



UNIVERSITÀ
DI PAVIA

Department of Economics and Management

Master's degree in Finance

Data-Driven Finance: Artificial Intelligence and Machine Learning under Global Uncertainty

Supervisor:

Prof. Paolo Giudici

Candidate

Salvatore De Marco

Matr. n. 556847

Academic Year 2025-2026

ABSTRACT

Over the last decade, financial markets have experienced a profound transformation driven by digitalisation, technological innovation and the unprecedented expansion of information ecosystems. In an increasingly interconnected world, characterised by geopolitical instability, economic uncertainty and rapidly evolving market conditions, the ability to access, process and interpret vast amounts of information has become a fundamental determinant of financial decision-making. As a result, the traditional sources of competitive advantage in financial markets are progressively being complemented, and in some cases surpassed, by the capacity to transform data into actionable knowledge through advanced analytical tools and Artificial Intelligence.

Against this backdrop, this thesis investigates the relationship between financial market dynamics and information-based indicators, with the aim of exploring how uncertainty may influence market behaviour and investors' expectations over time. Particular attention is devoted to major geopolitical events, including Brexit and the Israel– Hamas conflict, whose repercussions have extended well beyond their immediate political dimensions, affecting risk perception, market sentiment and asset price dynamics on a global scale. Within this context, information is no longer merely a reflection of market activity; it increasingly represents a strategic resource capable of shaping expectations and influencing financial outcomes.

The study further examines the contribution of Artificial Intelligence and Machine Learning techniques to the analysis and prediction of financial market behaviour. By adopting a data-driven perspective, the research seeks to evaluate how advanced predictive methodologies can enhance the interpretation of complex market relationships, identify hidden patterns and support decision-making processes in environments characterised by high levels of uncertainty. Rather than replacing traditional financial analysis, these approaches complement

existing methodologies by providing new instruments through which market dynamics can be observed, interpreted and anticipated.

In conclusion, this thesis argues that contemporary financial markets are undergoing a structural shift in which information has become one of the most valuable economic resources. In an era where the ability to collect, process and exploit data increasingly determines strategic advantage, competition extends far beyond capital itself. The new frontier of financial power is no longer defined solely by who owns capital, but by who controls information. In this sense, the most significant battle of our time is no longer fought over resources, territories or market: it is fought over data.

ASTRATTO

Nell'ultimo decennio, i mercati finanziari hanno vissuto una profonda trasformazione guidata dalla digitalizzazione, dall'innovazione tecnologica e dall'espansione senza precedenti degli ecosistemi informativi. In un mondo sempre più interconnesso, caratterizzato da instabilità geopolitica, incertezza economica e condizioni di mercato in rapida evoluzione, la capacità di accedere, elaborare e interpretare enormi quantità di informazioni è diventata un fattore determinante fondamentale del processo decisionale finanziario. Di conseguenza, le fonti tradizionali di vantaggio competitivo nei mercati finanziari vengono progressivamente affiancate, e in alcuni casi superate, dalla capacità di trasformare i dati in conoscenza operativa attraverso strumenti analitici avanzati e l'Intelligenza Artificiale.

In questo contesto, la presente tesi indaga la relazione tra le dinamiche dei mercati finanziari e gli indicatori basati sull'informazione, con l'obiettivo di esplorare come l'incertezza possa influenzare il comportamento del mercato e le aspettative degli investitori nel corso del tempo. Particolare attenzione è dedicata ai grandi eventi geopolitici, tra cui la Brexit e il conflitto Israele–Hamás, le cui ripercussioni si sono estese ben oltre la loro immediata dimensione politica, incidendo sulla percezione del rischio, sul sentiment di mercato e sulle dinamiche dei prezzi delle attività finanziarie su scala globale. In questo quadro, l'informazione non è più un semplice riflesso dell'attività di mercato; essa rappresenta sempre più una risorsa strategica capace di plasmare le aspettative e di influenzare i risultati finanziari.

Lo studio esamina inoltre il contributo delle tecniche di Intelligenza Artificiale e di Machine Learning all'analisi e alla previsione del comportamento dei mercati finanziari. Adottando una prospettiva data-driven, la ricerca si propone di valutare come le metodologie predittive avanzate possano migliorare l'interpretazione di relazioni di mercato complesse, individuare pattern nascosti e

supportare i processi decisionali in contesti caratterizzati da elevati livelli di incertezza. Aniché sostituire l'analisi finanziaria tradizionale, questi approcci integrano le metodologie esistenti, fornendo nuovi strumenti attraverso i quali le dinamiche di mercato possono essere osservate, interpretate e anticipate.

In conclusione, questa tesi sostiene che i mercati finanziari contemporanei stanno attraversando un cambiamento strutturale in cui l'informazione è diventata una delle risorse economiche di maggior valore. In un'epoca in cui la capacità di raccogliere, elaborare e sfruttare i dati determina in misura crescente il vantaggio strategico, la competizione si estende ben oltre il capitale stesso. La nuova frontiera del potere finanziario non è più definita soltanto da chi possiede il capitale, ma da chi controlla l'informazione. In questo senso, la battaglia più significativa del nostro tempo non si combatte più per le risorse, i territori o i mercati: si combatte per i dati.

Table of Contents

1. INTRODUCTION	8
2. FINANCIAL MARKETS IN THE AGE OF DATA	9
2.1 Big Data as a strategic resource for financial analysis and forecasting .9	
2.1.1. The Microstructure of Modern Data Landscapes	10
2.1.2. Advanced Forecasting and the DDDM	11
2.1.3. Artificial Intelligence and Automation in Risk Management	12
2.1.4. Blockchain Integration and Technological Convergence	12
2.1.5. Systemic Risks, Ethics, and Evolving Global Standards	13
2.2. Information, uncertainty and investors' behaviour	15
2.3. Macroeconomic and geopolitical drivers of financial market volatility	17
2.3.1. Brexit and European Market Fragmentation.....	19
2.3.2. Covid-19 and Global Market Volatility	21
2.3.3. The Russia-Ukraine War and Market Shocks.....	22
2.3.4. The Israel-Hamas War and Commodity Shocks.....	24
3. DATA CONSTRUCTION AND EMPIRICAL FRAMEWORK.....	25
3.1. Dataset Construction and Scope	25
3.2. Preprocessing and Missing Values.....	26
3.3. Variables and Event Modelling.....	27
3.4. Financial Market Dynamics	29
4. AI AND MACHINE LEARNING FOR FINANCIAL FORECASTING .31	
4.1 Problem definition and predictive objectives	33
4.2 Data partitioning, tuning protocol and evaluation metrics	40
4.3 Econometric benchmarks: ARIMAX, GARCH and EGARCH	44
4.3.1 ARIMAX	46
4.3.2 GARCH(1,1)	47
4.3.3 EGARCH(1,1)	50
4.4 Penalized linear models: Ridge, Lasso and Logistic Regression.....	52
4.4.1 Ridge regression.....	53
4.4.2 Lasso regression.....	55
4.4.3 Logistic Regression.....	57
4.5 Non-linear Machine Learning: ensembles and kernel methods	60
4.5.1 Random Forest	60
4.5.2 Gradient Boosting	62
4.5.3 XGBoost	64
4.5.4 Support Vector Machine	65

4.6 Deep learning architectures: ANN and LSTM.....	67
4.6.1 Artificial Neural Network	67
4.6.2 Long Short-Term Memory network.....	70
4.7 Comparative analysis and interpretation.....	72
4.8 Implications for data-driven financial decision making.....	81
5. CONCLUSION	87
6. BIBLIOGRAPHY	95

1. INTRODUCTION

In recent years, financial markets have undergone a profound transformation driven not only by globalization, technological acceleration, and geopolitical instability, but also by the growing centrality of data as a strategic asset. In an environment characterized by volatility, information overload, and global uncertainty, the ability to collect, process, and interpret large volumes of structured and unstructured data has become a decisive factor in shaping market behavior. Big Data is no longer simply a support for financial analysis; it increasingly represents the infrastructure through which power is exercised within contemporary markets. Institutions capable of accessing and exploiting data more efficiently than others are often able to strengthen their market position, reduce informational asymmetries, and gain a competitive advantage in decision-making processes.

Alongside the expansion of Big Data, the sentiment expressed by investors, consumers, media outlets, and online communities has emerged as a crucial variable in understanding and anticipating market dynamics. Financial markets are not influenced solely by macroeconomic indicators, corporate fundamentals, or traditional valuation models, but also by collective emotions, perceptions, expectations, and narratives that circulate rapidly through digital platforms. In this sense, sentiment analysis offers a valuable perspective for examining how public opinion and behavioral reactions can affect price movements, trading volumes, and the distribution of influence among financial actors. The increasing relevance of social media, digital news flows, and alternative data sources has therefore redefined the relationship between information and market power, making the interpretation of sentiment an essential dimension of modern finance.

Within this framework, artificial intelligence and machine learning have become essential tools for transforming raw data into predictive knowledge. Their use in finance extends beyond automation or efficiency gains, as these technologies

allow market participants to detect hidden patterns, classify signals, forecast trends, and respond to uncertainty with greater speed and precision. Under conditions of global instability, the capacity to anticipate market positioning through data-driven models becomes a key source of strategic advantage. This thesis aims to explore precisely this intersection between Data Driven Finance, sentiment, and intelligent technologies, by analyzing how the control and interpretation of data contribute to the consolidation of power within financial markets, and how AI and machine learning can be used to anticipate market behavior. In this perspective, the idea that "today's war is fought through data" effectively captures the logic of a financial system in which information is no longer merely a resource, but a field of competition in itself.

2. FINANCIAL MARKETS IN THE AGE OF DATA

2.1 Big Data as a strategic resource for financial analysis and forecasting

The transition from traditional financial market structures to a mature, data-driven economy is fundamentally altering the epistemology of investment decision-making. Historically, financial analysis was a retrospective discipline, relying heavily on manual interventions, periodic corporate disclosures, and the econometric evaluation of historical time-series. While these descriptive, backward-looking models provided a solid foundation for understanding post-market behaviors, they inherently lacked the agility required to navigate modern economic dynamics.

Today, this legacy framework is being rapidly superseded by intelligent, real-time, and predictive systems powered by Big Data architectures and advanced Artificial Intelligence (AI). The core catalyst of this evolution is the ability to process exabytes of structured and unstructured information instantaneously. Consequently, investors and institutions are no longer constrained by the latency

of traditional reporting cycles; they can now execute data-driven strategies based on immediate market signals. This shift from reactive analysis to proactive, predictive intelligence represents a cornerstone of modern financial theory and practice, redefining how value is assessed and captured.

2.1.1. The Microstructure of Modern Data Landscapes

To truly grasp why contemporary datasets require advanced architectural frameworks, we must evaluate how their core properties challenge traditional relational databases within an environment structurally defined by the traditional "4 V's" paradigm, which has elevated data from a mere operational byproduct to a premier strategic asset and a primary factor of production. This structural transformation is driven by Volume, representing the overwhelming, exponential scale of information generated every single second that expands far beyond what standard, centralized storage systems can computationally handle, alongside Variety, which encompasses the diverse and fragmented nature of incoming formats where clean financial tables must be dynamically synthesized with messy, unstructured alternative data streams.

The critical real-time pressure of this ecosystem is dictated by Velocity, the incredible speed at which new information streams in, demanding rapid, near-zero-latency processing to capture its immediate economic value before the window of opportunity closes, all while being inherently constrained by Veracity, which dictates the structural trustworthiness and statistical accuracy of the data to ensure that the insights extracted are resilient enough to guide high-stakes capital allocation. By harmonizing this endless variety of modern touchpoints, financial institutions can now look far beyond traditional fundamentals such as balance sheets, P/E ratios, and lagging macroeconomic indicators to transform peripheral information into actionable alpha, allowing investors to identify opportunities above market benchmarks through the real-time monitoring of public perception via social media sentiment analysis, the tracking of international supply chain

activity via satellite imagery, and the immediate assessment of consumer spending behavior through geolocation and transaction-level data before official corporate earnings are published.

2.1.2. Advanced Forecasting and the DDDM

Enhanced financial forecasting leverages this influx of multi-dimensional data by substituting rigid, fixed models with machine learning algorithms. Instead of making static, periodic projections, predictive analytics models are continuously updated with new data inputs, making forecasts adaptive to shifting volatile markets. This predictive intelligence enables organizations to run dynamic "what-if" scenario analyses, assessing how fluid variables, such as fluctuations in interest rates or commodity prices, may impact ultimate financial outcomes, thereby optimizing resource allocation.

This infrastructure underpins Data-Driven Decision-Making, a practice that replaces institutional intuition and executive guesswork with empirical data interpretation. In fast-moving markets, having immediate access to real-time cash flow, revenue tracking, and expense metrics allows companies to remain highly agile. Furthermore, backing corporate choices with concrete evidence fosters an internal culture of transparency and accountability, mitigating cognitive biases and personal subjectivity in leadership decisions.

The pinnacle of this analytical maturity is achieved by combining predictive and prescriptive analytics. While predictive tools use statistical modeling and artificial intelligence to forecast what is likely to happen, prescriptive analytics goes a step further. By utilizing optimization algorithms and simulation techniques under strict corporate constraints, prescriptive models recommend the optimal course of action among various alternatives, transforming raw foresight into automated, strategic execution.

2.1.3. Artificial Intelligence and Automation in Risk Management

Beyond high-level strategy, Artificial Intelligence (AI) and Machine Learning (ML) are structurally redefining the operational mechanics of accounting and risk management. In big data accounting, AI algorithms automate repetitive, high-latency tasks such as data entry and ledger reconciliation, drastically reducing human error and freeing human capital for higher-value activities like financial planning. Simultaneously, these advanced analytics tools translate massive, granular financial datasets into intuitive, multi-dimensional visualizations, enabling stakeholders to rapidly grasp key performance metrics and identify cost-saving opportunities.

In the realm of risk management and fraud detection, big data analytics has staged a transformative impact by replacing slow, manual auditing processes and predefined, rigid rules. By processing vast internal and external data streams simultaneously, machine learning models can identify emerging credit risks, forecast potential liquidity disruptions, and monitor broad economic factors in real time. For fraud mitigation, these algorithms continuously analyze transaction metadata, instantly flagging anomalies that deviate from established behavioral norms, such as multiple transactions executed from conflicting geolocations within unrealistic timeframes. Because these systems continuously learn from new data inputs, they dynamically adapt to evolving financial crime techniques, increasing detection accuracy while significantly reducing the false positives common in legacy systems.

2.1.4. Blockchain Integration and Technological Convergence

When big data architectures converge with blockchain technology, the financial ecosystem gains an unprecedented layer of data security, transparency, and trust. Blockchain's decentralized, distributed, and immutable ledger ensures that shared

financial information is entirely tamper-proof, eliminating traditional data silos and ensuring absolute data integrity across multiple cross-border stakeholders.

This technological integration is further supercharged by the deployment of smart contracts, self-executing agreements hardcoded directly into the blockchain.

When paired with real-time big data inputs, smart contracts can fully automate complex workflows without manual intervention, such as instantly triggering corporate payments the moment a global supply chain delivery is digitally verified. This automated synchronization vastly accelerates operational velocity and enhances strategic agility in an increasingly interconnected corporate world.

2.1.5. Systemic Risks, Ethics, and Evolving Global Standards

However, the creation of this highly efficient, data-centric scaffolding is not without its severe architectural flaws. As the velocity and complexity of automated financial systems increase, they give rise to critical systemic risks and threats to global economic security that were largely contained within human-intermediated markets.

First, the hyper-digitization of financial platforms drastically heightens institutional exposure to sophisticated cybersecurity threats and malicious data breaches, compromising sensitive corporate intelligence. Second, a profound structural concern stems from the "explainability" problem in complex, deep-learning artificial intelligence models; when algorithms operate as a "black box," it becomes incredibly difficult to audit the rationale behind their financial choices. If the underlying historical training data contains systemic biases, the AI will inevitably execute severely skewed decisions. Moreover, the highly interconnected nature of modern digital platforms creates precarious network vulnerabilities, where an isolated operational failure or data disruption in a single node can rapidly cascade across global networks, triggering automated, market-wide algorithmic sell-offs and flash crashes within microseconds.

Consequently, navigating this precarious landscape depends entirely on the strategic development of robust cyber resilience, ethical data governance, and adaptive oversight. Traditional, reactive regulatory models consistently struggle to keep pace with the exponential velocity of technological change. This necessitates a profound evolution in global standards. International frameworks, such as the International Financial Reporting Standards (IFRS) and the International Organization for Standardization (ISO), are actively revising their protocols to address cross-border data privacy, interoperability, and the ethical deployment of automation¹.

Ultimately, the mass adoption of big data is shifting the entire financial profession from a reactive, historical discipline to a predictive, proactive, and strategic function. The long-term financial health and resilience of the global economy will depend on achieving a delicate balance: harnessing the immense predictive potential of machine efficiency while retaining the indispensable human judgment and transparency required to safeguard ethical and systemic foundations.

¹ See Georgios L.Thanasas et al., The role of Big Data Analytics in Financial Decision-Making and Strategic Accounting.

2.2. Information, uncertainty and investors' behaviour

While the technological shift toward big data architectures offers unprecedented analytical capabilities, it simultaneously introduces a critical paradox: the sheer volume of modern market data often collides with the inherent cognitive and processing limits of human investors. Because market participants cannot possibly digest the vast quantity of information present in high-frequency environments, their decision-making processes adapt through specific behavioral shortcuts. In this context, investors frequently place a disproportionate weight on public signals over private ones, particularly when they believe other participants are observing the same data. This reliance on shared information serves as a strategic tool to predict the average market opinion rather than the intrinsic value of an asset. Within this dynamic, transaction history acts as a dominant catalyst; the mere observation of past transaction volumes is often more influential in explaining herding behavior than any other form of public data. When this collective observation intensifies, it frequently triggers information cascades. These cascades occur when the observed actions of peers become so overwhelmingly informative that individuals entirely ignore their own private signals, choosing instead to copy the majority with absolute certainty through a mechanism of selective attention, where only the most appealing or salient public signals dictate choice.

This structural vulnerability to data overload means that deep-seated uncertainty regarding an asset's true value inevitably drives investors to look at their peers for behavioral cues. Under high-volatility and high-uncertainty conditions, herding behavior ceases to be merely an irrational impulse and instead becomes a deliberate protective mechanism against the perceived risks of an unstable environment. This pervasive uncertainty exerts a powerful leveling effect across the market, effectively suppressing individual cognitive differences and forcing collective imitation regardless of an investor's personal psychological profile.

This phenomenon is particularly acute during specific market stages, such as market openings, where previous transaction history is scarce and initial uncertainty is at its peak. In these early phases, the introduction of additional public data often fails to alleviate the baseline feeling of insecurity among participants, thereby accelerating the reliance on mutual imitation rather than independent analysis.

Consequently, an investor's unique cognitive profile defined by a specific combination of psychological biases and beliefs becomes a decisive factor in behavioral outcomes only when baseline market uncertainty is relatively low. When the market stabilizes enough for individual psychology to resurface, certain biases actively promote herding dynamics. For instance, the Illusion of Control (ILC) leads investors to assume that the successful choices of others are driven by superior skill rather than random chance, prompting emulation, while the Hot-Hand Fallacy (HHF) falsely convinces them that a recent trend of buy or sell orders will inevitably persist. Conversely, other psychological biases act as inhibitors to herding; Self-Attribution (SA) bias causes individuals to credit successes to their own unique skills, reducing their inclination to copy peers, while the Gambler's Fallacy (GF) prompts investors to trade against a recent trend in the mistaken belief that random sequences must naturally self-correct. Navigating this psychological landscape is further complicated by Overconfidence (OC), a bias that typically reduces herding by making individuals over-reliant on their own data, but which can paradoxically fuel speculative bubbles. In low-uncertainty environments, overconfident investors may misinterpret the prevailing collective market mood as a mathematically ideal indicator, leading them to override their independent judgment and join the herd. Ultimately, market microstructure is shaped by a complex interplay between environmental uncertainty and individual psychology, where extreme uncertainty effectively masks personal biases and locks the market into states of collective imitation².

² See Fernàndez et al., Herding, information uncertainty and investors' cognitive profile.

2.3. Macroeconomic and geopolitical drivers of financial market volatility

The investigation of financial market volatility cannot be restricted solely to endogenous technological shifts or the cognitive biases of market participants; it must necessarily integrate the exogenous macroeconomic and geopolitical forces that structurally disrupt global capital markets. While traditional economic literature has frequently conceptualized geopolitical risk as a smooth, linear, and gradually evolving process, modern empirical frameworks require a more nuanced identification strategy focused on "geopolitical turning points." These turning points are defined as sudden, unanticipated inflection points in bilateral state-to-state relations that represent an abrupt departure from prior historical trajectories. Unlike conventional, aggregated global risk indicators, these dyadic shocks function as macroeconomic "news shocks" to which financial markets respond with heightened sensitivity and immediate structural re-pricing. By adopting this bilateral lens, the analytical focus shifts from generic "global uncertainty" to a precise evaluation of how strategic diplomatic deteriorations or reconciliations between major systemic powers, act as primary drivers of market instability. Methodologically, these volatile shocks are captured through the second difference of the Political Relationship Index , an econometric metric designed to isolate the non-linear acceleration or deceleration of political tensions, effectively filtering out persistent trends to reveal the true, high-frequency triggers of market asset volatility.

The transmission mechanism through which these geopolitical turning points destabilize financial markets, particularly within the highly sensitive commodity and energy sectors, operates through two distinct and temporally asymmetric channels. In the short run, an exogenous improvement in bilateral geopolitical relations induces a "risk compression" effect, leading to an immediate reduction in market volatility and spot prices, such as those of crude oil. This immediate cooling is driven by a sharp contraction in "precautionary demand," as the easing

of international tensions diminishes the necessity for market participants to aggressively hedge against future supply disruptions. Conversely, the medium-run economic outlook introduces a opposing dynamic: approximately 32 months post-shock, a positive geopolitical turning point can paradoxically lead to an upward trend in asset prices and volatility. This secondary wave occurs as market pricing begins to reflect forward-looking expectations of expanded international trade, heightened macroeconomic activity, and increased aggregate energy demand resulting from the newly stabilized geopolitical landscape.

