



University of Essex

School of Mathematics, Statistics
and Actuarial Science

MA981 DISSERTATION

A Statistical Mechanics Approach to Stock Market Dynamics: Modeling Price Fluctuations and Volatility

Md Ashraful Alam Gazi

Supervisor: **Dr. Alastair Litterick**

September 25, 2025
Colchester

Contents

1	Introduction	5
2	Literature Review	8
3	Methodology	11
3.1	Overview of the Approach	11
3.2	Data Collection and Preprocessing	12
3.2.1	Data Sources	12
3.2.2	Time Frame	13
3.2.3	Data Cleaning	13
3.3	Baseline Financial Models	15
3.3.1	Geometric Brownian Motion (GBM)	15
3.3.2	GARCH(1,1) & GARCH with Drift	16
3.4	Physics-Inspired Models	19
3.4.1	Langevin Equation — drift + noise (from particle physics to prices)	19
3.4.2	Ornstein–Uhlenbeck (OU) Process – Mean Reversion Mechanics	23
3.4.3	Fokker–Planck Equation – Probability Distribution Evolution .	25
3.5	Entropy-Based Market Diagnostics	27
3.5.1	Shannon Entropy – measuring uncertainty in returns	27
3.5.2	Tsallis Entropy – generalisation for heavy-tailed distributions .	28
3.5.3	Rolling Window Implementation	28
3.6	Simulation Framework	29
3.6.1	Practical Steps	29
3.6.2	Parameter Estimation	30
3.6.3	Simulation Steps for Path-Based Models (GBM, GARCH, Langevin, OU)	31

3.6.4	Numerical Simulation Method: Euler–Maruyama	32
3.6.5	Simulation Steps for the Fokker–Planck Model	32
3.6.6	Post-Simulation Processing	33
3.7	Model Evaluation Metrics	34
3.7.1	RMSE – Root Mean Square Error	34
3.7.2	AIC & BIC – Balancing Fit and Complexity	34
3.7.3	Kolmogorov–Smirnov (KS) Test	35
3.7.4	Wasserstein Distance	35
3.7.5	Jarque–Bera (JB) Test for Normality	36
3.7.6	Augmented Dickey–Fuller (ADF) Test for Stationarity	36
3.7.7	ARCH Test for Volatility Clustering	37
3.7.8	Skewness and Kurtosis (Distributional Moments)	37
3.8	Comparative Analysis Procedure	37
3.9	Tools and Libraries	38
4	Results & Discussion	39
4.1	Descriptive Statistics and Correlation	40
4.1.1	Core Statistics	40
4.1.2	Distributional Moments	41
4.1.3	Tails and Statistical Tests	42
4.1.4	Correlation Matrices	43
4.2	Volatility Patterns and Return Distributions	46
4.2.1	Price and Return Patterns	46
4.2.2	Rolling Volatility	48
4.2.3	Return Distributions	49
4.3	Entropy Analysis	51
4.3.1	Entropy and Volatility	51
4.3.2	Lead–Lag Analysis	52
4.3.3	Cross-Market Comparisons	53
4.4	Model Parameter Calibration	54
4.4.1	Cross-Market Comparison	55
4.5	Empirical Cumulative Distribution Function (ECDF) Analysis	56
4.5.1	Comparison of Model Fit	57

4.5.2	Strengths and Weaknesses	58
4.6	Model Evaluation Metrics	58
4.6.1	Comparative Ranking of Models	60
4.6.2	Comparison of Entropy Fit	61
5	Conclusions	64
5.1	Limitations and Market Structure Considerations	65
5.2	Future Work	66
5.2.1	Market Regimes and Time Horizons	66
5.2.2	Hybrid and Extended Models	66
5.2.3	Broader Scope	67
A	Supplementary Figures	68
A.1	Extended Result Visualisations	68
A.1.1	Bitcoin (BTC)	69
A.1.2	S&P 500 (GSPC)	72
A.1.3	QQQ (NASDAQ-100)	76
A.1.4	FTSE 100	79
A.1.5	Nikkei 225	83
B	Methodology–Code Alignment Statement	87

Introduction

Financial markets are often thought of as unpredictable arenas where prices can swing dramatically in response to news, policy changes, or shifts in investor mood. Traditional models in finance, such as the well-known Geometric Brownian Motion (GBM), assume that prices evolve smoothly, with a fixed level of risk, and that returns follow a tidy “bell curve.” These models are mathematically elegant and simple to use, but they frequently fall short when compared with reality. In practice, extreme rises and crashes occur far more often than a bell curve would suggest, calm periods alternate with turbulence, and price changes tend to cluster rather than appear in isolation [1].

Because of these shortcomings, researchers have turned to physics for inspiration. In physics, systems such as gases, fluids, or magnets are made up of countless interacting parts. Although the behaviour of each particle seems random, the system as a whole often follows clear, sometimes surprising, patterns. Markets can be seen in a similar way: instead of particles, we have traders, institutions, and algorithms, all interacting in ways that generate feedback loops and collective behaviour. This perspective has given rise to the field of *econophysics* [1, 4].

One of the striking findings from econophysics is that financial markets exhibit so-called “stylized facts” that echo behaviours in physics:

- **Fat tails:** extreme events like crashes or rallies are much more common than predicted by a bell curve [1].
- **Volatility clustering:** big changes in price tend to follow other big changes, just as

calm periods tend to follow calm [4].

- **Scaling laws and memory effects:** market fluctuations show repeating patterns across different time scales.

To capture such features, this dissertation explores both classical and physics-inspired models side by side. Along with GBM and GARCH (a finance model that accounts for changing volatility), we test models borrowed from physics:

- the **Langevin equation**, which balances steady trends with random shocks [8];
- the **Ornstein–Uhlenbeck process**, which models the tendency for prices to revert toward a long-term average [8];
- and the **Fokker–Planck equation**, which focuses on how the entire probability distribution of returns evolves over time [9].

