary
print(results.summary())

#THREE DAY
df_threeday2_ALLFIN = pd.read_excel('Regression.xlsx',
sheet_name='AllFin', usecols=[0, 1, 4, 9], skiprows=1)
df_threeday2_ALLFIN = df_threeday2_ALLFIN.sort_values(by='Date',
ascending=True)
df_threeday2_ALLFIN.set_index('Date', inplace=True)
df_threeday2_ALLFIN.index = pd.to_datetime(df_threeday2_ALLFIN.index)
print(df_threeday2_ALLFIN)

# Extract variables for regression
X = df_threeday2_ALLFIN[['All financials net hedging pressure','3 day
hist log return']]
y = df_threeday2_ALLFIN['3 day log return']

X = X.reindex(y.index)


# Add a constant to the independent variable (Total HP)
X = sm.add_constant(X)

# Fit the regression model
model = sm.OLS(y, X)
results = model.fit()
```

```
# Print regression summary
print(results.summary())

#FOUR DAY
df_fourday2_ALLFIN = pd.read_excel('Regression.xlsx',
sheet_name='AllFin', usecols=[0, 1, 5, 10], skiprows=1)
df_fourday2_ALLFIN = df_fourday2_ALLFIN.sort_values(by='Date',
ascending=True)
df_fourday2_ALLFIN.set_index('Date', inplace=True)
df_fourday2_ALLFIN.index = pd.to_datetime(df_fourday2_ALLFIN.index)
print(df_fourday2_ALLFIN)

# Extract variables for regression
X = df_fourday2_ALLFIN[['All financials net hedging pressure','4 day
hist log return']]
y = df_fourday2_ALLFIN['4 day log return']

X = X.reindex(y.index)


# Add a constant to the independent variable (Total HP)
X = sm.add_constant(X)

# Fit the regression model
model = sm.OLS(y, X)
results = model.fit()

# Print regression summary
print(results.summary())

#FIVE DAY
df_fiveday2_ALLFIN = pd.read_excel('Regression.xlsx',
sheet_name='AllFin', usecols=[0, 1, 6, 11], skiprows=1)
df_fiveday2_ALLFIN = df_fiveday2_ALLFIN.sort_values(by='Date',
ascending=True)
df_fiveday2_ALLFIN.set_index('Date', inplace=True)
df_fiveday2_ALLFIN.index = pd.to_datetime(df_fiveday2_ALLFIN.index)
print(df_fiveday2_ALLFIN)

# Extract variables for regression
X = df_fiveday2_ALLFIN[['All financials net hedging pressure','5 day
hist log return']]
y = df_fiveday2_ALLFIN['5 day log return']

X = X.reindex(y.index)


# Add a constant to the independent variable (Total HP)
X = sm.add_constant(X)

# Fit the regression model
model = sm.OLS(y, X)
results = model.fit()

# Print regression summary
print(results.summary())
```

```
###########MODEL 3##############
##TOTAL HP
#ONE DAY + VIX
df_onedayVIX_HP = pd.read_excel('Regression.xlsx',
sheet_name='TOTALHP', usecols=[0, 1, 2, 7, 12], skiprows=1)
df_onedayVIX_HP = df_onedayVIX_HP.sort_values(by='Date',
ascending=True)
df_onedayVIX_HP.set_index('Date', inplace=True)
df_onedayVIX_HP.index = pd.to_datetime(df_onedayVIX_HP.index)
# Drop rows with any NaN values
df_onedayVIX_HP.dropna(inplace=True)
print(df_onedayVIX_HP)


# Extract variables for regression
X = df_onedayVIX_HP[['Total net hedging pressure','1 day hist log
return','VIX']]
y = df_onedayVIX_HP['1 day log return']

X = X.reindex(y.index)