Crucially, the magnitude and directional intensity of these transmission channels are not uniform across time, but are highly dependent on the pre-existing state of the market, exhibiting a profound asymmetric sensitivity. Incorporating a sophisticated quantile analysis reveals that the initial macroeconomic conditions of a financial market act as an operational multiplier for the volatility triggered by a geopolitical turning point. In low-price environments, financial markets display a significantly higher sensitivity to short-run geopolitical improvements, as the sudden relaxation of precautionary demand carries a disproportionate mathematical weight during market troughs. On the other hand, the medium-run demand-driven expansions are far more pronounced when the market is initially in a high-price quantile, where expectations of future economic growth exert a stronger marginal impact on an already tight, highly active market.

Ultimately, these geopolitical turning points do not remain isolated within the specific dyad of the nations involved; rather, they reverberate globally, transforming into standalone causes of structural terms-of-trade shifts that redefine international macroeconomic equilibria. By altering the underlying costs of strategic imports, geopolitical shocks fundamentally destabilize the saving-investment balance of both exporting and importing nations, thereby reshaping global trade balances and current accounts. For commodity-importing economies, these shocks generate immediate supply-side inflationary pressures that

necessitate aggressive monetary interventions and interest rate adjustments. Conversely, for exporting nations, the resulting price fluctuations directly impact fiscal revenue sustainability, capital flow patterns, and real exchange rate configurations. Geopolitical risk thus emerges not merely as an external political variable, but as a core macroeconomic determinant that dictates asset price distributions and amplifies volatility across the global financial architecture³.

In order to rigorously evaluate the empirical validity of these theoretical transmission channels and their direct impact on modern market instability, the subsequent sections of this chapter will analyze four major contemporary geopolitical disruptions. Specifically, the analysis will focus on the macroeconomic and financial ramifications triggered by the institutional friction of Brexit, the systemic global shock of the COVID-19 pandemic, the structural energy and commodity re-pricing caused by the Russia-Ukraine war, and the heightened regional and supply-chain volatility stemming from the Israel-Hamas conflict.

2.3.1. Brexit and European Market Fragmentation

In evaluating the empirical manifestations of geopolitical turning points, the institutional friction of Brexit represents a seminal case study, serving as the first major de-globalization and economic disintegration event in post-war Europe. Rather than acting as a single, isolated macroeconomic shock, the withdrawal of the United Kingdom from the European Union functioned as a prolonged, volatile process of structural negotiations and profound political disagreements that fundamentally altered the continent's economic landscape. By systematically introducing novel institutional barriers to the exchange of goods, services, and human capital, this disintegration process forced global financial markets to fundamentally recalibrate for an era of reduced regional integration, triggering immediate and severe asset re-pricing.

³ See Jamel Saadaoui et al., Geopolitical turning points and macroeconomic volatility: a bilateral identification strategy.

The core operational mechanism through which Brexit generated this systemic instability was cross-border financial contagion, whereby localized political shocks within the United Kingdom were rapidly transmitted across European financial systems. Given the UK's position as a dominant hub in global finance, any significant political or regulatory shift inherently destabilizes international market linkages. This contagion mechanism became acutely evident following the unexpected outcome of the 2016 referendum, an event that triggered an unprecedented collapse in financial integration within the EU-15 zone. This plummeting interdependence was statistically reflected in a sharp decline in both the speed and magnitude of volatility spillovers, signaling that the European financial architecture was undergoing a process of structural fragmentation.

This regional volatility was further exacerbated by the highly asymmetric roles played by different European economies in transmitting and receiving market risk. Large, highly liquid capital markets, acted as the primary drivers of regional volatility networks. Throughout the extended Brexit timeline, France consistently emerged as a key net transmitter of volatility within the Union. Conversely, the United Kingdom's macroeconomic role evolved dynamically: it functioned as a major exporter of systemic volatility during the early, highly uncertain stages of political negotiation, but gradually shifted toward becoming a net receiver of volatility as the institutional process matured and the focus transitioned toward final trade cooperation agreements. In contrast to these dominant hubs, smaller and more open economies, such as Ireland, Portugal, and Spain, were forced to continuously absorb these cross-border shocks, highlighting their heightened structural sensitivity to the political turbulence originating from the continent's primary financial centers. And so, this persistent volatility was sustained by deteriorating macroeconomic expectations and pervasive policy uncertainty that overshadowed the entire withdrawal timeline. Financial markets reacted sharply to the anticipated long-term structural damage to the British economy, which included a predicted 22% reduction in foreign direct investment (FDI) inflows

and significant, compounding losses in aggregate productivity⁴. These structural headwinds, combined with the sharp currency depreciation of the British pound and the ongoing threat of regulatory divergence, created a self-reinforcing feedback loop of market instability. In this manner, Brexit demonstrated that domestic political and institutional processes can act as standalone drivers of market volatility, permanently reshaping investor sentiment, altering transmission channels, and elevating systemic risk across entire geographic regions.

2.3.2. Covid-19 and Global Market Volatility

Concurrently, the COVID-19 pandemic represented a monumental shift in global financial stability, mutating rapidly from a localized health emergency into a comprehensive macroeconomic catastrophe. Following its official declaration as a global pandemic by the World Health Organization on March 11, 2020, the crisis plunged the global economy into a severe recession comparable in magnitude to the Great Depression of 1929 and the 2008 Financial Crisis. The primary structural catalyst for the ensuing asset price volatility was the synchronized implementation of stringent lockdown protocols and international travel controls by global governments. These interventions triggered massive, simultaneous supply and demand shocks, severely interrupting global supply chains due to production halts in central manufacturing hubs like China, and creating an environment of radical uncertainty that severely complicated policy responses and heightened investor anxiety.

Empirical evidence rigorously confirms a statistically significant positive relationship between the epidemiological progression of the pandemic and market turbulence. The exponential trajectory of newly confirmed cases and mortality rates acted as immediate risk signals that drove volatility upward across both developed and emerging markets. This period was characterized by an

⁴ See Marwan Izzeldin et al., Brexit and Its Impact on EU Financial Markets.

unprecedented level of network "connectedness" between global exchanges, meaning that the systemic shocks of COVID-19 were transmitted with greater velocity and severity than in previous financial crises. Institutional investors abruptly adjusted their capital allocation strategies in response to the perceived existential threat to global value chains, accelerating cross-border volatility spillovers⁵. Furthermore, the pandemic serves as a definitive empirical manifestation of behavioral finance, where high levels of panic, fear, and collective anxiety overrode rational market mechanisms. This psychological strain resulted in historic volatility spikes and fueled extreme speculative bubbles, as investors engaged in a massive influx of financial transactions driven by emotional contagion rather than fundamental asset analysis. Behavioral biases, including herding behavior and severe loss aversion, were vastly amplified, rendering market microstructures highly unpredictable. Ultimately, the COVID-19 outbreak demonstrated that in a highly globalized world, a non-economic health emergency can instantaneously destabilize global capital markets and become a primary driver of systemic financial risk.

2.3.3. The Russia-Ukraine War and Market Shocks

Subsequently, the escalation of the Russia-Ukraine war into a full-scale military conflict in February 2022 represents one of the most profound geopolitical shocks of the post-Cold War era, fundamentally re-pricing systemic risk across global financial architectures. This conflict has functioned as a primary empirical driver of structural uncertainty, precisely captured by the Geopolitical Risk (GPR) Index, which tracks news-based indicators of interstate tensions, military escalations, and the unilateral imposition of unprecedented economic sanctions. The sheer abruptness of the invasion created an exogenous "news shock" that rapidly reverberated through international capital markets, abruptly altering investor behavior and global capital allocation patterns. As a consequence, market participants were forced to instantaneously reassess the long-term

⁵ See Sabeeh Ullah et al., COVID-19 pandemic and financial market volatility: a quantile regression approach.

stability of the rules-based global political order and the operational security of cross-border trade networks.

At the macroeconomic level, the transmission mechanism of this geopolitical disruption operates primarily through the severe interruption of global trade flows and critical energy supply chains. The extensive, multi-layered sanctions imposed on the Russian Federation, combined with the global market's structural dependence on Russian commodities, triggered massive, non-linear shifts in the demand and supply dynamics of foreign exchange markets, thereby directly amplifying exchange rate volatility. This currency instability has exhibited an asymmetric sensitivity determined by individual national trade structures; specifically, economies that are deeply integrated into global value chains or those maintaining a high marginal dependence on Russian energy and agricultural imports have experienced the most pronounced currency depreciations and volatility spikes. These structural disruptions have heavily distorted import-export pricing and complicated the transmission of national monetary policies, highlighting how geopolitical warfare translates directly into tangible macroeconomic vulnerability⁶.

Furthermore, this heightened geopolitical instability has activated a powerful "flight-to-safety" transmission channel, prompting an aggressive reallocation of international portfolios away from high-risk environments and toward established safe-haven assets. This capital flight has generated a highly heterogeneous and polarized response across the global financial landscape. Conversely, while global safe-haven currencies experienced compressed volatility due to intense capital inflows, emerging markets and commodity-dependent economies faced compounding risks of sudden capital reversals and structural financial instability. Ultimately, the ongoing war has sustained an elevated geopolitical risk baseline, underscoring the deep vulnerability of international financial markets to sudden diplomatic breakdowns and military escalations.

⁶ See Zulfikar Ali Mir et al., *Geopolitical Risks and Exchange Rate Volatility: Evidence from Russia-Ukraine War*.

2.3.4. The Israel-Hamas War and Commodity Shocks

Adding to these systemic international disruptions, the outbreak of the 2023 conflict between Hamas and Israel emerged as an abrupt and highly unpredictable driver of asset price volatility, occurring without any prior structural signals or overt threats detectable by market participants. At a macroeconomic and geopolitical level, this intense escalation functioned as a powerful negative news shock that instantly destabilized the global economic order and sharply elevated uncertainty across international capital networks. Because financial investors exhibit an asymmetric sensitivity to geopolitical events, responding far more aggressively to negative political developments than to positive ones, the conflict triggered immediate downward corrections in leading global equity indices, such as the S&P 500 and the MSCI World, as sentiment-driven uncertainty vastly increased the conditional variance of market returns.

From a macroeconomic perspective, the primary transmission channel of this geopolitical shock was concentrated within the pricing dynamics of strategic commodities, most notably crude oil and gold. Energy markets faced intense, immediate speculative pressure due to forward-looking anxieties regarding potential structural disruptions in production, refining, and maritime distribution routes within the Middle East, a historically vital choke point for global oil supply; this translated into notable spot price surges and elevated volatility across major regional exporters, including the United Arab Emirates, Saudi Arabia, and Qatar. Simultaneously, gold reaffirmed its classic empirical role as a premier safe-haven asset, experiencing sharp capital inflows and price appreciations across nearly all major international hubs as global investors scrambled to diversify their portfolios to hedge against systemic tail risk and regional structural instability. The critical importance of this conflict as a catalyst for market volatility is further highlighted by the significant and highly divergent

reactions observed across global capital markets and foreign exchange architectures. Major sovereign equity markets, including those of the United States, China, Indonesia, and Saudi Arabia, exhibited pronounced anomalies in abnormal returns, return variability, and aggregate trading volumes, signaling a synchronized operational rush by institutional participants to digest the incoming geopolitical information. In the foreign exchange domain, while the U.S. dollar maintained its structural stability due to its global reserve status, other major and emerging currencies, such as the Chinese Yuan and the Indonesian Rupiah, displayed immediate signs of weakening⁷.

Ultimately, the conflict activated a powerful signaling mechanism that forced global policymakers and institutional investors to fundamentally recalibrate their strategic asset allocations in the face of persistent financial fragility and heightened geopolitical friction.

3. DATA CONSTRUCTION AND EMPIRICAL FRAMEWORK

3.1. Dataset Construction and Scope

For the development of this project, I began by formulating an extensive longitudinal quantitative inquiry, specifically engineered to map a decade of global financial instability through the structural lens of the United States equity market. The construction of the dataset deliberately transcends a conventional compilation of historical price series; instead, it establishes a multidimensional informational architecture that establishes a direct, quantifiable correlation between standardized macroeconomic indicators and discrete, high-impact geopolitical shocks. By shifting the analytical focus toward this interdisciplinary synthesis, the dataset provides a rigorous empirical canvas to observe how exogenous political and social crises are absorbed by endogenous market microstructures.

Chronologically, the scope of the analysis is comprehensively vast, spanning a continuous ten-year horizon from June 20, 2016, to June 17, 2026. This decade-

⁷ See Ainun Rochimah et al., Capital Market and Macroeconomic Reaction to The Hamas-Israel Conflict 2023.

long window is intentionally selected to encompass a series of radically heterogeneous market regimes, including late-cycle expansionary phases, acute black-swan contractions, post-crisis monetary interventions, and inflationary regime shifts. This extensive temporal depth allows the study to move decisively beyond the inherent limitations of short-term or localized observations, which frequently suffer from sample selection biases or temporary cyclical distortions. Consequently, the research offers a granular, long-term perspective on macroeconomic market resilience and the conditional speed at which asymmetric information is discounted into asset prices. To capture these dynamics without suffering from temporal aggregation bias, the design integrates high-frequency daily data. This ensures that the analytical framework remains sensitive not only to long-term secular trends and structural breaks but also to the instantaneous, daily volatility spikes and liquidity dislocations that define investor behavior and capital flight during moments of systemic panic.

3.2. Preprocessing and Missing Values

The validity and statistical reliability of any sophisticated econometric analysis depend critically on the mathematical rigor applied to the initial stages of data preparation. Financial time series are notoriously plagued by structural irregularities; thus, this dataset underwent a meticulous data cleaning and harmonization process designed to neutralize the temporal asymmetries inherent in global financial markets. Primary among these challenges are the asynchronous gaps caused by national holiday closures, weekend trading suspensions, and unexpected market halts. Rather than employing crude deletion methods that risk undermining the continuity of the time series, sophisticated interpolation and alignment protocols were implemented to preserve the statistical properties of the underlying distributions.

Through advanced temporal mapping techniques, each discrete observation_date was uniquely and flawlessly aligned with the concurrent status of the exogenous

geopolitical variables. This precise synchronization ensures that the activation of any specific binary dummy variable (0 or 1) precisely reflects the real-world operational onset and subsequent duration of the respective geopolitical event. Furthermore, the preprocessing phase required the rigorous normalization of structurally disparate time series. Integrating absolute price data, bounded implied volatility indices, and unbounded policy uncertainty metrics within a single corporate framework necessitates mathematical transformations, such as log-differencing and relative scaling, to eliminate non-stationarity and prevent severe statistical distortions like spurious regressions. This rigorous foundational work alters the signal-to-noise ratio of the dataset, ensuring that the final econometric model can clearly distinguish between routine, idiosyncratic market "noise" and genuine, statistically significant systemic crisis signals. Ultimately, this clean dataset provides a robust empirical basis for partitioning the temporal horizon into specific, highly localized event windows.

3.3. Variables and Event Modelling

Feature engineering serves as the analytical core of this investigation, providing the mathematical apparatus required to translate qualitative, non-linear geopolitical shifts into discrete, quantitative variables suitable for advanced econometric testing. The structural components engineered for this research framework are systematically consolidated into a comprehensive, multi-dimensional informational structure:

- **The SP500 Index (Performance Dimension):** This variable tracks the daily closing value of the Standard & Poor's 500 index. Utilized as the primary market anchor, the SP500 acts as a macro-barometer for global economic health, institutional asset valuation, and large-cap corporate stability. By analyzing the continuous log-returns of this index, the model

calculates capital destruction trajectories, dynamic market adjustments, and post-crisis recovery velocities across varying shock regimes.

- **The VIXCLS (Psychological Hedging Dimension):** The CBOE Volatility Index functions as the definitive gauge of near-term equity risk, calculating the market's expectation of 30-day forward-looking volatility implied by S&P 500 index options. It isolates immediate, speculative investor fear and sudden tactical demand for downside hedging, capturing sharp, short-term psychological and behavioral reactions to exogenous disruptions.
- **The WLEMUINDXD (Structural Macro-Uncertainty Dimension):** The Equity Market Uncertainty Index provides a broader, macro-level diagnostic of structural anxiety specifically tied to capital markets. By tracking deep-seated policy, economic, and institutional ambiguities, this metric allows the econometric framework to separate pure, short-term derivatives-market volatility VIXCLS from deeply rooted, long-term macroeconomic instability, capturing systemic anxieties that standard implied volatility metrics might fail to register.
- **The BREXIT Structural Indicator (Geopolitical Dummy):** A binary variable engineered to capture the profound institutional and macroeconomic decoupling caused by the UK referendum. The variable is mathematically modeled to capture the structural break and subsequent market adjustments following the unexpected voting outcome:^[17]_[SEP]

$$BREXIT = \begin{cases} 1 & \text{on or after June 24, 2016} \\ 0 & \text{otherwise} \end{cases}$$

- **The COVID Exogenous Shock Indicator (Geopolitical Dummy):** A binary variable designed to encapsulate the unprecedented systemic crisis, global lockdown protocols, and severe macroeconomic contractions

induced by the health emergency. The activation threshold is formalized on the day of the official multilateral crisis declaration:

$$COVID = \begin{cases} 1 & \text{on or after March 11, 2020} \\ 0 & \text{otherwise} \end{cases}$$

- **The UKRAINEWAR Geopolitical Disruption Indicator (Geopolitical Dummy):** A binary variable mapping the extensive structural shock caused by the military invasion. This indicator accounts for prolonged macroeconomic vulnerabilities, multi-channel energy supply chain disruptions, commodity price inflation, and the implementation of international sanction regimes:

$$UKRAINEWAR = \begin{cases} 1 & \text{on or after February 24, 2022} \\ 0 & \text{otherwise} \end{cases}$$

- **The ISRAELHAMAS Regional Instability Indicator (Geopolitical Dummy):** A binary variable configured to model the capital market impact, localized geopolitical risk contagion, and macro-financial uncertainty stemming from the escalation of conflict in the Middle East:

$$ISRAELHAMAS = \begin{cases} 1 & \text{on or after October 7, 2023} \\ 0 & \text{otherwise} \end{cases}$$

3.4. Financial Market Dynamics

An initial exploratory analysis of the compiled dataset reveals an immediate, intense, and often violent reactivity of financial markets to the exogenous shocks mapped by the binary features. The statistical distributions of both asset prices and volatility indices exhibit massive departures from normality during these transition phases, structurally characterized by heavy-tailed behavior, negative skewness, and abrupt regime breaks. For instance, the unexpected outcome of the

UK referendum in June 2016 caused an instantaneous, severe contraction of the SP500, which plummeted in a single trading session from 2113.32 to 2037.41, representing a massive, abrupt evaporation of equity value across global capital markets. Simultaneously, the options market reacted with immediate turbulence, causing the VIXCLS to surge from a complacent baseline of 17.25 to an aggressive peak of 25.76 in a single session, highlighting a sudden spike in hedging costs and systemic panic.

The onset of the COVID-19 pandemic generated the most extreme, unprecedented market dynamics recorded in modern financial history, shifting the baseline distributions into uncharted territory. On March 16, 2020, amid accelerating global lockdowns, liquidity contractions, and systemic supply chain disruptions, the VIXCL experienced an unprecedented, vertical escalation, reaching a historic closing peak of 82.69. This extraordinary level signaled a near-total paralysis of rational investment sentiment and an absolute corporate premium on liquidity and cash preservation. Concurrently, the Equity Market Uncertainty Index WLEMUINDXD experienced extreme, non-linear surges, shattering all historical benchmarks to exceed a value of 900 as the global macroeconomic engine faced sudden halts.

In a similar vein, the geopolitical shock of the Ukraine War in February 2022 introduced deep, persistent structural volatility into capital markets. Unlike the rapid, mean-reverting "impulse" shock seen during the Brexit referendum, the invasion induced prolonged uncertainty; the VIXCL remained elevated well above the 30 threshold for several consecutive weeks, reflecting deep structural anxieties regarding global energy supply chains, commodity price inflation, and the macro-financial implications of international sanction regimes. This was further compounded by the geopolitical escalation in October 2023, where the activation of the Middle Eastern conflict dummy triggered immediate capital reallocations toward safe-haven assets, embedding a persistent risk premium into international equity options.

The deep heterogeneity of these shocks demonstrates the necessity of a layered, multi-stage econometric approach. To properly decode these market dynamics, classical forecasting methodologies, encompassing both standard linear models and foundational non-linear financial specifications, provide the essential baseline framework. These traditional econometric techniques are crucial for isolating structural breaks, identifying long-term trends, and establishing a rigorous, interpretable statistical foundation. However, as markets transition rapidly from short-term impulses to highly intricate, prolonged regime shifts, these benchmark models can be significantly enhanced. Rather than replacing traditional methods, the analytical objective becomes an empirical convergence: leveraging the structural clarity of classical econometrics alongside the advanced pattern-recognition capabilities of machine learning algorithms. This integration allows for the deployment of advanced computational architectures, turning the focus toward non-linear AI systems and neural frameworks designed to complement traditional methods and transform real-time information into dynamic, predictive financial intelligence.

4. AI AND MACHINE LEARNING FOR FINANCIAL FORECASTING

The preceding chapters have established the conceptual and empirical foundations of this research. Chapter 2 argued that contemporary financial markets have entered an era in which data constitute both a strategic resource and an instrument of competitive power, while sentiment, uncertainty and digital narratives increasingly shape the formation of asset prices. Chapter 3 translated this theoretical framework into a concrete empirical architecture, constructing a decade-long, daily-frequency dataset that aligns the performance of the United States equity market with two complementary measures of collective psychology, the CBOE Volatility Index (VIXCLS) and the Equity Market Uncertainty index

(WLEMUINDXD), and with four discrete geopolitical shocks: the Brexit referendum, the COVID-19 pandemic crash, the Russian invasion of Ukraine and the outbreak of the Israel– Hamas conflict.

This chapter represents the analytical core of the thesis. Its purpose is to assess, in a rigorous and fully out-of-sample setting, whether the informational content embedded in volatility, sentiment-driven uncertainty and geopolitical event indicators can be exploited to anticipate the short-term direction of the S&P 500 index. Twelve predictive architectures are implemented and compared under identical experimental conditions, spanning the entire methodological arc of quantitative finance: three classical econometric models that constitute the discipline's traditional toolkit, an ARIMAX specification for the conditional mean and the GARCH and EGARCH models for the conditional variance; three penalised linear methods at the frontier between econometrics and statistical learning, Ridge, Lasso and Logistic Regression; four non-linear Machine Learning algorithms, Random Forest, Gradient Boosting, XGBoost and Support Vector Machines; and two Artificial Intelligence architectures designed for hierarchical and sequential patterns, a feed-forward Artificial Neural Network and a Long Short-Term Memory recurrent network. The confrontation between these four generations of models on identical data is itself a central contribution: it measures, holding information constant, what each successive wave of methodological innovation actually adds when the target is as hostile as the daily direction of an efficient market.

