UNIL | Université de Lausanne
HEC Lausanne

Data Science & Advanced Programming

Fall 2025

Data-Driven Risk Analysis in Global Commodity Market

Gessica Gagliardini

Table of Contents

1. ABSTRACT 2

2. INTRODUCTION 2

3. RESEARCH QUESTION & LITERATURE 3

4. METHODOLOGY 4

5. IMPLEMENTATION 6

6. CODEBASE & REPRODUCIBILITY 7

7. RESULTS AND DISCUSSION 8

8. CONCLUSION 11

REFERENCES 12

1. Abstract

This report develops a quantitative and data-driven framework for modelling, forecasting, and visualising volatility regimes in global commodity markets. Given the inherently stochastic nature of commodity prices, accurate volatility estimation is a key component of effective risk management across major sectors, including metals (gold, copper, aluminium), agricultural products (cocoa, coffee, cotton), and energy markets (crude oil and natural gas).

The empirical strategy relies on a hybrid modelling approach that contrasts classical econometric techniques with modern machine-learning methods. The analysis is based on a ten-year dataset covering the period from January 2015 to December 2024, allowing for both robust model estimation and meaningful out-of-sample evaluation. A standard GARCH(1,1) specification is used as a benchmark and is compared with three machine-learning models: Ridge regression, XGBoost, and a multilayer perceptron neural network. To enhance predictive performance, the models are supplied with a rich set of engineered features, including rolling volatility measures, exponentially weighted moving averages (EWMA), and lagged returns.

The results indicate that machine-learning approaches, and Ridge regression in particular, often outperform the GARCH benchmark in terms of root mean squared error when evaluated on unseen data. Finally, the study is complemented by a two-level interactive visualisation system, consisting of a global dashboard for analysing cross-market correlations and a dedicated "deep-dive" interface that enables detailed, asset-specific comparisons of volatility forecasts across models.

2. Introduction

Commodity markets play a central role in the global economy, as they directly affect inflation, production costs, and food security. In contrast to equity markets, commodity prices are exposed to particularly strong and sudden supply-and-demand shocks arising from factors such as geopolitical tensions, extreme climatic events, and disruptions in transportation and logistics. These structural characteristics generate heavy-tailed return distributions and pronounced volatility clustering, which complicate both portfolio allocation and hedging decisions.

The motivation for this study stems from the growing inadequacy of traditional risk-measurement tools in such an environment. Although econometric models such as GARCH (Generalized Autoregressive Conditional Heteroskedasticity) have long served as a reference framework for modelling volatility persistence, they rely on restrictive assumptions, including distributional stability and stationarity, that are often violated during periods of abrupt regime shifts. Given the increasing availability of high-frequency data and the complexity of market interdependencies, it is therefore essential to assess whether modern data-driven methods, particularly those based on machine learning, can provide more accurate and adaptable volatility forecasts than classical approaches.

To address these issues, this project proposes an integrated computational framework that combines insights from financial econometrics with contemporary data-science techniques. The specific objectives are threefold.

First, the study conducts a systematic comparison between the standard GARCH(1,1) specification and several machine-learning models, namely Ridge regression, XGBoost, and neural networks, in order to determine which approach delivers the most accurate volatility predictions.

Second, the project evaluates the contribution of advanced feature engineering by incorporating technical indicators such as exponentially weighted moving averages, rolling volatility measures, and lagged returns, with the aim of improving the models' ability to capture market shocks and regime changes.

Finally, the analysis is complemented by an interactive Python-based dashboard that allows users to explore cross-asset correlations, assess potential contagion effects, and compare model performance in terms of forecasting errors across different commodity classes.

3. Research Question & Literature

3.1 Theoretical Context

Traditional financial risk management has long been grounded in the assumptions of constant variance and normally distributed returns. Yet a substantial body of empirical research has shown that financial time series rarely conform to these hypotheses. Instead, returns display heteroskedasticity, meaning that their variance changes over time. Early contributions by Mandelbrot (1963) and Fama (1965) documented that periods of large price movements tend to cluster together, as do periods of relative calm, a phenomenon commonly referred to as volatility clustering.

To model this behaviour, Engle (1982) proposed the Autoregressive Conditional Heteroskedasticity (ARCH) framework, which was subsequently generalised by Bollerslev (1986) into the GARCH class of models. Among these, the GARCH(1,1) specification became the benchmark in both academic research and industry practice, as it offers a parsimonious representation of volatility dynamics by jointly capturing the immediate impact of shocks and the persistence of variance over time.

