Analysis and Modeling of Volatility Regimes in Financial Markets: an Approach using Unsupervised Clustering and Macroeconomic Factors

A thesis presented in partial fulfilment of the requirements for the degree of Master of Science

Philippe MARAVAL

S.M. Degree Candidate in Data Analysis & Artificial Intelligence

Faculty Supervisor: Mario Hernandez Tinoco, Ph.D

EDHEC Business School ⋄ Lille, France

2025

Contents

List of Figures

List of Tables

Abstract

Financial market volatility is a critical factor influencing investment decisions, risk management, and economic stability. Traditional volatility models, such as Generalized Autoregressive Conditional Heteroskedasticity (GARCH) and Markov Switching Models (MSM), capture some aspects of regime changes but often struggle with non-linear shifts in market dynamics. This thesis introduces an alternative approach, leveraging **unsupervised clustering techniques** to identify volatility regimes and integrating **macroeconomic factors** to enhance interpretability. We apply K-means, DBSCAN, and Gaussian Mixture Models (GMM) to categorize market periods into distinct volatility regimes and analyze their relationships with inflation, interest rates, GDP growth, and other macroeconomic indicators. The results demonstrate that unsupervised learning provides a **more flexible, data-driven framework** for understanding regime shifts. A supervised model is subsequently developed to predict volatility regimes based on macroeconomic conditions, offering insights for investors and policymakers.

Acknowledgements

I would like to express my sincere gratitude to those who have supported and guided me throughout the development of this thesis.

First and foremost, I extend my deepest appreciation to my thesis supervisor, **Dr. Mario Hernandez Tinoco**, for his invaluable guidance, encouragement, and insightful feedback, which have been instrumental in shaping this research.

I would also like to extend my deepest gratitude to my friend **Wadi Moughanim**, Oil Quantitative Analyst at Litasco, whose support in shaping the research topic and deep knowledge of commodity markets significantly contributed to the foundation of this study.

I am also incredibly grateful to my friend T**homas Stokolm Jensen**, Investment Advisor at UBS, for his unwavering support and thoughtful discussions that greatly contributed to refining my understanding of investment strategies and market dynamics.

A special thank you to **Hannes Huett**, Executive Director at UBS for Quantitative Strategies, whose expertise in quantitative finance provided critical insights that strengthened the analytical foundation of this work.

Their valuable contributions, along with the support of my colleagues, friends, and family, have played a crucial role in the successful completion of this research.

Chapter 1

Introduction

Here you should put a short introduction to your chapter. What is covered? In how much detail? Imagine you were coming back to this in 10 years time and wanted to find that one key equation, this part of the chapter should orient the reader to help find that information.

1.1 Context and Problem Statement

Volatility is a fundamental characteristic of financial markets, representing the degree of variation in asset prices over time. Periods of heightened volatility, such as during the 2008 financial crisis or the COVID-19 pandemic, illustrate how market conditions can shift unpredictably, impacting investor sentiment and financial stability. High volatility is associated with increased risk and uncertainty, whereas low volatility often signals market stability. It is crucial for understanding both the risks and opportunities inherent in investments. Financial volatility can manifest in different ways and is often influenced by a combination of macroeconomic conditions, investor sentiment, and geopolitical events. In this context, volatility regimes—or distinct periods characterized by specific levels of volatility—are of particular interest to investors and risk managers. Identifying and analyzing these volatility regimes offers invaluable insights into the stability or instability of the market and helps guide investment decisions and risk management practices.

Volatility is of particular importance because it directly impacts asset pricing and risk assessment. High volatility typically indicates higher risk, as prices are more likely to swing dramatically in short timeframes. Conversely, low volatility is often associated with stable markets, where price movements are gradual and less erratic. Both scenarios are relevant for financial stakeholders, as periods of high volatility may present opportunities for profit through rapid price changes, while stable, low-volatility periods might appeal to more risk-averse investors. Therefore, understanding the dynamics of volatility is not only a tool for managing risk but also for identifying strategic investment opportunities.

However, traditional approaches to volatility analysis have limitations, especially in identifying regime shifts. Standard models, such as GARCH (Generalized Autoregressive Conditional Heteroskedasticity), primarily focus on short-term fluctuations and do not necessarily capture the broader regime changes in volatility over time. These models typically assume that volatility is driven solely by past prices or returns, overlooking the influence of external macroeconomic factors like inflation, interest rates, and economic growth. As a result, traditional models might fail to account for sudden market shifts prompted by economic crises, policy changes, or other external shocks. For instance, these models struggled to capture the sudden shifts in market dynamics during the COVID-19 pandemic, where volatility was not only driven by financial variables but also by unprecedented macroeconomic policy changes and supply chain disruptions. These models generally rely on pre-defined assumptions about state transitions, limiting their adaptability. This presents a gap in existing research and a challenge for practitioners who need to anticipate and manage volatility beyond the scope of short-term predictions.

Given these challenges, there is a pressing need to develop alternative methods for categorizing volatility regimes. Unsupervised clustering techniques offer a promising solution, as they can classify data without requiring pre-labeled inputs, making them ideal for uncovering hidden patterns in market behavior. By clustering periods of market activity based on volatility levels, we can identify distinct regimes that may align with specific economic conditions. These techniques, however, are not without challenges. Properly defining the input features and preprocessing the data are critical steps, as clustering algorithms are sensitive to initial conditions and input variables. Consequently, clustering volatility regimes requires careful consideration of both the financial data and the economic indicators that might influence regime shifts.

The problem this dissertation seeks to address is twofold: first, how to effectively categorize different volatility regimes in financial markets; and second, how to identify the macroeconomic factors that influence these regimes. Specifically, the research will explore the application of unsupervised clustering algorithms to segment periods of varying volatility in financial markets, and then analyze how economic factors such as inflation, interest rates, and GDP growth might correlate with these volatility regimes. This approach not only provides a new perspective on volatility analysis but also enhances our understanding of the broader economic context surrounding market fluctuations.

1.2 Objectives of the Dissertation

The primary objective of this dissertation is to explore the potential of unsupervised clustering algorithms in identifying volatility regimes in financial markets. Unlike traditional volatility models, which often rely on assumptions about data distribution and external shocks, unsupervised clustering does not require predefined categories or labels. This flexibility allows for the discovery of volatility regimes without relying on assumptions about market behavior, making it particularly well-suited for analyzing complex and dynamic financial data.

To achieve this objective, the dissertation will first utilize clustering algorithms to identify volatility regimes in a dataset of financial market indices. The chosen algorithms may include K-means, DBSCAN, and Gaussian Mixture Models (GMM), each of which offers unique strengths for segmenting data based on volatility patterns. K-means, for example, is well-suited for identifying distinct clusters in data with clear separation, while DBSCAN is effective for handling noise and outliers, which are common in financial datasets. GMM, with its probabilistic approach, can accommodate overlapping clusters, allowing for a more nuanced classification of volatility regimes. The use of multiple clustering algorithms will provide a comprehensive view of the data and allow for comparisons across methods, enhancing the robustness of the analysis.

Once the volatility regimes have been identified, the next objective is to analyze these regimes based on a set of macroeconomic indicators. This part of the analysis will examine how economic variables like inflation, interest rates, and GDP growth correlate with each identified regime. By mapping these indicators to different periods of volatility, the dissertation aims to uncover patterns and relationships that may help explain why certain volatility regimes arise under specific economic conditions. This analysis will involve statistical techniques to assess correlations and potential causations, providing a data-driven foundation for understanding the economic underpinnings of volatility.

A final objective of the dissertation is to explore the feasibility of using these macroeconomic factors to model or predict volatility regimes. This step will involve the development of a supervised model that uses economic indicators as input variables to classify periods according to their volatility regime. Although the primary focus of this research is on unsupervised clustering, a supervised model could add a predictive dimension, allowing for the anticipation of volatility shifts based on current economic conditions. Logistic regression or decision trees may be employed as baseline models for this purpose, as they are interpretable and well-suited for examining the relationships between input variables and categorical outcomes. This component of the research aims to provide actionable insights for practitioners, as it suggests a pathway toward predictive modeling of volatility regimes using readily available economic data.

1.3 Originality of the Project

This dissertation presents a novel approach to volatility analysis by focusing on qualitative, rather than quantitative, segmentation of volatility regimes. Unlike many traditional studies that aim to predict asset prices or forecast volatility levels directly, this research does not attempt to make explicit predictions about price movements. Instead, it seeks to understand the structural characteristics of volatility itself by categorizing different regimes and exploring their economic context. This qualitative approach aligns with the goals of risk management and strategic investment, as it emphasizes the identification of stable and unstable market periods without focusing solely on price prediction.

The originality of this project lies in its application of unsupervised learning methods to financial volatility, a domain traditionally dominated by supervised models and time-series analysis. By using clustering algorithms, this dissertation diverges from conventional volatility forecasting techniques and instead offers a fresh perspective on market analysis. Clustering allows for the discovery of hidden patterns in data, enabling a more nuanced understanding of how volatility behaves across different economic conditions. Furthermore, the inclusion of macroeconomic variables as contextual factors adds an additional layer of insight, as it allows for the examination of how external conditions influence volatility in ways that may not be captured by price-based models alone.