# Add a constant to the independent variable (Total HP)
X = sm.add_constant(X)

# Fit the regression model
model = sm.OLS(y, X)
results = model.fit()

# Print regression summary
print(results.summary())

#ONE DAY + BDI
df_onedayBDI_HP = pd.read_excel('Regression.xlsx',
sheet_name='TOTALHP', usecols=[0, 1, 2, 7, 13], skiprows=1)
df_onedayBDI_HP = df_onedayBDI_HP.sort_values(by='Date',
ascending=True)
df_onedayBDI_HP.set_index('Date', inplace=True)
df_onedayBDI_HP.index = pd.to_datetime(df_onedayBDI_HP.index)
# Drop rows with any NaN values
df_onedayBDI_HP.dropna(inplace=True)
print(df_onedayBDI_HP)


# Extract variables for regression
X = df_onedayBDI_HP[['Total net hedging pressure','1 day hist log
return','BDI']]
y = df_onedayBDI_HP['1 day log return']

X = X.reindex(y.index)


# Add a constant to the independent variable (Total HP)
X = sm.add_constant(X)

# Fit the regression model
```

```
model = sm.OLS(y, X)
results = model.fit()

# Print regression summary
print(results.summary())

#ONE DAY + PMI
df_onedayPMI_HP = pd.read_excel('Regression.xlsx',
sheet_name='TOTALHP', usecols=[0, 1, 2, 7, 14], skiprows=1)
df_onedayPMI_HP = df_onedayPMI_HP.sort_values(by='Date',
ascending=True)
df_onedayPMI_HP.set_index('Date', inplace=True)
df_onedayPMI_HP.index = pd.to_datetime(df_onedayPMI_HP.index)
# Drop rows with any NaN values
df_onedayPMI_HP.dropna(inplace=True)
print(df_onedayPMI_HP)


# Extract variables for regression
X = df_onedayPMI_HP[['Total net hedging pressure','1 day hist log
return','PMI']]
y = df_onedayPMI_HP['1 day log return']

X = X.reindex(y.index)


# Add a constant to the independent variable (Total HP)
X = sm.add_constant(X)

# Fit the regression model
model = sm.OLS(y, X)
results = model.fit()

# Print regression summary
print(results.summary())


#ONE DAY + SPX
df_onedaySPX_HP = pd.read_excel('Regression.xlsx',
sheet_name='TOTALHP', usecols=[0, 1, 2, 7, 15], skiprows=1)
df_onedaySPX_HP = df_onedaySPX_HP.sort_values(by='Date',
ascending=True)
df_onedaySPX_HP.set_index('Date', inplace=True)
df_onedaySPX_HP.index = pd.to_datetime(df_onedaySPX_HP.index)
# Drop rows with any NaN values
df_onedaySPX_HP.dropna(inplace=True)
print(df_onedaySPX_HP)


# Extract variables for regression
X = df_onedaySPX_HP[['Total net hedging pressure','1 day hist log
return','SPX']]
y = df_onedaySPX_HP['1 day log return']

X = X.reindex(y.index)
```

```
# Add a constant to the independent variable (Total HP)
X = sm.add_constant(X)

# Fit the regression model
model = sm.OLS(y, X)
results = model.fit()

# Print regression summary
print(results.summary())

##COMMERCIAL UNDERTAKINGS
#ONE DAY + VIX
df_onedayVIX_COMHP = pd.read_excel('Regression.xlsx',
sheet_name='COMMUND', usecols=[0, 1, 2, 7, 12], skiprows=1)
df_onedayVIX_COMHP = df_onedayVIX_COMHP.sort_values(by='Date',
ascending=True)
df_onedayVIX_COMHP.set_index('Date', inplace=True)
df_onedayVIX_COMHP.index = pd.to_datetime(df_onedayVIX_COMHP.index)
# Drop rows with any NaN values
df_onedayVIX_COMHP.dropna(inplace=True)
print(df_onedayVIX_COMHP)


# Extract variables for regression
X = df_onedayVIX_COMHP[['Commercial Undertakings Net Hedging
Pressure','1 day hist log return','VIX']]
y = df_onedayVIX_COMHP['1 day log return']

X = X.reindex(y.index)