The chapter proceeds as follows. Section 4.1 formalises the prediction problem and the construction of the feature space. Section 4.2 describes the chronological data-partitioning scheme, the hyperparameter-tuning protocol and the battery of evaluation metrics. Sections 4.3 to 4.6 present, family by family, the methodological foundations and the empirical results of the econometric, penalised-linear, non-linear and deep-learning models respectively. Section 4.7 conducts a systematic comparative analysis of the twelve architectures, and

Section 4.8 draws the implications of the evidence for data-driven financial decision making, connecting the results back to the central thesis of this work: that in modern finance, informational advantage, not merely capital, is the decisive battleground.

4.1 Problem definition and predictive objectives

The first methodological decision of any predictive exercise concerns the precise definition of the quantity to be forecast. Forecasting the level of an equity index is notoriously ill-posed: price series are non-stationary, dominated by stochastic trends, and their levels carry little economic meaning for a trader whose profit depends on changes rather than levels. Following standard practice in empirical asset pricing, the raw closing prices of the S&P 500 are therefore transformed into continuously compounded daily returns. Denoting by P_t the closing value of the index on trading day t , the daily log-return is defined as

$$r_t = \ln(P_t / P_{t-1}) \quad (4.1)$$

The logarithmic transformation removes the unit-root behaviour of the price level, makes multi-period returns additive, and renders the distribution of daily changes approximately symmetric around zero. Figure 4.4 examines this distribution directly: the histogram towers above the fitted normal density at the centre and extends far beyond it in the tails, and the quantile–quantile plot bends sharply away from the reference line at both extremes. The numbers behind the picture are in Table 4.1: the daily return has a standard deviation of 1.14 per cent but a minimum of -12.77 per cent and an excess kurtosis of 16.9, the COVID-19 crash of March 2020 produced single days more than eleven standard deviations from the mean, events that a Gaussian world would rule out over horizons longer than the age of the universe. Fat tails, negative skewness and the volatility clustering visible in Figure 4.3 are the defining stylised facts of the data; they motivate, respectively, the Student-t-robust design of the evaluation, the asymmetric EGARCH specification of Section 4.3.3, and the inclusion of volatility itself among the

predictors.

Rather than attempting to predict the magnitude of the next-day return, a task in which even sophisticated models rarely outperform a naive constant forecast, the problem is cast as a binary classification exercise. The target variable indicates whether the market will close higher or lower on the following trading day:

$$y_{t+1} = 1 \text{ if } r_{t+1} > 0, \quad y_{t+1} = 0 \text{ otherwise} \quad (4.2)$$

This directional formulation has a direct operational interpretation, a signal of one corresponds to a long position opened at the close of day t , a signal of zero to staying flat, and a solid theoretical foundation. Christoffersen and Diebold (2006) demonstrate that sign predictability and mean predictability are logically distinct properties: even when the conditional mean of returns is entirely unforecastable, as the efficient-market benchmark implies, the probability of a positive return remains forecastable whenever volatility is, because that probability depends on the ratio between the (small, positive) drift and the (highly persistent) conditional standard deviation. Since volatility is by far the most robustly predictable object in empirical finance, directional forecasting is the channel through which the well-documented dynamics of risk can be converted into statements about the likely sign of tomorrow's move. This insight also dictates the composition of the model set: the GARCH family enters the comparison precisely because, under the Christoffersen–Diebold logic, a pure volatility model is already, implicitly, a direction model, and Section 4.3.2 will show both the power and the subtle limits of that implication.

The classification framing also has an empirical pedigree independent of its theoretical one. In a comparison that has become canonical, Leung, Daouk and Chen (2000) evaluated level-forecasting and direction-forecasting versions of the same model families on the major international indices and found that the directional versions systematically produced both better statistical performance and higher trading profits: the sign of the return is a coarser but more learnable object than its magnitude, because the discretisation absorbs a large share of the

irreducible noise that magnitude models must fight observation by observation. The choice also aligns the estimation criterion with the economic decision, a long/flat investor acts on the sign, not the size, of the expected move, and it equips the study with the full, mature toolkit of classification metrics deployed in Section 4.2, several of which (precision on bullish calls, the threshold-free AUC) have no clean counterpart in the regression setting. Casting direction as the target, finally, puts all twelve models on strictly common ground: the econometric variance models, the penalised regressions, whose continuous forecasts are converted through their sign, and the native classifiers are scored by identical yardsticks on identical days, which is the precondition for the generational comparison at the heart of the chapter.

The predictive information set is deliberately restricted to variables that capture the three dimensions at the heart of this research: market memory, collective fear and structural geopolitical uncertainty, all measured at or before the close of day t so that the exercise is free of look-ahead contamination. Market memory enters through lagged log-returns at horizons of one, two, three and five trading days, two realised-volatility measures (the five- and ten-day rolling standard deviations of returns) and a five-day rolling mean acting as short-term momentum. Collective fear enters through the level and daily change of the VIX and of the Equity Market Uncertainty index; the latter, whose distribution is extremely right-skewed (skewness 4.05, maximum 1,922 against a median of 59), is stabilised through the transformation

$$EMU_t^* = \ln(1 + WLEMUINDXD_t) \quad (4.3)$$

together with lags of both gauges at the same four horizons. Structural uncertainty enters through the four binary dummies BREXIT, COVID, UKRAINEWAR and ISRAELHAMAS, which mark the acute windows of the geopolitical shocks identified in Chapter 3. The complete set comprises twenty-seven predictors. The two fear gauges deserve a theoretical remark, because they measure conceptually different objects: the VIX is a market-implied quantity, extracted by the CBOE

from the cross-section of S&P 500 option prices, and represents the risk-neutral expectation of the index's volatility over the following thirty days, the priced fear of investors who commit capital to their views; the EMU index of Baker, Bloom and Davis is a text-based measure that counts, day by day, the frequency of newspaper articles jointly discussing the economy, uncertainty and equity markets, the diffused fear of the public narrative, whether or not it is yet embedded in option prices. Feeding the models with both equips them with a stereoscopic view of collective psychology, operationalising the argument of Chapter 2 that market dynamics emerge from the interaction between trading infrastructures and information ecosystems.

Table 4.1, Descriptive statistics of the core variables, daily observations, June 2016 – June 2026.

	Mean	Median	Std. dev.	Min	Max	Skew.	Kurt.
S&P 500	4006.75	3900.96	1397.32	2085.18	7609.78	0.64	-0.57
VIX	18.57	16.88	7.33	9.14	82.69	2.56	12.56
EMU index	104.66	59.04	139.67	4.94	1922.1	4.05	27.37
Log-return (%)	0.05	0.08	1.14	-12.77	9.09	-0.69	16.88

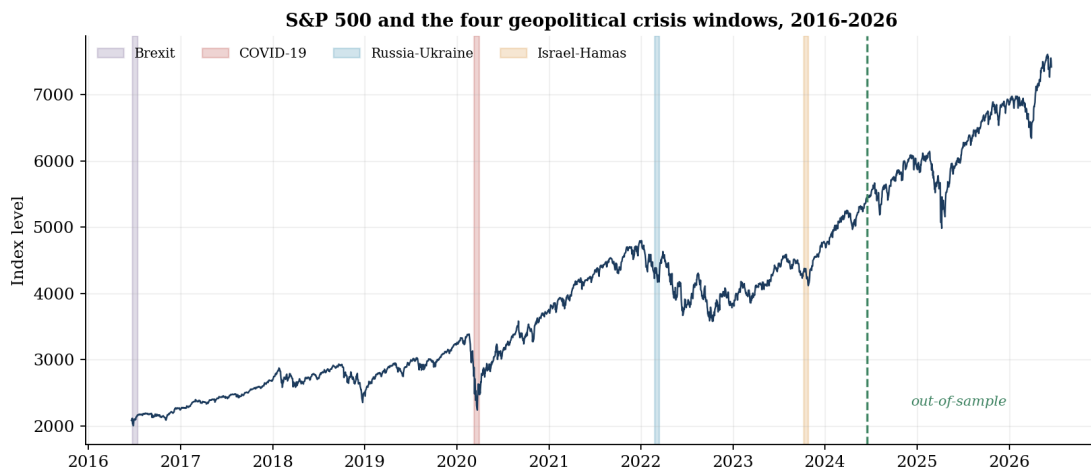


Figure 4.1, S&P 500 daily closing price with the four geopolitical crisis windows shaded; the

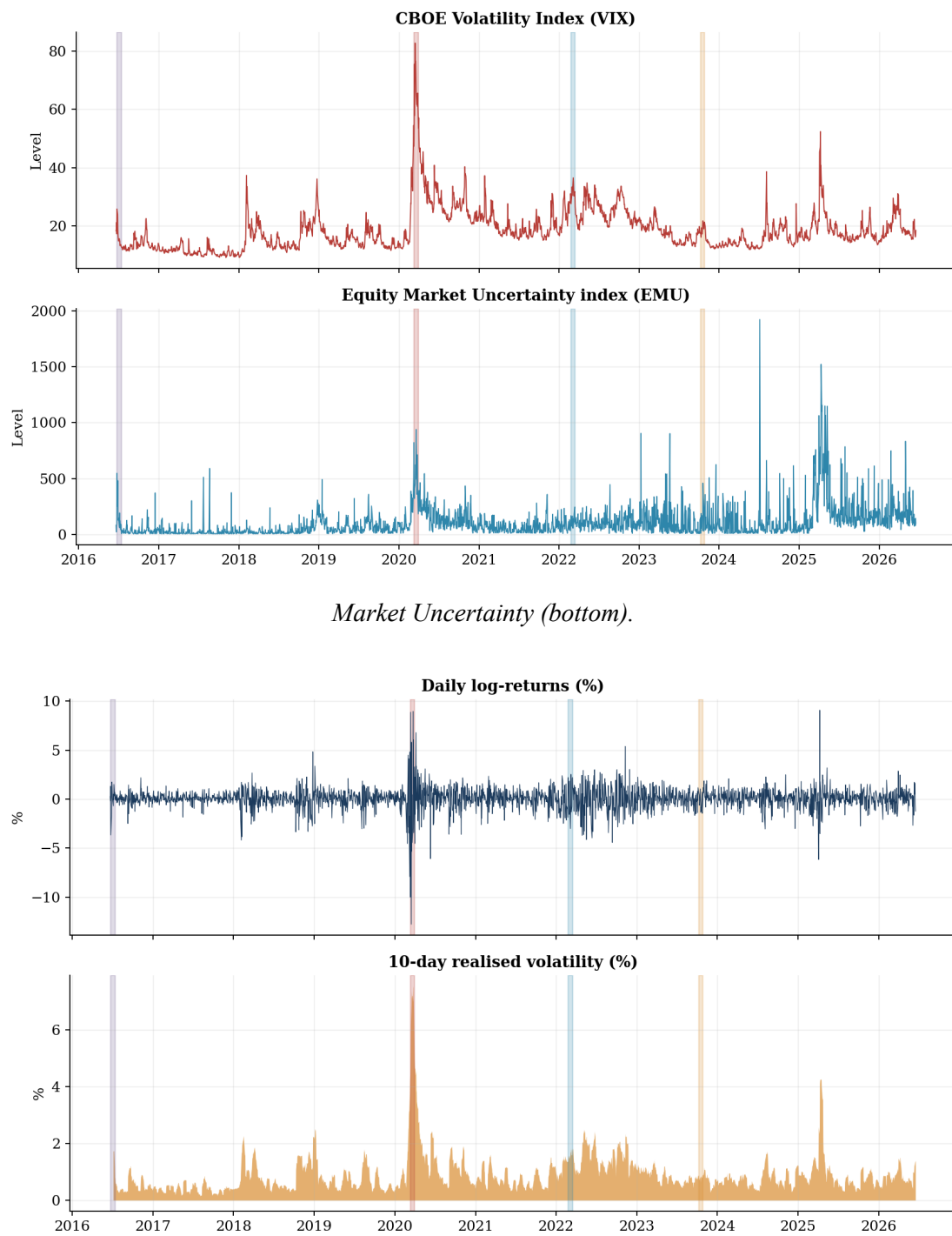
dashed line marks the beginning of the out-of-sample period.

Figure 4.1 places the four event windows on the price trajectory of the index and makes their heterogeneity immediately visible, while Table 4.2 quantifies it. During the fifteen COVID trading days the index lost on average 0.73 per cent per day with a daily volatility of 6.51 per cent, more than six times the unconditional level, while the average VIX of 65.6 and the average uncertainty index of 472 dwarf their non-crisis means of 18.2 and 102. The Ukraine window combines elevated volatility (1.71 per cent) with a mildly positive average return, reflecting the violent two-sided swings of that period; the Israel– Hamas window displays a milder but persistent negative drift; and the Brexit window that survives lag construction captures the sharp rebound that followed the referendum shock. This heterogeneity is analytically important: the dummies do not proxy a single homogeneous 'crisis effect' but four distinct uncertainty regimes, whose informational signatures the models must disentangle, or, as the results will show, whose content the continuous fear gauges have typically already absorbed.

Table 4.2, Market behaviour inside the geopolitical crisis windows versus ordinary trading days (modelling sample).

Event window	Start	Days	Mean ret. (%)	Vol. (%)	Mean VIX	Mean EMU
Brexit referendum	2016-07-05	8	0.36	0.65	13.9	85.2
COVID-19 crash	2020-03-11	15	-0.73	6.51	65.6	472.4
Russia-Ukraine war	2022-02-24	15	0.21	1.71	31.2	121.8
Israel-Hamas conflict	2023-10-09	15	-0.3	0.83	19.0	140.0
Non-crisis days	-	2449	0.06	1.03	18.2	102.2

Figure 4.2, The two fear gauges: option-implied volatility (VIX, top) and news-based Equity M



Market Uncertainty (bottom).

Figure 4.3, Daily log-returns (top) and ten-day realised volatility (bottom): volatility clustering and crisis spikes.

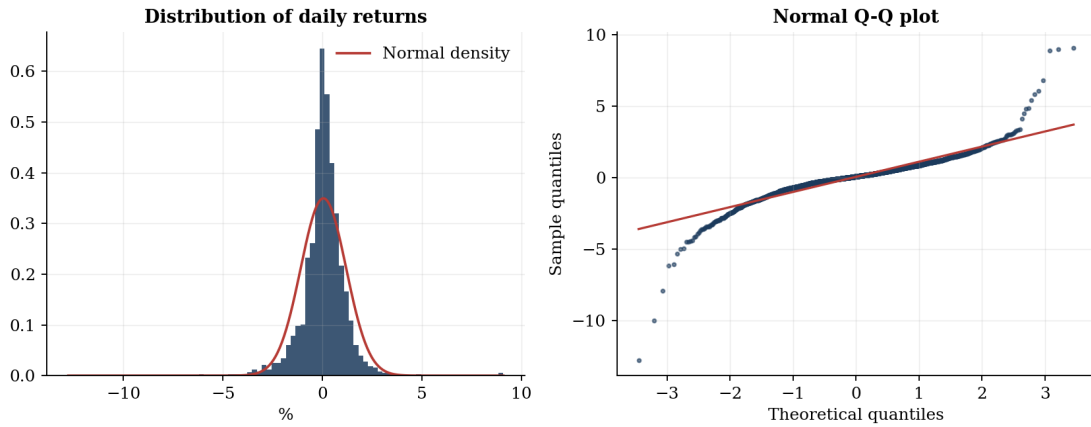


Figure 4.4, Empirical distribution of daily returns against the normal benchmark (left) and normal Q - Q plot (right): fat tails on both sides.

A final look at the dependence structure completes the exploratory analysis. Figure 4.5 reports the correlation matrix of the core contemporaneous predictors. Three features anticipate the difficulties ahead. First, the contemporaneous correlation between the daily return and the change in the VIX is strongly negative (-0.79): fear rises when prices fall, the asymmetric volatility effect that the EGARCH model of Section 4.3.3 is designed to capture parametrically. Second, the volatility block is heavily collinear (the two realised measures correlate at 0.90, and both correlate above 0.8 with the VIX level), a configuration that destabilises unpenalised regression coefficients and constitutes the classic habitat of the Ridge estimator. Third, and most importantly, the correlations between lagged information and the future return are minute, the faint trace of weak-form efficiency. The predictive signal, if it exists, is small, partially non-linear and buried in noise; extracting it under these conditions is precisely the comparative test to which the twelve models are subjected.

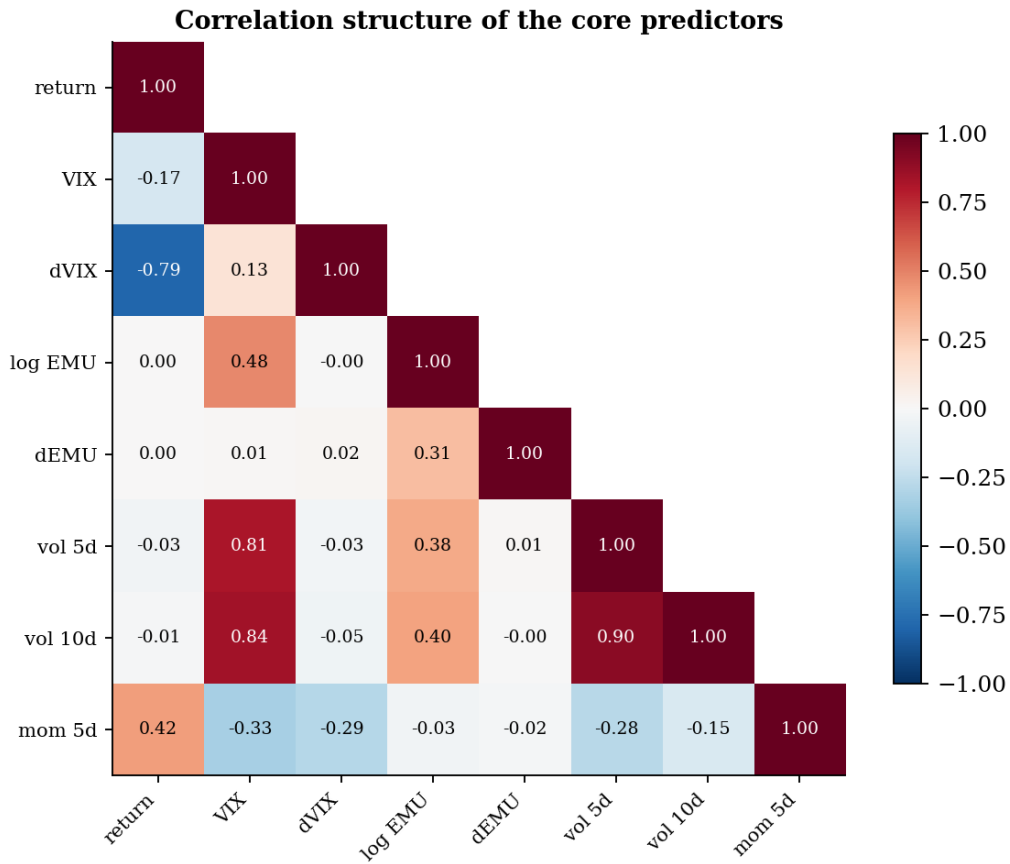


Figure 4.5, Correlation structure of the core predictors (lower triangle).

4.2 Data partitioning, tuning protocol and evaluation metrics

The credibility of any claim of predictability rests entirely on the design of the validation protocol. Financial time series impose a constraint that generic Machine Learning practice often ignores: the arrow of time. A random shuffling of observations into training and test folds would allow the models to learn from the future when predicting the past, generating a look-ahead bias that inflates measured performance and destroys operational relevance. The dataset is therefore partitioned chronologically: the first 80 per cent of the sample, 2,001 trading days from July 2016 to June 2024, constitutes the training set, while the most recent 20 per cent, 501 trading days from June 2024 to June 2026, is held out as a strictly out-of-sample test set that no model sees during estimation. Figure 4.1 already marked this frontier on the price trajectory; Table 4.3 reports the composition of

the partition. Two properties deserve emphasis. First, all four geopolitical crisis windows fall inside the training period: the models learn how markets behave under acute stress and are evaluated on an unseen two-year future, mimicking the predicament of a risk manager who calibrates on past crises and deploys on an unknown tomorrow. Second, the class distribution drifts across the frontier, upward days are 54.3 per cent of the training set but 57.3 per cent of the test set, reflecting the strong bull market of 2024–2026, a drift that is handled through balanced class weights in the classifiers and through evaluation metrics that remain informative under imbalance.

Table 4.3, Chronological train/test partition.

Partition	Obs.	From	To	Up-days
Training set	2001	2016-07-05	2024-06-14	54.3%
Test set	501	2024-06-17	2026-06-16	57.3%

Within the training window, hyperparameters are selected by five-fold expanding-window cross-validation: the training sample is split into five consecutive blocks, and each candidate configuration is estimated on the first k blocks and evaluated on block $k+1$, respecting the temporal ordering inside the tuning loop exactly as the outer protocol respects it between training and test. This is the time-series analogue of standard cross-validation, and it is preferred to the random k -fold scheme for the two reasons developed at length in the statistical-learning literature: serial dependence lets information leak across randomly drawn folds through the persistence of volatility, and structural change makes randomly shuffled folds answer a question, how well would the model predict a randomly drawn day of the past?, that no investor faces. Table 4.4, at the opening of the model sections, documents the grid explored for every model and the settings held fixed; the selected values are reported in the text of each subsection. Models without tunable regularisation (the GARCH family, the neural architectures with fixed topology)

are estimated once with the conventions stated in the table.

A second source of potential contamination is the standardisation of the predictors. Distance-based and gradient-based learners, the Support Vector Machine, the penalised regressions, the neural networks, are sensitive to the scale of the inputs. Each feature is therefore transformed into a z-score,

$$\tilde{x}_{ij} = (x_{ij} - \mu_j) / \sigma_j \quad (4.4)$$

where the mean and standard deviation of each predictor are estimated exclusively on the training set and applied unchanged to the test observations: estimating the scaling on the full sample would leak the future distribution of volatility into the training phase, a subtle and well-documented form of data snooping that this protocol rules out by construction.