Beyond these, we also borrow concepts from information theory. In particular, *entropy*—a measure of uncertainty or disorder—is used as a diagnostic tool. Unlike volatility, which only measures the size of price movements, entropy captures their unpredictability and changing patterns. Rising entropy may act as an early-warning signal of instability, echoing research that shows similar warning signs in both natural and financial systems [10].

Another important thread of research concerns reflexivity—the idea that market behaviour is shaped by traders reacting not just to fundamentals, but also to each other’s actions. Hawkes process models, for instance, have been used to quantify how much today’s trades trigger tomorrow’s activity, offering insights into flash crashes and liquidity crises [11]. More recently, hybrid approaches that combine classical models with deep learning have emerged. Neural networks have been trained to calibrate the Log-Periodic Power Law Singularity (LPPLS) model, providing promising tools for forecasting the timing of market crashes and other critical transitions [16].

The study analyses over three decades of daily data (1992–2025) from major global markets: the S&P 500 and Nasdaq (United States), the FTSE 100 (United Kingdom), the Nikkei 225 (Japan), and Bitcoin as a highly volatile digital asset. Together, these markets provide a testing ground across both stable and turbulent environments.

The main aim of this dissertation is not only to evaluate which models reproduce real market patterns most faithfully but also to highlight the usefulness of physics-based approaches for financial risk assessment. By framing markets as complex adaptive systems, this research suggests that tools such as entropy could complement traditional finance metrics and potentially help detect instability before crises fully unfold.

Literature Review

Classical finance has long relied on mathematical models that assume markets behave in neat, predictable ways. One of the most widely used is the *Geometric Brownian Motion* (*GBM*), which treats prices as if they drift steadily upward or downward with a fixed level of risk and random day-to-day noise. In this framework, returns should follow a normal bell-curve distribution. Real markets, however, look very different. Extreme crashes and rallies happen far more often than the bell curve suggests, calm periods are often followed by storms, and scaling laws emerge that simple models cannot explain.

In the 1990s, Mantegna and Stanley [1] showed that financial markets share features with complex physical systems such as fluids or magnets. They demonstrated that price changes follow irregular patterns shaped by hidden scaling laws and non-equilibrium dynamics. Farmer and Joshi [5] went further, showing that traders themselves actively generate patterns: when investors follow trends or buy undervalued assets, they amplify small bits of noise and create bursts of volatility. Similarly, Lux and Marchesi [4] built agent-based models illustrating how herding behaviour and feedback loops can produce crashes and fat-tailed returns, much like physical systems approaching critical points.

These insights support the use of models borrowed from physics. The Langevin and Ornstein–Uhlenbeck processes, for instance, describe systems where outcomes fluctuate around an average but are constantly disturbed by random shocks. Hänggi and Thomas [8] introduced the Fokker–Planck framework as a way of modelling both drift (long-term tendency) and diffusion (random shocks). Later work adapted this to finance, allowing researchers to model the entire distribution of returns—including the

likelihood of extreme events—rather than just average price paths [9].

Other approaches study unpredictability through the lens of *entropy*, a concept from physics that measures disorder. Applied to finance, entropy captures how uncertain or unstable markets are. Zunino et al. [10] reviewed entropy-based methods, cautioning that some can be biased, but still concluded that entropy is a valuable way to detect turbulence and inefficiency. Shannon entropy provides a basic measure of unpredictability, while Tsallis entropy generalises the idea to give more weight to rare, extreme events. When calculated in rolling windows, these measures reveal how market complexity builds before and during crises.

Recent studies continue to strengthen these links between physics and finance. Das and Fong [9] showed that the Fokker–Planck equation can handle systems with memory effects, making it especially suitable for financial time series. Filimonov and Sornette [11] warned that some advanced models risk detecting “false” signs of instability if they ignore structural shifts, highlighting the need for robust evaluation with metrics such as AIC, BIC, and RMSE. At the same time, log-periodic power law (LPPL) models describe bubbles and crashes as critical phenomena, and new work has combined them with deep learning to improve predictive power [16].

Physics-inspired models each bring certain strengths but also clear limits. For instance, classical Geometric Brownian Motion (GBM) is simple and elegant but fails to capture real-world fat tails and sudden crashes [1]. GARCH models handle volatility clustering much better [3], though they cannot reliably predict the timing of extreme moves. Agent-based and Langevin-style models explain how trader behaviour or feedback loops create patterns [5, 7], but they often rely on strong assumptions or simplifications that limit their realism.

More specialised tools, such as the Log-Periodic Power Law (LPPL) and entropy measures, attempt to spot bubbles or early-warning signals of crises [12, 13, 14]. These approaches are innovative and sometimes effective, yet they can be sensitive to parameter choices or give mixed results across different events. Taken together, the literature suggests no single “best” model; rather, each captures a piece of market complexity. This has led many researchers to argue that hybrid approaches, combining insights from finance and physics, may offer the most promising path forward.

Physics-Inspired or Physics-Adapted?

A key debate in econophysics is whether markets should be described by the exact same equations used in physics, or whether those equations need to be adapted. Classical tools such as Brownian motion or Langevin dynamics are useful starting points, but they assume smooth and Gaussian behaviour that real markets often do not show. This has led researchers to argue that the best progress comes not from copying physics directly, but from creating finance-specific versions that capture the same spirit while fitting the data better.

For example, Lin, Ren and Sornette [15] reshaped the log-periodic power law model so that it more accurately described financial bubbles, while Nielsen, Raissi and Sornette [16] extended it further using modern machine learning techniques. Both cases show how ideas borrowed from physics can be adjusted to the special conditions of markets. In this dissertation the focus is on comparing existing models rather than creating new ones, but these works highlight an important point: physics-inspired models are most useful when they are treated as analogies to be adapted, not as laws to be applied wholesale.

Overall, the literature suggests that markets are best viewed as **complex adaptive systems**, more like ecosystems or weather patterns than like coin tosses. Traditional Gaussian-based models such as GBM and GARCH remain useful benchmarks, but they miss deeper dynamics such as feedback, instability, and fat tails. Physics-inspired models—including Langevin, Ornstein–Uhlenbeck, Fokker–Planck, and LPPL—offer richer ways to capture market behaviour. Entropy adds an extra layer of insight, potentially serving as an early-warning signal of instability. This dissertation builds on these ideas by comparing classical and physics-inspired models and exploring whether entropy can help identify hidden risks in global markets.