# Add a constant to the independent variable (Total HP)
X = sm.add_constant(X)

# Fit the regression model
model = sm.OLS(y, X)
results = model.fit()

# Print regression summary
print(results.summary())

#ONE DAY + BDI
df_onedayBDI_COMHP = pd.read_excel('Regression.xlsx',
sheet_name='COMMUND', usecols=[0, 1, 2, 7, 13], skiprows=1)
df_onedayBDI_COMHP = df_onedayBDI_COMHP.sort_values(by='Date',
ascending=True)
df_onedayBDI_COMHP.set_index('Date', inplace=True)
df_onedayBDI_COMHP.index = pd.to_datetime(df_onedayBDI_COMHP.index)
# Drop rows with any NaN values
df_onedayBDI_COMHP.dropna(inplace=True)
print(df_onedayBDI_COMHP)


# Extract variables for regression
X = df_onedayBDI_COMHP[['Commercial Undertakings Net Hedging
Pressure','1 day hist log return','BDI']]
y = df_onedayBDI_COMHP['1 day log return']
```

```
X = X.reindex(y.index)


# Add a constant to the independent variable (Total HP)
X = sm.add_constant(X)

# Fit the regression model
model = sm.OLS(y, X)
results = model.fit()

# Print regression summary
print(results.summary())

#ONE DAY + PMI
df_onedayPMI_COMHP = pd.read_excel('Regression.xlsx',
sheet_name='COMMUND', usecols=[0, 1, 2, 7, 14], skiprows=1)
df_onedayPMI_COMHP = df_onedayPMI_COMHP.sort_values(by='Date',
ascending=True)
df_onedayPMI_COMHP.set_index('Date', inplace=True)
df_onedayPMI_COMHP.index = pd.to_datetime(df_onedayPMI_COMHP.index)
# Drop rows with any NaN values
df_onedayPMI_COMHP.dropna(inplace=True)
print(df_onedayPMI_COMHP)


# Extract variables for regression
X = df_onedayPMI_COMHP[['Commercial Undertakings Net Hedging
Pressure','1 day hist log return','PMI']]
y = df_onedayPMI_COMHP['1 day log return']

X = X.reindex(y.index)


# Add a constant to the independent variable (Total HP)
X = sm.add_constant(X)

# Fit the regression model
model = sm.OLS(y, X)
results = model.fit()

# Print regression summary
print(results.summary())

#ONE DAY + SPX
df_onedaySPX_COMHP = pd.read_excel('Regression.xlsx',
sheet_name='COMMUND', usecols=[0, 1, 2, 7, 15], skiprows=1)
df_onedaySPX_COMHP = df_onedaySPX_COMHP.sort_values(by='Date',
ascending=True)
df_onedaySPX_COMHP.set_index('Date', inplace=True)
df_onedaySPX_COMHP.index = pd.to_datetime(df_onedaySPX_COMHP.index)
# Drop rows with any NaN values
df_onedaySPX_COMHP.dropna(inplace=True)
print(df_onedaySPX_COMHP)


# Extract variables for regression
```

```
X = df_onedaySPX_COMHP[['Commercial Undertakings Net Hedging
Pressure','1 day hist log return','SPX']]
y = df_onedaySPX_COMHP['1 day log return']

X = X.reindex(y.index)


# Add a constant to the independent variable (Total HP)
X = sm.add_constant(X)

# Fit the regression model
model = sm.OLS(y, X)
results = model.fit()

# Print regression summary
print(results.summary())

##INVESTMENT FIRMS
#ONE DAY + VIX
df_onedayVIX_INVFIR = pd.read_excel('Regression.xlsx',
sheet_name='InvFirms', usecols=[0, 1, 2, 7, 12], skiprows=1)
df_onedayVIX_INVFIR = df_onedayVIX_INVFIR.sort_values(by='Date',
ascending=True)
df_onedayVIX_INVFIR.set_index('Date', inplace=True)
df_onedayVIX_INVFIR.index = pd.to_datetime(df_onedayVIX_INVFIR.index)
print(df_onedayVIX_INVFIR)

# Extract variables for regression
X = df_onedayVIX_INVFIR[['Investment Firms or Credit Institution’s net
hedging pressure','1 day hist log return','VIX']]
y = df_onedayVIX_INVFIR['1 day log return']

