FINANCE AREA

INTERNSHIP REPORT

FINANCIAL RESEARCH INTERN

By:

Abhijeet Avhale

Academic Supervisor:

Dr. Mohammad Shameen Jawed

Institution:

Indian Institute of Management, Visakhapatnam

Period of Internship:

12th June 2023 - 6th August 2023

8 weeks(2 months)

6 weeks offline, 2 weeks remote*

APPROVAL

Approved By

Academic Supervisor

Name	Signature	Date

Institute Supervisor

Name	Signature	Date

DECLARATION

I certify that I am the author of this project and that any assistance I received in its preparation is fully acknowledged and disclosed in this project. I have also cited any source from which I used data, ideas, or words, either quoted or paraphrased. Further, this report meets all the rules of quotation and referencing, as well as adheres to the fraud policies listed in the honor code.

No portion of the work referred to in this study has been submitted in support of an application for another degree or qualification to this or any other university or institution of learning.

Abhijeet Avhale

5th August 2023

WORK TERM RELEASE

I hereby state and verify by my signature that I have reviewed this report. I hereby affirm that the report contains

no confidential data/information, and I authorize it to be released.

confidential data/information, and I do not authorize it to be released.

Supervisor

Name	Signature	Date

ABSTRACT

This research paper investigates the correlation between commodity prices across three sectors - Agriculture, Metals, and Energy - during various significant events and crises, aiming to assess whether these events lead to increased correlations among commodities, potentially diminishing the effectiveness of diversified portfolios during times of crisis. The study analyzes a comprehensive dataset encompassing major commodities such as wheat, soybean, crude oil, gold, aluminum, etc. The events and crises considered include the COVID-19 pandemic, the 2008 housing market crash, the Ebola outbreak, the European crisis, and other major geopolitical and economic upheavals. The primary hypothesis suggests that during these events and crises, the correlation between different commodities increases, impacting the benefits of traditional diversification strategies for risk mitigation.

To examine the behavior of commodity prices in response to these events, the research employs statistical methods, including correlation analysis and time-series modeling, utilizing the Detrended Cross-Correlation Analysis (DCCA) from Fathon. This paper studies the changes in correlations during the timeline of various events and finds generalized patterns between the change in correlations. By gaining insights into the dynamics of commodity correlations, investors and asset managers can make informed decisions to mitigate risks and potentially improve their performance during crisis periods.

Keywords: Correlation, Diversification, Risk Management

ACKNOWLEDGEMENTS

I would like to express my heartfelt gratitude and appreciation to all those who have contributed to the successful completion of my internship and the preparation of this report.

First and foremost, I am immensely thankful to Prof. Mohammad Shameen Jawed for his invaluable guidance, unwavering support, and insightful feedback throughout the duration of my internship. His mentorship has not only enriched my understanding of the subject matter but has also inspired me to strive for excellence in all my endeavors.

I extend my sincere thanks to my friends and family for their constant encouragement and belief in my abilities. Their unwavering support provided me with the strength and motivation to overcome challenges and pursue this internship with dedication.

I am grateful to the PhD students who generously shared their knowledge and expertise, offering their assistance whenever needed. Their inputs and discussions greatly enhanced the quality of my work and broadened my perspectives.

I would also like to acknowledge the efficient and cooperative administration of IIM Visakhapatnam. Their seamless coordination and support played a pivotal role in ensuring a smooth workflow throughout the internship period.

Lastly, I am thankful to all those individuals who have directly or indirectly contributed to this internship report. Your collective efforts have significantly enriched my learning experience and have been instrumental in shaping the outcomes presented here.

In conclusion, I am truly fortunate to have been surrounded by such a supportive and inspiring network of individuals. This internship journey has been a valuable learning experience, and I am thankful for every person who has played a part in it.

TABLE OF CONTENTS

1 INTRODUCTION

The research focuses on analyzing the Energy, Metals, and Agriculture Sectors within the commodity markets during a significant global event. While various commodities across these sectors offer diversification prospects for portfolio managers, our hypothesis suggests that this diversified portfolio loses its diversification benefits amidst a major crisis. Our study delves into the interrelationships among these commodities during the crisis period, comparing these correlations with those before and after the event.