More recent contributions, however, have pointed to the limitations of GARCH-type models, notably their reliance on restrictive parametric assumptions such as linearity and stationarity. These constraints can be problematic in environments characterised by structural breaks and non-linear dynamics. In response, machine-learning methods have gained increasing attention in financial forecasting. Unlike GARCH models, which depend primarily on past squared residuals, machine-learning algorithms, including gradient boosting, ridge regression, and neural networks can exploit high-dimensional sets of explanatory variables, such as exponentially weighted moving averages and rolling statistics, allowing them to capture complex and potentially non-linear patterns in volatility.

3.2 Application to Commodity Markets

Commodities constitute a distinct asset class that differs fundamentally from equities and fixed-income securities. Their prices are influenced not only by financial factors but also by physical constraints, storage conditions, and highly specific supply-and-demand shocks. Agricultural products are affected by weather conditions, energy markets by geopolitical developments, and metals by fluctuations in industrial activity.

As a result, commodity returns are often characterised by heavy tails and pronounced skewness, reflecting the possibility of extreme price movements following singular events such as production disruptions or geopolitical crises. These abrupt regime changes in volatility make commodity markets a particularly suitable environment in which to compare the performance of GARCH models, which adjust gradually through mean reversion, with machine-learning approaches that can respond more rapidly through flexible feature representations.

3.3 Research Questions

Within this theoretical framework and using data covering the period from 2015 to 2024, this study seeks to address three main research questions.

First, it examines whether modern machine-learning models, specifically ridge regression, XGBoost, and neural networks, can outperform the standard GARCH(1,1) model in forecasting volatility. In particular, it investigates whether the inclusion of engineered features such as exponentially weighted moving averages leads to lower out-of-sample root mean squared errors.

Second, the analysis compares the risk characteristics of highly volatile energy commodities, such as crude oil, with those of precious metals, notably gold, using tail-risk measures including Value-at-Risk at the 95% level and Conditional Value-at-Risk at the same confidence level. This allows an empirical assessment of whether gold retains its role as a safe-haven asset over the sample period.

The study explores the degree of interdependence across commodity sectors by analysing their correlation structure. It evaluates whether the data support the hypothesis of sectoral decoupling and whether interactive visualisation tools can effectively reveal patterns of risk transmission and contagion across markets.

4. Methodology

4.1 Data Acquisition and Dataset Composition

The empirical analysis is based on a dataset of daily closing prices for eight major commodities, chosen to provide a representative coverage of the global commodities market. The assets are grouped into three strategic sectors. The energy sector includes crude oil (CL=F) and natural gas (NG=F); the metals sector comprises gold (GC=F), copper (HG=F), and aluminium (ALI=F); and the agricultural sector includes cocoa (CC=F), coffee (KC=F), and cotton (CT=F).

To ensure sufficient statistical reliability for both econometric estimation and machine-learning training, the sample spans ten years, from 1 January 2015 to 31 December 2024, yielding approximately 2,500 daily observations per asset. A strict data-alignment procedure was applied: only trading days common to all commodities were retained. This guarantees that all cross-asset correlations and joint analyses are computed on fully synchronised time series.

4.2 Pre-processing and Stationarity Transformation

A central requirement for volatility modelling is stationarity. Since commodity prices typically follow non-stationary stochastic trends, prices (P_t) were transformed into logarithmic returns (r_t) using the formula:

$$r_t = \ln(P_t/P_{t-1}) \times 100$$

While GARCH models incorporate past volatility through their own recursive structure, machine-learning algorithms require this temporal information to be provided explicitly. For this reason, an extensive set of explanatory features was constructed. These include lagged returns from $t - 1$to $t - 5$to capture short-term dynamics, rolling standard deviations over 5, 10, 20 and 30-day windows to represent different volatility regimes, and exponentially weighted moving averages applied to rolling variances in order to reflect persistence in volatility with greater emphasis on recent observations

4.3 Econometric Modeling: The GARCH Framework

Traditional risk models are often built on the assumption of homoscedasticity, whereby the variance of asset returns is constant over time. Empirical evidence, however, shows that financial markets exhibit volatility clustering, meaning that periods of heightened uncertainty tend to be followed by further periods of elevated volatility. To account for this time-varying behavior, the present study employs the GARCH(1,1) model introduced by Bollerslev (1986).