X = X.reindex(y.index)


# Add a constant to the independent variable (Total HP)
X = sm.add_constant(X)

# Fit the regression model
model = sm.OLS(y, X)
results = model.fit()

# Print regression summary
print(results.summary())

#ONE DAY + BDI
df_onedayBDI_INVFIR = pd.read_excel('Regression.xlsx',
sheet_name='InvFirms', usecols=[0, 1, 2, 7, 13], skiprows=1)
df_onedayBDI_INVFIR = df_onedayBDI_INVFIR.sort_values(by='Date',
ascending=True)
df_onedayBDI_INVFIR.set_index('Date', inplace=True)
df_onedayBDI_INVFIR.index = pd.to_datetime(df_onedayBDI_INVFIR.index)
print(df_onedayBDI_INVFIR)

# Extract variables for regression
X = df_onedayBDI_INVFIR[['Investment Firms or Credit Institution’s net
hedging pressure','1 day hist log return','BDI']]
```

```
y = df_onedayBDI_INVFIR['1 day log return']

X = X.reindex(y.index)


# Add a constant to the independent variable (Total HP)
X = sm.add_constant(X)

# Fit the regression model
model = sm.OLS(y, X)
results = model.fit()

# Print regression summary
print(results.summary())

#ONE DAY + PMI
df_onedayPMI_INVFIR = pd.read_excel('Regression.xlsx',
sheet_name='InvFirms', usecols=[0, 1, 2, 7, 14], skiprows=1)
df_onedayPMI_INVFIR = df_onedayPMI_INVFIR.sort_values(by='Date',
ascending=True)
df_onedayPMI_INVFIR.set_index('Date', inplace=True)
df_onedayPMI_INVFIR.index = pd.to_datetime(df_onedayPMI_INVFIR.index)
# Drop rows with any NaN values
df_onedayPMI_INVFIR.dropna(inplace=True)
print(df_onedayPMI_INVFIR)


# Extract variables for regression
X = df_onedayPMI_INVFIR[['Investment Firms or Credit Institution’s net
hedging pressure','1 day hist log return','PMI']]
y = df_onedayPMI_INVFIR['1 day log return']

X = X.reindex(y.index)


# Add a constant to the independent variable (Total HP)
X = sm.add_constant(X)

# Fit the regression model
model = sm.OLS(y, X)
results = model.fit()

# Print regression summary
print(results.summary())

#ONE DAY + SPX
df_onedaySPX_INVFIR = pd.read_excel('Regression.xlsx',
sheet_name='InvFirms', usecols=[0, 1, 2, 7, 15], skiprows=1)
df_onedaySPX_INVFIR = df_onedaySPX_INVFIR.sort_values(by='Date',
ascending=True)
df_onedaySPX_INVFIR.set_index('Date', inplace=True)
df_onedaySPX_INVFIR.index = pd.to_datetime(df_onedaySPX_INVFIR.index)
# Drop rows with any NaN values
df_onedaySPX_INVFIR.dropna(inplace=True)
print(df_onedaySPX_INVFIR)
```

```
# Extract variables for regression
X = df_onedaySPX_INVFIR[['Investment Firms or Credit Institution's net
hedging pressure','1 day hist log return','SPX']]
y = df_onedaySPX_INVFIR['1 day log return']

X = X.reindex(y.index)


# Add a constant to the independent variable (Total HP)
X = sm.add_constant(X)

# Fit the regression model
model = sm.OLS(y, X)
results = model.fit()

# Print regression summary
print(results.summary())


##INVESTMENT FUNDS
#ONE DAY + VIX
df_onedayVIX_INVFUN = pd.read_excel('Regression.xlsx',
sheet_name='InvFunds', usecols=[0, 1, 2, 7, 12], skiprows=1)
df_onedayVIX_INVFUN = df_onedayVIX_INVFUN.sort_values(by='Date',
ascending=True)
df_onedayVIX_INVFUN.set_index('Date', inplace=True)
df_onedayVIX_INVFUN.index = pd.to_datetime(df_onedayVIX_INVFUN.index)
print(df_onedayVIX_INVFUN)