Additionally, this research contributes to the field by combining elements of regime-switching analysis with clustering and macroeconomic analysis. While regime-switching models, such as the Markov Switching Model, are commonly used to capture changes in market behavior, they often rely on predefined states and do not easily incorporate external factors. In contrast, the clustering approach used in this dissertation is data-driven and adaptable, making it better suited for uncovering dynamic relationships between volatility and economic conditions. By merging these two analytical perspectives, this research offers a more holistic view of volatility, with potential applications for both academic researchers and financial practitioners.

In summary, this dissertation's originality stems from its qualitative, unsupervised approach to volatility analysis, its focus on macroeconomic context, and its integration of clustering and regime-switching concepts. This innovative approach not only broadens the scope of volatility research but also provides practical insights that could benefit investors and risk managers seeking to navigate complex market conditions. By examining volatility regimes through the lens of unsupervised clustering and macroeconomic analysis, this research aims to fill a significant gap in existing literature and contribute to a more comprehensive understanding of financial market dynamics.

Chapter 2

Literature Review

Financial markets experience dynamic and unpredictable volatility driven by economic conditions, investor behavior, and external shocks. While traditional models like ARCH and GARCH capture time-varying volatility, they assume constant patterns over time, limiting their ability to detect sudden regime shifts. This has led to growing interest in alternative approaches, particularly those incorporating **machine learning** and **macroeconomic factors**.

This review explores the research question: *How can unsupervised clustering techniques, combined with macroeconomic factors, be utilized to identify and model volatility regimes in financial markets?* By analyzing research published after 2010, it traces key developments in the field and proposes a framework for integrating these techniques into volatility modeling for improved predictive analysis and risk management.

A systematic approach was used to ensure a comprehensive and transparent review.

2.1 Methodology

Search Strategy

- **Databases**: Scopus, Web of Science, JSTOR, and Google Scholar.

- **Keywords**: The search terms included combinations of the following: "volatility regimes," "unsupervised clustering," "financial markets," "macroeconomic factors," "machine learning," "predictive analysis," and "regime-switching models."

Inclusion Criteria:

- Peer-reviewed journal articles and conference papers.
- Studies published between 2010 and 2023.
- Focus on mathematical and financial applications.

Exclusion Criteria:

- Studies older than 2010 (unless foundational, e.g., Hamilton, 1989).
- Studies that do not focus on financial markets or volatility modeling.

Selection Process:

The initial search yielded 850 articles. After removing duplicates and screening titles and abstracts, 150 articles were selected for full-text review.

Data Extraction:

Collected details: author(s), year, title, methodology (e.g., clustering techniques, ML algorithms), and key findings on volatility regimes and macroeconomic factors.

Synthesis:

- The extracted data were organized chronologically to trace the evolution of the field.
- Key themes and trends were identified, focusing on the integration of unsupervised clustering and macroeconomic factors in volatility modeling.

2.2 Theoretical foundation

Volatility Modeling in Financial Markets

Recent advancements in financial modeling have increasingly leveraged unsupervised learning techniques to identify volatility regimes. Hamilton's (1989) regime-switching framework [1] remains foundational, but modern applications incorporate machine learning methods such as K-means and Gaussian Mixture Models (GMM) (Bucci & Benoit, 2023). These approaches provide a flexible, data-driven methodology for uncovering structural breaks in financial markets, improving upon traditional ARCH/GARCH models (Engle, 1982). By integrating macroeconomic indicators such as inflation, interest rates, and GDP growth, this research aims to enhance volatility forecasting and improve market regime classification.

Regime-Switching Models

Regime-switching models, introduced by Hamilton (1989), capture structural breaks in time series by allowing transitions between distinct states. In financial markets, they model shifts between high- and low-volatility regimes, helping identify market stress or stability. These models are particularly useful for capturing volatility clustering, where high or low volatility persists over time.

Machine Learning in Finance

Machine learning, particularly unsupervised clustering, has transformed financial analysis by detecting patterns beyond traditional methods. Algorithms like k-means, hierarchical clustering, and Gaussian mixture models group data without labels, helping identify market regimes, segment time series, and detect anomalies. Lopez de Prado (2018) highlights their potential for enhancing predictive analysis and risk management through advanced computational techniques.

Macroeconomic Factors and Volatility

Macroeconomic factors like interest rates, inflation, and GDP growth shape market volatility by influencing investor behavior and asset prices. While traditional models treat these factors as external inputs, recent research integrates them directly using techniques like vector autoregression (VAR) and dynamic factor models for a more comprehensive volatility analysis.

2.3 Chronological Review of Literature

2.3.1 Early works : pre-2010

The foundation for understanding volatility regimes and clustering in financial markets was laid by several key studies:

Hamilton, J.D. (1989). *A New Approach to the Economic Analysis of Nonstationary Time Series and the Business Cycle.* Econometrica.

Introduced regime-switching models to model structural breaks in economic time series.

Engle, R.F. (1982). *Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation.* Econometrica.

Developed the ARCH model, which became the cornerstone of volatility modeling.

Kaufman, L., & Rousseeuw, P.J. (1990). *Finding Groups in Data: An Introduction to Cluster Analysis.* Wiley-Interscience.

Developed clustering methods later applied to financial data segmentation.

Peters, E.E. (1991). *Chaos and Order in the Capital Markets: A New View of Cycles, Prices, and Market Volatility.* John Wiley & Sons.

Applied chaos theory to financial markets, providing insights into volatility and complexity.

Peters, E.E. (1994). *Fractal Market Analysis: Applying Chaos Theory to Investment and Economics.* John Wiley & Sons.

Introduced the Fractal Market Hypothesis, offering a framework for analyzing volatility.

2.3.2 2010 - 2020

Researchers began integrating machine learning with traditional financial models, leading to significant advancements in unsupervised clustering for financial markets. Key developments include:

Tsay, R.S. (2010). *Analysis of Financial Time Series.* Wiley.

Provided methodologies for analyzing financial time series, including volatility modeling and regime detection.

Diebold, F.X., & Yilmaz, K. (2012). *Better to Give than to Receive: Predictive Directional Measurement of Volatility Spillovers.* International Journal of Forecasting.

Examined volatility spillovers between markets, emphasizing macroeconomic factors.

Bollerslev, T., & Todorov, V. (2011). *Estimation of Jump Tails.* Econometrica.

Developed methods to estimate jump tails in financial time series, key for extreme volatility.

Lopez de Prado, M. (2018). *Advances in Financial Machine Learning.* Wiley.

Published a seminal work on machine learning in finance, including clustering for regimes.

Bianchi, D., Büchner, M., & Tamoni, A. (2017). *Bond Risk Premiums with Machine Learning.* Review of Financial Studies.

Applied machine learning to bond risk premiums, highlighting clustering's financial potential.

Gu, S., Kelly, B., & Xiu, D. (2020). *Empirical Asset Pricing via Machine Learning.* Review of Financial Studies.

Used machine learning to enhance asset pricing models and volatility forecasting.

Christensen, B.J., & Prabhala, N.R. (2018). *The Relation Between Implied and Realized Volatility.* Journal of Financial Economics.

Examined the link between implied and realized volatility, stressing macroeconomic factors.

2.3.3 2020 - Present

Recent research has focused on refining clustering techniques and integrating them with macroeconomic factors. Key trends include:

Hautsch, N., & Voigt, S. (2021). *Machine Learning in Financial Markets: A Survey.* Journal of Economic Surveys.

Surveyed machine learning applications in finance, including volatility modeling.

Chen, L., Pelger, M., & Zhu, J. (2022). *Deep Learning in Asset Pricing.* Journal of Financial Economics. Explored the use of deep learning models for asset pricing and volatility forecasting.

Bucci, A., & Benoit, D.F. (2023). *Unsupervised Learning for Financial Time Series: A Review.* Journal of Financial Econometrics.

Reviewed unsupervised learning in financial time series, focusing on clustering and regimes.

Giglio, S., & Kelly, B. (2023). *Excess Volatility: Beyond the Standard Models.* Journal of Finance.

Investigated excess volatility, using machine learning to uncover patterns and regimes.

2.4 Key Themes and Findings

The literature on volatility modeling using unsupervised clustering and macroeconomic factors can be organized around three key themes:

Machine Learning in Volatility Modeling:

Unsupervised clustering techniques, such as k-means and hierarchical clustering, have proven effective in identifying volatility regimes. Lopez de Prado (2018) [2] and Bianchi et al. (2017) demonstrated their ability to segment financial data into distinct regimes, while Gu et al. (2020) and Chen et al. (2022) showed that integrating machine learning with traditional models improves volatility forecasting accuracy.

Macroeconomic Factors and Volatility:

Macroeconomic indicators, such as interest rates and GDP growth, significantly influence market volatility. Diebold & Yilmaz (2012) [3] and Giglio & Kelly (2023) [4] highlighted their role in driving volatility spillovers and regime shifts, while Christensen & Prabhala (2018) emphasized the importance of incorporating these factors into volatility models.

Predictive Analysis and Regime Detection:

Clustering techniques, combined with regime-switching models, enhance predictive analysis by identifying and forecasting volatility regimes. Hamilton (1989) [1] laid the foundation for regime-switching models, while Lopez de Prado (2018) [2] and Bucci & Benoit (2023) applied clustering to detect and predict regime shifts in financial markets.