Methodology

3.1 Overview of the Approach

The main goal of this research is to test whether models inspired by physics can describe how stock markets behave more accurately than the traditional finance models used for decades. Financial markets, much like physical systems, are made up of many interacting parts—investors, institutions, and trading algorithms—that together create unpredictable patterns.

This is similar to gases or fluids in physics, where countless particles interact under both regular forces and random shocks. Statistical mechanics, the branch of physics that studies such systems, has long been applied in econophysics and provides the tools to capture real market features such as volatility clustering, heavy tails in returns, and sudden crashes.

Traditional Financial Stochastic Models include Geometric Brownian Motion (GBM), GARCH(1,1), and GARCH(1,1) with drift. Physics-inspired models include the Langevin Equation, Ornstein–Uhlenbeck (OU) Process, and Fokker–Planck Equation. More about them will be discussed in Section 3.3 and Section 3.4.

The key question of this research is whether borrowing tools from physics can help us describe markets more accurately, spot risks earlier, and capture extreme events better than standard finance models. Since no single model can perfectly represent a market, several models will be tested, each judged not only on how closely it matches

real price movements but also on whether it reproduces key market patterns like heavy tails, volatility clustering, and skewness.

3.2 Data Collection and Preprocessing

Before any model can be tested, reliable historical market data are essential, as poor data can produce misleading results regardless of model sophistication.

3.2.1 Data Sources

The dataset was obtained using the Yahoo Finance API, accessed via the `yfinance` Python library. The key advantages of the Yahoo Finance API are:

- It is free and accessible, so the data of extensive historical market can be downloaded without licensing fees.
- It provides reliable coverage of major global indices, ETFs, and cryptocurrencies in a consistent format.
- It automatically accounts for things like stock splits and dividends, so the price data stays accurate and easy to compare over time.

The selected tickers for this study are:

- SPY – S&P 500 ETF
- QQQ – Nasdaq 100 ETF
- FTSE – FTSE 100 Index
- N225 – Nikkei 225 Index
- BTC-USD – Bitcoin (USD)

The chosen assets were selected to provide both variety and reliability. They cover different regions—the U.S., UK, and Japan—as well as Bitcoin, which adds a decentralised, non-traditional asset. They also span sectors, with Nasdaq being technology-focused, the FTSE 100 including energy, mining and finance, and the Nikkei reflecting strong industrial and export components.

The set also captures different volatility levels too: Bitcoin is highly volatile, the FTSE 100 is relatively stable, and the S&P 500 sits in between, making it possible to test models across market conditions. Finally, all of these assets are highly liquid and widely traded, reducing problems with missing data or distortions from low activity.

3.2.2 Time Frame

The time frame used in this study spans from **January 1, 1992** to **January 1, 2025**. This time period was chosen because it includes many different market conditions, such as::

- The early 1990s economic recovery,
- The dot-com boom and subsequent crash,
- The 2008 global financial crisis,
- The prolonged bull market of the 2010s,
- The sharp COVID-19 market drop and rapid rebound,
- The inflation-driven volatility of 2022–2023.

This mix of assets makes it possible to test the models in both calm and turbulent market conditions, showing how well they hold up across different environments. With more than three decades of daily data and thousands of observations for each asset, the dataset is large enough to provide a strong statistical basis for reliable model calibration and testing.

3.2.3 Data Cleaning

Once downloaded, the raw data required several preparation steps before modelling. The main steps were as follows:

1. **Handling Missing Data:** The raw data contained gaps due to market holidays and occasional missing entries. These were filled using a forward-fill method, carrying forward the previous day's closing price. This avoids introducing sudden artificial jumps that could distort return calculations.

2. **Adjustment for Corporate Actions:** Prices were adjusted for dividends, stock splits, and similar events using Yahoo Finance's *adjusted closing prices*. This ensures that price changes reflect genuine market movements rather than technical adjustments.
3. **Log Return Calculation:** Adjusted prices were converted into daily logarithmic returns using:

$$r_t = \ln \left(\frac{P_t}{P_{t-1}} \right)$$

where P_t is the adjusted closing price on day t . Log returns were preferred over simple returns because:

- They sum naturally over time,
- They tend to be closer to normally distributed,
- They treat upward and downward moves symmetrically,
- They naturally account for differences in price scale between assets.

4. **Outlier and Stationarity Checks:** Outliers were examined visually and via descriptive statistics (mean, standard deviation, skewness, kurtosis). No unrealistic spikes (e.g., a hundredfold price jump in a single day) were detected, so no manual corrections were required.

A preliminary stationarity check indicated that, while volatility rises in crises, daily returns are sufficiently stable for modelling. These observations are substantiated later in Chapter 4: Tables 4.1 and 4.2 report summary statistics (including skewness and kurtosis) that capture heavy tails and asymmetry, and Table 4.3 reports Augmented Dickey–Fuller (ADF) tests confirming return stationarity alongside significant ARCH effects evidencing volatility clustering.

After preprocessing, the dataset comprised five clean and consistent daily log return series, each spanning more than three decades. These were ready for calibration in both traditional stochastic finance models and physics-inspired approaches. Every step—from ticker selection to return computation—was carried out with a focus on accuracy, comparability, and strong statistical integrity.

3.3 Baseline Financial Models

3.3.1 Geometric Brownian Motion (GBM)

Geometric Brownian Motion (GBM) is one of the most widely used models in finance and serves as the foundation for the Black–Scholes option pricing framework. It models stock prices as a type of random walk, but with two key components. The first is a drift term, which captures the asset's average long-term growth or decline. The second consists of random shocks, representing unpredictable day-to-day fluctuations driven by news, market sentiment, and other unforeseen events.