Model quality is assessed through five complementary metrics, all computed on the out-of-sample test set. Let TP, TN, FP and FN denote the numbers of true positives (correctly predicted up-days), true negatives, false positives and false negatives. The four threshold-based metrics are

$$Accuracy = (TP + TN) / (TP + TN + FP + FN) \quad (4.5)$$

$$Precision = TP / (TP + FP) \quad (4.6)$$

$$Recall = TP / (TP + FN) \quad (4.7)$$

$$F1 = 2 \cdot (Precision \cdot Recall) / (Precision + Recall) \quad (4.8)$$

computed at the conventional cut-off of one half: a day is classified as 'up' whenever the model assigns it a probability, or, for the regression-based models, a predicted return, above the neutral point. Precision measures the reliability of the bullish signals issued; recall the share of actual rallies captured; the F1-score their harmonic mean. The fifth metric, the area under the Receiver Operating Characteristic curve, evaluates the quality of the predictive scores themselves, independently of any threshold: the ROC curve traces the true-positive rate against the false-positive rate as the decision threshold sweeps its range, and the area under

it equals the probability that a randomly chosen up-day receives a higher score than a randomly chosen down-day,

$$AUC = P(\hat{s}(x^+) > \hat{s}(x^-)) \quad (4.9)$$

An AUC of one half denotes zero ranking ability; in daily equity-direction forecasting, where the efficient-market hypothesis predicts near-unforecastability, even values of 0.52–0.55 can be economically meaningful if stable. The joint use of threshold-dependent and threshold-free metrics is not redundancy but a designed safeguard: in an imbalanced sample, a degenerate model that always predicts the majority class attains deceptively high accuracy, recall and F1 while possessing no discriminative power whatsoever, a pathology that only the AUC exposes. The safeguard will prove essential: three of the twelve models, including the two with the highest headline accuracy, will turn out to be exactly of this kind. In financial applications the costs of the two errors are, moreover, rarely symmetric, a false bullish call costs capital, a missed rally only foregone gain, and decision theory prescribes in such cases a threshold equal to the cost ratio rather than one half; the fitted scores serve any such threshold without re-estimation, which is why the threshold-free AUC, rather than accuracy at the symmetric operating point, is adopted as the primary ranking criterion throughout the chapter.

The protocol also guards against a subtler enemy than look-ahead bias: selection among models after observing test results. When many configurations are evaluated on the same sample, the best of them will look impressive by chance alone, the multiple-testing problem that the financial literature knows as data snooping, and that inflates a nontrivial share of published anomalies. The discipline adopted here is therefore twofold. Within each model class, exactly one configuration reaches the test set: the one selected by the expanding-window cross-validation inside the training window, so that the test data are touched once per model and never inform any choice. Across model classes, all twelve pre-specified architectures are reported in full, the failures alongside the successes, the degenerate solutions alongside the informative ones, rather than the flattering

subset a results-driven narrative would curate. This completeness is what gives the comparative conclusions of Section 4.7 their evidential force: the claim will not be that some model beats chance (a claim twelve attempts would eventually produce by luck), but that a specific, theoretically motivated ordering of model families emerges, and that the ordering is exactly the one the bias–variance analysis predicts *ex ante*.

One last remark concerns the architecture of the comparison itself. Because all twelve models receive the same predictors, the same training window, the same tuning discipline and the same evaluation battery, the design functions as a controlled experiment in which the only treatment that varies is the model class. Differences in outcomes are therefore attributable to differences in inductive bias, the assumptions each architecture imposes on the shape of the predictable component, and not to differences in information, preprocessing or protocol, the confounders that make cross-study comparisons in this literature notoriously unreliable. The twelve treatments span four methodological generations and, more importantly, four orders of magnitude of effective capacity: from the five estimated parameters of a GARCH(1,1) to the several thousand weights of the recurrent network. Holding information constant while sweeping capacity across this range is what allows the chapter to answer not only the applied question, which model should a practitioner use?, but the deeper one that motivates the thesis: in a market near informational efficiency, is the binding constraint on prediction the sophistication of the algorithm or the content of the data? The comparative results of Section 4.7 will return an unambiguous answer.

4.3 Econometric benchmarks: ARIMAX, GARCH and EGARCH

Any assessment of what Machine Learning adds to financial forecasting must begin from what it seeks to replace. This section estimates the three workhorses of classical financial econometrics on the same data and under the same protocol as the learning algorithms of the later sections: an ARIMAX model for the

conditional mean of returns, and the GARCH and EGARCH models for their conditional variance. Their inclusion serves a double purpose. Methodologically, they are the disciplined baseline, parametric, parsimonious, estimated by (quasi-)maximum likelihood, interpretable coefficient by coefficient, against which the flexibility of the algorithmic models can be priced. Substantively, they embody two competing hypotheses about where the predictable component of market direction resides: in the serial structure of returns themselves (the ARIMAX view) or in the dynamics of risk (the GARCH view, activated through the sign-probability logic of Christoffersen and Diebold introduced in Section 4.1). Table 4.4 summarises the estimation conventions for all twelve models of the study.

Table 4.4, Model specifications: tuned parameters and fixed settings; all tuning by 5-fold expanding-window cross-validation on the training window.

Model	Tuned parameters (grid)	Fixed settings
ARIMAX	(p, q) = (1, 1); CSS estimation	6 exogenous regressors, d = 0
GARCH(1,1)	-	Gaussian MLE, constant mean
EGARCH(1,1)	-	Gaussian MLE, leverage term
Ridge	lambda: 30 values, 10^{-1} - 10^4	standardised features
Lasso	lambda: 30 values, 10^{-6} - $10^{-2.5}$	standardised features
Logistic	C in {0.01, 0.1, 1, 10}	balanced class weights
Random forest	min leaf in {5, 20, 50}	400 trees, depth 6, m = sqrt(p)
Gradient boosting	depth in {2, 3}	300 trees, nu = 0.05
XGBoost	nu in {0.01, 0.05}; depth in {2, 3}	300 rounds, L2 = 1, rebalanced
SVM	C in {0.1, 1, 10}	RBF kernel, Platt calibration
ANN (MLP)	-	64-32-16, ReLU, Adam, early stop
LSTM	-	32 units, 10-day window, Adam

4.3.1 ARIMAX

The ARIMAX model extends the classical autoregressive–moving-average representation of a stationary series with exogenous regressors. Since log-returns are already stationary (the integration order is zero), the specification adopted is an ARMAX(1,1): the next-day return is explained by its own most recent value, by the most recent forecast error, and by a vector of six exogenous market-state variables, the VIX level and its daily change, the log uncertainty index and its change, the five-day realised volatility and the five-day momentum, all observed at the close of day t :

$$r_{t+1} = c + \phi r_t + \theta \varepsilon_t + \gamma^T x_t + \varepsilon_{t+1} \quad (4.10)$$

The autoregressive coefficient ϕ captures the direct persistence of returns; the moving-average coefficient θ captures the propagation of yesterday's surprise; the vector γ measures the marginal predictive content of the fear and momentum variables for the conditional mean. Estimation proceeds by conditional least squares: the innovations are computed recursively from the start of the training sample and the parameters chosen to minimise the sum of squared one-step-ahead errors,

$$(\hat{c}, \hat{\phi}, \hat{\theta}, \hat{\gamma}) = \operatorname{argmin} \sum_t \varepsilon_{t+1}^2 \quad (4.11)$$

a criterion asymptotically equivalent to Gaussian maximum likelihood. On the training window the estimation delivers $\hat{\phi} = -0.44$ and $\hat{\theta} = +0.35$: a negative autoregressive root partially offset by a positive moving-average term, the classical ARMA signature of short-horizon mean reversion contaminated by microstructure noise. The two coefficients largely cancel, their sum, which governs the persistence of the predictable component, is close to -0.09 , so the model's forecasts oscillate in a narrow band around the unconditional mean, tilted day by day by the exogenous block. Out of sample, the directional translation of these forecasts (long when $\hat{r}_{t+1} > 0$) achieves an accuracy of 51.5 per cent and an AUC of 0.512: a genuine but frail edge, statistically indistinguishable from the stronger linear

learners of Section 4.4 in ranking terms, and an instructive first datum, the oldest model in the study already extracts roughly half of the total predictability that any model will manage to find.

The specification search follows the Box–Jenkins methodology that has organised applied time-series work since the 1970s: identification of the orders from the correlation structure of the data, estimation, and diagnostic checking of the residuals. Identification here is undemanding, the autocorrelations of daily returns die out essentially at the first lag, pointing to low orders, and an information-criterion scan over $(p, q) \in \{0, 1, 2\}^2$ confirms the parsimonious $(1, 1)$ choice, penalising the marginal likelihood gains of richer dynamics with their additional parameters exactly as the bias–variance logic of this chapter prescribes in its econometric dialect. Two technical conditions discipline the estimates: stationarity requires $|\phi| < 1$ and invertibility $|\theta| < 1$, both comfortably satisfied, the latter guaranteeing that the unobserved innovations can be recovered recursively from the observed data, which is what makes the conditional-least-squares recursion of (4.11) well defined. Diagnostics on the training residuals complete the classical loop with an instructive verdict: the residual autocorrelations are negligible (the mean equation has absorbed what linear memory exists) but the autocorrelations of the squared residuals are large and slowly decaying. The linear model is correctly specified for the mean and violently misspecified for the variance, which is precisely the empirical observation that motivated the ARCH revolution, and the natural bridge to the models of the next two subsections.

4.3.2 GARCH(1,1)

The GARCH family shifts the object of the model from the mean to the variance of returns. In the benchmark GARCH(1,1) of Bollerslev, the return is decomposed into a constant mean and an innovation whose conditional variance evolves as a weighted combination of its own past and of the squared surprise:

$$r_t = \mu + \varepsilon_t, \quad \varepsilon_t = \sigma_t z_t, \quad z_t \sim N(0, 1) \quad (4.12)$$

$$\sigma_t^2 = \omega + \alpha \varepsilon_{t-1}^2 + \beta \sigma_{t-1}^2 \quad (4.13)$$

The parameter α measures the immediate impact of news (yesterday's squared shock), β the memory of the variance process, and their sum the persistence of volatility; the unconditional variance exists when $\alpha + \beta < 1$ and equals

$$\sigma^2 = \omega / (1 - \alpha - \beta) \quad (4.14)$$

Gaussian maximum likelihood on the training returns yields $\hat{\alpha} = 0.180$, $\hat{\beta} = 0.797$ and a persistence of 0.977, the canonical fingerprint of daily equity data, in which a volatility shock decays with a half-life of about a month. Figure 4.6 displays the fitted conditional volatility over the full decade against the ten-day realised measure: the model tracks every stress episode of the sample, exploding to annualised values close to 90 per cent in March 2020 and mean-reverting through the calm of 2017 and 2024–2025, with the parameters frozen at their training-window estimates throughout the out-of-sample segment. As a description of risk dynamics, the sixty-dollar econometric model is essentially unimprovable, a fact worth registering before the far more expensive architectures ahead.

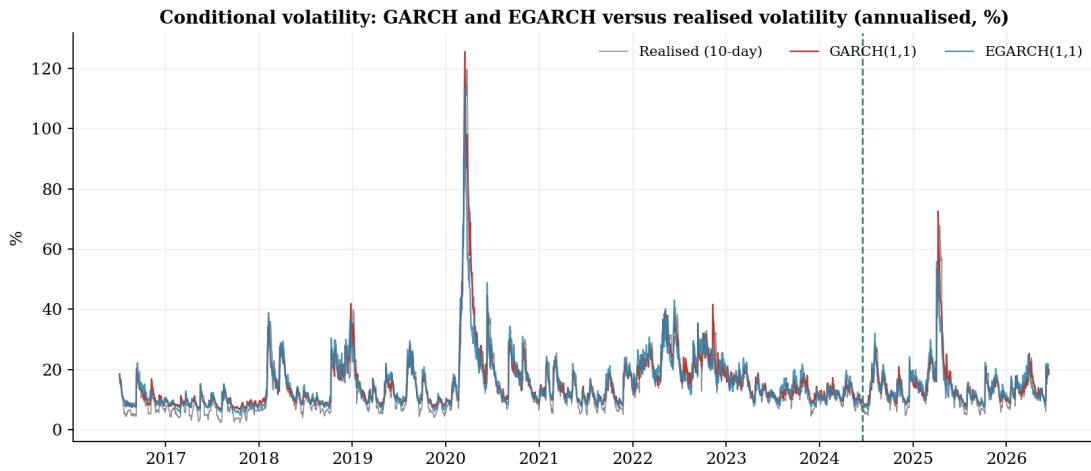


Figure 4.6, Conditional volatility from $GARCH(1,1)$ and $EGARCH(1,1)$ versus ten-day realised volatility (annualised). Parameters are estimated on the training window only; the dashed line marks the out-of-sample frontier.

The intellectual history of the model is worth a sentence, because it marks the moment econometrics accepted that risk, not return, is the forecastable dimension

of financial markets. Engle's original ARCH specification of 1982, conceived for inflation, and honoured with the Nobel Prize two decades later, made the conditional variance a function of past squared shocks alone; Bollerslev's 1986 generalisation added the lagged variance term, achieving with two parameters the long memory that ARCH needed many lags to mimic. The (1,1) order adopted here is not a simplification of convenience: an extensive comparison literature has established that the plain GARCH(1,1) is extraordinarily difficult to beat for one-step-ahead volatility forecasting of equity indices, and higher orders add parameters without adding accuracy. Two further properties recommend it as a benchmark. First, robustness: the Gaussian likelihood maximised here retains consistency for the variance parameters even when the true innovations are fat-tailed, the quasi-maximum-likelihood property, so the estimates are trustworthy despite the kurtosis documented in Table 4.1. Second, the model endogenously generates fat tails: even with normal innovations, the mixing of high- and low-variance regimes produces an unconditional distribution with excess kurtosis, reproducing qualitatively the empirical shape of Figure 4.4. A model of five parameters that captures volatility clustering, tail thickness and mean reversion of risk is the appropriate yardstick for everything more complicated.

The conversion of a variance model into a direction forecast follows the Christoffersen–Diebold logic exactly. Under the Gaussian assumption, the probability of an upward day is

$$P(r_{t+1} > 0 \mid \Omega_t) = \Phi(\mu / \sigma_{t+1}) \quad (4.15)$$

a quantity that rises in calm markets, when the small positive drift looms large relative to the noise, and falls towards one half as volatility explodes. The formula exposes with complete transparency both the strength and the structural limit of the volatility route to direction. The strength: the probability in (4.15) is genuinely time-varying and highly persistent, inheriting the forecastability of σ_{t+1} . The limit: since the estimated drift is positive ($\hat{\mu} > 0$), the probability never crosses below one half, and the model, classified at the symmetric threshold, predicts an upward day

unconditionally. Its test-set confusion matrix is the degenerate corner (287 true positives, 214 false positives, zero negatives of either kind), its accuracy of 57.3 per cent is exactly the base rate, and its recall of 100 per cent is vacuous. The AUC, immune to thresholds, delivers the substantive verdict on the ranking: 0.463, significantly below one half. The inversion is economically meaningful: the model ranks calm days as the most bullish, but in this particular test period the largest positive returns clustered in the turbulent phases (the sharp rebounds of the 2025 corrections), so the volatility-based ordering of bullishness ran systematically backwards. GARCH knows, superbly, how risky tomorrow will be; it does not know which side of zero tomorrow will fall on.

4.3.3 EGARCH(1,1)

The exponential GARCH of Nelson removes two rigidities of the symmetric model: by driving the logarithm of the variance, it dispenses with positivity constraints on the parameters, and by including the signed standardised shock alongside its absolute value, it allows negative and positive surprises of equal size to move volatility differently, the leverage effect anticipated by the strong negative return– Δ VIX correlation of Figure 4.5. The variance equation becomes

$$\ln \sigma_t^2 = \omega + \alpha (|z_{t-1}| - E|z|) + \gamma z_{t-1} + \beta \ln \sigma_{t-1}^2 \quad (4.16)$$

where z_{t-1} is the lagged standardised innovation and a negative γ implies that bad news raises tomorrow's volatility more than good news of the same magnitude. The training-window estimates are unambiguous: $\hat{\gamma} = -0.141$, with $\hat{\alpha} = 0.259$ and a log-variance persistence $\hat{\beta} = 0.955$. The asymmetry is displayed in Figure 4.7 through the news impact curves of the two models: the GARCH parabola treats a -3 and a $+3$ standardised shock identically, while the EGARCH curve is visibly rotated, assigning a markedly higher next-day variance to the negative shock. This is the econometric portrait of the fear asymmetry that runs through this entire thesis, markets do not merely react to news, they react roughly twice as strongly to threatening news, and it is captured here by a single interpretable parameter

estimated in seconds.

Two economic mechanisms compete to explain the asymmetry that γ measures, and the distinction matters for interpretation. The original 'leverage' story of Black is mechanical: a fall in the equity price raises the firm's debt-to-equity ratio, making the equity a claim on a more leveraged, hence riskier, asset. The 'volatility feedback' story runs the causality the other way: if anticipated volatility rises, required returns rise, and prices must fall today to deliver them, so bad volatility news causes negative returns rather than the reverse. At the daily frequency and index level of this study, the behavioural reading developed in Chapter 2 is at least as compelling: losses trigger fear, fear triggers de-risking, forced selling and hedging flows, and those flows generate further volatility, an amplification loop through the psychology and the plumbing of markets that a symmetric model cannot represent. The EGARCH parameter does not adjudicate among these mechanisms; it measures their combined footprint, and its estimated magnitude here, negative shocks raising next-day variance roughly as much as positive shocks twice their size, is consistent with the international evidence for equity indices. It is also the parametric twin of a model-free observation already in the data: the -0.79 correlation between daily returns and VIX changes in Figure 4.5 is the same asymmetry seen through the options market rather than through a filter.

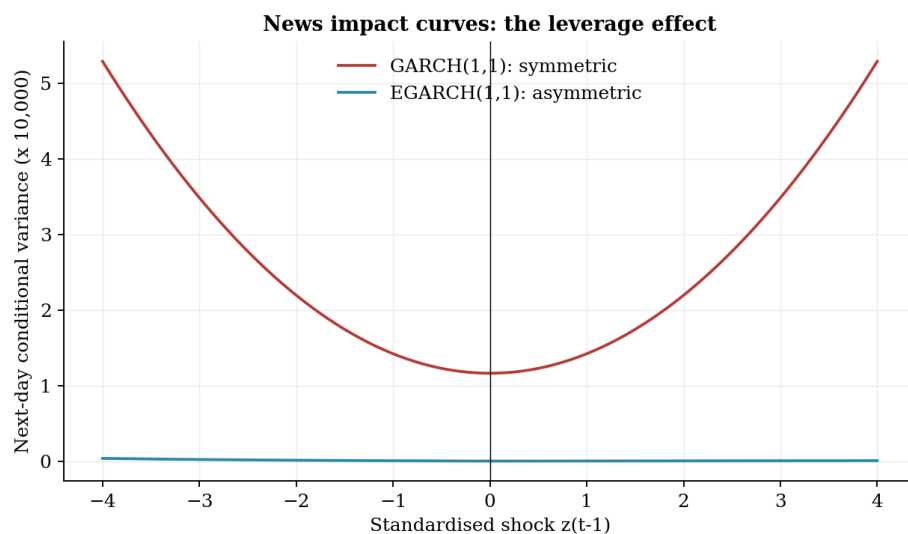


Figure 4.7, News impact curves: next-day conditional variance as a function of the standardised shock. The EGARCH asymmetry (leverage effect, $\gamma < 0$) assigns higher volatility to negative surprises.

For directional purposes, however, the refinement changes nothing structural: the sign probability remains $\Phi(\mu/\sigma_{t+1})$ with a positive estimated drift, the classification remains degenerately bullish, and the test-set AUC (0.463) coincides with that of the symmetric model to the third decimal. The twin failure of the two variance models is the cleanest single finding of the chapter, because it separates two concepts that casual discussion conflates: predicting risk and predicting direction. The GARCH family solves the first problem almost perfectly and contributes to the second only the unconditional drift. Everything the remaining nine models attempt can be read as the search for conditioning information, in past returns, in sentiment, in cross-variable interactions, capable of moving the probability in (4.15) to the correct side of one half on the correct days.

4.4 Penalized linear models: Ridge, Lasso and Logistic Regression

The second family occupies the methodological frontier where econometrics and statistical learning meet: linear models estimated not by unconstrained least squares or maximum likelihood, but by penalised criteria that deliberately bias the coefficients towards zero in exchange for a reduction in variance. The exchange is governed by the bias–variance decomposition of prediction error,

$$E[(y - \hat{f}(x))^2] = \text{Bias}^2[\hat{f}(x)] + \text{Var}[\hat{f}(x)] + \sigma^2$$

whose logic dominates this entire study: in an environment where the irreducible noise σ^2 dwarfs the systematic component, the variance of a flexible estimator is the binding constraint, and shrinkage, accepting a small systematic distortion to stabilise the estimate, is not a compromise but the optimal strategy. The twenty-seven predictors of this study, with their heavily collinear volatility block, are the natural habitat of these methods. Two members of the family attack the regression version of the problem, predicting the next-day return itself and converting its sign

into the directional call; the third, Logistic Regression, attacks the classification problem directly.

A word on the two routes the family takes to the common directional target. The regression route, predict the return, act on its sign, uses strictly more information from the training data than the classification route, because the magnitude of past returns disciplines the fit: a model that explains a +3 per cent day and one that explains a +0.03 per cent day are rewarded differently, whereas a classifier sees both simply as 'up'. The cost is a mismatch between the estimation loss (squared error on magnitudes, dominated by the largest moves) and the evaluation loss (directional correctness, in which all days weigh equally); the classification route accepts the information loss in exchange for optimising the object actually scored. Which trade prevails is an empirical question, and the twelve-model design answers it cleanly, since both routes are represented within the same penalised-linear family and can be compared holding everything else fixed. The answer, anticipated here and documented below, is that the regression route wins the ranking metric, the magnitude information sharpens the ordering of days by bullishness, while the classification route wins the calibrated operating point, a division entirely consistent with what each criterion optimises. For the practitioner the lesson is that the choice of target formulation is itself a modelling decision of the same rank as the choice of algorithm, and a considerably cheaper one to get right.