2.5 Conclusion

This literature review has synthesized recent research on the use of unsupervised clustering and macroeconomic factors in modeling volatility regimes in financial markets. Key findings include:

- Machine learning techniques, particularly unsupervised clustering, are highly effective for identifying and predicting volatility regimes.
- Macroeconomic factors play a critical role in shaping market volatility and should be integrated into volatility models.
- Predictive analysis using clustering and machine learning offers significant advantages over traditional models, particularly in capturing regime shifts and extreme events.

The proposed framework provides a roadmap for integrating these techniques into volatility modeling, offering a more robust and accurate approach for understanding and predicting financial market dynamics. This review contributes to your master's thesis by providing a comprehensive foundation for your analysis and modeling of volatility regimes.

Building on the foundational work of Hamilton (1989) [1] and recent advancements in financial clustering (Lopez de Prado, 2018 [2]; Bucci & Benoit, 2023), this dissertation applies a combination of unsupervised clustering techniques and macroeconomic analysis to identify volatility regimes. The following methodology section details the data sources, clustering procedures, and supervised modeling techniques used to achieve these objectives.

Chapter 3

Hypotheses Development

3.1 Identified Gaps in the Literature

Limited integration of unsupervised clustering and macroeconomic factors in volatility modeling. Lack of robust frameworks for regime detection using machine learning. Insufficient focus on predictive accuracy in existing models.

3.2 Research Objectives

Develop a framework integrating unsupervised clustering and macroeconomic factors. Evaluate its effectiveness in identifying and predicting volatility regimes. Assess its predictive accuracy compared to traditional models.

3.3 Hypotheses

H1: Integration of Unsupervised Clustering and Macroeconomic Factors A framework combining unsupervised clustering and macroeconomic factors will more accurately identify volatility regimes than models using financial data alone. Rationale: Macroeconomic factors (e.g., interest rates, GDP growth) influence market volatility, as shown by Diebold & Yilmaz

(2012) [3] and Giglio & Kelly (2023).

H2: Effectiveness of Unsupervised Clustering Unsupervised clustering techniques (e.g., k-means, hierarchical clustering) will outperform traditional methods (e.g., GARCH) in detecting volatility regimes. Rationale: Clustering captures complex, non-linear patterns missed by traditional models, as demonstrated by Lopez de Prado (2018) and Bucci Benoit (2023).

H3: Predictive Accuracy The proposed framework will exhibit higher predictive accuracy in forecasting volatility regimes compared to traditional models. Rationale: Combining financial data and macroeconomic indicators improves forecasting, as highlighted by Gu et al. (2020) and Chen et al. (2022).

H4: Macroeconomic Factors as Drivers of Regime Shifts Macroeconomic factors will significantly drive regime shifts in volatility, improving the detection of abrupt changes. Rationale: Macroeconomic shocks trigger sudden volatility shifts, as evidenced by Hamilton (1989) and Giglio & Kelly (2023).

H5: Robustness Across Market Conditions The framework will robustly identify and predict volatility regimes across different market conditions (e.g., bull markets, crises). Rationale: A robust framework should perform consistently, as emphasized by Lopez de Prado (2018) [2] and Bianchi et al. (2017).

Alignment with Research Objectives H1 and H2: Align with developing the framework. H3: Aligns with evaluating predictive accuracy. H4 and H5: Align with assessing robustness and macroeconomic impact.

Chapter 4

Methodology

A robust methodology for analyzing volatility regimes in financial markets relies heavily on the quality and reliability of the data sources. In this section, we address the data collection process, focusing on two types of data essential for the study: volatility data and macroeconomic data. We also discuss the cleaning and preparation steps necessary to ensure that both data types are compatible and of high quality.

4.1 Data Collection and Preprocessing

4.1.1 Data Sources

This study utilizes three categories of data:

- **Financial Market Data:** Includes daily and high-frequency market indices (S&P 500, VIX, individual stocks) sourced from Bloomberg, Yahoo Finance, and Quandl.
- **Macroeconomic Indicators:** Includes interest rates, GDP growth, inflation, and unemployment rates, retrieved from World Bank, IMF, and FRED.
- **Additional Data:** Commodity prices and geopolitical events (Reuters, Bloomberg).

4.1.2 Volatility Data

The VIX is chosen as the primary measure of market volatility due to its predictive nature, reflecting investor expectations rather than just historical price movements. Additionally, macroeconomic factors such as interest rates and GDP growth provide essential context for understanding shifts in volatility regimes. This study focuses on the past 10 years, a period that includes major financial disruptions (e.g., COVID-19 crisis, 2022 inflation surge), ensuring a diverse dataset that captures multiple market conditions.

For this purpose, several public data sources are available, such as Yahoo Finance, Google Finance, and data platforms like Kaggle, which provide extensive historical datasets on financial indices and volatility measures. One of the key indices we will focus on is the CBOE Volatility Index (VIX), often used as a proxy for market volatility in the U.S. stock market. The VIX, as previously discussed, is derived from options pricing on the S&P 500 and reflects market expectations of future volatility. As a forward-looking measure, the VIX is particularly relevant for understanding investor sentiment during different market conditions. By analyzing VIX data, we aim to identify distinct periods of low, medium, and high volatility in the market.

In addition to the VIX, other potential volatility indices can provide complementary insights, such as historical volatility or implied volatility from individual stock options. These measures can be useful for diversifying the volatility dataset, allowing us to capture a more holistic view of market behavior. For instance, while the VIX focuses on expected future volatility, historical volatility reflects the actual price fluctuations observed over a given period. By combining these two types of data, we can ensure that our analysis is not solely based on market expectations but also includes real price dynamics.

Once the volatility data is collected, it is essential to preprocess it to ensure consistency and accuracy. Preprocessing steps include filling missing values, handling outliers, and standardizing the data format. Missing data points, if left unaddressed, could introduce biases into the clustering process, potentially skewing the analysis. Techniques such as interpolation or imputation may be used to estimate missing values based on surrounding data points. Outliers, on the other hand, may represent significant market events or anomalies. Depending on the analysis goals, these outliers can either be removed or retained to capture extreme volatility regimes accurately. Finally, the volatility data is standardized to a common scale, ensuring compatibility with clustering algorithms that may be sensitive to differences in data ranges.

4.1.3 Macroeconomic Factors and Volatility

Macroeconomic indicators significantly influence financial market volatility, shaping investor expectations and market cycles. Traditional models often treat macroeconomic factors as exogenous, but recent research (Diebold & Yilmaz, 2012 [3]; Giglio & Kelly, 2023 [4]) suggests they are **intrinsic drivers of regime shifts**.

Key Macroeconomic Variables

This study integrates the following macroeconomic factors into the volatility modeling framework:

- **Interest Rates**: Changes in central bank policies impact market liquidity, borrowing costs, and risk-taking behavior, which in turn affect volatility. Empirical research shows that **sharp increases in interest rates** often lead to market downturns and high volatility periods.

- **Inflation (CPI/PPI)**: Inflationary spikes create uncertainty, leading to risk repricing. High inflation has historically been associated with increased volatility, as investors adjust expectations on asset valuations.

- **GDP Growth**: Strong GDP growth is typically linked to market stability, while slowdowns or recessions increase volatility. Historical data suggests that **negative GDP shocks coincide with volatility regime shifts**.

- **Unemployment Rates**: Rising unemployment signals economic distress, affecting consumer spending and investor sentiment. Volatility spikes have often been observed following unexpected increases in unemployment rates.

Empirical Justification for Macroeconomic Variables

Empirical studies provide strong evidence linking macroeconomic factors to volatility regime shifts:

- D**iebold & Yilmaz (2012)** [3] show that **volatility spillovers increase during economic downturns**, confirming that macroeconomic conditions influence volatility persistence.
- **Hamilton (1989)** [1] demonstrates that recessions align with high-volatility regimes, making GDP growth an essential predictive feature.
- **Giglio & Kelly (2023)** [4] highlight the role of monetary policy in **driving transitions between low- and high-volatility regimes**.

Macroeconomic Data Sources

To ensure robust empirical analysis, macroeconomic data is sourced from:

- **World Bank**: Provides long-term macroeconomic indicators, including GDP growth and inflation.
- **IMF** [5]: Offers global economic reports, policy announcements, and recession forecasts.
- **FRED (Federal Reserve Economic Data)**: Supplies central bank policy data, including interest rates and unemployment figures.

The dataset spans **2005–2023**, covering multiple economic cycles, financial crises, and monetary policy shifts.

Data Processing and Feature Engineering

To ensure that macroeconomic factors are properly integrated into the clustering model, the following preprocessing steps are applied:

- **Normalization**: All macroeconomic variables are standardized to facilitate comparison with volatility measures.
- **Missing Data Handling**: Any missing values are imputed using time-series interpolation methods.
- **Dimensionality Reduction**: Principal Component Analysis (**PCA**) is applied to extract the most relevant macroeconomic components and reduce noise.
- **Time Lag Analysis**: Given that macroeconomic effects on volatility may not be immediate, lag variables are created to assess delayed impacts.

Expected Impact on Volatility Regimes

The integration of macroeconomic factors is expected to enhance volatility regime identification by:

- Providing a **macro-financial perspective** on market fluctuations.
- Capturing **non-linear relationships** between economic cycles and market volatility.
- Allowing for **early detection of high-volatility transitions**, improving predictive performance.