To depict the global asset prices central to our study, we employ commodity subindices. Specifically, we utilize the following commodities: Bloomberg Commodity Index (BCOM), BCOM Energy (BCOMEN), BCOM Industrial Metals (BCOMIN), BCOM Agriculture (BCOMAG), BCOM Brent Crude (BCOMCO), BCOM WTI Crude Oil (BCOMCL), BCOM Natural Gas (BCOMNG), BCOM Unleaded Gasoline (BCOMRB), BCOM Gold (BCOMGC), BCOM Platinum (BCOMPL), BCOM Palladium (BCOMPA), BCOM Aluminium (BCOMAL), BCOM Silver (BCOMSI), BCOM Copper (BCOMHG), BCOM Zinc (BCOMZS), BCOM Lead (BCOMPB), BCOM Nickel (BCOMNI), BCOM Wheat (BCOMWH), BCOM Corn (BCOMCN), BCOM Soybeans (BCOMSY), BCOM Coffee (BCOMKC), BCOM Sugar (BCOMSB), BCOM Cocoa (BCOMCC), and BCOM Cotton (BCOMCT).

For our analysis, we leverage Python's "faton" package, specifically utilizing the Detrended Cross-Correlation Analysis(DCCA) model to evaluate the correlation among the diverse commodities. We use this model over the timeline of the major global crisis. Specifically, COVID, 2008 Housing Market Crash, Ebola Outbreak, European Crisis, and Asian Crisis.

2 METHODOLOGY

This section outlines the methodology employed in this research to investigate the correlation between commodity prices across different sectors during significant events and crises. The approach combines statistical techniques, correlation analysis, and time-series modeling, utilizing the Detrended Cross-Correlation Analysis (DCCA) method implemented in the Fathon library in Python.

Data Collection and Preprocessing

A comprehensive dataset of major commodities across three sectors - Agriculture, Metals, and Energy - was assembled. This dataset includes prices of commodities such as wheat, soybean, crude oil, gold, aluminum, and more. The data spans a specific time frame that encompasses significant events and crises, including the COVID-19 pandemic, the 2008 housing market crash, the Ebola outbreak, the European crisis, and other major geopolitical and economic upheavals.

Before analysis, the raw price data underwent preprocessing steps to ensure consistency and reliability. This included handling missing values by duplicating previous values, outlier detection, and normalization to minimize the impact of differing price scales among commodities.

Correlation Analysis

To assess the relationships between commodity prices during different events and crises, correlation analysis was performed. Correlation coefficients were calculated for pairs of commodities within each sector (Agriculture, Metals, and Energy). These coefficients provide insight into the strength and direction of linear or monotonic relationships between commodity prices.

Detrended Cross-Correlation Analysis (DCCA)

To explore potential nonlinear dependencies and long-range correlations between commodity prices, the Detrended Cross-Correlation Analysis (DCCA) method was employed. DCCA extends the conventional cross-correlation analysis by considering both short-term and long-term dependencies between time series, making it suitable for capturing complex relationships that may exist among commodity prices.

The DCCA involves the following steps:

Detrending: Each commodity price time series is detrended to remove any linear trends. This step helps focus on the underlying correlation structure rather than the overall trend.

Cross-Correlation Profile Calculation: The cross-correlation profile between two detrended commodity price series is computed. This profile quantifies the similarity between the fluctuations of the two series at different time scales.

Profile Fluctuation Analysis: The fluctuation of the cross-correlation profile is analyzed to identify potential long-range correlations. This involves fitting a power-law relationship to assess the scaling behavior of the fluctuations.

A rolling period of 20 days was selected and events with a timeframe significantly greater than 20 days were selected.

Statistical Analysis and Interpretation

The correlation coefficients obtained from the traditional correlation analysis and the DCCA results were analyzed statistically to identify patterns and trends across different sectors and events. Heatmaps of correlations were generated to visually represent the relationships between commodities before and during specific crises and events.

Timeframes for analysis were: Pre-event: 2 years before the event until and unless an event overlaps, in that case, the amount of time without any ongoing event. Event: Timeframe of the underlying Event. Post-event: 2 years after the event until and unless an event overlaps or data gets limited due to the event being too recent, in this case, the available timeframe is used.