4.4.1 Ridge regression

Ridge regression augments the least-squares criterion with a penalty proportional to the squared Euclidean norm of the coefficient vector:

$$\beta^R = \operatorname{argmin} \{ \sum_t (r_{t+1} - \beta_0 - \beta^T x_t)^2 + \lambda \sum_j \beta_j^2 \} \quad (4.17)$$

The tuning parameter λ traces a continuous path between ordinary least squares ($\lambda = 0$) and the degenerate constant forecast ($\lambda \rightarrow \infty$); its value is selected here over a thirty-point logarithmic grid by the expanding-window cross-validation of

Section 4.2, which chooses $\hat{\lambda} = 0.149$, a light but non-trivial shrinkage. The distinctive virtue of the L2 penalty is its treatment of correlated predictors: where ordinary least squares reacts to the near-collinearity of the volatility block by inflating offsetting coefficients of opposite sign, Ridge shrinks correlated variables towards a common, moderate contribution, stabilising the fitted signal. The estimator never sets coefficients exactly to zero, every predictor retains a voice, however faint, which is appropriate when, as the exploratory analysis suggested, the information is diffused in small amounts across many weakly informative variables rather than concentrated in a few strong ones.

The mechanics of the stabilisation are transparent in the closed-form solution, $\beta^R = (X^T X + \lambda I)^{-1} X^T y$: the penalty adds λ to every eigenvalue of the predictor covariance matrix before inversion. Collinearity is, in this algebra, the presence of near-zero eigenvalues, directions of the feature space along which the data contain almost no independent variation, and along which ordinary least squares therefore estimates with enormous variance, producing the notorious pattern of huge offsetting coefficients on correlated regressors. The ridge correction is surgical: directions with large eigenvalues (well-identified combinations, such as the general level of fear) are barely shrunk, while the ill-identified directions (the difference between two nearly identical volatility measures) are shrunk drastically. The estimator thus allocates its scepticism exactly where the data are least informative, a property with a Bayesian reading, since the ridge solution is the posterior mean under independent Gaussian priors centred at zero, the formal statement of the belief, eminently reasonable in efficient markets, that most predictors have effects close to nil. From the perspective of this study the eigenvalue view explains why the method is so well matched to the feature set of Section 4.1: a block of fear gauges that largely measure the same latent state is precisely a covariance matrix with a few dominant directions and many near-degenerate ones, the configuration in which ridge regression enjoys its largest advantage over unpenalised estimation.

The out-of-sample result is the first genuine surprise of the study: an accuracy of 53.9 per cent and, decisively, a ROC-AUC of 0.538, which no other model in the comparison will exceed. A shrunk linear regression of tomorrow's return on today's fear and momentum variables, a model estimable in milliseconds and interpretable at a glance, provides the best probability ranking of the entire twelve-model field. The composition of its signal anticipates the Lasso analysis of the next subsection, with which it agrees closely: negative loadings on recent momentum and short-run volatility, positive loadings on the lagged uncertainty index, the mean-reversion-modulated-by-fear narrative that will recur, with variations, in every successful model of this chapter.

4.4.2 Lasso regression

The Lasso replaces the squared norm of the Ridge penalty with the sum of absolute values:

$$\beta^L = \operatorname{argmin} \{ \sum_i (r_{t+1} - \beta_0 - \beta^T x_t)^2 + \lambda \sum_j |\beta_j| \} \quad (4.18)$$

The geometric consequence of this apparently minor change is qualitative: the L1 constraint region has corners on the coordinate axes, and the solution path hits those corners, setting coefficients exactly to zero. The Lasso therefore performs continuous variable selection, it does not merely shrink the predictor set, it prunes it, and the sequence in which variables enter as the penalty relaxes constitutes an importance ranking generated by the optimisation itself. Figure 4.8 displays this regularisation path: each curve tracks one coefficient as λ decreases from the fully sparse solution on the left towards least squares on the right. The order of entry is economically eloquent: the five-day momentum enters first and grows steadily negative, followed by the short-run realised volatility (negative) and by the lagged returns and lagged uncertainty index, the same hierarchy that the cross-validated selection ultimately retains. At the selected penalty ($\lambda = 5.3 \times 10^{-6}$, dashed vertical line) the model keeps twenty-four of the twenty-seven predictors, discarding three of the four crisis dummies, a data-driven confirmation of the redundancy argument

of Section 4.7: once the continuous fear gauges are observed, the binary event markers add almost nothing.

The selection mechanism is easiest to see in the orthonormal special case, where the lasso solution has a closed form: each least-squares coefficient is moved towards zero by exactly λ and truncated there, the soft-thresholding operator, $\beta_j^L = \text{sign}(\beta_j) \cdot \max(|\beta_j| - \lambda, 0)$. The formula exposes the division of labour between the two penalties with perfect clarity: ridge rescales every coefficient proportionally and never annihilates one, while the lasso subtracts a fixed toll that small coefficients cannot pay. Predictors whose marginal contribution falls below the toll are removed from the model entirely; the survivors are estimated with a uniform shrinkage bias. In the correlated, non-orthonormal world of actual financial features the closed form no longer holds, the solution is computed by coordinate descent along the path of Figure 4.8, but the qualitative behaviour persists, with one caveat the literature has documented carefully: among a group of highly correlated predictors the lasso tends to select one representative and discard the rest, so the identity of the chosen volatility measure, for instance, is partly arbitrary even when the importance of the volatility block is not. This is why the interpretation offered below leans on blocks of variables and on the agreement across models, not on the survival of any individual regressor. Within those limits, the pruning has a distinctly economic value in the context of this thesis: in a discipline where each additional data stream has a price, a procedure that identifies which streams earn their place in the information set, and does so as a by-product of estimation, is itself a data-valuation technology.

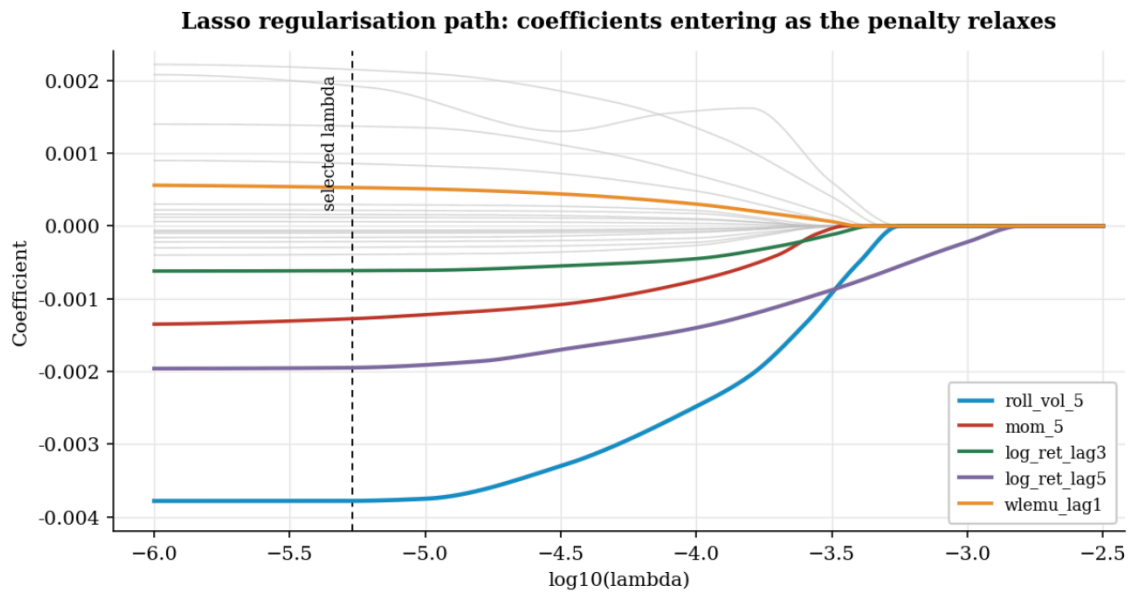


Figure 4.8, Lasso regularisation path on the training window: coefficients as functions of the penalty. Highlighted: the first and most persistent entrants; the dashed line marks the cross-validated λ .

Out of sample the Lasso edges past its L2 sibling on every threshold metric while matching its AUC of 0.538, and with an accuracy of 54.3 per cent and precision of 58.8 per cent it is ranked first overall. That the two penalised regressions share the summit is itself informative: their agreement indicates that the result is a property of the shrunk linear signal, not an artefact of a particular penalty geometry. The pattern of the selected coefficients tells the same story told by the path: after a week of abnormal gains the odds of a further rise decline (mean reversion); after turbulent days they decline further (volatility persistence); after days of elevated media uncertainty they rise marginally (uncertainty resolution, prices rebound as feared scenarios fail to materialise). It is a faint signal, but three independent estimation principles have now converged on it, and the strongest models of the remaining sections will do no better than rediscover it.

4.4.3 Logistic Regression

Logistic Regression attacks the classification problem directly, modelling the conditional probability of an upward day as a monotone transformation of a linear

score. Given the feature vector $x_t \in \mathbb{R}^{27}$, the model postulates

$$P(y_{t+1} = 1 \mid x_t) = \sigma(\beta_0 + \beta^T x_t) = 1 / (1 + e^{-(\beta_0 + \beta^T x_t)}) \quad (4.19)$$

where the sigmoid function maps the real line into a valid probability, and the parameters maximise the (penalised) log-likelihood of the training sample,

$$\ell(\beta) = \sum_t [y_{t+1} \ln p_t + (1 - y_{t+1}) \ln(1 - p_t)] - (1/2C) \sum_j \beta_j^2 \quad (4.20)$$

a strictly concave criterion whose inverse regularisation strength C is tuned on the grid of Table 4.4 (selected: $C = 1$, moderate shrinkage). Estimation employs balanced class weights, which rescale the two classes to equal aggregate mass so that the drifting base rate documented in Table 4.3 does not bias the estimator towards perpetual optimism. Because the predictors are standardised, the coefficients are directly comparable: each β_j measures the change in the log-odds of a rise induced by a one-standard-deviation move in predictor j . Figure 4.9 ranks the twelve largest. The picture is by now familiar, and its familiarity is the point: the largest coefficient is the five-day momentum (negative), followed by the short-run realised volatility (negative) and by the lagged uncertainty and return terms, replicating in the language of odds the signal that the penalised regressions expressed in the language of expected returns. The geopolitical dummies receive negligible weights, their rare activations absorbed by the continuous gauges that spike simultaneously.

Two properties explain the model's enduring centrality in finance beyond its role here. The first is the odds interpretation: rearranging (4.19) shows that the log-odds of a rise, $\ln[p/(1-p)]$, is exactly linear in the predictors, so that exponentiating a coefficient yields the multiplicative effect of a one-standard-deviation shock on the odds, a form of statement that translates directly into the language of risk committees. The second is institutional: logistic regression is the historical backbone of credit scoring and default prediction, the domains through which statistical learning first entered banking, and it remains the reference model wherever regulation demands that automated decisions be decomposable attribute

by attribute and auditable ex post. Its estimation is also statistically benign in the present design, a strictly concave likelihood with a unique optimum, two thousand observations against twenty-eight parameters, far from the separation pathologies of wide datasets, and deliberately spartan: no interactions, no polynomial terms, no stepwise search, so that the fitted model is an uncontaminated linear benchmark against which every non-linear architecture of the next sections must justify its complexity.

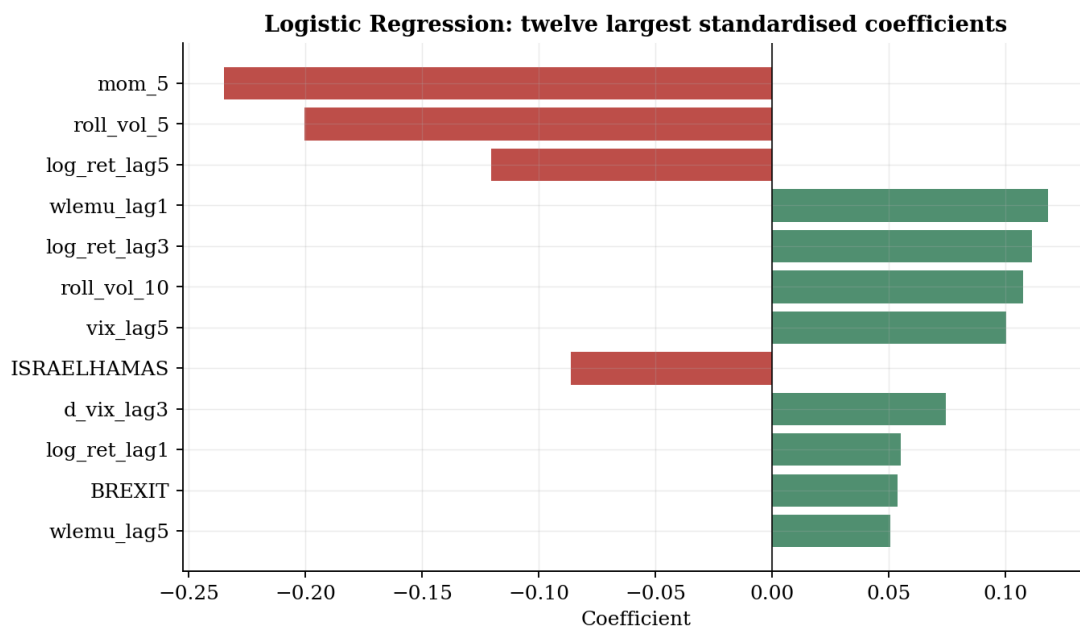


Figure 4.9, Logistic Regression: the twelve largest standardised coefficients. Green bars raise, red bars lower, the probability of an upward day.

Out of sample the logistic model posts an accuracy of 55.3 per cent, a precision of 58.6 per cent, an F1 of 0.657 and an AUC of 0.521, third among the well-behaved models and within two hundredths of the penalised regressions. The classification formulation trades a little ranking ability against a better calibrated operating point at the symmetric threshold: the logistic identifies 62 of the 214 down-days while keeping its bullish precision above the 57.3 per cent base rate, meaning that its 'up' calls are genuinely more reliable than an unconditional bet. Taken together, the three linear models of this section define the empirical frontier of the study, and they do so with a total of at most twenty-eight estimated parameters each.

Everything that follows asks whether thousands more parameters, and arbitrarily non-linear functional forms, can push beyond it.

4.5 Non-linear Machine Learning: ensembles and kernel methods

The third family abandons linearity altogether. Its four members, two ensembles of decision trees built on opposite statistical philosophies, their regularised industrial evolution, and a kernel-based maximum-margin classifier, can represent interactions and thresholds that no linear score can express: a momentum effect that operates only in high-volatility regimes, a sentiment signal that switches sign inside crisis windows, a non-monotone response to the level of the VIX. Whether such structures exist in the data, and whether they are stable enough to survive out of sample, is the question this section answers empirically. The bias–variance logic set out in Section 4.4 frames the stakes: every increment of expressive power is purchased with estimation variance, and in a low signal-to-noise environment the purchase is profitable only if the non-linear structure is real and persistent.

4.5.1 Random Forest

Decision trees partition the feature space through recursive binary splits, each chosen to maximise the purity of the resulting subsamples, measured for classification by the Gini impurity of node m ,

$$G(m) = 1 - \hat{p}_m^2 - (1 - \hat{p}_m)^2 \quad (4.21)$$

where $\hat{p}_m$ is the share of up-days among the observations reaching the node. A single deep tree is a low-bias, extremely high-variance estimator; the Random Forest of Breiman attacks the variance through double randomisation, growing B trees on bootstrap resamples of the training data and restricting each split to a random subset of $m = \sqrt{p}$ predictors, then averaging the ensemble:

$$\hat{p}(x) = (1/B) \sum_{\beta} \hat{p}_{\beta}(x) \quad (4.22)$$

The forest is best understood as the refinement of an older idea, bootstrap

aggregation, which already averaged trees grown on resampled data but left every tree free to use any predictor at any split. When a few variables dominate, as momentum and volatility do here, bagged trees all choose the same early splits and remain highly correlated, and the variance of an average of correlated estimators is bounded below by their common covariance no matter how many are aggregated: averaging a thousand near-copies of the same opinion sharpens nothing. Restricting each split to a random subset of candidates is the innovation that breaks the correlation, and it explains the counterintuitive strength of the method, each individual tree is made worse on purpose so that the ensemble becomes better. The bootstrap mechanism also yields a free internal diagnostic: each tree is estimated on roughly two thirds of the observations, and its accuracy on the excluded third, the out-of-bag error, aggregated over the forest, approximates the generalisation error without touching any test data. In this study the out-of-bag estimate plays only a confirmatory role, since all models are disciplined by the common chronological test set; but no amount of decorrelated aggregation can concentrate a signal that is not there, and the test set will shortly say exactly that.

The feature subsampling is the decisive ingredient, it decorrelates the trees, and the variance of an average of nearly independent estimators falls approximately as $1/B$, while the constraints selected by cross-validation (four hundred trees of depth six with at least twenty observations per leaf) prevent the individual learners from memorising crisis-specific noise. Out of sample the forest achieves an accuracy of 52.5 per cent and an AUC of 0.484: below the random-ranking frontier, and eight AUC points behind the penalised regressions estimated on identical information. The diagnosis is the textbook one: the flexible partitions learned on 2016–2024 encode regime-specific patterns that the 2024–2026 market did not repeat. The internal information hierarchy of the ensemble remains instructive, however. Figure 4.10 ranks the predictors by mean decrease in impurity: the momentum signal, the lagged returns and the VIX dynamics dominate, the same variables

selected by the Lasso path, with the crisis dummies at the bottom. One caveat disciplines the reading of such rankings: impurity importance mechanically favours continuous variables offering many split points over binary ones, and distributes credit almost arbitrarily among correlated predictors, so the figure is a map of how the forest spends its splits, not a causal decomposition; its value lies in the convergence with the rankings produced by the linear models under a completely different mathematical principle.

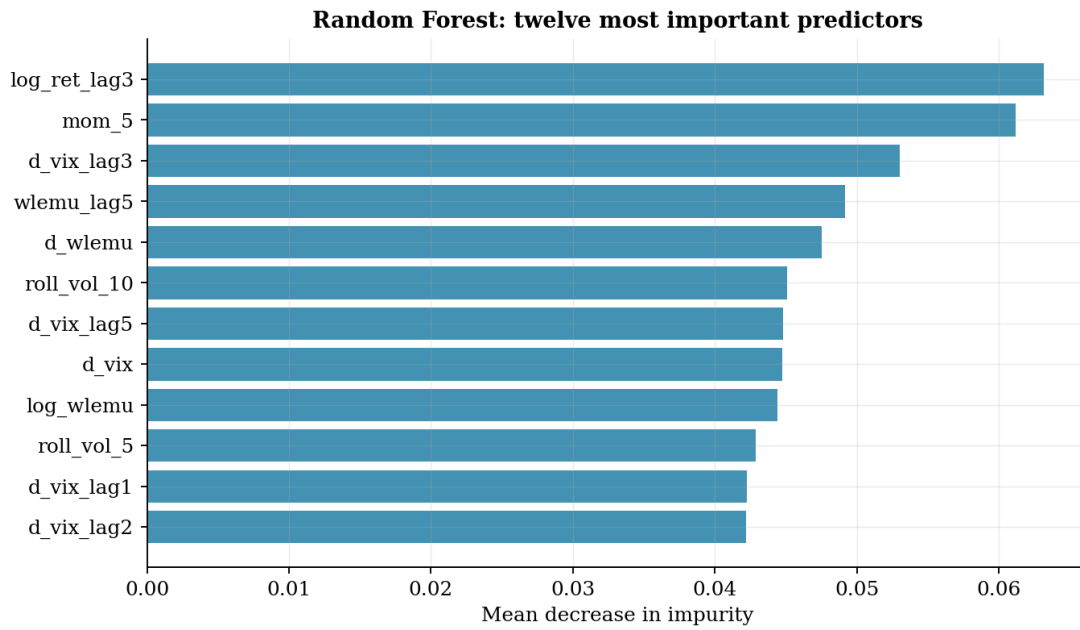


Figure 4.10, Random Forest: the twelve most important predictors by mean decrease in impurity.

4.5.2 Gradient Boosting

Where the forest reduces variance by averaging independent deep learners, Gradient Boosting reduces bias by chaining dependent shallow ones. The model is built stage-wise as an additive expansion of M small trees,

$$F_M(x) = \sum_m v \cdot h_m(x) \quad (4.23)$$

in which, following the functional gradient descent interpretation of Friedman,

each new tree is fitted to the negative gradient of the loss at the current ensemble, which for the binomial deviance is simply the gap between outcome and predicted probability:

$$h_m \approx \text{fit to } y_t - \hat{p}_t^{(m-1)} \quad (4.24)$$

The idea descends from a theoretical question posed in computational learning theory, can a committee of 'weak learners', each barely better than chance, be combined into an arbitrarily strong one?, answered affirmatively by the AdaBoost algorithm of Freund and Schapire, which reweights observations to force each successive learner to concentrate on the mistakes of its predecessors. Friedman's statistical reformulation, adopted here, revealed AdaBoost to be a special case of a general principle: gradient descent performed not in a parameter space but in the space of functions, with each stage taking a step in the direction that most steeply reduces the chosen loss. The reformulation matters because it exposes both the source of boosting's power and the precise location of its financial Achilles heel. The power: by choosing the loss (here the binomial deviance) and the base learner's depth, the analyst controls exactly what is being optimised and what order of interactions can be represented, depth two, selected by the cross-validation, permits pairwise interactions such as momentum conditioned on the volatility regime. The heel: the algorithm has no internal concept of irreducible noise. Whatever residual remains after each stage is treated as error to be pursued, so in an environment where the residual is overwhelmingly noise, precisely the daily-direction setting, the machinery of relentless improvement becomes machinery for memorising the training decade, and only external devices (shrinkage, early termination, shallow trees) stand between the model and its own diligence.

The learning rate v scales each contribution and is the principal regularisation device: a small v forces the algorithm to advance in many cautious steps, each correcting a fraction of the remaining error. Cross-validation selects stumps of depth two, the shallowest interaction structure available, with three hundred stages at $v = 0.05$, and the choice of minimal depth is already a verdict: offered the

possibility of three-way interactions, the tuning procedure declines it. Out of sample the boosted ensemble achieves 55.1 per cent accuracy but an AUC of 0.501, no ranking ability. The juxtaposition of the two figures is diagnostic: boosting calibrates the level of its predictions well (its accuracy benefits from the drift), but the sequential architecture, obliged at every stage to chase the observations the ensemble currently misclassifies, spends its capacity pursuing what is overwhelmingly irreducible noise. With weak signals, variance-reducing ensembles degrade gracefully towards randomness; bias-reducing ensembles work diligently towards it.