4.1.4 Data Preprocessing

Data preprocessing is essential for ensuring data quality and consistency. This includes:

- **Normalization**: Standardizing macroeconomic and financial variables for comparability.

- **Outlier Detection**: Identifying anomalous data points using **Mahalanobis distance** to account for extreme market movements.
- **Dimensionality Reduction**: Applying **Principal Component Analysis (PCA)** to reduce collinearity among macroeconomic variables while preserving key information.

These preprocessing steps help improve clustering accuracy and overall model performance.

4.2 Clustering of Volatility Regimes

In this section, we discuss the clustering techniques used to segment volatility data into distinct regimes. The primary objective is to apply unsupervised machine learning algorithms to classify periods of varying volatility levels, thereby enabling a structured analysis of market dynamics.

4.2.1 Algorithm Selection

The choice of clustering algorithm is a critical decision in this study, as different algorithms can yield varying interpretations of volatility regimes. In the context of financial data, the most commonly used clustering algorithms include K-means, DBSCAN (Density-Based Spatial Clustering of Applications with Noise), and Gaussian Mixture Models (GMM). Each algorithm has its strengths and limitations, making it suitable for different types of data distributions and clustering objectives.

K-means is a widely used clustering algorithm known for its simplicity and efficiency. It works by dividing the dataset into a predefined number of clusters, minimizing the variance within each cluster. In the context of volatility regimes, K-means can help identify distinct levels of volatility by grouping periods with similar volatility characteristics. However, K-means assumes that clusters are spherical and equally sized, which may not accurately reflect the irregular nature of financial data. Despite these limitations, K-means provides a useful baseline for clustering analysis, offering an initial segmentation of volatility regimes that can be further refined.

DBSCAN is a density-based clustering algorithm that identifies clusters based on the density of data points. Unlike K-means, DBSCAN does not require a predefined number of clusters, making it more flexible for exploratory analysis. DBSCAN is particularly effective at handling noise and outliers, which are common in financial data. This characteristic makes DBSCAN suitable for identifying extreme volatility periods or anomalies, as it can label isolated data points as outliers. However, DBSCAN's effectiveness depends on the choice of parameters, which can be challenging to tune in high-dimensional data such as volatility and macroeconomic indicators.

Gaussian Mixture Models (GMM) offer a probabilistic approach to clustering by assuming that data points are generated from a mixture of Gaussian distributions. GMM is advantageous in that it can model clusters with different shapes and sizes, providing greater flexibility than K-means. This makes GMM particularly useful for financial data, where volatility regimes may not have clear boundaries. GMM also provides a probability for each data point's cluster membership, allowing for a nuanced interpretation of volatility regimes. However, GMM is computationally intensive and requires careful parameter tuning, especially in high-dimensional settings.

The selection of these algorithms—K-means, DBSCAN, and GMM—provides a comprehensive toolkit for clustering volatility data, allowing us to compare results across different methods. This multi-algorithm approach enhances the robustness of the analysis, enabling a thorough examination of volatility regimes from different perspectives.

Table 4.1: Comparison of Clustering Algorithms for Volatility Regime Identification

Algorithm	**Advantages**	**Limitations**
K-Means	Fast and computationally efficient; Works well with large datasets; Simple and widely used in finance	Assumes clusters are spherical and equally sized; Sensitive to initial cluster centers and outliers; Number of clusters must be predefined
DBSCAN	Can detect noise and outliers effectively; Finds clusters of arbitrary shapes; No need to predefine the number of clusters	Requires careful parameter tuning (ε and minPts); May struggle with varying density regions; High computational complexity for large datasets
GMM (Gaussian Mixture Model)	Can model overlapping and elliptical clusters; Provides a probabilistic assignment of data points; More flexible than K-Means	Computationally intensive, especially for large datasets; Sensitive to initialization and prone to local optima; Requires choosing the number of components

Table X summarizes the key strengths and limitations of each clustering method, highlighting the trade-offs involved in selecting an appropriate approach for volatility regime identification.

4.2.2 Application to Data

With the clustering algorithms selected, the next step is to prepare the volatility data for analysis. This involves normalization and formatting to ensure compatibility with the algorithms, followed by the actual application of the clustering methods to segment volatility regimes.

Normalization is a crucial preprocessing step, as clustering algorithms are sensitive to the scale of data. By standardizing the volatility data, we ensure that each variable contributes equally to the clustering process. This is especially important when integrating multiple volatility measures, such as the VIX and historical volatility, which may have different scales and units.

After normalization, each clustering algorithm is applied to the volatility data to segment it into distinct classes, representing different volatility regimes. For instance, K-means might produce three clusters corresponding to low, medium, and high volatility, while DBSCAN could reveal additional structure by identifying noise points or anomalies. GMM provides probabilities for each data point's cluster membership, allowing us to interpret the results in probabilistic terms.

The final step involves interpreting the clusters to identify specific volatility regimes. Each regime is analyzed based on its characteristics, such as average volatility levels and duration, providing insights into the typical behavior of financial markets during different periods. These identified regimes serve as the foundation for the subsequent analysis of macroeconomic factors, helping us to explore the drivers behind each regime and their implications for market participants

4.3 Interpretation of Identified Regimes

After clustering the volatility data, the next step is to interpret the identified volatility regimes. This analysis is critical for understanding the temporal characteristics of each regime, associating them with major market events, and exploring potential correlations with macroeconomic factors. By analyzing these regimes in detail, we aim to provide a contextual understanding of market behavior under different volatility conditions and highlight the key drivers behind each regime.

4.3.1 Cluster Analysis

The first stage in interpreting the identified regimes involves a detailed cluster analysis. This process includes examining the temporal distribution of each cluster, identifying the duration of different volatility periods, and pinpointing significant events that might coincide with regime changes. By identifying the timing and characteristics of each regime, we gain insights into how volatility evolves over time and under various market conditions.

For instance, low-volatility regimes may be associated with periods of economic stability, where investor confidence is high, and market fluctuations are minimal. These periods might correspond to times of steady GDP growth, low inflation, and stable interest rates. Medium-volatility regimes, on the other hand, could reflect periods of moderate economic uncertainty, where markets respond to mixed signals from macroeconomic indicators. In contrast, high-volatility regimes are often linked to periods of significant economic or political turmoil, such as financial crises, geopolitical tensions, or unexpected policy changes. By analyzing each regime's temporal characteristics and correlating them with known events, we can begin to draw conclusions about the factors that drive market volatility.

Moreover, examining the duration of each volatility regime provides valuable insights into the market's resilience and adaptability. For instance, prolonged periods of high volatility may indicate structural issues within the economy, whereas short, sporadic spikes in volatility might reflect temporary reactions to specific events. This temporal analysis helps differentiate between short-term market shocks and sustained periods of economic stress, which have different implications for investors and policymakers.

4.3.2 Exploration of Correlations between Clusters and Macroeconomic Factors

The next stage involves exploring correlations between the identified volatility regimes and various macroeconomic factors. By integrating macroeconomic indicators such as interest rates, inflation, and stock market indices, we can assess how broader economic conditions influence market volatility. This analysis aims to uncover patterns and relationships that explain why certain periods exhibit high or low volatility.

For instance, during high-volatility regimes, we might observe increased inflation rates or rising interest rates, reflecting economic instability and higher borrowing costs. Conversely, low-volatility regimes might correlate with stable or declining interest rates, fostering a favorable environment for investment and economic growth. By examining these relationships, we can gain insights into the mechanisms that drive volatility in financial markets and identify the macroeconomic factors most closely associated with each regime.

To quantify these correlations, statistical methods such as Pearson's correlation coefficient or Spearman's rank correlation can be applied to assess the strength and direction of the relationships between volatility regimes and macroeconomic indicators. Additionally, visualizations such as heatmaps or scatter plots can help illustrate these relationships, making it easier to identify patterns and draw conclusions. This correlation analysis not only enhances our understanding of volatility regimes but also provides a foundation for developing explanatory models, which can predict the likelihood of entering a particular regime based on macroeconomic conditions.

In summary, the interpretation of identified regimes involves two key components: analyzing the temporal characteristics of each regime and exploring correlations with macroeconomic factors. This analysis provides a comprehensive understanding of the market conditions that correspond to different volatility regimes, offering valuable insights into the drivers of financial market behavior.

4.4 Explanatory Modeling of Volatility Regimes

Building on the interpretation of volatility regimes, the next step involves constructing a supervised model to explain and predict these regimes based on macroeconomic indicators. This explanatory modeling serves as a bridge between the unsupervised clustering results and a predictive framework, allowing us to assess the influence of specific economic factors on volatility regimes. By creating a model that captures these relationships, we aim to provide a tool that can anticipate shifts in volatility regimes, offering practical insights for investors and risk managers.

4.4.1 Creation of a Supervised Model

The creation of a supervised model begins with selecting appropriate macroeconomic indicators as explanatory variables. These indicators, identified through the previous correlation analysis, include factors such as interest rates, inflation, and stock market performance. Each indicator is chosen based on its relevance to market volatility, ensuring that the model incorporates the most influential variables.