# Extract variables for regression
X = df_onedayVIX_INVFUN[['Investment funds net hedging pressure','1
day hist log return','VIX']]
y = df_onedayVIX_INVFUN['1 day log return']

X = X.reindex(y.index)


# Add a constant to the independent variable (Total HP)
X = sm.add_constant(X)

# Fit the regression model
model = sm.OLS(y, X)
results = model.fit()

# Print regression summary
print(results.summary())


#ONE DAY + BDI
df_onedayBDI_INVFUN = pd.read_excel('Regression.xlsx',
sheet_name='InvFunds', usecols=[0, 1, 2, 7, 13], skiprows=1)
df_onedayBDI_INVFUN = df_onedayBDI_INVFUN.sort_values(by='Date',
ascending=True)
df_onedayBDI_INVFUN.set_index('Date', inplace=True)
df_onedayBDI_INVFUN.index = pd.to_datetime(df_onedayBDI_INVFUN.index)
print(df_onedayBDI_INVFUN)
```

```
# Extract variables for regression
X = df_onedayBDI_INVFUN[['Investment funds net hedging pressure','1
day hist log return','BDI']]
y = df_onedayBDI_INVFUN['1 day log return']

X = X.reindex(y.index)


# Add a constant to the independent variable (Total HP)
X = sm.add_constant(X)

# Fit the regression model
model = sm.OLS(y, X)
results = model.fit()

# Print regression summary
print(results.summary())

#ONE DAY + PMI
df_onedayPMI_INVFUN = pd.read_excel('Regression.xlsx',
sheet_name='InvFunds', usecols=[0, 1, 2, 7, 14], skiprows=1)
df_onedayPMI_INVFUN = df_onedayPMI_INVFUN.sort_values(by='Date',
ascending=True)
df_onedayPMI_INVFUN.set_index('Date', inplace=True)
df_onedayPMI_INVFUN.index = pd.to_datetime(df_onedayPMI_INVFUN.index)
# Drop rows with any NaN values
df_onedayPMI_INVFUN.dropna(inplace=True)
print(df_onedayPMI_INVFUN)

# Extract variables for regression
X = df_onedayPMI_INVFUN[['Investment funds net hedging pressure','1
day hist log return','PMI']]
y = df_onedayPMI_INVFUN['1 day log return']

X = X.reindex(y.index)


# Add a constant to the independent variable (Total HP)
X = sm.add_constant(X)

# Fit the regression model
model = sm.OLS(y, X)
results = model.fit()

# Print regression summary
print(results.summary())

#ONE DAY + SPX
df_onedaySPX_INVFUN = pd.read_excel('Regression.xlsx',
sheet_name='InvFunds', usecols=[0, 1, 2, 7, 15], skiprows=1)
df_onedaySPX_INVFUN = df_onedaySPX_INVFUN.sort_values(by='Date',
ascending=True)
df_onedaySPX_INVFUN.set_index('Date', inplace=True)
df_onedaySPX_INVFUN.index = pd.to_datetime(df_onedaySPX_INVFUN.index)
# Drop rows with any NaN values
df_onedaySPX_INVFUN.dropna(inplace=True)
print(df_onedaySPX_INVFUN)
```

```
# Extract variables for regression
X = df_onedaySPX_INVFUN[['Investment funds net hedging pressure','1
day hist log return','SPX']]
y = df_onedaySPX_INVFUN['1 day log return']

X = X.reindex(y.index)


# Add a constant to the independent variable (Total HP)
X = sm.add_constant(X)

# Fit the regression model
model = sm.OLS(y, X)
results = model.fit()

# Print regression summary
print(results.summary())


##OTHER FINANCIAL INSTITUTIONS
#ONE DAY + VIX
df_onedayVIX_OTHFIN = pd.read_excel('Regression.xlsx',
sheet_name='OtherFin', usecols=[0, 1, 2, 7, 12], skiprows=1)
df_onedayVIX_OTHFIN = df_onedayVIX_OTHFIN.sort_values(by='Date',
ascending=True)
df_onedayVIX_OTHFIN.set_index('Date', inplace=True)
df_onedayVIX_OTHFIN.index = pd.to_datetime(df_onedayVIX_OTHFIN.index)
print(df_onedayVIX_OTHFIN)