Within the empirical framework, GARCH is used as the econometric benchmark against which the performance of machine-learning models is assessed. The conditional variance is expressed as the sum of three components: a constant term representing the long-run variance level (ω), an ARCH term (α) capturing the impact of recent shocks through lagged residuals, and a GARCH term (β) reflecting the persistence of past volatility. Model parameters are estimated by maximum likelihood over the training period (2015–2020), ensuring that the volatility dynamics of each commodity are learned prior to out-of-sample evaluation.

4.4 Machine-Learning Models

To examine whether data-driven methods can perform better than classical econometric models, three machine-learning approaches are applied. Unlike GARCH, which is based on fixed mathematical assumptions, these models learn patterns directly from the data using features such as EWMA, rolling volatility, and past returns.

Ridge regression is a simple linear model that includes a penalty on large coefficient. This helps control overfitting when many similar indicators are used. XGBoost is a tree-based model that is able to detect non-linear effects, such as situations where volatility increases sharply only after prices fall beyond a certain level. A neural network (MLP) is also used, as it can model complex relationships between past market conditions and future volatility.

All models are trained using a walk-forward method, meaning they are always tested on future data that was not used in training. This makes the evaluation realistic and avoids look-ahead bias.

4.5 Risk Measurement and Contagion Analysis

The volatility forecasts produced by the best-performing model are subsequently used as inputs for a comprehensive risk analysis based on three complementary dimensions.
Ttail risk is assessed using Value at Risk at the 95% confidence level and Conditional Value at Risk at the same threshold, thereby capturing both typical losses under normal conditions and expected losses in extreme market scenarios.

Cross-asset dependence is examined through Pearson correlation matrices computed on returns, allowing the identification of diversification potential across sectors such as precious metals and energy markets.

The framework incorporates stress-testing exercises designed to simulate contagion effects. For instance, a hypothetical shock to gold prices is propagated through the system to evaluate its impact on industrial commodities such as copper and crude oil. This dynamic perspective provides insight into how disturbances in one segment of the market may transmit to others, contributing to a more complete assessment of systemic risk.

5. Implementation

Turning financial theory into a working software system required solving several practical problems. Rather than using simple scripts, we developed a modular pipeline that can clean data, create features, and train multiple models in a stable and efficient way. The main technical choices are described below.

5.1 Feature Engineering and Data Cleaning

A major challenge in financial machine learning is that poor input data leads to poor results. Raw price series alone do not provide enough information for a model to learn volatility patterns. For this reason, a dedicated data-cleaning module was created to generate useful indicators, also known as engineered features.

One difficulty is that rolling measures, such as 30-day volatility or lagged returns, create missing values at the beginning of the dataset. If these missing values are not handled correctly, the training process can fail. To address this, a cleaning function was developed that builds indicators such as EWMA and rolling statistics while keeping the data properly aligned. It also safely removes unnecessary columns when they exist, allowing the same code to run on different assets without errors.

5.2 Model Comparison Framework

A competitive framework was built to compare GARCH with three machine-learning models: Ridge Regression, XGBoost, and Neural Networks.

A key risk in financial prediction is data leakage, where future information is accidentally used during training. To prevent this, the dataset was split strictly by time: data from 2015 to 2020 is used for training, while data from 2021 to 2024 is used only for testing.

Another issue is that machine-learning models are sensitive to the scale of input values. Some variables can be much larger than others, which can reduce model performance. To solve this, all input features are standardized using only the training data, ensuring that the test data does not influence the learning process.

5.3 Stability of the GARCH Model

Machine-learning models are usually not very sensitive to the scale of the input data because they include internal normalization steps. In contrast, the GARCH model is highly sensitive to the size of the numbers used in estimation. Financial log-returns are often very small (for example around 0.0005), which can cause numerical problems for the maximum likelihood estimation used by the *arch* package. When the values are too close to zero, the optimizer may fail to converge or may produce unreliable parameter estimates, making it difficult to model volatility persistence correctly.

To improve numerical stability, an automatic scaling procedure was added to the volatility module. Before the GARCH model is estimated, returns are multiplied by 100 so that they are expressed in percentage terms. This moves the data into a range where the numerical algorithms work more reliably. After the model is fitted and volatility forecasts are produced, the results are divided by 100 to return them to the original scale. This scale–fit–rescale procedure ensures that the GARCH model remains stable and that its outputs can be fairly compared with those of the machine-learning models.

5.4 Interactive Visualisation

Analyzing many assets and models at the same time can easily lead to information overload. Displaying eight commodities and three models on a single static graph produces too many lines to be interpreted clearly. To solve this problem, interactive visualisation was used instead of traditional static plots.