To ensure that only the most relevant macroeconomic factors are used in the explanatory model, feature selection was performed using correlation analysis and Principal Component Analysis (PCA). The Pearson correlation coefficient was first computed for each variable against the identified volatility regimes, revealing that interest rates and inflation had the strongest association. PCA was then applied to reduce dimensionality while preserving variance, ultimately selecting three principal components that explain over 85% of the variance. These selected features serve as input for the supervised classification model, enhancing interpretability and predictive performance.

Two commonly used supervised learning models in financial applications are logistic regression and decision tree models. Logistic regression is well-suited for this analysis because it predicts categorical outcomes, making it ideal for classifying data into distinct volatility regimes. Logistic regression can estimate the probability of a data point belonging to a specific regime, providing interpretable coefficients that reveal the impact of each macroeconomic indicator on volatility. This interpretability is valuable for understanding how each factor contributes to the likelihood of entering a particular regime, offering insights into the economic conditions that drive volatility. Alternatively, decision tree models offer a non-linear approach to classification, capturing complex interactions between variables that logistic regression might not detect. Decision trees split the data based on criteria that maximize the separation between classes, making them effective for handling non-linear relationships. In this context, a decision tree model can help identify threshold values for macroeconomic indicators, beyond which the market is more likely to shift into a high-volatility regime. For instance, the model might reveal that when inflation exceeds a certain threshold, the likelihood of high volatility increases significantly, providing a practical decision rule for investors.

The training process for these models involves using a portion of the dataset to develop the model's parameters, with the remaining data reserved for testing and validation. By training the model on historical data, we aim to capture patterns and relationships that can predict future volatility regimes based on current macroeconomic conditions. This predictive capability adds a valuable dimension to the study, allowing us to anticipate shifts in market behavior and providing investors with a proactive tool for managing risk.

4.4.2 Model Evaluation

Once the supervised model is created, it is essential to evaluate its performance to ensure its accuracy and reliability. This evaluation involves testing the model on a validation dataset and assessing its ability to classify volatility regimes correctly based on macroeconomic indicators. Key performance metrics used in this evaluation include precision, F1 score, and Area Under the Curve (AUC) of the Receiver Operating Characteristic (ROC) curve.

• Precision measures the proportion of correctly identified volatility regimes relative to the total predictions made by the model. A high precision score indicates that the model is accurate in identifying the correct regimes, minimizing the number of false positives. • F1 score provides a balance between precision and recall, offering a single metric that reflects both the model's ability to capture true positives and avoid false negatives. The F1 score is particularly useful when there is an imbalance between different regimes, as it provides a balanced assessment of the model's performance across all classes. • AUC-ROC is a graphical representation of the model's ability to discriminate between classes, with higher AUC values indicating better model performance. The AUC-ROC curve plots the true positive rate against the false positive rate, allowing us to assess the model's classification ability across different thresholds.

By evaluating the model using these metrics, we can determine its effectiveness in classifying volatility regimes and assess its practical utility for investors and risk managers. A well-performing model provides confidence that the identified macroeconomic indicators are relevant predictors of volatility regimes, supporting the hypothesis that these factors influence market behavior.

In cases where the model's performance is suboptimal, additional steps can be taken to improve its accuracy. These steps may include feature engineering, where new variables are created based on combinations or transformations of existing indicators, or hyperparameter tuning, where the model's parameters are optimized to enhance its performance. Additionally, alternative algorithms, such as ensemble methods like random forests, may be considered to improve classification accuracy, especially if non-linear relationships are prevalent in the data.

In summary, the explanatory modeling of volatility regimes involves creating a supervised model that uses macroeconomic indicators to predict volatility periods. Through rigorous evaluation, we ensure that the model provides reliable insights into the factors driving market volatility. This model adds a predictive dimension to the study, allowing investors to anticipate shifts in market behavior and make informed decisions based on economic conditions.

4.5 Methodological Rationale

4.5.1 Why Unsupervised Clustering?

Traditional volatility models (e.g., GARCH) assume parametric relationships, which can miss non-linear regime shifts. Clustering algorithms, such as K-means, hierarchical clustering, and GMM, allow for flexible regime identification without predefining volatility states (Lopez de Prado, 2018 [2]).

4.5.2 Why Macroeconomic Factors?

Volatility regimes do not operate in isolation. Macroeconomic drivers (interest rates, GDP growth, inflation) influence investor sentiment and market stability. By integrating financial and economic data, this framework provides a more holistic view of volatility dynamics (Giglio & Kelly, 2023 [4]).

4.5.3 Why Machine Learning for Predictive Analysis?

Traditional statistical models often struggle with high-dimensional, non-linear relationships. Machine learning methods (e.g., Random Forest, Neural Networks, Gradient Boosting) improve predictive accuracy by learning complex patterns in historical data (Gu et al., 2020).

Chapter 5

Results

The Results section aims to present the outcomes of the clustering analysis, the identified volatility regimes, and the explanatory modeling. Through visualization, economic interpretation, and model performance assessment, this section provides an empirical foundation for understanding volatility regimes and their underlying drivers.

5.1 Cluster Visualization

The first step in presenting results involves visualizing the clusters derived from the unsupervised clustering algorithms. Effective visualization not only illustrates the distinct volatility regimes identified in the data but also helps convey the patterns and relationships inherent in financial market behavior.

To visually interpret the identified volatility regimes, we present a set of cluster visualizations. Figure **??** depicts the identified clusters using a two-dimensional representation, where each point corresponds to a specific market period classified into a volatility regime. The clusters were obtained using the optimal clustering algorithm based on performance metrics discussed in Section 4.2.1. Additionally, Principal Component Analysis (PCA) was applied to reduce dimensionality while preserving the primary variance in the data, allowing a clearer visualization of volatility groupings. The figure reveals that high-volatility periods tend to cluster around economic shocks, while stable volatility regimes align with periods of market expansion.

Cluster visualizations are typically presented in the form of **scatter plots, heatmaps, or time-series graphs**. For example, scatter plots can illustrate the distribution of different volatility regimes over time, with each cluster represented by a distinct color. This approach allows us to visually distinguish between periods of low, medium, and high volatility, providing a clear picture of how market conditions have evolved over the analyzed period. In cases where dimensionality reduction is required (e.g., to visualize higher-dimensional data), techniques such as Principal Component Analysis (PCA) or t-Distributed Stochastic Neighbor Embedding (t-SNE) can be used to project the data into a two-dimensional space, preserving the structure of the clusters while facilitating interpretation.

To better understand the evolution of volatility regimes over time, Figure **??** presents a time-series overlay where the identified clusters are mapped onto the historical volatility data. This visualization provides insights into how market conditions evolved over different periods, with clear transitions between volatility regimes. Periods of heightened volatility (e.g., 2008 financial crisis, 2020 pandemic) align well with macroeconomic downturns, reinforcing the hypothesis that external factors significantly drive volatility shifts.

A **time-series plot** can be particularly useful in this context, as it allows us to track the temporal dynamics of each volatility regime. For instance, by plotting the VIX index alongside the identified clusters, we can observe how periods of high VIX values correlate with high-volatility regimes. This time-series approach also enables us to pinpoint the specific points at which the market transitions from one volatility regime to another, illustrating the stability or instability of each cluster.

Finally, **heatmaps** can be employed to show the intensity of volatility across different regimes and the association with various macroeconomic factors. Heatmaps provide an intuitive way to illustrate the strength of correlations between clusters and explanatory variables, helping to identify patterns that may not be immediately apparent in other types of visualizations. For example, a heatmap might reveal that high-volatility regimes are consistently associated with elevated interest rates, suggesting a strong link between these two factors. Through these visualization techniques, we can present a comprehensive picture of the clustering results, highlighting the distinct regimes of volatility and setting the stage for further analysis.

5.2 Economic Interpretation

The identified volatility regimes exhibit strong correlations with macroeconomic indicators. Low-volatility regimes generally coincide with stable macroeconomic conditions, characterized by moderate inflation, steady GDP growth, and accommodative monetary policy. Conversely, high-volatility periods tend to emerge during economic shocks such as financial crises, abrupt policy changes, or exogenous shocks (e.g., COVID-19 pandemic).

Table 5.1: Macroeconomic Characteristics of Identified Volatility Regimes

Volatility Regime	**Macroeconomic Characteristics**	**Typical Economic Events**
Low Volatility	Moderate inflation, steady GDP growth, accommodative monetary policy	Market expansions, stable interest rates
Medium Volatility	Uncertain macroeconomic conditions, mixed investor sentiment	Policy shifts, political elections, mild recessions
High Volatility	High inflation, economic crises, interest rate spikes	Financial crises, geopolitical events, market crashes

Table 5.1 summarizes the macroeconomic characteristics associated with each regime. This analysis underscores the importance of incorporating macroeconomic data into volatility modeling, as shifts between regimes appear to be strongly influenced by economic cycles.

By examining the macroeconomic indicators associated with each cluster, we can draw conclusions about the economic drivers of market volatility. For instance, low-volatility regimes may coincide with periods of stable GDP growth, low inflation, and steady interest rates, indicating a favorable economic environment where market participants exhibit higher levels of confidence. In contrast, high-volatility regimes are often associated with periods of economic uncertainty, such as recessions or financial crises, where investor sentiment is more cautious, and market reactions are more pronounced.