The GBM model is described by a stochastic differential equation:

$$dS_t = \mu S_t dt + \sigma S_t dW_t$$

Where:

- S_t = asset price at time t
- μ = drift rate (average growth rate)
- σ = volatility (standard deviation of returns)
- W_t = a Wiener process (a mathematical representation of pure randomness, often called Brownian motion)
- dt = small time step
- dW_t = random shock in that time step

Solving the GBM equation gives:

$$S_t = S_0 \exp \left(\left(\mu - \frac{\sigma^2}{2} \right) t + \sigma W_t \right)$$

This tells us:

Price = Starting price \times exponential of (drift + randomness)

The $-\frac{\sigma^2}{2}$ term corrects for the fact that volatility skews the average.

From our historical log returns:

Drift (μ) is estimated as the average daily log return.

Volatility (σ) is estimated as the standard deviation of daily log returns.

From the historical log returns, the drift (μ) was calculated as the average daily log return, while the volatility (σ) was measured as the standard deviation of daily log returns. These parameters were estimated separately for each asset—SPY, QQQ, FTSE, Nikkei, and BTC—using the complete 1992–2025 dataset. By relying on the full historical range, the estimates capture the influence of all market conditions, from periods of rapid growth to major crashes and more stable phases. Using log returns for these calculations is also consistent with the mathematical foundations of the GBM model, ensuring theoretical alignment as well as practical accuracy.

Once the drift (μ) and volatility (σ) are estimated, the GBM model can be used to generate synthetic price paths. The process begins with the first observed price (S_0), and for each subsequent day, a random number is drawn from a standard normal distribution to represent the daily market shock. The GBM formula is then applied to calculate the next day's price, and this step is repeated until the full time period is simulated.

To gain a more complete picture, multiple simulations are run rather than relying on a single path. This is because one simulation could happen to follow an unusually lucky or unlucky trajectory, while many simulations together reveal the full range of outcomes the model predicts. This distribution of possible price paths helps illustrate both the expected trend and the inherent uncertainty captured by GBM.

Limitations of GBM

Although GBM is elegant, it has well-known shortcomings:

- Constant volatility assumption – in reality, volatility changes over time (especially during crises).
- No volatility clustering – turbulent days tend to follow turbulent days, but GBM treats each day's volatility as independent.
- No heavy tails – extreme events (e.g., 1987 crash, 2020 COVID-19 drop) occur more often in reality than GBM predicts.

3.3.2 GARCH(1,1) & GARCH with Drift

In the real world, market volatility is not constant — it comes in bursts. Periods of high volatility (large daily price changes) often cluster together, and periods of calm can

persist for weeks or months. This behaviour is called volatility clustering and is one of the “stylised facts” of financial markets. GBM cannot capture this: it assumes that daily volatility is constant and independent from one day to the next. To overcome this, GARCH models are used.

GARCH, short for Generalised Autoregressive Conditional Heteroskedasticity, is a model designed to capture how market volatility changes over time. In simple terms, heteroskedasticity means that volatility is not constant but varies from one period to another, autoregressive means that today’s volatility is influenced by yesterday’s, and conditional means that the volatility estimate is continually updated using the most recent data.

It assumes that today’s volatility comes from two main sources: the level of volatility observed yesterday and the size of yesterday’s “surprise” move—how much the market shifted in an unexpected way.

Mathematical form of GARCH(1,1)

The GARCH(1,1) model has two parts:

Return equation:

$$r_t = \mu + \epsilon_t$$

Where:

- r_t = return at time t (log return in our case)
- μ = mean return
- ϵ_t = shock or innovation at time t (the unpredictable part of the return)

Volatility equation:

$$\sigma_t^2 = \omega + \alpha \epsilon_{t-1}^2 + \beta \sigma_{t-1}^2$$

Where:

- σ_t^2 = variance (volatility squared) at time t
- ω = baseline variance (long-term average volatility)
- α = weight given to yesterday’s shock (new information)
- β = weight given to yesterday’s volatility estimate (old information)

In a GARCH(1,1) model, the notation “(1,1)” means the model uses one lag of the squared shock (ϵ_{t-1}^2) and one lag of the previous variance (σ_{t-1}^2) to forecast current volatility. This is the simplest yet highly effective form of GARCH for capturing volatility clustering, which is why it is widely used in both academic studies and industry applications.

In its basic form, GARCH without drift, the mean return (μ) is assumed to be zero or constant over time. This version is useful when the focus is solely on modelling volatility patterns rather than the long-term growth or decline of the market. In GARCH with drift, μ is allowed to be a non-zero constant, introducing an average daily return into the model. This adjustment is important for markets like equities, where long-term historical data shows a consistent upward trend, making the simulated price paths more realistic.

Model Parameters

Model parameters— ω , α , β , and μ (for drift models)—are estimated using maximum likelihood estimation (MLE) on the historical log returns. MLE is a standard econometric technique that finds the parameter values most likely to produce the observed data, and it is readily implemented in Python’s `arch` package. Once parameters are estimated, the simulation process involves generating random shocks from a normal distribution, updating daily volatility using the GARCH equation, calculating simulated returns from these volatilities, and then compounding them from the starting price to obtain simulated price paths.

GARCH plays a key role in this study because it overcomes a major limitation of GBM by allowing volatility to vary over time. It also captures the “memory” effect seen in markets, where periods of high or low volatility tend to persist rather than changing abruptly—meaning when markets are very calm or very turbulent, that state usually continues for a while rather than shifting suddenly from one day to the next.

Limitations

- GARCH assumes shocks are normally distributed, but real market returns have fat tails — extreme events are more common than the normal distribution predicts.
- It doesn’t directly model mean reversion in prices (only in volatility).
- It can lag behind sudden changes — if volatility spikes abruptly, the model adjusts

only gradually.

Despite its limitations, GARCH remains one of the most successful and widely used volatility models in practice. Its strong empirical performance makes it an essential benchmark in this study, setting a high standard that any physics-based model must exceed in order to demonstrate superior volatility forecasting ability.

3.4 Physics-Inspired Models

3.4.1 Langevin Equation — drift + noise (from particle physics to prices)

In statistical mechanics, the Langevin equation describes the motion of a particle influenced by two simultaneous forces: a systematic “drift” that pushes it in a specific direction, and a random “noise” caused by countless small collisions with surrounding molecules.