4.5.3 XGBoost

Extreme Gradient Boosting, introduced by Chen and Guestrin, is the evolution of the boosting principle that has dominated applied Machine Learning for a decade, and its distinctive contribution is the incorporation of regularisation into the boosting objective itself. Each stage minimises a second-order approximation of

$$Obj = \sum_t L(y_t, \hat{y}_t) + \sum_m \Omega(h_m) \quad (4.25)$$

with a complexity penalty on every tree,

$$\Omega(h) = \gamma T + (\lambda/2) \sum_j w_j^2 \quad (4.26)$$

charging a fixed cost per leaf T and an L2 cost on the leaf weights, the Ridge idea of Section 4.4.1 transplanted into function space. Second-order (Newton-type) optimisation makes splits and leaf values exact given the penalty; histogram-based split finding makes the search scale with quantile bins rather than observations; and stochastic subsampling of rows and columns (both at 80 per cent) injects the decorrelating randomness of the forest into the boosting architecture.

One element of the configuration deserves separate mention because it addresses the pathology that will disable the LSTM: the treatment of class imbalance. In a boosted classifier the pull towards perpetual optimism operates through the gradients, with 54 per cent of training days positive, the aggregate gradient

pressure of the majority class outweighs that of the minority, and an unconstrained learner drifts towards the base rate. The rebalancing weight applied here scales the gradients of the down-days by the negative-to-positive ratio, equalising the aggregate pressure of the two classes on the loss landscape; it is the boosting-space counterpart of the balanced class weights used in the logistic, forest and SVM estimators, and it is the reason the model's confusion matrix, alone in the field, approaches symmetry. The broader point generalises beyond this study: in financial classification, where the classes are almost never balanced and the imbalance itself drifts over time, the rebalancing choice is not a technical detail but a modelling decision about what the classifier is for, a drift-follower or a conditioning device, and it should be made explicitly, as here, rather than inherited silently from a library default.

The cross-validated configuration, three hundred rounds, depth three, $v = 0.05$, class rebalancing by the negative-to-positive ratio, delivers an out-of-sample accuracy of 51.1 per cent and an AUC of 0.503. The regularised objective repairs the pathology of its classical ancestor, the AUC returns to the random frontier instead of sinking below it, but it manufactures no predictability: XGBoost is, on these data, an expensively well-behaved coin. Its confusion matrix is nonetheless the most balanced of the twelve (88 true negatives, the maximum of the study), a property that Section 4.7 will revisit when the models are read as risk postures rather than as competitors on a single metric.

4.5.4 Support Vector Machine

The Support Vector Machine approaches classification geometrically, seeking the hyperplane that separates the classes with the widest possible margin:

$$\min (1/2)\|w\|^2 + C \sum_i \xi_i \quad s.t. \quad y_i(w^T x_i + b) \geq 1 - \xi_i, \quad \xi_i \geq 0 \quad (4.27)$$

where the slack variables tolerate violations and C prices them. The optimisation is solved through its dual, whose solution depends only on the support vectors, the ambiguous days lying on or inside the margin, so that the model summarises a

decade of history through its frontier cases, assigning zero weight to the unambiguous mass: a sparse, frontier-focused representation with no counterpart among the other eleven models.

The maximum-margin principle is not a heuristic but the operational form of a deep result in statistical learning theory: the generalisation error of a classifier is controlled not by the dimension of the space in which it operates but by the width of the margin it achieves relative to the spread of the data. This is what licenses the apparent extravagance of the construction that follows, kernels corresponding to implicit feature spaces of arbitrarily high dimension, in which overfitting would seem inevitable, yet where the margin-maximising solution remains statistically well behaved because its effective capacity is governed by the margin and by C , not by the dimension count. The two hyperparameters thus form the model's complete complexity dial: C prices margin violations, large values force the boundary to accommodate nearly every training day, producing a convoluted, high-variance frontier, while small values buy a wide, stable margin at the cost of training errors, and the kernel width γ sets the resolution of the similarity measure, from sharply local matching of near-identical historical days to smooth global comparisons, with the scale heuristic adopted here adapting the resolution to the dimensionality and variance of the standardised features. The kernel trick itself provides the non-linearity: since the dual involves the data only through inner products, these are replaced by the Radial Basis Function kernel

$$K(x, x') = \exp(-\gamma \|x - x'\|^2) \quad (4.28)$$

which classifies a new trading day by its similarity to the most closely comparable historical days in the standardised feature space. Cross-validation selects $C = 10$, a relatively hard margin, indicating that the tuning procedure found structure worth fitting in-sample, and Platt scaling converts the decision values into probabilities for the ROC analysis. Out of sample, the verdict is identical to the forest's: accuracy 52.5 per cent, AUC 0.484. The kernel machine, granted arbitrarily flexible frontiers, generalises no better than random ranking, while its linear

cousins of Section 4.4 sit at the top of the field. The four models of this section thus return a unanimous answer to the question that opened it: whatever non-linear structure connects fear, momentum and next-day direction in the training decade, it is either too faint or too unstable to survive two years out of sample. The last hope for complexity lies with the architectures built explicitly for hierarchical and sequential representation, to which the study now turns.

4.6 Deep learning architectures: ANN and LSTM

The models of the preceding sections, however diverse, share a static view of the market: each trading day is an independent vector of features. Deep learning relaxes this constraint along two dimensions. Feed-forward networks introduce hierarchical representations, learned compositions of the raw variables that can express abstract states such as 'stress building despite calm prices'. Recurrent networks add explicit temporal processing, ingesting an ordered window of days and maintaining an internal state that can, in principle, track a regime as it forms. The empirical literature calibrates expectations precisely: the successes of deep learning in finance, the return-prediction gains of Gu, Kelly and Xiu, the LSTM results of Fischer and Krauss, arise in settings with thousands of assets and hundreds of thousands of training examples, and the same literature reports that the advantage of depth shrinks toward zero as samples thin and horizons shorten. One index observed daily for a decade is the unfavourable extreme of that spectrum, which makes it the sharpest available test of where the boundary of usefulness lies.

4.6.1 Artificial Neural Network

A feed-forward network is a composition of layers, each applying an affine transformation followed by a non-linear activation:

$$a^{(l)} = g(W^{(l)} a^{(l-1)} + b^{(l)}) \quad (4.29)$$

with the Rectified Linear Unit in the hidden layers,

$$\text{ReLU}(z) = \max(0, z) \quad (4.30)$$

whose piecewise linearity avoids the vanishing gradients of saturating activations while preserving the universal-approximation property. A sigmoid output produces the probability of a rise, and the parameters minimise the binary cross-entropy

$$L = -(1/N) \sum_t [y_t \ln \hat{p}_t + (1 - y_t) \ln(1 - \hat{p}_t)] \quad (4.31)$$

The theoretical warrant for the architecture is the universal-approximation property: a network with even a single sufficiently wide hidden layer can approximate any continuous function of the inputs to arbitrary precision, and depth achieves the same expressiveness with exponentially fewer units when the target function is compositional. But the warrant must be read with the distinction that governs this entire chapter, the distinction between approximation and estimation. Universality guarantees that some configuration of the weights can represent virtually any relationship between the twenty-seven predictors and the probability of a rise; it says nothing about whether that configuration can be identified from a finite, noisy sample. With several thousand trainable parameters against seventeen hundred effective training observations, the estimation error, not the approximation error, is the binding constraint, and every element of the training protocol described below is a response to that imbalance: the network's practical capacity must be forced far below its nominal one, or the universality that is its theoretical glory becomes the mechanism of its empirical ruin. Mini-batch stochasticity contributes to the same end through a subtler channel, the gradient noise of small batches discourages convergence to sharp, brittle minima of the loss surface, acting as an implicit regulariser on top of the explicit ones.

by backpropagation with the Adam optimiser, which adapts each coordinate's step size through running estimates of the first and second moments of its gradient. The architecture, three hidden layers of 64, 32 and 16 units, carries several thousand parameters against seventeen hundred effective training observations, a ratio at which estimation error, not approximation error, binds; the training protocol

responds with the full arsenal of implicit and explicit regularisation: L2 weight decay, mini-batches of thirty-two, and above all early stopping, which monitors a fifteen per cent validation split and halts the descent, rolling the weights back to their best recorded value, when validation performance stagnates for ten epochs. Early stopping is approximately equivalent to a ridge penalty whose strength decreases with training time, so the stopping epoch selects the network's effective complexity along a continuous regularisation path; Figure 4.11 shows the mechanism at work, the loss collapsing over the first epochs and the validation accuracy oscillating around its plateau until the criterion fires.

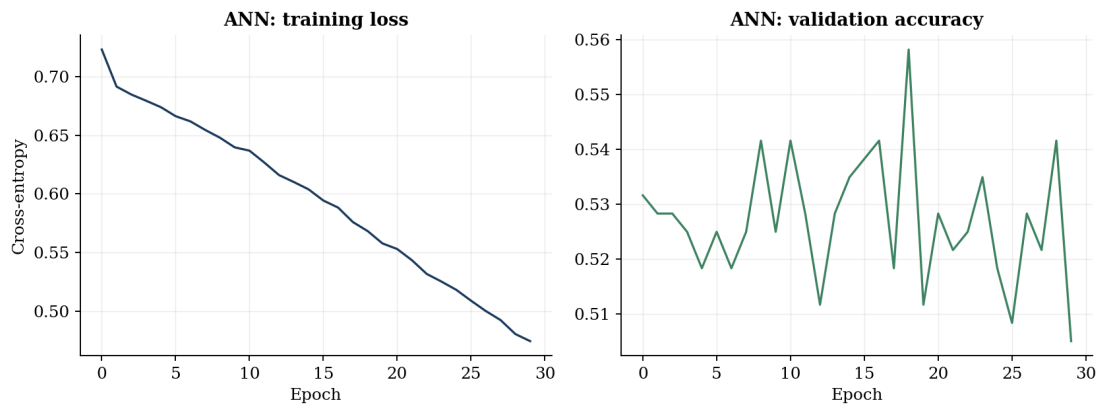


Figure 4.11, ANN training dynamics: cross-entropy loss (left) and validation accuracy (right) under early stopping.

Out of sample the network posts an accuracy of 55.7 per cent, the best of any non-degenerate model, with an F1 of 0.662 and an AUC of 0.523, statistically level with the logistic regression and marginally behind the penalised regressions. Its confusion matrix reveals the strategy: 217 true positives against 62 true negatives, a recall of 75.6 per cent, the network leans on the upward drift and modulates it mildly with the fear variables. The hierarchical representation, in other words, converges to a gently non-linear version of the same signal the linear models found; several thousand additional parameters buy a rounding error of ranking ability. It is a respectable, honest performance, and a precise measurement of how little non-linear structure the data contain.

4.6.2 Long Short-Term Memory network

The Long Short-Term Memory network of Hochreiter and Schmidhuber is the canonical architecture for sequences. For each prediction it ingests the full ordered window of the previous ten trading days and processes it through a memory cell regulated by three multiplicative gates:

$$f_t = \sigma(W_f x_t + U_f h_{t-1} + b_f) \quad (4.32)$$

$$i_t = \sigma(W_i x_t + U_i h_{t-1} + b_i) \quad (4.33)$$

$$o_t = \sigma(W_o x_t + U_o h_{t-1} + b_o) \quad (4.34)$$

$$c_t = f_t \odot c_{t-1} + i_t \odot \tanh(W_c x_t + U_c h_{t-1} + b_c) \quad (4.35)$$

$$h_t = o_t \odot \tanh(c_t) \quad (4.36)$$

The gates admit a suggestive economic reading. The cell state c_t is the model's running assessment of the market regime; the forget gate f_t controls its persistence, preserving the assessment in stable climates and flushing it at breaks; the input gate i_t filters how much of today's configuration of volatility, sentiment and returns deserves writing into memory; the output gate o_t modulates how much of the internal view is exposed to the prediction layer. An ideally trained LSTM would implement, end to end, the cognitive routine of a discretionary macro trader, maintain a view, update it on salient news, discard it at regime changes. The additive path that the cell state provides for the gradient solves the vanishing-gradient problem of plain recurrent networks, enabling dependencies spanning the full window; training proceeds by backpropagation through time with Adam, gradient clipping, a forget-gate bias initialised at one, and the same early-stopping protocol as the ANN, on sequences built strictly with the training-window scaler. The ten-day window itself is a compromise: long enough to contain the build-up and decay of the uncertainty episodes documented in Table 4.2, short enough to keep the sequence count, and the parameter-to-data ratio, defensible.

Training a recurrent model adds a dimension of difficulty that the feed-forward

case does not know. Backpropagation through time unrolls the recurrence into a chain of ten replicated cells and applies the chain rule along it, so every gradient traverses a product of ten Jacobians; in a plain recurrent network that product shrinks or explodes geometrically, the vanishing and exploding gradient problem, making dependencies beyond a few steps unlearnable. The LSTM's additive cell-state path is the architectural cure for the vanishing half of the problem, and gradient clipping the standard prophylactic for the exploding half; both are employed here. The window length itself is then a genuine hyperparameter with a bias–variance character: a window must be long enough to contain the dynamics of interest, the two-week build-up and decay of the uncertainty episodes in Table 4.2, yet every added timestep lengthens the computational graph, dilutes the learning signal of each sequence, and reduces the number of independent sequences the sample can supply, a doubly unfavourable trade in two thousand days of data. Ten steps of twenty-seven features keep the parameter count of the recurrent layer, four gate matrices over inputs and hidden state, roughly eight thousand weights, in a defensible ratio to the data, which is as much as design can do; whether the data can fill even that reduced capacity is the empirical question the results now answer.

The training curves of Figure 4.12 tell the story before the test set does: the validation loss reaches its minimum within a handful of epochs and early stopping intervenes almost immediately, the sequential structure contains little that the network can extract. The out-of-sample outcome is the most instructive failure of the study. On the surface, the LSTM posts the highest accuracy of any learning model (57.1 per cent) and the second-highest F1 (0.725); underneath, its confusion matrix records 284 true positives against 2 true negatives, a recall of 99.0 per cent, and an AUC of exactly 0.500. The network has collapsed onto the unconditional climatology of the market, the same corner solution the GARCH models occupied by theoretical necessity, reached here by empirical exhaustion. Confronted with 27 features over 10 timesteps and an evanescent signal, the most expressive

architecture of the study rationally concluded that the best compression of the data is 'markets usually rise'. The contrast with Fischer and Krauss quantifies the diagnosis: their successful LSTM harvested hundreds of thousands of training sequences from the cross-section of five hundred stocks; this design supplies fewer than two thousand. Deep sequential models are not unsuited to finance; they are unsuited to small-sample finance, and the degenerate solution is the signature of the mismatch, invisible to accuracy and exposed only by the threshold-free metric that the evaluation protocol was designed to include.

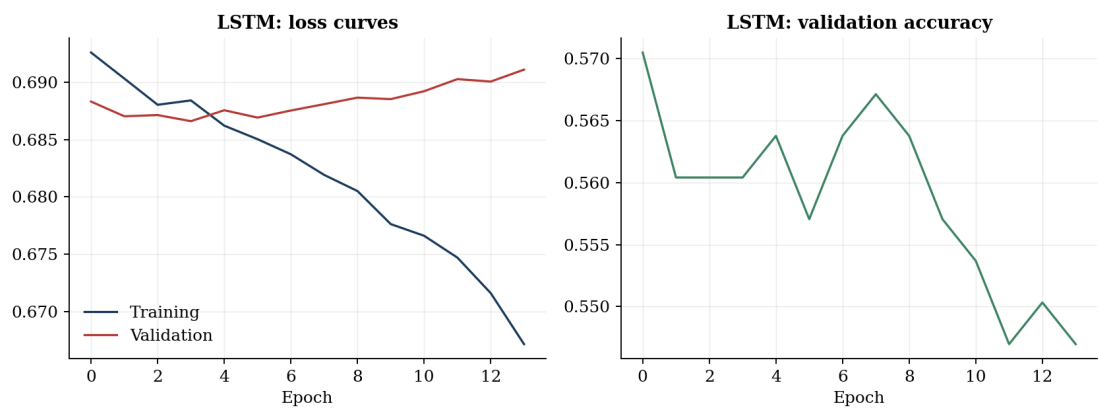


Figure 4.12, LSTM training dynamics: loss curves (left) and validation accuracy (right). Early stopping intervenes within few epochs.

4.7 Comparative analysis and interpretation

Having examined the twelve architectures family by family, this section confronts them on the common ground of the 501-day test period. Table 4.5 collects the five metrics for every model, ranked by ROC-AUC; Figure 4.13 visualises the comparison; Figure 4.14 superimposes the twelve ROC curves; Figure 4.15 and Table 4.6 decompose each model's behaviour into its confusion matrix; and Figure 4.16 tracks the stability of the leading models through time. Read jointly, these exhibits answer the central empirical question of the thesis with unusual precision: how much of tomorrow's market direction can four generations of quantitative methodology actually anticipate, and which informational channels do the work?

Table 4.5, Out-of-sample performance of the twelve models, test set June 2024 – June 2026,

ranked by ROC-AUC.

Model	Accuracy	Precision	Recall	F1	ROC_AUC
Lasso	0.543	0.588	0.672	0.628	0.538
Ridge	0.539	0.585	0.669	0.624	0.538
ANN (MLP)	0.557	0.588	0.756	0.662	0.523
Logistic Regression	0.553	0.586	0.749	0.657	0.521
ARIMAX	0.515	0.564	0.672	0.614	0.512
XGBoost	0.511	0.571	0.585	0.578	0.503
Gradient Boosting	0.551	0.582	0.77	0.663	0.501
LSTM	0.571	0.573	0.99	0.725	0.5
Random Forest	0.525	0.572	0.679	0.621	0.484
SVM (RBF)	0.525	0.574	0.662	0.615	0.484
EGARCH(1,1)	0.573	0.573	1.0	0.728	0.463
GARCH(1,1)	0.573	0.573	1.0	0.728	0.463

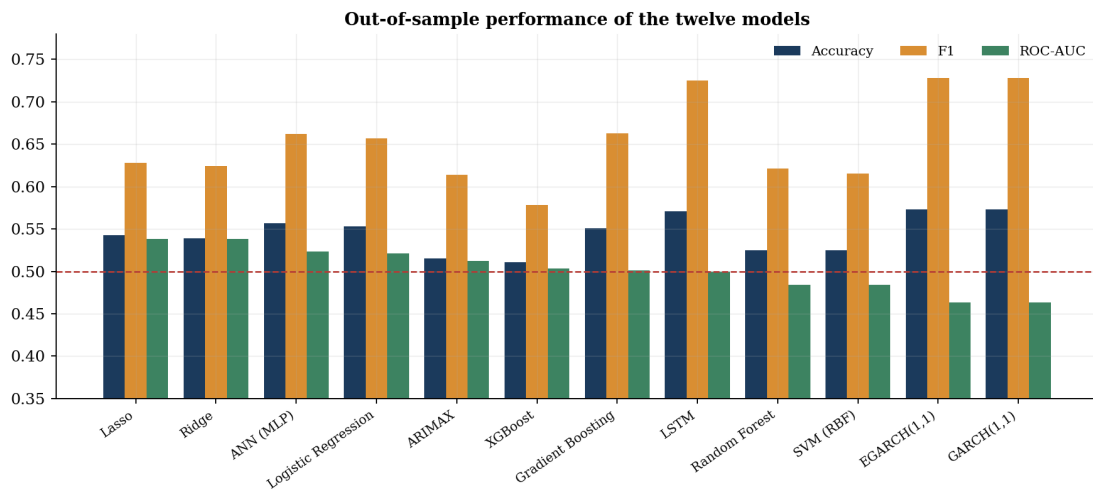


Figure 4.13, Accuracy, F1 and ROC-AUC of the twelve models; the dashed line marks the 0.5 no-information frontier.

The headline result is a hierarchy that inverts eight decades of methodological sophistication. The summit belongs to the penalised linear regressions, Lasso and Ridge, tied at an AUC of 0.538, followed by the two directly probabilistic learners that lean on the same linear signal, the ANN (0.523) and the Logistic Regression (0.521), with the 1970s-vintage ARIMAX (0.512) ahead of the entire non-linear Machine Learning block. The flexible architectures, Random Forest, SVM, Gradient Boosting, XGBoost, cluster at or below the random-ranking frontier, and the two most complex objects in the study, the LSTM and the GARCH family, occupy the degenerate corner where high accuracy coexists with zero or negative ranking ability. The performance band is, moreover, compressed: across twelve models the AUC spans eight hundredths and accuracy six percentage points. The compression is itself the finding. It indicates that the binding constraint is not model capacity, capacity was varied across four orders of magnitude of parameter count with essentially no effect, but the informational content of the predictors: a ceiling imposed by the market, not by the econometrician.

One entry of the table deserves a technical footnote before the economic reading, because it looks paradoxical: the GARCH family's AUC of 0.463 lies significantly below one half, and an AUC below one half means the ranking carries information, inverted. A hypothetical contrarian who bought when the volatility models were most pessimistic would have ranked days better than chance. The temptation to exploit such inversions should be resisted, and the reasons are instructive. First, the inversion is regime-specific: it reflects the particular composition of this test period, in which the largest positive returns clustered in the turbulent rebound phases of the 2025 corrections, so that ranking calm days as bullish, the structurally correct implication of equation (4.15) in a market with positive drift, happened to run backwards; in a test window dominated by grinding calm rallies the same model would rank correctly. Second, trading an inversion is the purest form of the post-hoc selection that the protocol of Section 4.2 was built to exclude: the sign of the exploitation rule would be chosen after observing the test data. The correct

reading is subtler and more valuable, the below-chance AUC is evidence that volatility is genuinely informative about the distribution of next-day returns, but that the mapping from risk to direction is regime-dependent in a way the constant-drift specification cannot represent. It is, in miniature, the case for the conditioning variables that the winning models employ: fear matters for direction, but its meaning depends on context.

Why do the shrunken linear models win? The bias–variance decomposition introduced in Section 4.4 supplies the mechanism, and the twelve-model design turns it from interpretation into demonstration. Reading Table 4.5 downward from the summit, performance falls almost monotonically with model variance: heavily shrunken linear estimators first; the moderately regularised logistic and the early-stopped network, whose effective capacity is a fraction of its parameter count, immediately behind; the tree ensembles, whose variance the tuning could contain but not eliminate, at the frontier; and the highest-variance sequential architecture in degenerate collapse. Meanwhile bias, the price of rigidity, goes essentially unpunished: constraining the functional form to a linear score costs nothing detectable, which is precisely what one expects when the true predictable component is itself approximately linear and minute. In the language of the decomposition, the irreducible term dominates, the variance term discriminates among models, and the bias term never binds. The practical corollary for quantitative finance is significant and general: model selection should be governed by the signal-to-noise ratio of the problem, not by the sophistication frontier of the discipline.