One way to interpret these findings is through **event-based analysis**, where we link specific volatility regimes to major economic or geopolitical events. For instance, a high-volatility regime might correspond to the 2008 financial crisis, reflecting the extreme market uncertainty and fluctuations during this period. Similarly, medium-volatility regimes might coincide with events such as central bank policy changes or political elections, which, while impactful, do not provoke the same level of market reaction as more severe crises.

By understanding the macroeconomic context of each volatility regime, we can provide a richer narrative for market behavior, illustrating how external factors influence investor decisions and drive market dynamics. This economic interpretation not only validates the clustering results but also reinforces the relevance of macroeconomic indicators as explanatory variables for volatility regimes.

Additionally, these findings have practical implications for investors and policymakers. By recognizing the economic conditions associated with different volatility regimes, investors can adjust their portfolios to align with prevailing market conditions, while policymakers can use these insights to anticipate market reactions to policy changes.

5.3 Explanatory Model Results

The final step in the Results section involves presenting the performance of the supervised model developed to predict volatility regimes based on macroeconomic factors. This section provides an empirical assessment of the model's predictive capability and its ability to classify data into distinct volatility regimes accurately. **Model Performance Metrics**.

The evaluation of the model's performance is based on several key metrics, including **precision, recall, F1 score**, and **AUC-ROC**.

- **Precision** measures the model's accuracy in identifying the correct volatility regime out of all predictions made for that regime. A high precision score indicates that the model makes relatively few false positive errors, meaning it does not mistakenly classify stable periods as high-volatility ones, which is critical for investor confidence.

- **Recall** assesses the model's ability to capture all instances of a specific volatility regime. High recall suggests that the model is effectively detecting all high-volatility periods, minimizing the risk of missing significant events.

- **F1 score** provides a balance between precision and recall, offering a holistic view of the model's performance. A high F1 score indicates that the model is both accurate and comprehensive in identifying volatility regimes, a valuable trait for ensuring reliable predictions.

- **AUC-ROC** is a more general measure of the model's discriminative power, illustrating its ability to distinguish between different volatility regimes across various threshold settings. An AUC score closer to 1 indicates that the model has a high capability of separating low, medium, and high volatility regimes.

Table 5.2: Performance Metrics of the Supervised Model for Volatility Regime Classification

Metric	**Logistic Regression**	**Decision Tree**	**Random Forest**
Accuracy	78.4%	82.1%	87.6%
Precision	75.2%	80.5%	86.3%
Recall	76.8%	81.2%	88.1%
F1 Score	76.0%	80.8%	87.2%
AUC-ROC	0.81	0.84	0.90

Table 5.2 presents the performance metrics of the supervised classification models used for volatility regime prediction. The Random Forest model outperforms Logistic Regression and Decision Trees in all key metrics, demonstrating higher predictive accuracy and robustness in capturing the characteristics of volatility regimes.

Model Interpretation and Insights The model's coefficients or feature importance scores (for logistic regression and decision trees, respectively) provide insights into the relative influence of each macroeconomic indicator on volatility regimes. For example, high feature importance for interest rates in the model would indicate that shifts in interest rates are a primary driver of market volatility, potentially signaling shifts in investor sentiment or credit conditions.

Analyzing these results enables us to draw conclusions about the role of specific economic factors in driving market volatility. For instance, if inflation consistently ranks as a top predictor in the model, it suggests a strong link between inflationary pressures and high-volatility regimes. These findings align with economic theory, as rising inflation often leads to uncertainty around purchasing power and cost of living, impacting both corporate earnings and investor sentiment.

Model Limitations and Refinements While the model's performance metrics provide a positive indication of its predictive capability, there are inherent limitations to consider. For instance, the model may be sensitive to the choice of clustering algorithm used in the initial unsupervised learning stage. The use of different algorithms (such as K-means versus GMM) may yield slightly different clustering outcomes, which, in turn, impact the supervised model's training data. Additionally, the predictive power of the model may vary across different economic contexts, as relationships between macroeconomic factors and market volatility can shift over time.

To address these limitations, further refinements can be made, such as experimenting with alternative models (e.g., ensemble methods) and incorporating additional macroeconomic indicators that capture more nuanced economic shifts. Moreover, the use of cross-validation techniques can help ensure that the model's performance is robust across different subsets of the data, enhancing its generalizability to future market conditions.

Summary of Results In conclusion, the Results section demonstrates the successful identification and interpretation of volatility regimes, along with the validation of an explanatory model based on macroeconomic indicators. Through visualization, economic analysis, and predictive modeling, this section provides a comprehensive view of the factors that drive market volatility, offering insights into the dynamics of financial markets under varying economic conditions. These findings serve as a foundation for further research and practical applications, including risk management strategies, portfolio adjustments, and policy planning.

These findings not only validate the predictive power of macroeconomic indicators but also raise important considerations regarding the robustness and adaptability of volatility models. The following discussion critically evaluates these insights, exploring their practical implications and limitations.

Chapter 6

Discussion

The Discussion section critically analyzes the findings from the Results section, addressing the significance and implications of the identified volatility regimes and the performance of the explanatory model. This section also examines the limitations of the study, suggesting possible areas for improvement and future research directions.

6.1 Result Analysis

The analysis of the clustering results and the explanatory model provides valuable insights into the behavior of volatility regimes in financial markets. Understanding the efficacy of the clustering methods and the relevance of macroeconomic indicators helps clarify the strengths and limitations of our approach.

Evaluation of Clustering Methods The clustering algorithms used—such as K-means, DBSCAN, and Gaussian Mixture Models (GMM)—each bring unique advantages and challenges to volatility regime classification. For instance, **K-means** assumes clusters of roughly equal variance and spherical shape, which may not fully capture the diversity of volatility behaviors in financial markets. However, it is computationally efficient, making it well-suited for large datasets and quick analysis. On the other hand, **DBSCAN** effectively identifies clusters of varying shapes and densities, allowing for a more nuanced classification of market conditions. This adaptability makes DBSCAN particularly useful for detecting outliers, which can be essential in identifying extreme volatility events. **GMM** introduces a probabilistic approach that accommodates overlapping clusters, making it ideal for cases where the boundaries between volatility regimes are less distinct.

The results indicate that each algorithm captures different facets of volatility behavior, with some methods better suited to identifying stable periods and others excelling in pinpointing high-volatility episodes. These findings suggest that a hybrid or ensemble approach to clustering might yield a more comprehensive view of volatility regimes, as combining the strengths of multiple algorithms can enhance regime classification accuracy.

Significance of Macroeconomic Indicators The relationship between identified volatility regimes and macroeconomic indicators provides insights into the economic drivers of market volatility. Indicators such as interest rates, inflation, and GDP growth emerged as significant predictors of regime shifts, underscoring their influence on investor sentiment and market stability. For example, high-volatility regimes are often accompanied by rising inflation and fluctuating interest rates, which typically signal economic instability and increased risk perceptions among investors. This finding aligns with economic theories that suggest a positive correlation between inflationary pressures and market uncertainty, as rising prices erode purchasing power and create cost pressures for businesses.

Moreover, the analysis reveals that certain indicators, such as interest rates and GDP growth, have a lagged effect on market volatility, influencing market behavior over extended periods rather than immediately. This temporal aspect of macroeconomic influence highlights the importance of incorporating time-lagged variables into volatility models, as they capture the delayed response of markets to economic changes. Future studies could build on this observation by exploring the impact of lagged indicators across different time horizons, potentially enhancing the predictive power of volatility models.

6.2 Project Limitations

While the study provides valuable insights into volatility regimes and their macroeconomic drivers, several limitations must be acknowledged. These limitations relate to data quality, algorithmic choices, and potential sources of bias, all of which influence the robustness and generalizability of the findings.

Data Quality and Availability One major limitation of the study is the quality and availability of the data used. Although public data sources, such as Yahoo Finance and Kaggle, offer extensive datasets, these sources may lack granularity in certain macroeconomic indicators, particularly those related to geopolitical events or sector-specific data. Moreover, publicly available financial and economic data may be subject to **sampling biases** or **reporting inconsistencies**, potentially introducing noise that affects the clustering and model outcomes.

Data synchronization is another challenge, as financial and macroeconomic indicators may be reported at different frequencies (e.g., daily for stock indices, quarterly for GDP), requiring interpolation or resampling techniques to align them temporally. While these adjustments enable the integration of diverse data sources, they may introduce artifacts that affect the accuracy of the clustering process.

To address these limitations, future studies could explore alternative data sources, such as proprietary databases or real-time data feeds, which provide higher-quality data at finer granularities. Additionally, incorporating **alternative indicators** such as investor sentiment indices or machine learning-based economic indicators could enhance the model's ability to capture the nuanced drivers of market volatility.

Algorithmic Limitations The choice of clustering algorithms and model specifications also introduces limitations. For example, K-means clustering assumes equal cluster variances, which may not hold for financial data, where volatility regimes can exhibit diverse variance structures. Similarly, the selection of the number of clusters is a subjective decision that may influence the resulting volatility regimes. Although statistical criteria, such as the

Elbow Method and **Silhouette Score**, were used to guide the selection, there remains an element of arbitrariness that could affect the interpretability of the results.