Financial markets behave in a similar way, combining slow, structured forces such as economic growth, risk premia, policy changes, and liquidity with fast, unpredictable shocks from news events and trading activity. The Langevin framework provides a way to model both components explicitly, capturing the balance between underlying trends and sudden market fluctuations [7].

General form (continuous time)

The Langevin equation is a way to describe how something changes over time when it is influenced by both regular forces and random shocks. In this study, the variable X_t represents the **return** (the percentage change in price), rather than the price itself.

The general form is:

$$dX_t = F(X_t, t) dt + \sqrt{2D} dW_t$$

where:

- $F(X_t, t) dt$ is the **drift**, which represents the systematic or predictable influence (a gentle push in one direction).

- $\sqrt{2D} dW_t$ is the **random noise**. Here $D > 0$ measures the strength of the randomness, and W_t is a Wiener process, a mathematical way of modelling pure randomness.

We use **returns** instead of prices because returns add up neatly over time, capture realistic patterns like volatility changes, and prevent impossible outcomes such as negative prices. In practice, we first simulate returns using this equation, and then compound them to rebuild the price series. This approach ensures numerical stability and matches the way finance typically works with log returns.

How it is implemented?

We tested two practical drift choices within the Langevin framework for daily log-returns r_t :

Constant drift + white noise (baseline Langevin for returns):

$$dr_t = \mu_L dt + \sqrt{2D} dW_t$$

μ_L : average daily return (a constant).

D : diffusion level (randomness strength).

Weak stabilising drift + white noise (soft pull toward a typical level):

$$dr_t = -\gamma r_t dt + \sqrt{2D} dW_t$$

$\gamma > 0$: weak mean-reversion in returns (keeps returns from drifting too far for too long).

This model uses a Langevin form with a simple linear drift, but instead of applying it directly to prices, it is applied to returns (r_t). In this sense, it can be seen as the return-space counterpart of the Ornstein–Uhlenbeck (OU) process, while the OU model itself is kept for Section 3.4.2, where it is applied to prices.

The constant-drift version provides the most straightforward way to translate the “force + noise” idea into return dynamics [7], whereas the weak-reversion variant adds a small corrective pull that prevents simulated returns from producing unrealistically long upward or downward streaks. This approach allows us to capture realistic return behaviour without immediately moving to the full OU model on prices, which is discussed separately in Section 3.4.2.

Parameter estimation

Given the observed daily log-returns $\{r_t\}$ (1992–2025):

Constant-drift Langevin

$$\hat{\mu}_L = \text{mean}(r_t)$$

$$\hat{D} = \frac{1}{2} \times \text{var}(r_t)$$

(since $\text{var}(dr_t) = 2D dt$ with $dt = 1$).

This is the maximum likelihood estimator under Gaussian innovations and is consistent with the assumption of the model.

Weak-reversion Langevin

Regress r_{t+1} on r_t :

$$r_{t+1} = (1 - \gamma)r_t + \eta_t$$

$$\hat{\gamma} = 1 - \hat{\beta}$$

where $\hat{\beta}$ is the OLS slope.

$$\hat{D} = \frac{1}{2} \times \text{var}(\eta_t)$$

This linear-Gaussian state equation yields closed-form MLE via OLS; it's transparent and robust for daily data.

We estimate parameters separately for each instrument (SPY, QQQ, FTSE, N225, BTC-USD) so that each model reflects that asset's own average return and volatility scale.

Why use Langevin?

The Langevin model describes markets in terms of two main forces: a steady directional pull (the drift) and random disturbances (the noise). In this work, it is applied to returns rather than prices, which allows the random component to connect cleanly with the entropy measures introduced in Section 3.5. This makes Langevin a practical setting for testing early-warning indicators of market stress.

Beyond its simplicity, the model is valuable because it captures the basic “force + noise” idea from physics in its most direct form. It also provides a natural bridge to the Fokker–Planck framework in Section 3.4.3, which extends the analysis by modelling the evolution of the entire return distribution. At the same time, it offers an interpretable

baseline for entropy analysis: changes in the diffusion parameter D directly alter the level of uncertainty, making it possible to test whether entropy increases ahead of volatility spikes.

For the simulations, daily log returns from 1992–2025 are used together with an initial price P_0 . Parameters are estimated separately for each asset: (μ_L, D) for the constant-drift case and (γ, D) for the weak-reversion case. A fixed random seed ensures reproducibility. To capture the model’s average behaviour as well as its uncertainty, between 500 and 1,000 simulated paths are generated. The outputs include simulated prices (P_t), returns (r_t), evaluation metrics (RMSE, AIC, BIC, KS, Wasserstein), and rolling entropy measures (Shannon and Tsallis).

Limitations of the Langevin Approach

Although the Langevin model provides a clear “force + noise” framework, it also has several limitations. First, it assumes that shocks are Gaussian, following a bell-shaped distribution. In practice, financial returns often show much heavier tails, meaning extreme events occur more frequently than the model suggests. To check for this, we rely on statistical tests such as the Kolmogorov–Smirnov (KS) test (see Section 3.7.3) and the Wasserstein distance (Section 3.7.4), as well as entropy measures (Section 3.5), which highlight whether the model underestimates disorder in the data.

A second limitation is the absence of volatility clustering. Real markets often exhibit extended periods of calm followed by extended periods of turbulence. The plain Langevin formulation does not reproduce this effect. This omission is deliberate, as the aim is to isolate the basic physics mechanism of drift and noise, before comparing it fairly with the volatility dynamics of GARCH (Section 3.3.2) and the mean-reversion dynamics of the OU process on prices (Section 3.4.2).

Finally, there are technical considerations related to simulation. The model is implemented using the Euler–Maruyama discretisation scheme already described in Section 3.6.4. At the daily time step used here, the discretisation error is minimal. To ensure reproducibility, the step size Δt is kept fixed and random seeds, along with the number of runs, are documented consistently across simulations.