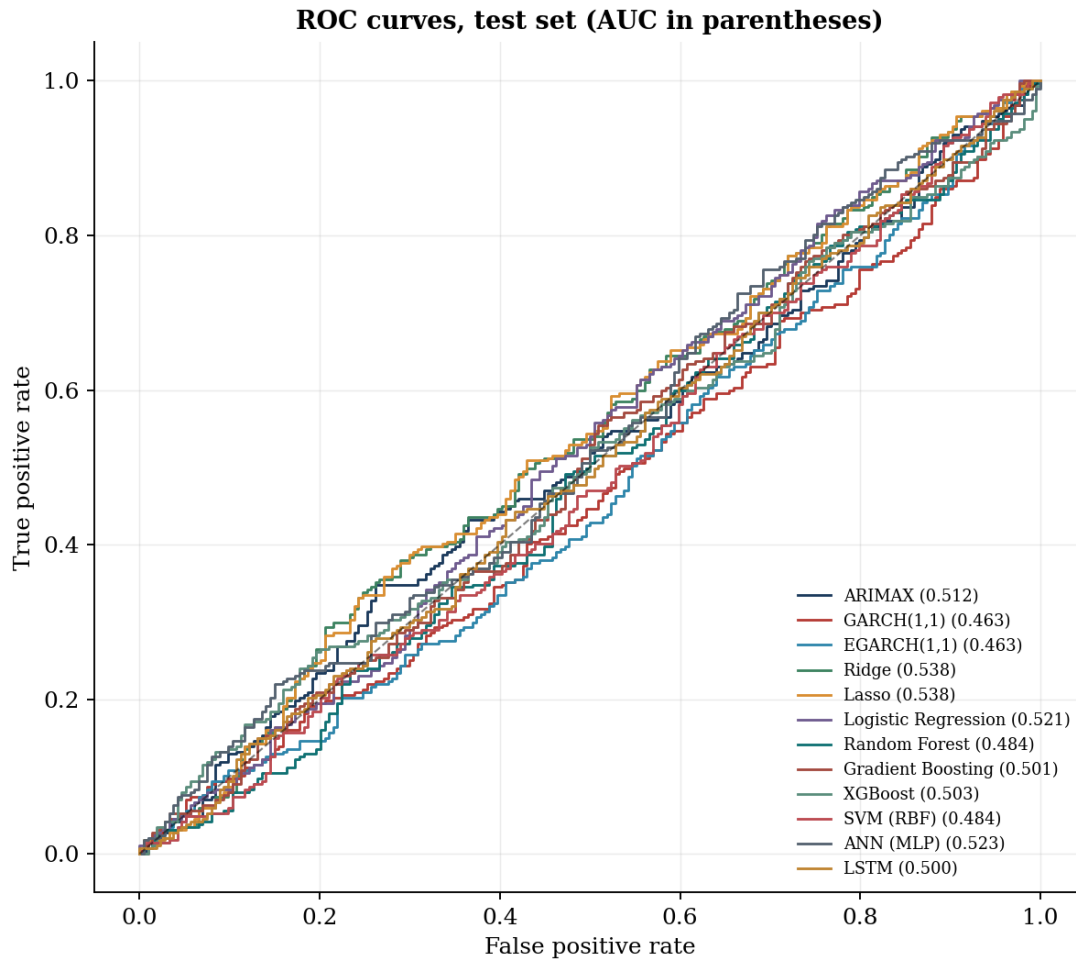


Figure 4.14, ROC curves of the twelve models. Curves hugging the diagonal indicate near-random ranking; the GARCH family lies below it.

Figure 4.14 rewards a closer geometric reading. The curves of the leading models separate from the diagonal in its central region, the regime of moderately confident predictions, while hugging it at the extremes: the models' strongest convictions are barely better than guesses, and their genuine edge lives in the unglamorous middle of the probability scale. The operational translation is that these outputs should be treated as a mild, continuous tilt on prior beliefs, never as a discrete trading trigger. The confusion matrices of Figure 4.15 and Table 4.6 add the anatomical view. Every model is tilted towards bullish calls, true positives exceed true negatives everywhere, but the degree of tilt spans a full spectrum: from XGBoost, the most balanced (88 true negatives against 168 true positives), through the linear block

(roughly 60–80 identified down-days), to the ANN's generous optimism and the total collapse of the LSTM and the GARCH pair (2 and 0 true negatives). What separates a useful model from a degenerate one is how much genuine conditioning it layers on top of the drift; the Lasso, which keeps its bullish precision at 58.8 per cent, a point and a half above the 57.3 per cent base rate, while still identifying 79 of the 214 down-days, layers the most.

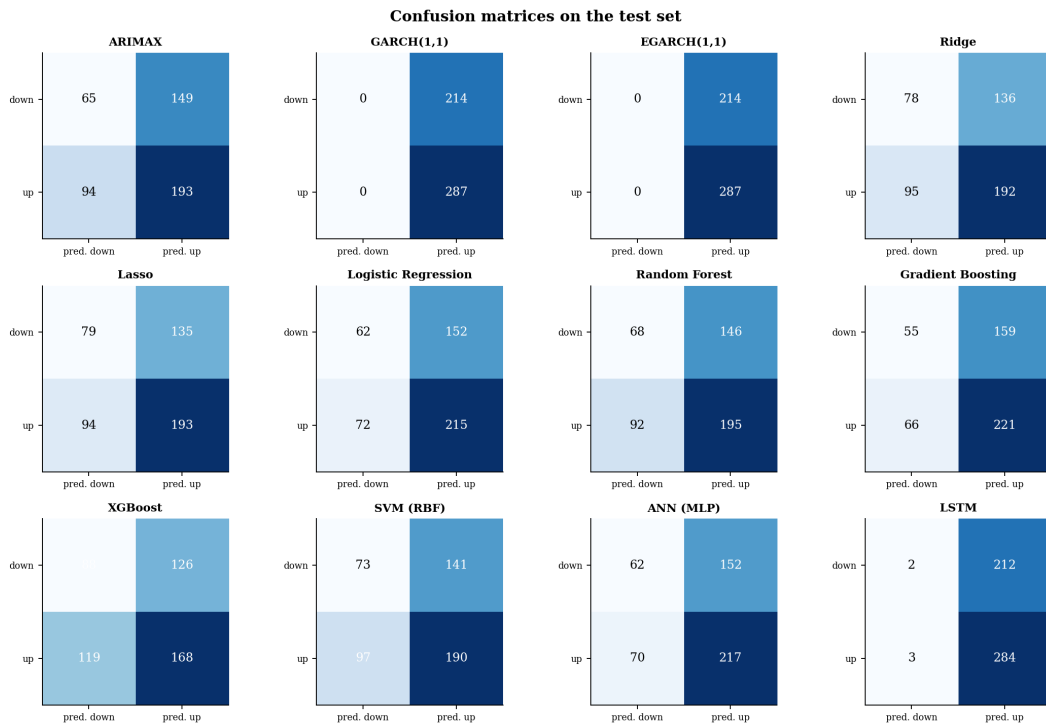


Figure 4.15, Confusion matrices of the twelve models on the test set.

Table 4.6, Confusion-matrix elements on the test set (TN, FP, FN, TP).

Model	TN	FP	FN	TP
ARIMAX	65	149	94	193
GARCH(1,1)	0	214	0	287
EGARCH(1,1)	0	214	0	287
Ridge	78	136	95	192
Lasso	79	135	94	193

Model	TN	FP	FN	TP
Logistic Regression	62	152	72	215
Random Forest	68	146	92	195
Gradient Boosting	55	159	66	221
XGBoost	88	126	119	168
SVM (RBF)	73	141	97	190
ANN (MLP)	62	152	70	217
LSTM	2	212	3	284

The joint reading of Table 4.5 and Table 4.6 also exposes a systematic trade-off along the precision–recall axis that a single ranking conceals. The models cluster into behavioural types. The conservative type, XGBoost above all, with the Ridge, Lasso and SVM close behind, issues bullish signals more sparingly (recalls of 59–67 per cent), accepting missed rallies in exchange for identifying more of the down-days. The aggressive type, Gradient Boosting, the ANN, the Logistic, blankets the sample more generously (recalls of 75–77 per cent), maximising participation in the drift at the price of blindness to reversals; the LSTM and the GARCH pair are this strategy taken to its degenerate limit. It is not a coincidence that the best AUC scores sit with models of intermediate-to-conservative posture: an operating profile that resists the pull of the base rate is where conditioning information, rather than unconditional drift, does the discriminating work, and the F1 column must be read with this in mind, since the harmonic mean mechanically rewards aggression in a bull-market test period, which is why the degenerate models top it. For portfolio use, the choice between types is an economic decision rather than a statistical one: a fully invested institution hedging tail risk values the conservative profile, which flags more of the dangerous days; a benchmark-tracking mandate fearful of missing rallies values the aggressive one. Table 4.5 is,

in this reading, not a leaderboard but a menu of risk postures, one more sense in which model selection in finance is inseparable from the economics of the user.

Are these magnitudes statistically meaningful? Against the coin-toss null, yes: over 501 trials the standard error of an accuracy estimate is about 2.2 percentage points, so the 54–56 per cent of the leading models lies two to two-and-a-half standard errors above chance. But the coin toss is the wrong benchmark in a drifting market: the relevant naive competitor is the climatological rule that unconditionally predicts a rise, exploiting the 57.3 per cent base rate, a rule that beats every model of the study on raw accuracy, and that the LSTM and GARCH replicate rather than beat. Three considerations organise the honest verdict. First, the climatological rule is not available *ex ante*: an investor standing in June 2024 knows the training base rate of 54.3 per cent, not the future one, and the drift between the two is itself unforecastable. Second, accuracy is the wrong yardstick for the comparison: the always-up rule has an AUC of exactly one half and a precision equal to the base rate, whereas the leading models push precision above it and sustain a positive ranking edge over every threshold, they do extract conditioning information beyond the drift. Third, Figure 4.16 shows that this extraction is not an artefact of one lucky quarter: the rolling 63-day accuracy of the three best models oscillates around the base-rate line, falling below it in the strongly trending phases, where nothing beats unconditional optimism, and rising above it in the turbulent ones, exactly where conditioning information should matter and where, as Section 4.8 will show, it is worth money.

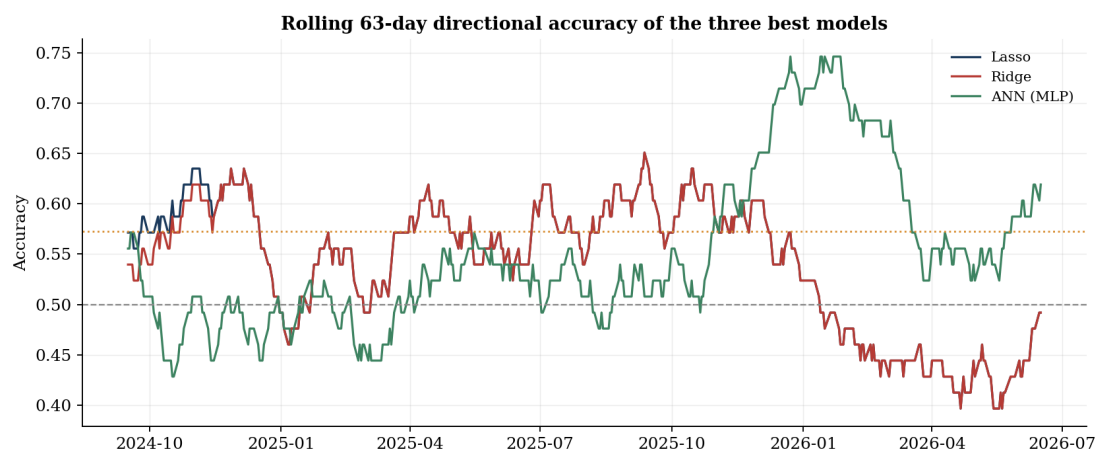


Figure 4.16, Rolling 63-day directional accuracy of the three best-ranked models against the up-day base rate (dotted line).

How should the absolute level of performance be judged against theory and literature? The natural benchmark is the Efficient Market Hypothesis in its weak form (Fama, 1970), which asserts that prices fully incorporate the information in past prices and public indicators, leaving future returns unforecastable from such data. The evidence is broadly, but not perfectly, consistent with it. Broadly, because the best AUC of 0.538 sits close to the efficiency frontier, and because the most powerful function approximators available gain nothing over a shrunken linear score. Not perfectly, because the edge of the leading models is not an isolated artefact: five architecturally unrelated estimators, two penalised regressions, a logistic classifier, a neural network and an ARMA filter, converge on the same economic narrative of short-horizon mean reversion modulated by fear, with elevated media uncertainty followed by marginal rebounds as feared scenarios fail to materialise. This convergence suggests a genuine, if tiny, violation of strict unforecastability, of the kind the adaptive-markets perspective (Lo, 2004) attributes to slowly decaying behavioural frictions: the collective emotions and uncertainty shocks that Chapter 2 placed at the centre of modern market dynamics leave a measurable, though faint and possibly transient, trace in next-day prices. The placement within the empirical literature confirms representativeness: four decades of published studies on daily directional prediction of major indices cluster

in the 50–60 per cent accuracy band, with the recurrent finding that results far above it dissolve, under scrutiny, into look-ahead bias, full-sample scaling, random splits or post-hoc model selection, the methodological sins that the protocol of Section 4.2 was engineered to exclude. A clean design lands where this one lands: a few points above chance, well below climatological certainty, in the equilibrium residue that a competitive market leaves behind after predictability has been consumed by the act of exploiting it.

A final comparative reading concerns the geopolitical channel. The event dummies received negligible weight in every model that could discard them, the Lasso pruned three of the four outright, yet Table 4.2 showed that the windows they mark exhibit radically abnormal return and volatility profiles. The reconciliation is that their information travels indirectly: a geopolitical shock moves prices through the fear channel, the VIX and the uncertainty index spike within hours, and by the time the binary indicator switches on, the continuous gauges have absorbed its content. In the language of this thesis, geopolitical power shocks are priced through the sentiment infrastructure of the market, and the redundancy of the explicit event marker is the measure of that infrastructure's speed. This sharpens, rather than weakens, the argument of Chapter 2: it is precisely because information diffuses so quickly through volatility and sentiment channels that controlling the speed and quality of information access constitutes a form of market power.

4.8 Implications for data-driven financial decision making

The final section translates the statistical evidence into implications for investors, risk managers and for the broader thesis of data as a competitive weapon. The first exercise probes where the best model's edge lives. Table 4.7 splits the test period at the median VIX and computes the Lasso's directional accuracy separately in the two regimes: 52.6 per cent on calm days, 56.0 per cent on turbulent ones. The asymmetry is the mirror image of the fragility that afflicts many quantitative strategies, calibrated on calm markets, failing under stress, and it is exactly the

profile that the composition of the signal would predict: a model built on volatility persistence, momentum reversal and uncertainty resolution has little to say when nothing is happening, and speaks most clearly when fear is moving. In the framework of this thesis, the finding has a precise meaning: the informational value of sentiment data is concentrated in the states of the world where information is scarcest and most valuable, the uncertainty regimes that Chapter 2 identified as the moments in which data-driven power is actually exercised.

Table 4.7, Regime analysis and economic evaluation of the best-ranked model (Lasso) on the test period; long/flat strategy, no transaction costs, zero cash rate.

Metric	Lasso strategy	Buy and hold
Accuracy, low-VIX days	0.526	-
Accuracy, high-VIX days	0.560	-
Total return (%)	34.6	35.6
Ann. volatility (%)	14.4	16.5
Sharpe ratio	1.04	0.93

The second exercise asks the bluntest economic question: could the edge have been monetised? A long/flat rule is simulated on the test period, hold the index tomorrow if the model predicts a rise, stay in cash otherwise, and Figure 4.17 plots the cumulative growth of one dollar against buy-and-hold, with the summary statistics in Table 4.7. The raw returns are a statistical draw: 34.6 per cent for the strategy against 35.6 per cent for the benchmark over two years. The risk profile is not. By stepping out of the market on predicted down-days, which, per the regime analysis, the model identifies disproportionately in turbulent phases, the strategy realises an annualised volatility of 14.4 per cent against the benchmark's 16.5, lifting the Sharpe ratio from 0.93 to 1.04. The shaded regions of Figure 4.17 localise the mechanism: the strategy concedes ground in the uninterrupted rallies,

where any deviation from full exposure is a cost, and recovers it in the drawdowns of the 2025 corrections, which it partially sidesteps. Two caveats discipline the reading, transaction costs on the strategy's frequent switches would erode part of the gap, and a two-year window is one draw from a distribution, but the qualitative conclusion is robust and important: the statistical edge documented in Section 4.7 is real enough to reshape the risk of an equity position, and too small to reliably beat the market's own drift on raw return. Machine Learning, on public daily data, is a risk-management technology, not an alpha machine.

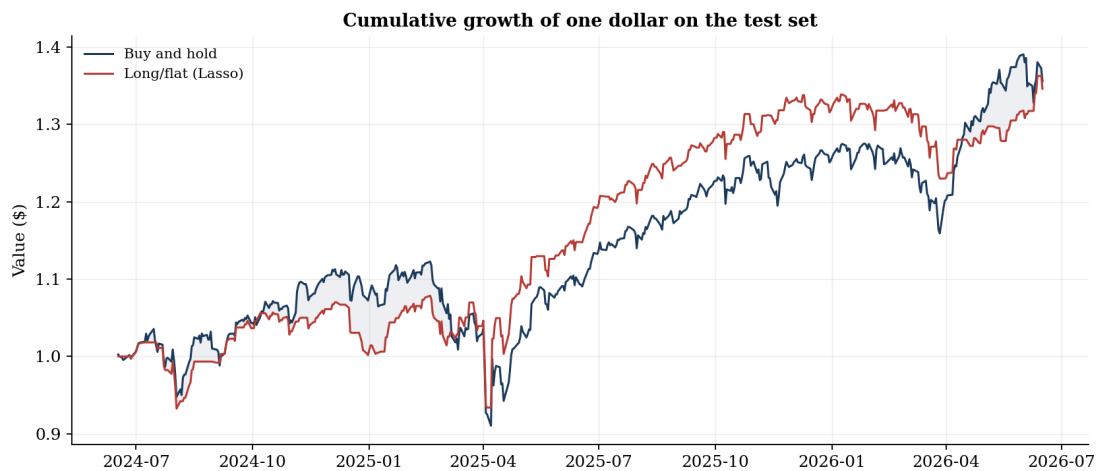


Figure 4.17, Cumulative growth of one dollar: Lasso-based long/flat strategy versus buy-and-hold; shaded areas mark the benchmark's advantage.

This reframing, from speculation to risk management, organises the practical value of every result in the chapter. The same probability output that fails to time the market succeeds at something subtler: it provides a daily, continuously updated, interpretable synthesis of the fear environment, and its exposure decisions systematically reduce variance where variance is most dangerous. For a pension fund, a corporate treasurer or a bank's market-risk function, an instrument that converts the torrent of volatility, sentiment and geopolitical data into a calibrated probability of adverse movement, one whose reliability rises with turbulence, is valuable even when, especially when, it promises no alpha. The GARCH results reinforce the same division of labour from the econometric side: the variance models describe the risk environment almost perfectly while contributing nothing

to direction, and a production architecture would naturally pair them, GARCH-family volatility forecasts sizing positions and stress scenarios, penalised-linear directional tilts modulating exposure at the margin, each doing the one thing the evidence shows it does well.

The results also discipline expectations about model complexity, and hence about where the marginal dollar of a quantitative budget should go. Across this study, every increment of sophistication beyond the penalised linear frontier either added nothing (SVM, ANN), merely repaired its own pathologies (XGBoost over Gradient Boosting), or actively harmed (Random Forest), in one case collapsing entirely (LSTM). For practitioners this inverts the fashionable hierarchy: in low signal-to-noise environments, the marginal investment is better spent on data, newer, faster, more exclusive informational sources, than on models. This is, restated operationally, the central thesis of this work. If twelve architectures spanning eight decades of methodology, fed with identical public information, converge within eight hundredths of AUC of the efficiency frontier, then the residual competitive space lies almost entirely in the information set itself: proprietary order flow, real-time textual sentiment, alternative data streams, and the infrastructure to process them earlier than competitors. The battle is not fought over algorithms, which are public goods, but over data. The finding also carries a governance corollary: the coincidence that the best-performing model is also the most interpretable one, auditable coefficient by coefficient, stress-testable analytically, explainable to a validator or a supervisor, is not a fortunate accident but a design criterion that regulated institutions, bound by model-risk-management expectations, should institutionalise: complexity must buy measurable, validated performance, or be refused.

These considerations converge on a realistic doctrine for the division of labour between human judgement and algorithmic prediction. The comparative evidence delineates what the machines demonstrably do well: they enforce discipline, evaluating twenty-seven informational channels every day without fatigue,

recency bias or narrative seduction; they quantify, converting the diffuse sense that 'markets are nervous' into a calibrated probability with a measurable track record; and they remember, retaining the statistical lessons of the pandemic and the invasion with a fidelity no committee can match. The same evidence delineates what they cannot do: they cannot anticipate structural novelty, because their knowledge is bounded by the regimes represented in the training window; they cannot interpret an unprecedented event until its consequences have accumulated into data; and, as the LSTM and GARCH episodes showed, they will quietly substitute statistical mimicry of the base rate for understanding whenever the data permit it, unless a multi-metric audit exposes the substitution. The resulting complementarity is precise rather than rhetorical: models should own the continuous, high-frequency layer of risk assessment, the daily probability tilt, the mechanical modulation of exposure, while human judgement owns the discontinuities, deciding when the world has changed enough that the training distribution no longer binds, and overriding the machine in exactly those moments when its confidence is least warranted. Institutions that architect this division explicitly convert the modest statistical edges documented here into durable organisational capability; institutions that blur it inherit the weaknesses of both parties.