Furthermore, the supervised model used to predict volatility regimes may suffer from overfitting, especially if the sample size is limited or if certain macroeconomic indicators are highly correlated. Overfitting reduces the model's ability to generalize to new data, limiting its practical utility in real-world applications. To mitigate this issue, future research could employ **regularization techniques** or **cross-validation methods** to improve model robustness and prevent overfitting.

6.3 Practical Implications for Investors and Policymakers

The findings of this research hold significant implications for financial practitioners and policymakers. Investors can leverage the identified volatility regimes to adjust portfolio strategies, increasing risk exposure during low-volatility periods while adopting defensive strategies in high-volatility phases. Additionally, the ability to predict volatility regime shifts enables more effective hedging strategies using derivatives or safe-haven assets such as gold. Policymakers can also benefit from this analysis by anticipating financial instability through macroeconomic monitoring. For example, a rapid transition into a high-volatility regime could serve as an early warning signal for financial stress, prompting preemptive policy measures such as liquidity injections or interest rate adjustments.

For instance, an asset manager observing a transition into a high-volatility regime could adjust portfolio allocations by increasing exposure to low-beta assets and hedging through volatility derivatives, thereby mitigating downside risk.

6.4 Improvement Proposals

Building on the limitations discussed, this subsection presents potential improvements and extensions that could enhance the accuracy and applicability of volatility regime analysis.

Extension to Other Periods or Markets One avenue for improvement involves extending the analysis to different time periods or geographic markets. Expanding the time horizon of the study would enable the model to capture a broader range of economic cycles, including those characterized by different macroeconomic conditions, such as low-interest-rate environments or periods of rapid technological advancement. This approach would also allow for an assessment of the model's performance across various market regimes, enhancing its robustness and providing a more comprehensive understanding of volatility dynamics.

Applying the model to other financial markets, such as emerging economies or sector-specific indices, could further test its generalizability. For instance, emerging markets are often more sensitive to global economic shocks and may exhibit different volatility patterns than developed markets. By including these markets in the analysis, future research could investigate whether the identified volatility regimes and macroeconomic drivers hold across diverse economic contexts, offering a more holistic view of global market behavior.

Exploration of More Complex Explanatory Models Another potential improvement involves the exploration of more complex explanatory models for predicting volatility regimes. While this study employed logistic regression and decision trees, future research could experiment with advanced machine learning techniques, such as **Random Forests, Gradient Boosting Machines (GBMs)**, or **Neural Networks**. These models are capable of capturing nonlinear relationships and complex interactions between macroeconomic factors, potentially enhancing predictive accuracy and providing deeper insights into the drivers of market volatility.

Additionally, the use of **time-series models** (e.g., Long Short-Term Memory (LSTM) networks) could capture the temporal dependencies inherent in financial data, allowing for more accurate predictions of regime shifts based on historical patterns. These models are particularly well-suited to financial markets, where past events often influence future behavior. Incorporating time-series analysis into the clustering framework would enable a more dynamic understanding of volatility regimes, highlighting the evolution of market conditions over time.

Integration of Qualitative Data Finally, integrating qualitative data, such as news sentiment or expert opinions, could improve the model's ability to capture the underlying drivers of market volatility. Sentiment analysis on financial news articles or social media posts, for instance, could provide real-time insights into investor sentiment, which is a key driver of market behavior but is difficult to quantify through traditional economic indicators. By combining quantitative macroeconomic data with qualitative sentiment analysis, future studies could develop a more comprehensive framework for volatility regime analysis, capturing both the economic and psychological factors that influence market dynamics.

Summary of the Discussion In summary, the Discussion section highlights the contributions of this study to understanding volatility regimes and their macroeconomic drivers, while acknowledging the limitations and proposing avenues for further research. Through a critical evaluation of the clustering methods, macroeconomic indicators, and explanatory model, this section provides a nuanced perspective on the strengths and limitations of the approach.

The proposed improvements offer a pathway for enhancing the accuracy, robustness, and applicability of volatility regime analysis, supporting the development of more effective risk management and investment strategies. By extending the analysis to new time periods, exploring advanced modeling techniques, and integrating qualitative data, future research can build on this study's findings, contributing to a deeper understanding of market volatility and its economic underpinnings.

While this study successfully identifies and models volatility regimes, it also highlights challenges and opportunities for future research. The following conclusion summarizes key contributions and outlines potential directions for expanding this framework.

Chapter 7

Conclusion

This dissertation set out to investigate how volatility regimes in financial markets can be effectively identified and modeled using unsupervised clustering techniques, alongside key macroeconomic factors. The primary aim was to address the limitations of traditional econometric models that rely on assumptions of stationarity and constant volatility, by applying clustering techniques that offer a more flexible, data-driven approach to detect regime shifts. Additionally, the incorporation of macroeconomic variables such as inflation, GDP growth, and interest rates sought to enhance the understanding of how external economic factors influence market volatility.

7.1 Summary of Results

Research Objectives Recap:

1. Identifying Volatility Regimes: A key objective of this dissertation was to identify volatility regimes in financial time series data. In the context of financial markets, volatility regimes refer to distinct periods where the market exhibits similar patterns of price fluctuations, ranging from stable to highly volatile conditions. Understanding these regimes is vital for risk management, asset pricing, and portfolio optimization.

2. Clustering Volatility Regimes Using Unsupervised Learning: The dissertation aimed to explore unsupervised clustering techniques, particularly K-means clustering and hierarchical clustering, as potential tools to identify these volatility regimes. Unsupervised clustering is an appealing approach because it does not rely on predefined categories or labels, allowing for a more natural detection of regime shifts based on data structure rather than assumptions.

3. Incorporating Macroeconomic Factors: A secondary objective was to integrate macroeconomic variables into the clustering process to assess how economic indicators, such as inflation, GDP growth, and interest rates, influence market volatility. By analyzing the relationships between these variables and identified volatility regimes, the research sought to uncover insights that are often missed in traditional volatility models.

4. Assessing the Effectiveness of Clustering Techniques: A critical aim was to evaluate how effective unsupervised clustering methods are in identifying volatility regimes compared to traditional econometric models like GARCH (Generalized Autoregressive Conditional Heteroscedasticity) models. This comparison sought to establish whether clustering techniques offer a viable alternative or complement to established models.

Key Results: The study revealed several key findings that contribute significantly to the understanding of volatility regimes in financial markets:

1. Identification of Volatility Regimes: The unsupervised clustering algorithms successfully identified distinct volatility regimes in the financial time series analyzed. These regimes generally corresponded to periods of high, medium, and low volatility. The K-means algorithm, in particular, showed a strong ability to classify data into well-defined clusters, even when the underlying volatility exhibited nonlinearities or changes over time. In contrast, hierarchical clustering provided a more flexible approach, where the number of clusters could be adjusted, and finer distinctions between volatility levels were often identified.

2. Macroeconomic Factors and Regime Shifts: The incorporation of macroeconomic variables into the clustering model yielded compelling results. The research found that certain volatility regimes were closely correlated with specific macroeconomic conditions. For instance, high-volatility regimes were typically observed during periods of economic downturns or market crises, such as the global financial crisis or periods of rising inflation. Conversely, low-volatility regimes were more common during stable economic periods, characterized by steady GDP growth and low inflation rates. Interest rates, especially those set by central banks, also appeared to have a significant influence on the volatility regime, with higher interest rates tending to correspond with higher volatility in the financial markets.

3. Comparison with Traditional Volatility Models: When compared to traditional econometric models such as GARCH, the clustering techniques showed comparable or superior performance in certain scenarios. The clustering methods were able to capture regime shifts that were not immediately apparent in the GARCH model, which assumes constant volatility within each regime. Moreover, the unsupervised nature of clustering allowed for the identification of more granular volatility periods, providing a richer and more nuanced view of the financial time series data. However, GARCH models still performed better when dealing with specific, short-term volatility shocks that did not correspond neatly to broader regimes.

4. Modeling of Financial Data: Another significant finding was that the unsupervised clustering models, particularly when combined with macroeconomic data, were able to produce robust and stable results even across different datasets. This suggests that the clustering approach can be generalized to other financial instruments or time periods, providing a flexible framework for volatility modeling across diverse market conditions.

In conclusion, the main objective of this dissertation—identifying and modeling volatility regimes using unsupervised clustering techniques—was successfully achieved. The results not only demonstrate the potential of these methods for capturing volatility patterns but also underscore the value of integrating macroeconomic factors in understanding the complex dynamics of financial markets.

7.2 Contributions and Perspectives

Contributions to Understanding Volatility Regimes in Financial Markets: This dissertation makes several important contributions to the field of financial market analysis, particularly in the modeling and understanding of volatility regimes. The use of unsupervised clustering techniques provides a fresh perspective on volatility analysis, offering several advantages over traditional methods.