3.4.2 Ornstein–Uhlenbeck (OU) Process – Mean Reversion Mechanics

The Ornstein–Uhlenbeck (OU) process is a standard model for mean reversion. Unlike Geometric Brownian Motion (GBM), which lets prices drift without any pull back, OU adds a restoring force that nudges the variable toward a long-run average. In market terms, this reflects the idea that prices (or log-prices) tend to hover around a “fair value,” with overpricing attracting sellers and underpricing attracting buyers.

Mathematical formulation

In continuous time, the OU process can be written as

$$dp_t = \kappa(\mu - p_t) dt + \sigma dW_t,$$

where p_t is (log) price, μ is the long-run level, $\kappa > 0$ controls the speed of reversion back to μ , σ sets the size of random shocks, and W_t is Brownian motion. Intuitively, when p_t is above μ the drift term is negative (a pull downward), and when p_t is below μ the drift is positive (a push upward), while the noise term keeps the path irregular.

Discrete implementation (what we actually simulate)

For simulation and estimation we use the OU process in its discrete-time form, which is equivalent to an AR(1) (autoregressive) model:

$$p_{t+1} = \alpha + \beta p_t + \varepsilon_t, \quad \varepsilon_t \sim \mathcal{N}(0, s^2).$$

We estimate α , β , and s^2 by ordinary least squares (OLS) on daily log-prices, then recover the continuous-time OU parameters via the standard mapping

$$\kappa = -\frac{\ln(\beta)}{\Delta t}, \quad \mu = \frac{\alpha}{1 - \beta}, \quad \sigma = \sqrt{s^2 \frac{2\kappa}{1 - \beta^2}}.$$

This approach is mathematically consistent with simulating the SDE by Euler–Maruyama and aligns with our general simulation setup (see Section 3.6.4). Practically, it is stable, easy to estimate from long daily series, and fully reproducible.

For completeness, the one-step **Euler–Maruyama** update for the OU, SDE is

$$p_{t+\Delta t} = p_t + \kappa(\mu - p_t) \Delta t + \sigma \sqrt{\Delta t} \varepsilon_t, \quad \varepsilon_t \sim \mathcal{N}(0, 1),$$

which coincides with the AR(1) update when parameters are mapped as above and Δt is fixed (daily in our case). In the code, we therefore simulate using the AR(1) update, while referring to Section 3.6.4 for the general numerical scheme.

Estimation, data, and simulation settings

We use daily data from 1992–2025 and fix $\Delta t = 1$ (trading day). Parameters are estimated per asset; random seeds and the number of runs are recorded for reproducibility. We simulate 500–1,000 paths to summarise both average behaviour and uncertainty.

Here, OU serves as a physics-inspired benchmark alongside GBM, GARCH, and Langevin/Fokker–Planck models. Performance is evaluated with multiple criteria: price-path accuracy (RMSE), fit-versus-complexity (AIC/BIC), and distributional similarity using the Kolmogorov–Smirnov (KS) test and the Wasserstein distance, with entropy measures used as complementary diagnostics where relevant. This allows us to test whether simple mean reversion improves replication of market behaviour relative to alternative mechanisms.

Limitations of the OU Approach

Although the Ornstein–Uhlenbeck model provides an elegant way to represent mean reversion, it also has important limitations. First, the assumption of Gaussian shocks implies light-tailed behaviour, whereas financial returns often display heavy tails with extreme events occurring more frequently than predicted. This limits OU’s ability to capture crises or sudden jumps.

Second, while the OU process enforces mean reversion around a fixed long-run average, real markets often undergo regime shifts, structural breaks, or changing fundamentals. In such cases, the assumption of a single stable equilibrium level μ may be too restrictive.

Third, the OU process assumes constant volatility. In reality, markets exhibit volatility clustering, with calm periods followed by turbulence. Models such as GARCH (Section 3.3.2) capture this feature more effectively.

Finally, although the discrete AR(1) approximation and Euler–Maruyama scheme (see Section 3.6.4) are mathematically consistent with the continuous SDE, they still introduce discretisation error(i.e. the small error that arises because a continuous process is simulated in discrete steps). At the daily time scale this error is small, but it should be noted when interpreting results.

3.4.3 Fokker–Planck Equation – Probability Distribution Evolution

Most of the models discussed so far—GBM, GARCH, Langevin, OU—simulate individual time series and compare them to historical data. The Fokker–Planck approach takes a different view, asking not “What will the price be tomorrow?” for a single path, but “What is the probability that the price or return will fall within a given range tomorrow, given its value today?”

This shifts the goal from forecasting single outcomes to modelling how the entire probability distribution of returns evolves over time, with the Fokker–Planck equation (or forward Kolmogorov equation) providing the standard framework from statistical physics to describe that evolution.

Mathematical formulation

If a system’s variable X_t follows a stochastic differential equation (SDE) of the form:

$$dX_t = A(X_t, t) dt + B(X_t, t) dW_t$$

where:

- $A(X_t, t)$ is the drift term (systematic tendency),
- $B(X_t, t)$ is the diffusion term (random noise amplitude),

then the Fokker–Planck equation tells us how the probability density $p(x, t)$ of X_t changes over time:

$$\frac{\partial p(x, t)}{\partial t} = -\frac{\partial}{\partial x} [A(x, t) p(x, t)] + \frac{1}{2} \frac{\partial^2}{\partial x^2} [B^2(x, t) p(x, t)]$$

In the Fokker–Planck equation, the drift term shifts the centre of the probability distribution over time—if $A > 0$, the mass drifts upward; if $A < 0$, it drifts downward. The diffusion term controls how the distribution spreads or contracts, with larger B causing faster spreading and greater uncertainty.

In physics, this describes processes like ink dispersing in water or the velocity distribution of gas particles. In finance, the “particles” are possible price or return states, where drift represents economic trends and diffusion represents volatility.

How is it implemented?