Going over a few heatmaps:

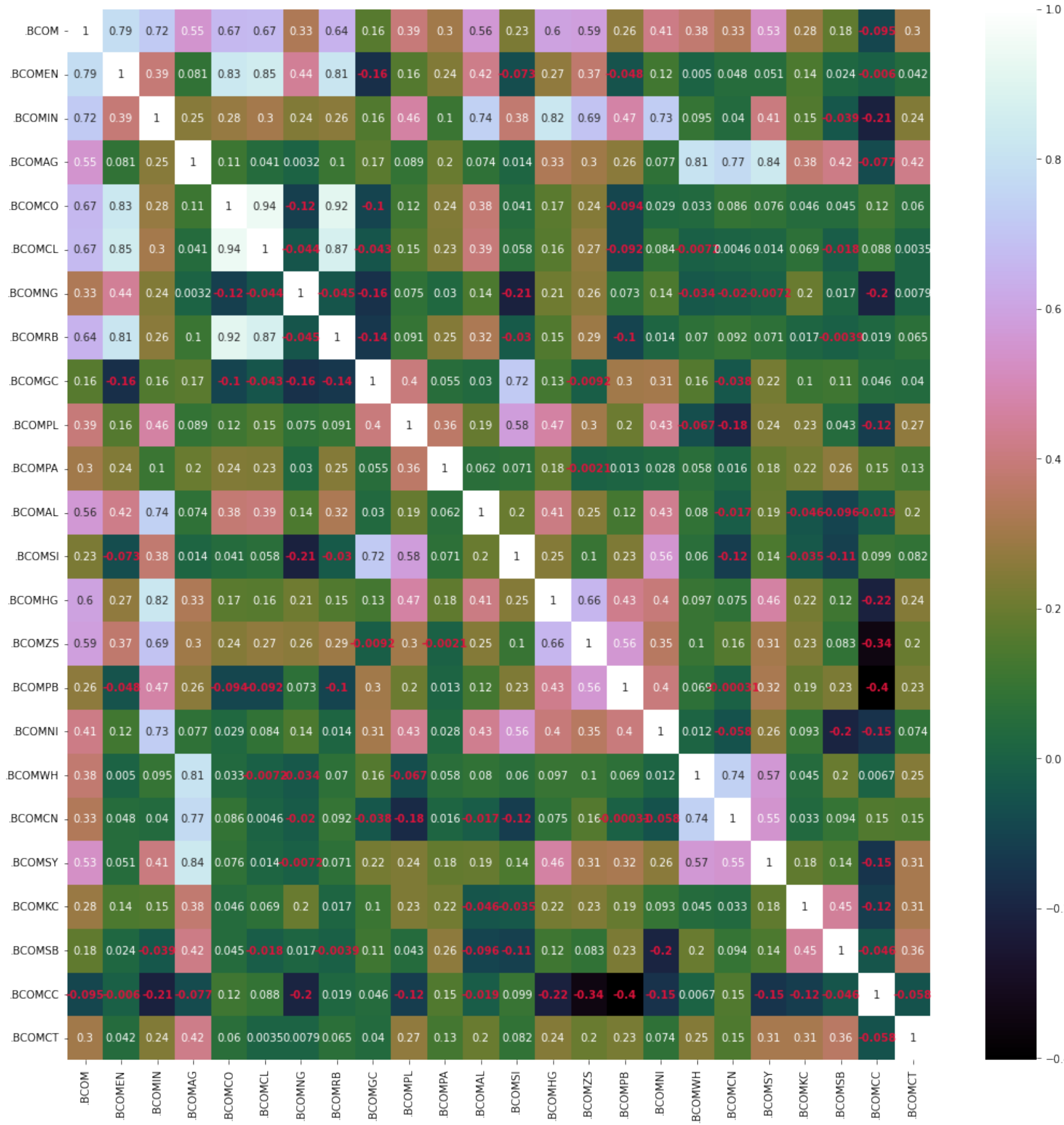

Figure 1: Pre COVID

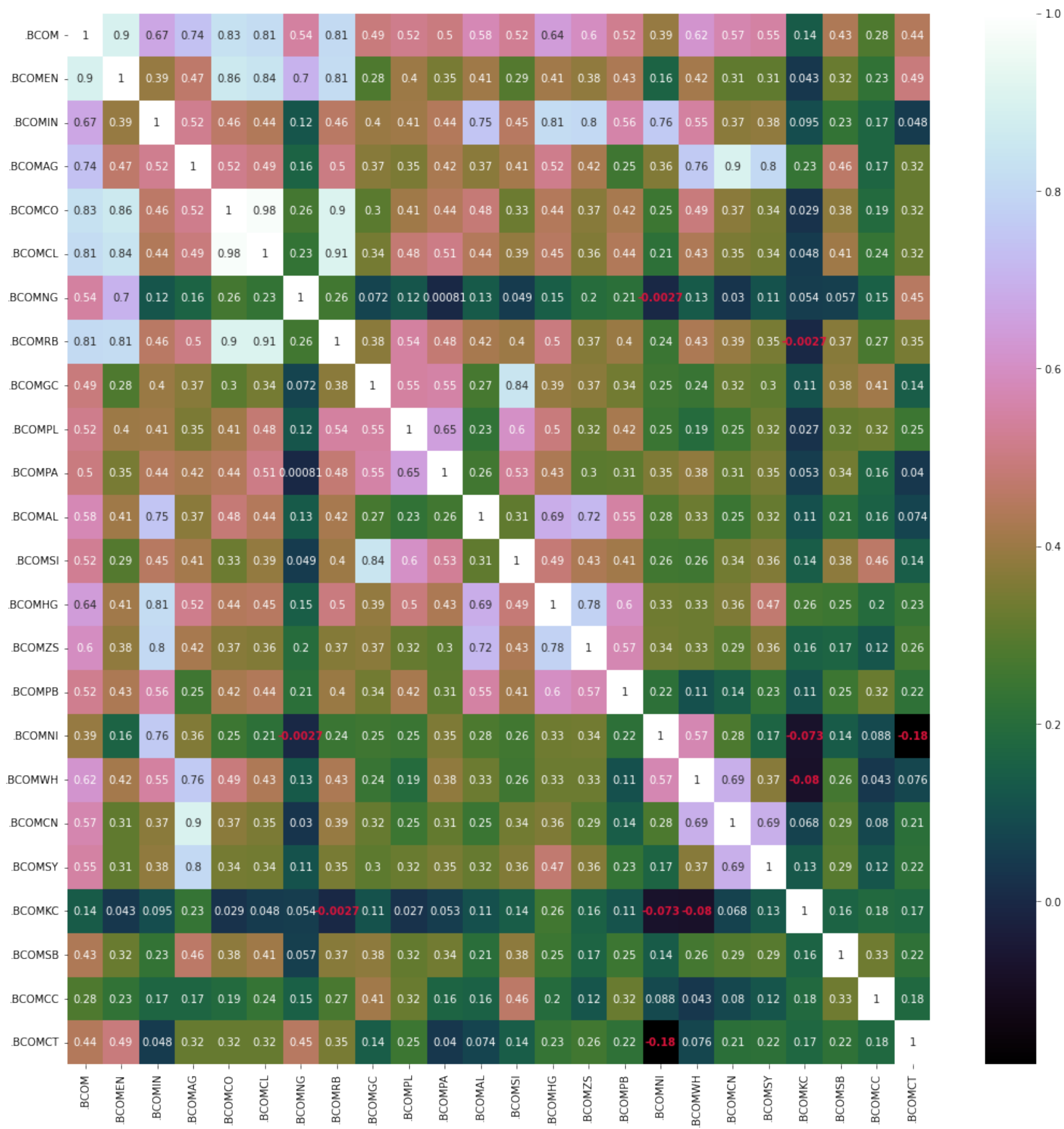

Figure 2: During COVID

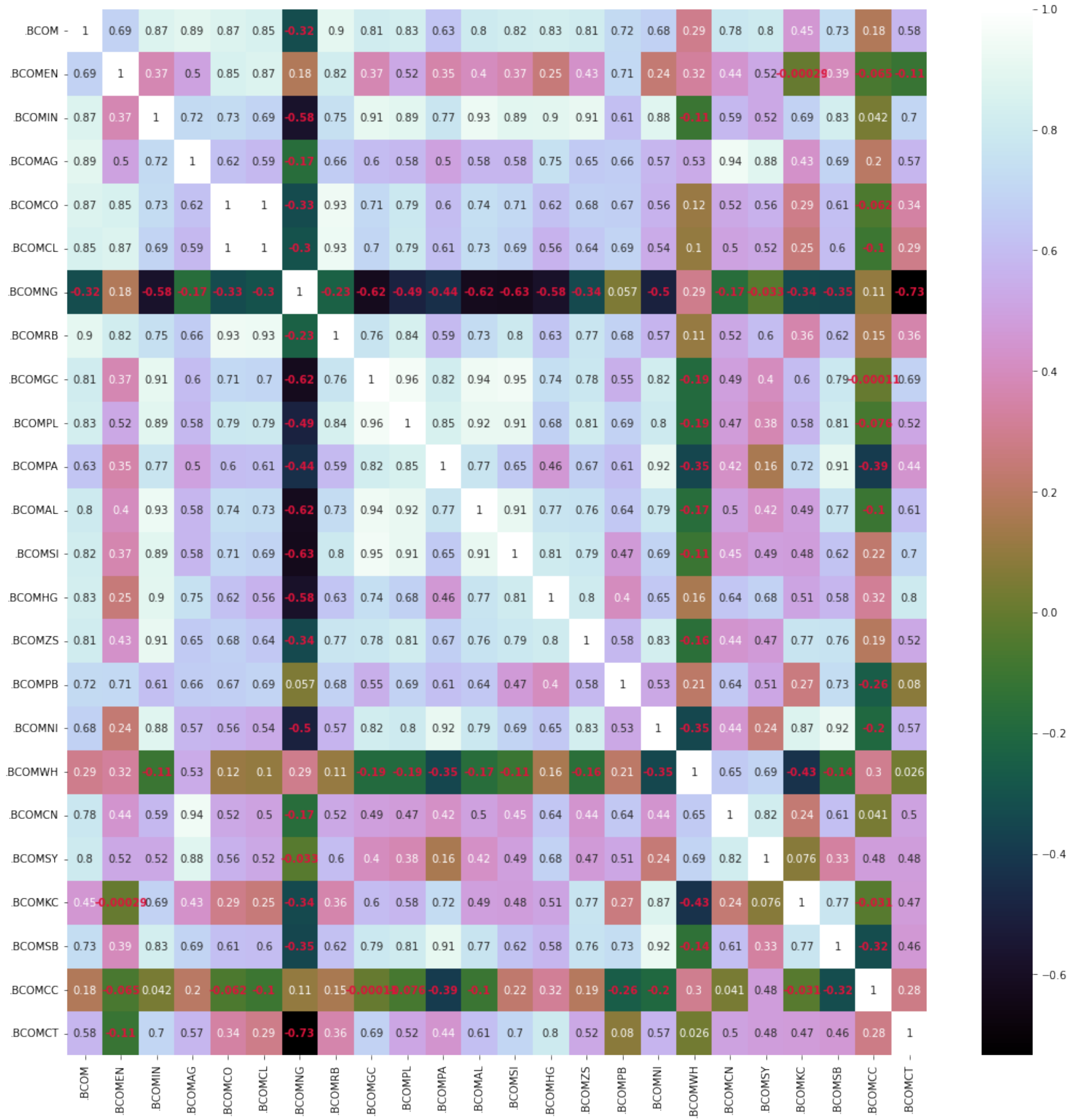

Figure 3: Post COVID

After getting correlation values during different timeframes, we can calculate the average correlations within sectors and cross-sectors. Reference the below code snippets for the mathematics used,

```python
# Calculating Average correlations within a sector
Sector = 0
for i in range(n):
    for j in range(n):
        Sector = Sector + rho_corr[m+i][m+j]
Sector = (Sector-n)/(n^2-n)
print(Sector)
```

Figure 4: Inter Sector

```python
# Calculating Average correlations across different sectors
EnergyMetal = 0
for i in range(n):
    for j in range(m):
        EnergyMetal = EnergyMetal + rho_corr[p+i][q+j]
EnergyMetal = (EnergyMetal)/(n*m)
print(EnergyMetal)
```

Figure 5: Cross Sector

Going over the results of average correlation values,

Table 1: Commodity Comparison

Commodity	PreCOVID	COVID	PostCOVID (2 months)	Pre2008	2008	Post2008 (6 Months)
Energy	0.4206	0.5899	0.3313	0.5230	0.6821	0.6590
Metal	0.2961	0.4487	0.7558	0.4371	0.4380	0.5154
Agri	0.1979	0.2241	0.3007	0.2034	0.4398	0.3569
EnergyMetal	0.1013	0.3253	0.3957	0.1660	0.3362	0.4189
MetalAgri	0.0772	0.2360	0.3789	0.1842	0.3079	0.3284
AgriEnergy	0.0323	0.2641	0.2156	0.1209	0.3016	0.3520