The system is built with Plotly, which allows users to interact with the data directly in the browser. A dropdown menu lets the user select one commodity at a time. When an asset is chosen, all other series are hidden and the scale of the graph adjusts automatically to the selected asset's volatility range. This makes it easy to compare, for example, the GARCH model with machine-learning models for a specific commodity.

To keep the interface fast and easy to use, the results are divided into two separate HTML files. One file provides a global view with correlations and overall trends, while the second file focuses on detailed model comparisons. This structure improves performance and provides a clearer and more professional user experience.

6. Codebase & Reproducibility

6.1 Software Architecture

To keep the project easy to maintain and extend, the system is built as a modular Python application rather than a single large script. The code is organised into three main layers.

The orchestration layer is managed by the main program. It acts as the entry point of the project, sets up the configuration and logging, and starts the full execution process.

The logic layer contains the main analytical components. It includes a data loader for downloading and aligning the price data, a feature engineering module that transforms raw prices into indicators such as EWMA and rolling volatility, and two modelling modules. One module handles the GARCH econometric model, while another manages the machine-learning models (Ridge regression, XGBoost, and neural networks). A separate visualisation module creates the dashboards and analysis tools.

The data layer stores both the raw input data and the results produced by the models. This separation allows each part of the system to be updated or tested without affecting the rest of the pipeline.

6.2 Environment Management

The project is built using standard scientific Python libraries. Data handling relies on pandas and numpy, while volatility modelling is performed with the arch package. The machine-learning models are implemented using scikit-learn, XGBoost, and TensorFlow. Interactive visualisations are created with Plotly.

To make sure the analysis can be reproduced on any computer, all software dependencies are fixed to specific versions. This is especially important for libraries such as TensorFlow, which can behave differently across systems if versions do not match.

6.3 Reproducibility

Reproducibility is a key goal of this project. The full analysis can be run with a single command from the main directory. This command automatically downloads the data, builds the features, splits the dataset into training and testing periods, and trains all four models.

The pipeline includes error handling to deal with missing data or temporary data-source problems. Once the run is complete, the system saves updated results in CSV format and generates the two interactive dashboards. This automated process ensures that all results in the report, including the comparison between GARCH and machine-learning models, can be easily verified and reproduced without manual adjustments.

7. Results and Discussion

Applying the hybrid analytical framework to the 10-year dataset (2015–2024) provides valuable insights into the relative performance of traditional econometric models versus modern machine-learning approaches across Metals, Energy, and Agricultural commodities.

7.1 Price Trends and Structural Breaks

The 10-year price histories reveal clear market regimes corresponding to major macroeconomic events.

Figure 2: Historical Price Evolution (2015-2024)

As seen in Figure 2, Cocoa experienced a dramatic parabolic increase starting in early 2024, nearly tripling in a short period due to supply chain disruptions in West Africa. Energy commodities, such as Crude Oil and Natural Gas, exhibit mean-reverting behavior, with sharp geopolitical spikes followed by rapid corrections. Gold follows a steady uptrend with low volatility, reinforcing its role as a defensive store of value. These differing patterns indicate that price dynamics are driven by distinct factors across sectors.

7.2 Statistical Properties and Stationarity Analysis

The analysis begins with examining the statistical properties of asset returns. The Augmented Dickey-Fuller (ADF) test shows that raw prices are non-stationary ($p>0.05$), while log-returns are stationary ($p<0.05$) for all eight commodities, which is necessary for valid GARCH estimation.