We start from the empirical drift and diffusion functions estimated from the data:

1. We divide historical returns into small bins based on their current value x .
2. For each bin, we compute:

- Drift estimate:

$$\hat{A}(x) = \frac{\langle r_{t+\Delta t} - r_t \mid r_t \approx x \rangle}{\Delta t}$$

- Diffusion estimate:

$$\hat{B}^2(x) = \frac{\langle (r_{t+\Delta t} - r_t)^2 \mid r_t \approx x \rangle}{\Delta t}$$

These empirical functions are then plugged into the Fokker–Planck equation.

We solve the PDE numerically using a finite-difference method with:

- Discretised price/return space (x grid)
- Small time steps (Δt)
- Initial distribution $p(x, 0)$ taken directly from observed market returns at the start of the simulation window.

The Fokker–Planck equation models the full evolution of the return distribution, capturing not just an average path but the entire range of outcomes, including tails. This enables early detection of instability through rapid distribution widening, integrates naturally with entropy-based diagnostics, and unifies models like GBM, OU, and Langevin under a single framework.

Binning is used because drift and diffusion can depend on the current return value, allowing asymmetries such as volatility spikes after large losses to be captured. Constant parameters are avoided, as real markets often show state-dependent volatility—calm near the mean and turbulent in the extremes.

In this study, Fokker–Planck complements the Langevin “single-path” perspective, allows direct calculation of Shannon and Tsallis entropy from $p(x, t)$, and supports richer model comparisons by evaluating how well each approach reproduces the full shape and evolution of the return distribution.

Limitations The approach comes with several limitations:

- Computationally more intensive than simulating a single path.

- Requires sufficient historical data to estimate drift and diffusion functions reliably — which is why we use 33 years of data (1992–2025).
- Numerical errors can accumulate in finite-difference solutions if grid and time step sizes are not chosen carefully (we mitigate this by stability checks and convergence testing).

3.5 Entropy-Based Market Diagnostics

Entropy is the measure of the degree of disorder or uncertainty in a system. In thermodynamics, it quantifies how disordered the particles in a gas are. In financial markets, entropy can be thought of as a measure of how unpredictable returns are.

Here, entropy is used as a diagnostic tool to spot market stress. The idea is that changes in entropy may reveal hidden instability before it shows up in volatility. In physics, sudden jumps in entropy often happen just before a system becomes chaotic. In finance, a similar rise could mean the market is close to a crisis or a big regime shift.

3.5.1 Shannon Entropy – measuring uncertainty in returns

Definition

For a discrete probability distribution p_i over N bins (possible return values), Shannon entropy is defined as:

$$H_S = - \sum_{i=1}^N p_i \log p_i$$

where:

- p_i = probability of a return falling in bin i ,
- the log is natural log (\log_e), so entropy is measured in nats.

In finance, Shannon entropy helps us see how unpredictable returns are. When Shannon entropy is high, it means returns are spread out in many directions, making the market look chaotic and uncertain. When it is low, returns stay within a narrow range. This means market is stable and more unpredictable.

It is the classic benchmark. It measures the “average surprise” in returns and is widely accepted, so results are easy to compare with existing finance and physics research. It’s reliable for the overall level of unpredictability.

3.5.2 Tsallis Entropy – generalisation for heavy-tailed distributions

Financial returns don’t always behave like a neat bell curve. They have fat tails, which means extreme events (big crashes or rallies) happen more often than standard models expect. Tsallis entropy introduces a parameter q that controls sensitivity to tail events:

$$H_T(q) = \frac{1 - \sum_{i=1}^N p_i^q}{q - 1}$$

If $q \rightarrow 1$, Tsallis entropy becomes Shannon entropy.

If $q > 1$, the measure is more sensitive to the tail structure of the distribution.

Shannon entropy doesn’t handle this well because it assumes things die off quickly (exponential decay) and events are mostly independent. That can make the market look calmer than it really is.

Tsallis entropy fixes this by adding a knob, called q , that adjusts how much attention we give to those rare extreme events.

$$\text{If } q \rightarrow 1, \quad H_T(q) \rightarrow H_{\text{Shannon}}$$

If $q > 1$, the measure pays more attention to fat tails and big surprises.

In this study, we use $q = 1.5$ because it is a common choice in econophysics for markets with fat tails [2, 6]. This value works well as a middle ground: it captures the effect of extreme events without letting a single outlier (for example, one sudden Bitcoin price spike) dominate the results. Tests on several major markets (SPY, QQQ, FTSE 100, Nikkei 225, and BTC) showed that $q = 1.5$ gives results that are consistent, stable, and easy to interpret.

3.5.3 Rolling Window Implementation

Markets evolve over time, and a single full-period entropy value cannot capture these shifts. To address this, entropy is computed in a rolling window of fixed length W

(e.g., 60 trading days). At each time t , the past W returns $\{r_{t-W+1}, \dots, r_t\}$ are used to estimate the empirical distribution of returns. From this distribution, probabilities $p_{i,t}$ are obtained (via histogram bins or kernel density estimation), and entropy is calculated. For Shannon entropy, this is given by

$$H_t = - \sum_{i=1}^N p_{i,t} \log(p_{i,t}),$$

while for Tsallis entropy the corresponding measure is

$$H_{T,t}(q) = \frac{1 - \sum_{i=1}^N p_{i,t}^q}{q - 1}.$$

Sliding the window forward step by step produces a time series of entropy values.

This rolling approach allows us to track how market unpredictability changes day by day. Calm periods typically show narrow, bell-shaped return distributions with low entropy, while pre-crisis phases exhibit widening distributions with heavy tails, causing entropy to rise. During recoveries, the distribution contracts and entropy falls.

Historical episodes such as the 2008 Global Financial Crisis and the 2020 COVID crash are expected to show that entropy increases sometimes preceded volatility spikes, suggesting predictive potential. Here, both Shannon entropy (overall unpredictability) and Tsallis entropy (sensitivity to fat tails) are applied to historical and model-simulated returns (GBM, GARCH, Langevin, OU, and Fokker–Planck). Comparing their entropy trajectories allows us to assess not only whether a model reproduces real-market entropy levels, but also whether it captures the timing of entropy rises, a key feature of instability build-up.