Table 2: Comparison of Different Periods

PreEbola	Ebola	PostEbola	PreEuro=Post2008	Euro	PostEuro	PreAsian	Asian	PostAsian
0.3586	0.5803	0.5606	0.6590	0.5881	0.5358	0.5931	0.5186	0.6043
0.6478	0.3190	0.4327	0.5154	0.6039	0.5807	0.2663	0.2323	0.1557
0.1667	0.2492	0.1352	0.3569	0.2689	0.3830	0.1057	0.1967	0.0996
0.2190	0.1131	0.0519	0.4189	0.5215	0.3815	-0.0423	0.1180	0.0638
0.0855	0.0503	0.0791	0.3284	0.2872	0.3580	0.0605	0.0275	0.0841
0.1262	0.0670	0.1527	0.3520	0.2105	0.3088	0.1305	0.0701	0.0399

Averaging the values for all events:

Table 3: Commodity Averages

Commodity	Average Pre Event	Average during Event	Average Post Event
Energy	0.5108	0.5918	0.5382
Metal	0.4325	0.4084	0.4881
Agri	0.2061	0.2757	0.2551
EnergyMetal	0.1726	0.2828	0.2624
MetalAgri	0.1472	0.1818	0.2457
AgriEnergy	0.1524	0.1826	0.2138
Net	0.2703	0.3205	0.3339

Plotting the Net correlations for inter and cross sectors,

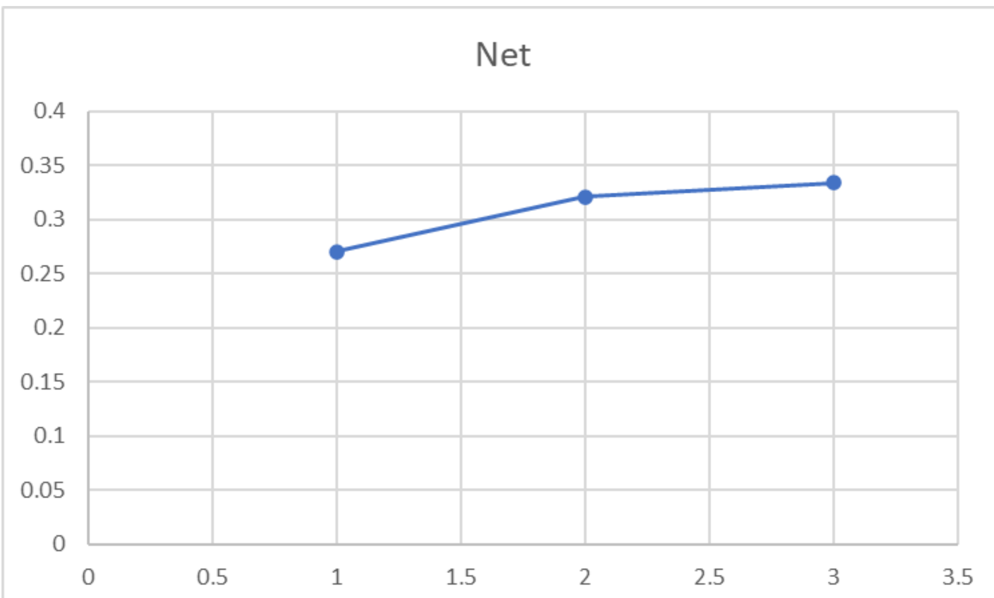

Figure 6: Net Correlations

3 DATA

Data Collection

For this research paper, a comprehensive dataset of daily commodity prices was collected from Reuters Refinitiv Eikon for Bloomberg commodity indices and subindices. The dataset covers the period from January 2, 1991, to July 21, 2023, allowing for a thorough analysis of commodity price behavior across various events and crises.

Commodity Subindices

To capture a diverse representation of the commodity markets, a selection of key commodity subindices was chosen. These subindices are part of the Bloomberg Commodity Index (BCOM) and cover three primary sectors: Agriculture, Metals, and Energy. The chosen subindices include:

Bloomberg Commodity Index (BCOM)

BCOM Energy (BCOMEN)

BCOM Industrial Metals (BCOMIN)

BCOM Agriculture (BCOMAG)

BCOM Brent Crude (BCOMCO)

BCOM WTI Crude Oil (BCOMCL)

BCOM Natural Gas (BCOMNG)

BCOM Unleaded Gasoline (BCOMRB)

BCOM Gold (BCOMGC)

BCOM Platinum (BCOMPL)

BCOM Palladium (BCOMPA)