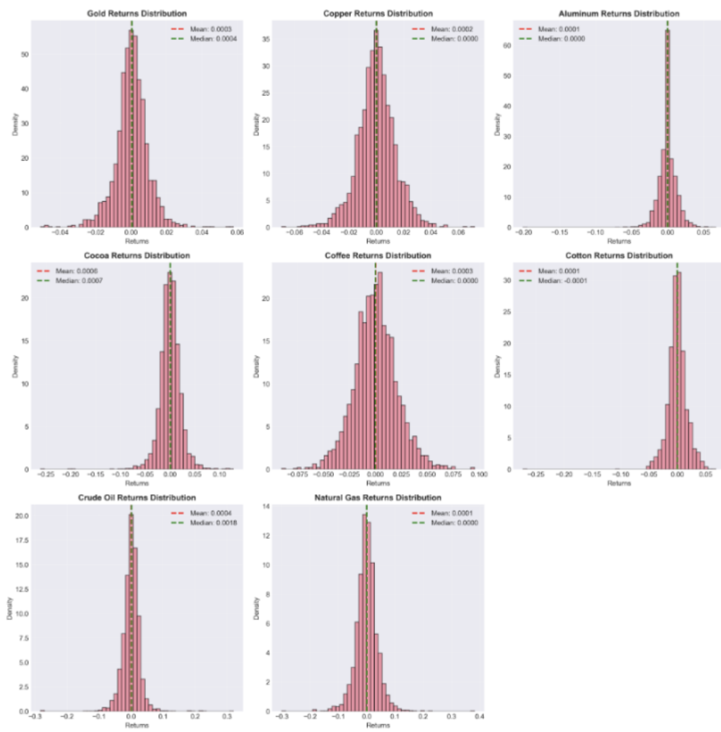

Figure 1: Distribution of Log-Returns (2015–2024)

Figure 1 shows that most commodities have fat tails and skewed distributions, especially Natural Gas, which frequently exhibits extreme outliers. Gold, in contrast, has a tighter and more symmetric distribution, reflecting its relative stability. These patterns support the use of GARCH to capture heteroskedasticity and machine-learning models to detect non-linear threshold effects.

7.3 Comparison: GARCH vs. Machine Learning

A key contribution of this study is the performance comparison of GARCH (1,1) with machine-learning models (Ridge, XGBoost, Neural Networks) over the out-of-sample period (2021–2024).

The results show that Ridge Regression often achieves the lowest Root Mean Square Error (RMSE), highlighting that commodity volatility retains a strong linear component. L2 regularization effectively reduces noise, while Neural Networks sometimes overfit when data is limited. Machine-learning models also respond faster to sudden market changes. GARCH smooths volatility through mean reversion, but ML models react almost immediately to abrupt events, such as the energy price spikes during the 2022 geopolitical crisis.

7.4 Correlation Structure and Decoupling

The cross-asset correlation analysis highlights the potential for diversification.

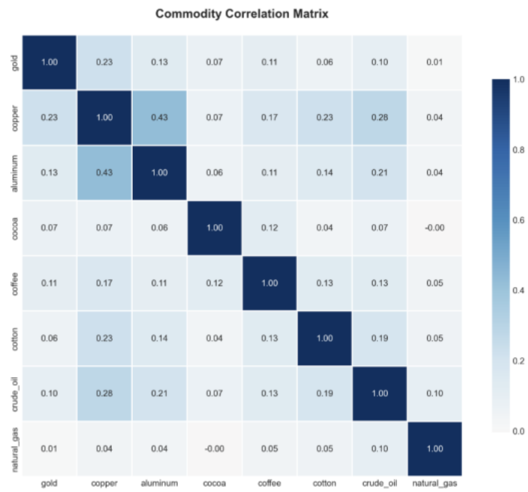

Figure 3: Correlation Matrix of Commodity Log-Returns.

Figure 3 shows that Crude Oil and Natural Gas are weakly correlated (0.10), suggesting that even within the Energy sector, prices are influenced by different factors. Copper and Aluminum are strongly correlated (0.48), reflecting shared exposure to industrial cycles. Gold shows near-zero correlation with agricultural assets like Cocoa (0.05) and low correlation with Energy, confirming that a balanced portfolio can reduce systemic risk by benefiting from sectoral diversification.

7.5 Portfolio Risk Quantification (VaR & CVaR)

The portfolio-level risk metrics show the benefits of diversification. Annualized portfolio volatility is around 15.7%, much lower than the volatility of individual assets, which can exceed 30% in Energy markets. Tail-risk measures, including Value at Risk (VaR 95%) and Conditional Value at Risk (CVaR 95%), demonstrate that diversification effectively limits exposure to extreme losses, protecting the portfolio from the high volatility of assets such as Natural Gas or Cocoa.