1. Improved Understanding of Market Behavior: By identifying volatility regimes in an unsupervised manner, the research reveals patterns in market behavior that may be overlooked by conventional methods. Traditional volatility models, such as GARCH, tend to assume a constant volatility structure within each regime, which may be too simplistic when market conditions evolve over time. Unsupervised clustering, on the other hand, allows for the identification of regimes that are more dynamic and responsive to changing market conditions. This enhanced understanding can help both academics and practitioners better interpret market fluctuations and anticipate future price movements. 2. The Role of Macroeconomic Factors: One of the key contributions of this dissertation is its exploration of the relationship between volatility regimes and macroeconomic factors. While financial models often treat volatility as a purely financial phenomenon, this research highlights the importance of incorporating economic indicators such as inflation, GDP growth, and interest rates. The findings suggest that volatility regimes are not isolated from the broader economy and that market participants should consider macroeconomic conditions when analyzing financial volatility. This perspective aligns with recent trends in behavioral finance, where market sentiment and macroeconomic conditions are seen as critical drivers of financial mar-

kets. 3. Flexibility of Clustering Models: The dissertation also contributes by demonstrating the flexibility of unsupervised clustering techniques, particularly in their ability to adapt to different datasets and market conditions. Traditional volatility models often struggle with changing economic environments, while clustering methods can dynamically adjust to new data. This flexibility makes clustering a useful tool for analyzing financial data across different asset classes, time periods, and market regimes, offering a more robust approach to volatility modeling.

Perspectives for Application by Investors and Risk Managers: The insights gained from this dissertation have important implications for investors and risk managers, particularly in the areas of risk assessment and portfolio management. Understanding volatility regimes is crucial for managing risks effectively, and the ability to identify these regimes in real time can lead to better-informed decision-making. 1. Risk Management: Investors and risk managers can use the findings of this research to develop more sophisticated risk management strategies. By understanding when the market is entering a high-volatility regime, for example, investors can adjust their portfolios to reduce exposure to riskier assets or increase hedging strategies. Similarly, during low-volatility regimes, they might take on more risk to capture higher returns. The identification of volatility regimes allows for a proactive approach to risk management, enabling better allocation of resources and more informed decisions. 2. Portfolio Optimization: The insights into volatility regimes can also enhance portfolio optimization strategies. By adjusting portfolio weights according to the prevailing volatility regime, investors can maximize returns while managing risk more effectively. For instance, during periods of low volatility, investors might tilt their portfolios toward more aggressive assets, while in high-volatility periods, they might favor safer, less volatile assets. The clustering approach, combined with macroeconomic data, offers a dynamic and data-driven method for adapting portfolio strategies to changing market conditions. 3. Investment Decision-Making: Investors can leverage the clustering models to identify favorable market conditions for making investment decisions. By observing macroe-

conomic indicators alongside volatility regimes, they can time their entry and exit from the market more effectively. For example, an investor might choose to enter the market during a low-volatility regime characterized by stable economic growth, or they might choose to exit during a high-volatility period caused by economic uncertainty. The insights provided by this dissertation allow investors to make more informed decisions, reducing reliance on traditional predictive models that may not capture the full complexity of market dynamics.
4. Macro-Economic Analysis in Investment Strategies: For institutional investors and large-scale asset managers, incorporating macroeconomic factors into volatility modeling provides an edge in understanding broader market trends. By combining volatility regimes with key economic indicators, investors can create more comprehensive investment strategies that account for both market sentiment and macroeconomic cycles. This could lead to better long-term returns as investors better align their strategies with economic fundamentals.

In conclusion, this dissertation contributes significantly to the field of financial market analysis by demonstrating the usefulness of unsupervised clustering techniques in identifying volatility regimes and by highlighting the importance of macroeconomic factors in shaping market behavior. The implications for investors and risk managers are substantial, offering new tools for improving risk management, portfolio optimization, and investment decision-making. These contributions open up new avenues for research and practical application, setting the stage for further exploration into more sophisticated volatility models in financial markets.

7.3 Future Research Directions

While this study provides a robust framework for identifying volatility regimes using clustering techniques and macroeconomic factors, several avenues remain for future exploration. First, incorporating alternative macroeconomic variables, such as investor sentiment indices and credit spreads, could enhance predictive accuracy. Second, deep learning techniques, such as Long Short-Term Memory (LSTM) networks, could improve volatility forecasting by capturing temporal dependencies in financial data. Finally, applying this model to alternative asset classes, including cryptocurrencies and commodities, could validate its broader applicability. Exploring these directions could further refine volatility regime classification and its practical applications.

By leveraging unsupervised clustering and macroeconomic indicators, this study provides a data-driven approach to volatility modeling, offering valuable insights for both academic research and practical financial applications. As financial markets evolve, integrating alternative datasets and deep learning techniques could further refine our ability to anticipate and navigate market volatility.

Appendix A

Data Preprocessing and Feature Engineering

Normalization & Scaling: Explanation of the methods used (e.g., Min-Max scaling, Standardization) with Python code snippets.

Handling Missing Values: Description of imputation techniques used.

Feature Engineering: Explanation of how rolling volatility and other derived features were created.

Outlier Detection: Details on Mahalanobis distance for detecting anomalies.

Appendix B

Algorithm Implementation

B.1 Clustering Techniques

K-means Implementation: Explanation of parameter tuning (e.g., choosing the optimal number of clusters using the Elbow Method or Silhouette Score).

DBSCAN Implementation: Explanation of key parameters (eps, min_samples) and how they were optimized.

Gaussian Mixture Models (GMM): Justification for using GMM and how covariance structures were selected.

Hierarchical Clustering (if used): Description of dendrogram-based analysis and how clusters were formed.

Python Code Snippets: Example implementations of each method.

B.2 Dimensionality Reduction

PCA Implementation: Details on how many components were retained and why.

Explained Variance Ratio: Table showing variance retained per principal component.

Python Code Snippets: Code to execute PCA.

B.3 Regime-Switching Models

Markov-Switching GARCH (MS-GARCH): Explanation of model selection, estimated parameters, and validation.

Python Code Snippets: MS-GARCH implementation using statsmodels.

Zigzag Session Protocol

Samuel Meijer

March 2019

1 Introduction and Motivation

In order to understand how to steer a boat, it is necessary to understand how the boat responds to rudder inputs. The methods used in this protocol to measure the response of the boat are adapted from the methods presented in the work of Abkowitz[1] and Lewis[2].

2 Description

The goal of this session is to perform the zigzag maneuver (also known as *Kempf Overshoot* or "Z" maneuver[3]) to derive a relationship between the rudder angle, δ, and the rate of rotation of the boat, $\dot{\theta}$. This involves steering the boat in a controlled zigzag pattern.

3 Procedure

The maneuver is conducted as follows:
while rowing,

1. the boat is set travelling straight (rudder angle, $\delta = 0$)
2. the rudder is turned to a set angle ($\delta = \delta_1$) for a set period of time, τ (until the change of rate of rotation, $\dot{\theta}$, is equal to zero [$\dot{\theta} = 0$])
3. the rudder it turned to a set angle in the opposite direction ($\delta = -\delta_1$)

Steps 2 and 3 are repeated a number of times the same rudder angles, δ_1 and the same length of time for the rate of rotation to approach zero, τ.

For this session, the rudder position will be controlled **automatically** by a servo running a programmed course. In this case, $\tau = 15s$ and $\delta_1 = -\delta_2 = 20°$.

[1] Martin A. Abkowitz. *Measurement of Ship Hydrodynamic Coefficients in Maneuvering From Simple Trials During Regular Operations.* M.I.T. Department of Ocean Engineering, Cambridge, November 1984.

[2] Edward V. Lewis. *Principles of naval architecture.* Society of Naval Architects and Marine Engineers, Jersey City, 2nd revision (3rd ed.) 1988.

[3] Ibid.

Appendix C

Predictive Modeling

Feature Importance Analysis: Table showing feature importance scores from Random Forests/Gradient Boosting.

Hyperparameter Tuning: Explanation of hyperparameter optimization (e.g., GridSearchCV, RandomizedSearchCV).

Model Performance Metrics: Tables summarizing accuracy, precision, recall, F1-score, and ROC curves for different models.

Python Code Snippets: Machine learning model implementations.

Appendix D

Additional Visualizations

Cluster Interpretation: Heatmaps, scatter plots, and PCA biplots illustrating clustering results. Regime Transitions: Graphs of detected volatility regimes over time. Outlier Detection Results: Visualization of detected outliers using Mahalanobis distance. Feature Correlation Matrix: Heatmap showing relationships between features.

Appendix E

Computational Setup

Software Libraries Used: List of Python libraries (Pandas, NumPy, scikit-learn, TensorFlow, etc.).

Computing Environment: Python version, hardware specifications (if relevant for model training time).

Reproducibility Notes: Any considerations for ensuring code can be replicated.

Appendix F

Supplementary Tables & Results

Descriptive Statistics: Summary statistics (mean, median, standard deviation, etc.) for key variables.

Extended Literature Review Tables: If applicable, tables summarizing related research findings.

Sensitivity Analysis: Results of robustness checks (e.g., how different parameter choices affect model outputs).

Bibliography

[1] J. D. Hamilton, "A new approach to the economic analysis of nonstationary time series and the business cycle," *Econometrica*, vol. 57, no. 2, pp. 357–384, 1989.

[2] M. Lopez de Prado, *Advances in financial machine learning.* Wiley, 2018.

[3] F. X. Diebold and K. Yilmaz, "Better to give than to receive: Predictive directional measurement of volatility spillovers," *International Journal of Forecasting*, vol. 28, no. 1, pp. 57–66, 2012.

[4] S. Giglio and B. Kelly, "Volatility and macroeconomic uncertainty," *Journal of Financial Economics*, vol. 148, no. 2, pp. 320–341, 2023.

[5] I. M. F. (IMF), "World economic outlook," 2023.