BCOM Aluminium (BCOMAL)

BCOM Silver (BCOMSI)

BCOM Copper (BCOMHG)

BCOM Zinc (BCOMZS)

BCOM Lead (BCOMPB)

BCOM Nickel (BCOMNI)

BCOM Wheat (BCOMWH)

BCOM Corn (BCOMCN)

BCOM Soybeans (BCOMSY)

BCOM Coffee (BCOMKC)

BCOM Sugar (BCOMSB)

BCOM Cocoa (BCOMCC)

BCOM Cotton (BCOMCT)

Data Preprocessing

The collected daily price data underwent thorough preprocessing to ensure its reliability and consistency. The preprocessing steps included handling missing data points, identifying and addressing outliers, and ensuring proper alignment of the data. To facilitate meaningful analysis and comparison, the commodity prices were normalized to a common scale.

Data Availability

Data is made available on request. The dataset spans several decades, covering various significant events and crises, such as the COVID-19 pandemic, the 2008 housing market crash, the Ebola outbreak, the European crisis, and other major geopolitical and economic upheavals. The extensive timeframe and diverse set of commodities provide a robust foundation for investigating the correlation dynamics during these critical periods.

4 CONCLUSION

In this research paper, we embarked on a comprehensive exploration of the correlation dynamics between commodity prices across the Agriculture, Metals, and Energy sectors during various significant events and crises. Our objective was to assess whether these pivotal events trigger heightened correlations among commodities, potentially challenging the efficacy of diversified portfolios in times of turmoil. Our primary hypothesis posited that during these exceptional events, the correlations among distinct commodities would escalate, thereby impacting the conventional benefits of risk-mitigating diversification strategies.

Leveraging statistical methods, including correlation analysis and time-series modeling, with the added insight of the Detrended Cross-Correlation Analysis (DCCA) from Fathon, we delved into the dynamics of commodity correlations. Our results illuminated intriguing patterns in the change of correlations over the course of these events. However, what stands out is the unexpected phenomenon that emerged. Contrary to our initial expectations, the average correlation of all assets increased by 18.597 percent during the crisis period. Even more surprisingly, this elevated correlation persisted with a 4.164 percent increase post-crisis, rather than reverting to pre-crisis levels.

One plausible explanation for this finding is that our analysis captured the immediate aftermath of the events, when the market may still have been in a state of flux, recovering from the upheaval. The lingering impact of the crises on market sentiment and behaviors might have contributed to this prolonged elevated correlation.

This outcome underscores the complexity of market responses to significant events and the importance of considering both immediate and longer-term effects. The interplay of factors beyond simple correlation, such as market sentiment, behavioral dynamics, and macroeconomic trends, cannot be understated. Investors and asset managers should exercise caution when extrapolating short-term trends to guide long-term strategies.

In conclusion, our research contributes to a nuanced understanding of how commodity correlations evolve during times of crisis. It highlights the need for adaptive and multifaceted portfolio management strategies that consider the shifting landscape of market relationships. As investors navigate the uncertainties of financial markets, our insights offer a valuable perspective for making informed decisions that can potentially enhance risk mitigation and overall performance during periods of economic turmoil.

OTHER WORK DONE DURING THE PERIOD OF INTERNSHIP

Overall I worked on 3 projects along with helping Ph.D. graduates with some of their projects mostly in the field of data analytics and data engineering using Pandas from Python.

The three projects are as follows:

1. Brief analysis of the correlation between commodities during several crises.
2. Designed an Experiment to analyze if an AI-assisted chatbox could positively impact investor decisions in developing portfolios and making investment decisions. This was done through a library called Streamlit in Python which provides an easy web interface to test machine learning models. Hosting was made possible by ngrok by having a local host connect to the internet. Chat-GPT API was connected at the backend to provide access to the chatbox.
3. Impacts of customer backlash on the underlying company and its competitors. In this project, a dataset of news was captured from Marketline with - The type of department the news is targeted at, the sentiment of the news, the company name, the date of publication, etc as its features. We aim to feed this data to a machine-learning model in order to understand how news impacts the company when it comes to its stock price. We filtered the news with keywords detecting customer backlash and similar public outcry situations. Considering a fair market hypothesis, we assume that this set of news impacts companies' cash flows, in turn impacting company valuations giving an advantage to competitors to expand in their respective markets and hence positively impacting their valuation in a higher relative timeframe.