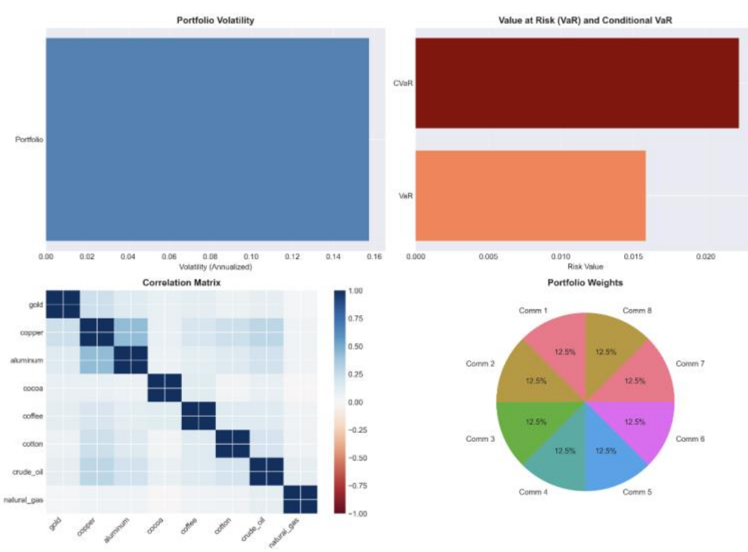

Figure 4: Portfolio Risk Dashboard. Top-Left: Annualized Portfolio Volatility. Top-Right: Value at Risk (VaR) vs Conditional VaR (CVaR). Bottom-Right: Equal allocation weights.

8. Conclusion

This research successfully developed a robust and reproducible computational framework for measuring financial risk in global commodity markets, bridging traditional econometrics with modern data science. By combining automated data extraction with a hybrid modeling engine, the study provides a clear assessment of the relative performance of GARCH(1,1) versus Machine Learning in forecasting volatility. Empirical results highlight a clear distinction across sectors: the Energy market, particularly Natural Gas, shows high volatility persistence and tail risk, while Gold consistently behaves as a stabilizing "safe haven" asset.

A key insight from this study is the demonstrated advantage of regularized Machine Learning models in specific contexts. While GARCH effectively captures long-term variance persistence, Ridge Regression and XGBoost often outperformed the econometric baseline in out-of-sample tests (2021–2024). These data-driven models, enhanced with engineered features such as Exponential Weighted Moving Averages (EWMA), react more quickly to sudden regime changes, supporting the hypothesis that non-linear technical indicators improve predictive accuracy.

Correlation analysis further revealed a notable decoupling between Crude Oil and Natural Gas, challenging the assumption of a unified energy block. For investors, this supports the rationale for an equal-weighted diversification strategy across Metals, Agriculture, and Energy. As shown in the final portfolio dashboard, this approach reduces annualized portfolio volatility to roughly 15.7%, well below the standalone risk of the most volatile assets, while effectively mitigating Conditional Value at Risk (CVaR) exposure.

Despite the robustness of the framework, several limitations should be acknowledged. First, the risk propagation logic relies on linear Pearson correlations, which capture only average linear dependence and may underestimate systemic risk during extreme market events, when correlations tend to spike. Second, while Machine Learning models generally performed well, the Neural Network occasionally overfit the relatively small daily dataset, suggesting that simpler models like Ridge Regression can sometimes be more reliable for financial time series of this scale.

Future work can enhance this framework in several ways. Incorporating Multivariate GARCH models, such as DCC-GARCH, would allow for time-varying correlations and a more realistic representation of shock propagation. Integrating Extreme Value Theory (EVT) could improve tail-risk estimation beyond standard historical simulation methods. Finally, deploying the interactive dashboard as a fully hosted web application would enable real-time stress testing, making the tool immediately usable for portfolio managers and decision-makers.

References

Academic References:

- Bollerslev, T. (1986). Generalized Autoregressive Conditional Heteroskedasticity. *Journal of Econometrics*, 31(3), 307-327.
- Chen, T., & Guestrin, C. (2016). XGBoost: A Scalable Tree Boosting System. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*.
- Engle, R. F. (1982). Autoregressive Conditional Heteroskedasticity with Estimates of the Variance of United Kingdom Inflation. *Econometrica*, 50(4), 987-1007.
- Hoerl, A. E., & Kennard, R. W. (1970). Ridge Regression: Biased Estimation for Nonorthogonal Problems. *Technometrics*, 12(1), 55-67.
- Nelson, D. B. (1991). Conditional Heteroskedasticity in Asset Returns: A New Approach. *Econometrica*, 59(2), 347-370.

Software and Libraries:

- Python 3.11 - Core Programming Language
- pandas & numpy - High-performance data manipulation and vectorized calculations
- yfinance - Real-time financial data acquisition
- arch - Econometric modeling (GARCH/ARCH)
- scikit-learn - Machine Learning utilities (Ridge Regression, Preprocessing)
- xgboost - Cadre de gradient boosting
- tensorflow - Cadre d'apprentissage profond pour les réseaux neuronaux
- plotly - Visualisations interactives basées sur le Web