# Extract variables for regression
X = df_onedayVIX_OTHFIN[['Other Financial institutions net hedging
pressure','1 day hist log return','VIX']]
y = df_onedayVIX_OTHFIN['1 day log return']

X = X.reindex(y.index)


# Add a constant to the independent variable (Total HP)
X = sm.add_constant(X)

# Fit the regression model
model = sm.OLS(y, X)
results = model.fit()

# Print regression summary
print(results.summary())

#ONE DAY + BDI
df_onedayBDI_OTHFIN = pd.read_excel('Regression.xlsx',
sheet_name='OtherFin', usecols=[0, 1, 2, 7, 13], skiprows=1)
df_onedayBDI_OTHFIN = df_onedayBDI_OTHFIN.sort_values(by='Date',
ascending=True)
df_onedayBDI_OTHFIN.set_index('Date', inplace=True)
df_onedayBDI_OTHFIN.index = pd.to_datetime(df_onedayBDI_OTHFIN.index)
print(df_onedayBDI_OTHFIN)
```

```
# Extract variables for regression
X = df_onedayBDI_OTHFIN[['Other Financial institutions net hedging
pressure','1 day hist log return','BDI']]
y = df_onedayBDI_OTHFIN['1 day log return']

X = X.reindex(y.index)


# Add a constant to the independent variable (Total HP)
X = sm.add_constant(X)

# Fit the regression model
model = sm.OLS(y, X)
results = model.fit()

# Print regression summary
print(results.summary())


#ONE DAY + PMI
df_onedayPMI_OTHFIN = pd.read_excel('Regression.xlsx',
sheet_name='OtherFin', usecols=[0, 1, 2, 7, 14], skiprows=1)
df_onedayPMI_OTHFIN = df_onedayPMI_OTHFIN.sort_values(by='Date',
ascending=True)
df_onedayPMI_OTHFIN.set_index('Date', inplace=True)
df_onedayPMI_OTHFIN.index = pd.to_datetime(df_onedayPMI_OTHFIN.index)
df_onedayPMI_OTHFIN.dropna(inplace=True)
print(df_onedayPMI_OTHFIN)

# Extract variables for regression
X = df_onedayPMI_OTHFIN[['Other Financial institutions net hedging
pressure','1 day hist log return','PMI']]
y = df_onedayPMI_OTHFIN['1 day log return']

X = X.reindex(y.index)


# Add a constant to the independent variable (Total HP)
X = sm.add_constant(X)

# Fit the regression model
model = sm.OLS(y, X)
results = model.fit()

# Print regression summary
print(results.summary())

#ONE DAY + SPX
df_onedaySPX_OTHFIN = pd.read_excel('Regression.xlsx',
sheet_name='OtherFin', usecols=[0, 1, 2, 7, 15], skiprows=1)
df_onedaySPX_OTHFIN = df_onedaySPX_OTHFIN.sort_values(by='Date',
ascending=True)
df_onedaySPX_OTHFIN.set_index('Date', inplace=True)
df_onedaySPX_OTHFIN.index = pd.to_datetime(df_onedaySPX_OTHFIN.index)
df_onedaySPX_OTHFIN.dropna(inplace=True)
print(df_onedaySPX_OTHFIN)
```

```
# Extract variables for regression
X = df_onedaySPX_OTHFIN[['Other Financial institutions net hedging
pressure','1 day hist log return','SPX']]
y = df_onedaySPX_OTHFIN['1 day log return']

X = X.reindex(y.index)