At the level of the market system, the same economics of information generates the concerns with which Chapter 2 opened. If algorithms are commoditised and data decide, then the concentration of exclusive data streams, internalised order flow, proprietary sentiment feeds, the vendor oligopoly of alternative data, translates directly into a concentration of what little predictability markets leave available, shifting returns from price discovery to privileged access. And as institutions converge on similar data, similar features and similar model families, a convergence this chapter has documented from the inside, with twelve models agreeing on the same handful of predictors, their reactions to shocks correlate, and risk systems designed to protect each institution individually can synchronise

selling pressure collectively, amplifying precisely the crisis dynamics recorded in the event windows of Table 4.2. The regulatory agenda that follows is not the prohibition of predictive technology but the governance of its informational substrate: transparency over data provenance, scrutiny of exclusive data arrangements, macroprudential attention to model homogeneity.

The limitations of the exercise trace the agenda for the concluding chapter. The feature set, though economically motivated, is restricted to public, daily-frequency indicators; textual sentiment extracted directly from news and social media, intraday volatility surfaces and cross-asset spillovers are the natural extensions that the data-as-power framework predicts should carry incremental information. The event windows are short and few, so crisis-conditional estimates rest on limited observations. The single chronological split, while methodologically clean, evaluates the models on one particular two-year future; a rolling-origin scheme would map the time variation of the small predictable component. None of these caveats overturns the central conclusion. Twelve models, four methodological generations, one decade of market, sentiment and geopolitical data, and one verdict returned unanimously: predictability exists, is minute, lives in the fear channel, is best extracted by disciplined and interpretable methods, resists conversion into speculative profit, and rewards instead the patient infrastructure of risk management and data acquisition. In a market this efficient, the frontier of advantage has moved decisively from models to information, and whoever controls the freshest and richest data controls what little surprise remains. Information is not merely an input to power: it is power.

The chapter thus completes the empirical arc of the thesis. Chapter 2 argued, on theoretical and institutional grounds, that data have become the decisive resource of contemporary finance; Chapter 3 assembled a measurement apparatus capable of observing a decade of interaction between prices, fear and geopolitics; and this chapter has subjected that apparatus to the most demanding test available, genuine out-of-sample prediction, across twelve architectures spanning the full history of

quantitative method, from the ARMA filters of the 1970s to the gated recurrent networks of the deep-learning era. The verdict is coherent on every level of the analysis: predictability exists, is minute, resides in the sentiment and volatility channels rather than in raw event indicators, is best extracted by disciplined and interpretable methods, resists translation into speculative profit, and rewards instead the patient infrastructure of risk management and data acquisition. The concluding chapter draws these threads together, returns to the research questions posed at the outset, and reflects on what the narrowness of the informational frontier documented here implies, for investors, for institutions, and for the regulation of markets in which intelligence, human and artificial, competes over an ever-shrinking residue of surprise.

5. CONCLUSION

This thesis set out to investigate one of the defining transformations of contemporary finance: the progressive displacement of capital, understood in its traditional sense, by information as the primary source of competitive advantage in financial markets. The guiding intuition, announced in the introduction and developed throughout the entire work, is that the decisive confrontation of our era is no longer fought over territories, physical resources or even market share, but over the capacity to collect, process and interpret data faster and more intelligently than one's competitors. Against this conceptual backdrop, the research pursued a twofold objective. On the theoretical level, it sought to clarify how Big Data, behavioural dynamics and geopolitical uncertainty jointly shape the formation of asset prices and the propagation of volatility across global markets. On the empirical level, it aimed to establish, through a rigorous and fully out-of-sample experiment, whether the informational content embedded in volatility, sentiment-driven uncertainty and geopolitical event indicators can be converted into anticipation of the short-term direction of the United States equity market, and

which of four generations of quantitative methodology performs that conversion best.

The theoretical investigation conducted in the second chapter yields a first, fundamental conclusion: the data-driven paradigm is not a mere technological upgrade of traditional financial analysis, but a structural reconfiguration of the way markets produce, distribute and exploit knowledge. The analysis demonstrated that the informational ecosystem defined by volume, variety, velocity and veracity has dissolved the boundary between financial and non-financial information, elevating alternative data streams to the same analytical dignity as balance sheets and macroeconomic releases. At the same time, the examination of Data-Driven Decision-Making, algorithmic risk management and blockchain convergence revealed that automation generates value precisely where human intermediation is slowest and most error-prone, but that this efficiency is purchased at the cost of new systemic fragilities: the opacity of complex models, the propagation of algorithmic errors at machine speed and the concentration of informational power in a restricted number of institutions emerged as the structural counterparts of the data revolution. The behavioural analysis added a second, complementary insight: the abundance of data does not automatically produce more rational markets. When the volume of available information exceeds the cognitive capacity of participants, imitation becomes a rational protective strategy, informational cascades propagate, and uncertainty acts as a levelling force that suppresses individual cognitive profiles and locks markets into synchronized collective dynamics. This insight proved decisive for the empirical part of the work, because it implies that indicators of fear and uncertainty are not accessory variables but primary carriers of predictive content.

The third pillar of the theoretical framework, devoted to the macroeconomic and geopolitical drivers of volatility, produced an equally clear result: geopolitical risk must be treated as a core macroeconomic determinant rather than as an exogenous disturbance. The comparative reading of Brexit, the COVID-19 pandemic, the

Russia-Ukraine war and the Israel-Hamas conflict revealed that, despite their radical heterogeneity in origin and nature, these events share a common grammar of market impact: each operated as an unanticipated news shock, each activated asymmetric transmission channels, and each demonstrated that the intensity of the market response depends decisively on the pre-existing state of the system. At the same time, the four case studies displayed profoundly different volatility signatures, from the sharp but mean-reverting impulse of the referendum to the systemic explosion of the pandemic and the persistent risk premia embedded by the two wars. The empirical framework of the third chapter translated these insights into a measurable architecture: a decade of daily observations, from June 2016 to June 2026, aligning the S&P 500 with the option-implied fear of the VIX, the news-based anxiety of the Equity Market Uncertainty index and four binary indicators marking the acute windows of the geopolitical shocks. The descriptive evidence confirmed the theoretical expectations with precision: heavy tails, negative skewness, volatility clustering and event windows in which daily volatility reached more than six times its unconditional level.

The comparative experiment of the fourth chapter delivers the central empirical message of this thesis, and it is a message that few would have predicted at the outset. Twelve architectures spanning the entire history of quantitative method, from the ARIMAX and GARCH filters of classical econometrics, through the penalised regressions at the frontier of statistical learning, to ensemble methods, kernel machines and deep recurrent networks, were confronted on identical data, identical features and an identical, strictly chronological validation protocol. The resulting hierarchy inverts eight decades of methodological sophistication: the summit belongs to the shrunken linear models, with the Lasso and Ridge regressions attaining the best out-of-sample ranking ability, followed by the neural network and the logistic regression that lean on the same linear signal, while the flexible non-linear learners cluster at the no-information frontier and the two most complex objects of the study, the recurrent network and the GARCH family,

collapse into degenerate solutions that replicate the bullish base rate without any discriminative power. Equally important is the compression of the performance band: across four orders of magnitude of model capacity, the measurable differences in predictive quality remain confined within a few hundredths of ROC-AUC and a few percentage points of accuracy. The binding constraint on prediction, the experiment demonstrates, is not the sophistication of the algorithm but the informational content of the data, a ceiling imposed by the market rather than by the econometrician.

The economic interpretation of this evidence constitutes the second major empirical conclusion. The performance of the leading models sits close to the frontier that the weak-form Efficient Market Hypothesis defines, and yet the small edge they extract is neither an artefact nor a fluke: five architecturally unrelated estimators converge on the same economic narrative of short-horizon mean reversion modulated by fear, in which spikes of media-driven uncertainty are followed by marginal rebounds as feared scenarios fail to materialise. This convergence is best read through the adaptive-markets lens: the behavioural frictions and collective emotions that the theoretical chapters placed at the centre of modern market dynamics leave a faint but measurable trace in next-day prices. Crucially, the edge is regime-dependent in exactly the direction the theory predicts: the best model is close to uninformative on calm days and reaches its highest directional accuracy on turbulent, high-VIX days, concentrating its informational value precisely in the states of the world where information is scarcest and most precious. The fate of the geopolitical dummies completes the picture: pruned by the very model that won the comparison, they proved individually redundant not because geopolitics is irrelevant, but because its content is absorbed within hours by the continuous fear gauges. Geopolitical shocks are priced through the sentiment infrastructure of the market, and the speed of that absorption is itself the measure of why controlling the quality and timeliness of information access constitutes a form of market power.

A third conclusion concerns the practical destination of predictive technology in finance. The simulated long/flat strategy built on the best model matched the raw return of the passive benchmark over the two-year test period but realised it with visibly lower volatility, improving the risk-adjusted performance by systematically stepping aside on a subset of the dangerous days. The lesson generalises well beyond the specific exercise: on public, daily-frequency data, Machine Learning is a risk-management technology rather than an alpha machine. Its genuine product is not the timing of the market but a continuously updated, calibrated and auditable synthesis of the fear environment, whose reliability rises exactly when turbulence makes it most valuable. The evidence also assigns a precise division of labour within the model set itself: the volatility models of the GARCH family describe the risk environment with near-perfection while contributing nothing to direction, and a production architecture would naturally pair their variance forecasts, used for position sizing and stress design, with penalised-linear directional tilts modulating exposure at the margin. Finally, the comparison delineates the complementarity between human judgement and algorithmic prediction with unusual sharpness: models should own the continuous, high-frequency layer of risk assessment, while human judgement must own the discontinuities, deciding when structural novelty has invalidated the training distribution and overriding the machine precisely when its confidence is least warranted.

These findings carry implications that extend well beyond the academic perimeter. For investors and asset managers, the results argue for a reallocation of the quantitative budget: in low signal-to-noise environments, the marginal investment is better spent on newer, faster and more exclusive informational sources than on additional model complexity, since every increment of sophistication beyond the penalised linear frontier either added nothing, merely repaired its own pathologies, or actively harmed performance. For risk managers, the study demonstrates that headline accuracy is a treacherous criterion in drifting markets, that multi-metric audits centred on threshold-free measures are indispensable for exposing

degenerate models that silently mimic the base rate, and that the interpretability of the winning model, auditable coefficient by coefficient, should be institutionalised as a design criterion rather than celebrated as a fortunate accident. For regulators and policymakers, the thesis identifies the true object of concern: not predictive technology itself, but the governance of its informational substrate. If algorithms are commoditised public goods and data decide, then the concentration of exclusive data streams translates directly into a concentration of what little predictability markets leave available, while the convergence of institutions on similar data and similar model families synchronises their reactions to shocks, transforming individually prudent risk systems into collectively procyclical amplifiers of the very crises documented in the event windows of this study.

Intellectual honesty requires an equally explicit recognition of the boundaries of this work. The empirical analysis concentrates on a single benchmark index of a single market, observed at daily frequency through public indicators; whether the identified patterns generalize to other asset classes, to emerging markets, to individual securities or to intraday horizons remains an open question. The binary modelling of geopolitical events, although methodologically transparent, compresses the internal complexity of each crisis into an on-off indicator that cannot distinguish escalation from de-escalation, and the crisis windows are short and few, so that crisis-conditional estimates rest on limited observations. The single chronological split, while immune to look-ahead contamination, evaluates the models on one particular two-year future, and the economic simulation abstracts from transaction costs that would erode part of the documented risk-adjusted advantage. Finally, as with any exercise in supervised learning on financial data, the relationships learned over the past decade may be altered by future structural transformations of the market itself, a possibility that no validation protocol can exclude and that the adaptive-markets perspective indeed predicts.

These limitations trace, symmetrically, the most promising directions for future research. The most natural extension consists in enriching the feature space with

textual sentiment extracted directly from news flows and social platforms through Natural Language Processing, transforming the qualitative notion of narrative, so central to the second chapter, into quantitative predictive signals measured at high frequency. A second line of development concerns the modelling of events themselves: replacing binary indicators with continuous intensity measures, derived for instance from daily geopolitical risk indices, would allow the models to track the internal dynamics of each crisis rather than merely its presence. On the methodological side, a rolling-origin evaluation scheme would map the time variation of the small predictable component, while attention-based architectures and transformer networks represent the logical successors of the recurrent framework tested here, provided that their additional capacity is disciplined by the same multi-metric audit that exposed the degeneracies of this study. Explainable Artificial Intelligence techniques, such as Shapley-value decompositions, offer a concrete path towards extending the interpretability that regulated finance demands to whatever architecture future evidence may justify. Finally, the extension of the framework to a multi-asset, cross-market setting would make it possible to study not only the prediction of individual market movements but the propagation of informational shocks across the global financial network.

In closing, it is worth returning to the thesis from which this work took its title and its direction. The decade analysed in these pages, opened by an unexpected referendum and traversed by a pandemic and two wars, has demonstrated with empirical force that financial markets have become the seismographs of a world in which information travels faster than policy and often faster than comprehension. The twelve-model experiment returned a verdict that is coherent on every level of the analysis: predictability exists, is minute, lives in the sentiment and volatility channels rather than in raw event indicators, is best extracted by disciplined and interpretable methods, resists conversion into speculative profit, and rewards instead the patient infrastructure of risk management and data acquisition. Artificial Intelligence and Machine Learning do not abolish uncertainty, and this

thesis has documented their limits with the same care with which it has documented their strength; but they redefine the frontier along which uncertainty can be measured, priced and managed. If twelve architectures spanning eight decades of methodology, fed with identical public information, converge within a few hundredths of the efficiency frontier, then the residual competitive space lies almost entirely in the information set itself, in proprietary flows, real-time sentiment and the infrastructure to process them earlier than competitors. The conclusion, then, is not that machines will replace financial judgment, but that financial judgment itself is being reconstituted around the mastery of data. Whoever controls information does not merely observe the market: increasingly, they shape it. It is in this precise sense that the most significant battle of our time is no longer fought over resources, territories or markets. It is fought over data.

6. BIBLIOGRAPHY

- Albulescu, C.T. (2021). COVID-19 and the United States financial markets' volatility. Finance Research Letters, 38, 101699.*
- Baker, S.R., Bloom, N., & Davis, S.J. (2016). Measuring economic policy uncertainty. The Quarterly Journal of Economics, 131(4), 1593-1636.*
- Baker, S.R., Bloom, N., Davis, S.J., Kost, K., Sammon, M., & Viratyosin, T. (2020). The unprecedented stock market reaction to COVID-19. The Review of Asset Pricing Studies, 10(4), 742-758.*
- Banerjee, A.V. (1992). A simple model of herd behavior. The Quarterly Journal of Economics, 107(3), 797-817.*
- Barberis, N., & Thaler, R. (2003). A survey of behavioral finance. In G.M. Constantinides, M. Harris & R.M. Stulz (Eds.), Handbook of the Economics of Finance (Vol. 1, pp. 1053-1128). Elsevier.*
- Baur, D.G., & Lucey, B.M. (2010). Is gold a hedge or a safe haven? An analysis of stocks, bonds and gold. Financial Review, 45(2), 217-229.*
- Belke, A., Dubova, I., & Osowski, T. (2018). Policy uncertainty and international financial markets: the case of Brexit. Applied Economics, 50(34-35), 3752-3770.*
- Bikhchandani, S., Hirshleifer, D., & Welch, I. (1992). A theory of fads, fashion, custom, and cultural change as informational cascades. Journal of Political Economy, 100(5), 992-1026.*
- Bikhchandani, S., & Sharma, S. (2000). Herd behavior in financial markets. IMF Staff Papers, 47(3), 279-310.*
- Bollen, J., Mao, H., & Zeng, X. (2011). Twitter mood predicts the stock market. Journal of Computational Science, 2(1), 1-8.*

- Bollerslev, T. (1986). *Generalized autoregressive conditional heteroskedasticity. Journal of Econometrics*, 31(3), 307-327.
- Boungou, W., & Yatié, A. (2022). *The impact of the Russia-Ukraine war on world stock market returns. Economics Letters*, 215, 110516.
- Box, G.E.P., & Jenkins, G.M. (1976). *Time Series Analysis: Forecasting and Control (Revised ed.)*. Holden-Day.
- Breiman, L. (2001). *Random forests. Machine Learning*, 45(1), 5-32.
- Caldara, D., & Iacoviello, M. (2022). *Measuring geopolitical risk. American Economic Review*, 112(4), 1194-1225.
- 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*, 785-794.
- Christoffersen, P.F., & Diebold, F.X. (2006). *Financial asset returns, direction-of-change forecasting, and volatility dynamics. Management Science*, 52(8), 1273-1287.
- Cortes, C., & Vapnik, V. (1995). *Support-vector networks. Machine Learning*, 20(3), 273-297.
- Dhingra, S., Ottaviano, G., Sampson, T., & Van Reenen, J. (2016). *The impact of Brexit on foreign investment in the UK. CEP Brexit Analysis No. 3, Centre for Economic Performance, London School of Economics*.
- Diebold, F.X., & Yilmaz, K. (2012). *Better to give than to receive: Predictive directional measurement of volatility spillovers. International Journal of Forecasting*, 28(1), 57-66.
- Engle, R.F. (1982). *Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. Econometrica*, 50(4), 987-1007.
- Fama, E.F. (1970). *Efficient capital markets: A review of theory and empirical work. The Journal of Finance*, 25(2), 383-417.

- Fischer, T., & Krauss, C. (2018). Deep learning with long short-term memory networks for financial market predictions. *European Journal of Operational Research*, 270(2), 654-669.
- Friedman, J.H. (2001). Greedy function approximation: A gradient boosting machine. *The Annals of Statistics*, 29(5), 1189-1232.
- Giudici, P. (2018). Fintech risk management: A research challenge for artificial intelligence in finance. *Frontiers in Artificial Intelligence*, 1, 1.
- Giudici, P., & Raffinetti, E. (2021). Shapley-Lorenz eXplainable Artificial Intelligence. *Expert Systems with Applications*, 167, 114104.
- Goldstein, I., Spatt, C.S., & Ye, M. (2021). Big data in finance. *The Review of Financial Studies*, 34(7), 3213-3225.
- Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep Learning*. MIT Press.
- Gu, S., Kelly, B., & Xiu, D. (2020). Empirical asset pricing via machine learning. *The Review of Financial Studies*, 33(5), 2223-2273.
- Hasan, M.M., Popp, J., & Oláh, J. (2020). Current landscape and influence of big data on finance. *Journal of Big Data*, 7, 21.
- Heaton, J.B., Polson, N.G., & Witte, J.H. (2017). Deep learning for finance: Deep portfolios. *Applied Stochastic Models in Business and Industry*, 33(1), 3-12.
- Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural Computation*, 9(8), 1735-1780.
- Hoerl, A.E., & Kennard, R.W. (1970). Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics*, 12(1), 55-67.
- Hosmer, D.W., Lemeshow, S., & Sturdivant, R.X. (2013). *Applied Logistic Regression* (3rd ed.). Wiley.
- Izzeldin, M., Muradoğlu, Y.G., Pappas, V., Petropoulou, A., & Sivaprasad, S. (2023). The impact of the Russia-Ukraine war on global financial markets.

- International Review of Financial Analysis*, 87, 102598.
- James, G., Witten, D., Hastie, T., & Tibshirani, R. (2021). *An Introduction to Statistical Learning with Applications in R* (2nd ed.). Springer.
- Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica*, 47(2), 263-291.
- Krauss, C., Do, X.A., & Huck, N. (2017). Deep neural networks, gradient-boosted trees, random forests: Statistical arbitrage on the S&P 500. *European Journal of Operational Research*, 259(2), 689-702.
- Leung, M.T., Daouk, H., & Chen, A.-S. (2000). Forecasting stock indices: A comparison of classification and level estimation models. *International Journal of Forecasting*, 16(2), 173-190.
- Lo, A.W. (2004). The Adaptive Markets Hypothesis: Market efficiency from an evolutionary perspective. *The Journal of Portfolio Management*, 30(5), 15-29.
- Nelson, D.B. (1991). Conditional heteroskedasticity in asset returns: A new approach. *Econometrica*, 59(2), 347-370.
- Pástor, L., & Veronesi, P. (2013). Political uncertainty and risk premia. *Journal of Financial Economics*, 110(3), 520-545.
- Ramiah, V., Pham, H.N.A., & Moosa, I. (2017). The sectoral effects of Brexit on the British economy: Early evidence from the reaction of the stock market. *Applied Economics*, 49(26), 2508-2514.
- Shiller, R.J. (2015). *Irrational Exuberance* (3rd ed.). Princeton University Press.
- Tetlock, P.C. (2007). Giving content to investor sentiment: The role of media in the stock market. *The Journal of Finance*, 62(3), 1139-1168.
- Tibshirani, R. (1996). Regression shrinkage and selection via the Lasso. *Journal of the Royal Statistical Society: Series B*, 58(1), 267-288.

Whaley, R.E. (2009). *Understanding the VIX*. *The Journal of Portfolio Management*, 35(3), 98-105.

World Bank (2023). *Commodity Markets Outlook, October 2023: Under the Shadow of Geopolitical Risks*. World Bank Group, Washington, DC.

Yermack, D. (2017). *Corporate governance and blockchains*. *Review of Finance*, 21(1), 7-31.

Zhang, D., Hu, M., & Ji, Q. (2020). *Financial markets under the global pandemic of COVID-19*. *Finance Research Letters*, 36, 101528.

Cboe Global Markets, Cboe Volatility Index (VIX):
https://www.cboe.com/tradable_products/vix/

Economic Policy Uncertainty project (Baker, Bloom & Davis):
<https://www.policyuncertainty.com/>

Federal Reserve Bank of St. Louis, FRED Economic Data, S&P 500 (SP500):
<https://fred.stlouisfed.org/series/SP500>

Federal Reserve Bank of St. Louis, FRED Economic Data, CBOE Volatility Index (VIXCLS): <https://fred.stlouisfed.org/series/VIXCLS>

Federal Reserve Bank of St. Louis, FRED Economic Data, Equity Market-related Economic Uncertainty Index (WLEMUINDXD):
<https://fred.stlouisfed.org/series/WLEMUINDXD>

Geopolitical Risk (GPR) Index (Caldara & Iacoviello):
<https://www.matteoiacoviello.com/gpr.htm>

IFRS Foundation, International Financial Reporting Standards:
<https://www.ifrs.org/>

International Organization for Standardization (ISO): <https://www.iso.org/>

Keras, Deep Learning API documentation: <https://keras.io/>

Scikit-learn, Machine Learning in Python documentation: <https://scikit-learn.org/>

*TensorFlow, An end-to-end open source machine learning platform:
<https://www.tensorflow.org/>*

*World Health Organization, WHO Director-General's opening remarks at the
media briefing on COVID-19, 11 March 2020: <https://www.who.int/>*