# Add a constant to the independent variable (Total HP)
X = sm.add_constant(X)

# Fit the regression model
model = sm.OLS(y, X)
results = model.fit()

# Print regression summary
print(results.summary())


 ##ALL FINANCIAL INSTITUTIONS
#ONE DAY + VIX
df_onedayVIX_ALLFIN = pd.read_excel('Regression.xlsx',
sheet_name='AllFin', usecols=[0, 1, 2, 7, 12], skiprows=1)
df_onedayVIX_ALLFIN = df_onedayVIX_ALLFIN.sort_values(by='Date',
ascending=True)
df_onedayVIX_ALLFIN.set_index('Date', inplace=True)
df_onedayVIX_ALLFIN.index = pd.to_datetime(df_onedayVIX_ALLFIN.index)
print(df_onedayVIX_ALLFIN)

# Extract variables for regression
X = df_onedayVIX_ALLFIN[['All financials net hedging pressure','1 day
hist log return','VIX']]
y = df_onedayVIX_ALLFIN['1 day log return']

X = X.reindex(y.index)


# Add a constant to the independent variable (Total HP)
X = sm.add_constant(X)

# Fit the regression model
model = sm.OLS(y, X)
results = model.fit()

# Print regression summary
print(results.summary())

#ONE DAY + BDI
df_onedayBDI_ALLFIN = pd.read_excel('Regression.xlsx',
sheet_name='AllFin', usecols=[0, 1, 2, 7, 13], skiprows=1)
df_onedayBDI_ALLFIN = df_onedayBDI_ALLFIN.sort_values(by='Date',
ascending=True)
df_onedayBDI_ALLFIN.set_index('Date', inplace=True)
df_onedayBDI_ALLFIN.index = pd.to_datetime(df_onedayBDI_ALLFIN.index)
print(df_onedayBDI_ALLFIN)

# Extract variables for regression
```

```
X = df_onedayBDI_ALLFIN[['All financials net hedging pressure','1 day
hist log return','BDI']]
y = df_onedayBDI_ALLFIN['1 day log return']

X = X.reindex(y.index)


# Add a constant to the independent variable (Total HP)
X = sm.add_constant(X)

# Fit the regression model
model = sm.OLS(y, X)
results = model.fit()

# Print regression summary
print(results.summary())

#ONE DAY + PMI
df_onedayPMI_ALLFIN = pd.read_excel('Regression.xlsx',
sheet_name='AllFin', usecols=[0, 1, 2, 7, 14], skiprows=1)
df_onedayPMI_ALLFIN = df_onedayPMI_ALLFIN.sort_values(by='Date',
ascending=True)
df_onedayPMI_ALLFIN.set_index('Date', inplace=True)
df_onedayPMI_ALLFIN.index = pd.to_datetime(df_onedayPMI_ALLFIN.index)
df_onedayPMI_ALLFIN.dropna(inplace=True)
print(df_onedayPMI_ALLFIN)

# Extract variables for regression
X = df_onedayPMI_ALLFIN[['All financials net hedging pressure','1 day
hist log return','PMI']]
y = df_onedayPMI_ALLFIN['1 day log return']

X = X.reindex(y.index)


# Add a constant to the independent variable (Total HP)
X = sm.add_constant(X)

# Fit the regression model
model = sm.OLS(y, X)
results = model.fit()

# Print regression summary
print(results.summary())

#ONE DAY + SPX
df_onedaySPX_ALLFIN = pd.read_excel('Regression.xlsx',
sheet_name='AllFin', usecols=[0, 1, 2, 7, 15], skiprows=1)
df_onedaySPX_ALLFIN = df_onedaySPX_ALLFIN.sort_values(by='Date',
ascending=True)
df_onedaySPX_ALLFIN.set_index('Date', inplace=True)
df_onedaySPX_ALLFIN.index = pd.to_datetime(df_onedaySPX_ALLFIN.index)
df_onedaySPX_ALLFIN.dropna(inplace=True)
print(df_onedaySPX_ALLFIN)

# Extract variables for regression
```

```
X = df_onedaySPX_ALLFIN[['All financials net hedging pressure','1 day
hist log return','SPX']]
y = df_onedaySPX_ALLFIN['1 day log return']

X = X.reindex(y.index)


# Add a constant to the independent variable (Total HP)
X = sm.add_constant(X)

# Fit the regression model
model = sm.OLS(y, X)
results = model.fit()

# Print regression summary
print(results.summary())
```