3.6 Simulation Framework

Once the models are set up and their parameters are estimated from historical data, the next step is to generate simulated price and return series from each model. In simple terms, this means creating model-generated data based on the model's rules. We can then compare these simulated datasets with the actual market data using the evaluation methods explained later in Section 3.7.

3.6.1 Practical Steps

Practical steps taken to ensure the simulations are:

- Fair (same conditions for each model),
- Statistically meaningful (enough simulations for stable results),
- Reproducible (same results if code is re-run),
- Directly comparable to real historical data.

All models are calibrated using the same 1992–2025 daily log returns for each asset to ensure fairness and avoid bias from unequal data. Simulations start from the same initial price as the first day of the test window so that any differences in results come from model dynamics rather than starting values.

The simulation horizon matches the length of the historical period (around 8,300 trading days for 33 years), enabling direct side-by-side comparisons with actual data. Each model is run at least 500 times per asset to average out the effects of random luck in individual runs, and a fixed random seed is used to guarantee reproducibility, ensuring the same input produces the same simulated output.

3.6.2 Parameter Estimation

Before simulation, model parameters are estimated from historical log returns:

- **GBM** – Drift (μ) is the sample mean of daily returns, and volatility (σ) is the sample standard deviation. In simple terms, we just take the average return as the drift and the typical fluctuation size as the volatility.
- **GARCH(1,1)** – Estimated using Maximum Likelihood Estimation (MLE) via Python’s `arch` package. MLE finds the parameters that make the observed data most likely under the GARCH model, which is designed to capture volatility clustering.
- **GARCH with drift** – Same as GARCH, but with an added constant mean parameter. This simply means the model assumes returns fluctuate around a fixed average, rather than around zero.
- **Langevin** – Drift (μ_L) or mean reversion (γ) is estimated from a regression of returns on past values; diffusion D is taken from the return variance. In plain

words, we check whether returns tend to drift in one direction or pull back, and we measure how noisy they are overall.

- **OU** – Long-term mean μ , reversion speed κ , and volatility σ are estimated by Ordinary Least Squares (OLS) regression on lagged log-prices, mapping the discrete AR(1) parameters into their continuous-time OU equivalents. OLS is the standard method for fitting a straight line; here it is used on today's price versus yesterday's price. The slope and intercept from that line are then translated into continuous-time OU parameters.
- **Fokker–Planck** – Empirical drift $A(x)$ and diffusion $B^2(x)$ are estimated via binning of returns. Binning means grouping returns into small ranges (like buckets). For each bucket we calculate the average change (drift) and the spread (diffusion), which provides a data-driven way of estimating how prices evolve.

Maximum likelihood estimation (MLE) is standard for probabilistic models like GBM and GARCH, while ordinary least squares (OLS) regression provides the most efficient estimates for linear drift models such as Langevin and OU. For the Fokker–Planck framework, binned estimation is required to capture state-dependent behaviour that constant-parameter approaches would miss.

3.6.3 Simulation Steps for Path-Based Models (GBM, GARCH, Langevin, OU)

1. Initialise variables: Starting price P_0 , number of days T , estimated parameters.

2. Loop over days:

- Draw random shock(s) $\epsilon_t \sim \mathcal{N}(0, 1)$.
- Update return r_t and/or volatility σ_t based on the model equation.
- Update price P_t from P_{t-1} using:

$$P_t = P_{t-1} \times e^{r_t}$$

3. Store results for each simulated path.

4. Repeat for 500 paths to get a distribution of possible outcomes.

3.6.4 Numerical Simulation Method: Euler–Maruyama

To simulate continuous-time stochastic differential equations (SDEs) such as GBM, Langevin, and Ornstein–Uhlenbeck, we use the Euler–Maruyama method. This approach discretises the continuous equations into small time steps, which makes them suitable for computer simulation. At the daily frequency used here, the discretisation error is minimal.

The time step Δt is fixed throughout, and random seeds are recorded to ensure reproducibility of results. Multiple simulation paths (typically 500–1,000) are generated, allowing us to capture both the average dynamics and the uncertainty around them. Subsequent subsections describe how this scheme is applied to each model in turn.

3.6.5 Simulation Steps for the Fokker–Planck Model

1. Initial distribution $p(x, 0)$ set to match the empirical histogram of returns at the start date.
2. Estimate local drift and diffusion directly from data using the Kramers–Moyal expansion. The return space is partitioned into bins, and for each bin the empirical drift $a(x)$ and diffusion $b^2(x)$ are calculated. Synthetic return paths are then generated via Euler–Maruyama integration:

$$r_{t+1} = r_t + a(r_t) \Delta t + b(r_t) \sqrt{\Delta t} z_t, \quad z_t \sim \mathcal{N}(0, 1).$$

This approach yields path-based simulations with state-dependent coefficients, providing a data-driven alternative to the fixed-parameter SDEs.

3. Convert distributions to synthetic returns by random sampling from $p(x, t)$ at each step.
4. Reconstruct prices by compounding sampled returns.

The distribution-based approach differs from single-path simulations by modelling the entire “cloud” of possible returns directly. Sampling from this evolving distribution makes it possible to compare its outcomes fairly with those of path-based models.

3.6.6 Post-Simulation Processing

For each model, asset, and simulation run:

- Align with real data – Simulation dates are matched to historical dates for RMSE computation.
- Compute evaluation metrics:
 - RMSE between simulated and actual price series.
 - AIC/BIC for model fit vs complexity.
 - KS test and Wasserstein distance for distribution matching.
 - Entropy tracking – Shannon and Tsallis entropy were computed from rolling windows of simulated returns, to compare with real market entropy patterns. For Tsallis entropy, we set $q = 2$, a standard choice in econophysics [2, 6], because it emphasises fat-tailed distributions and extreme market moves. In practice, using $q = 2$ downweights small fluctuations and highlights the contribution of large shocks, making it well suited for financial returns that often display heavy tails. The similarity between real and simulated entropy trajectories was quantified using RMSE, providing an objective measure of how well each model reproduces entropy dynamics.

This framework ensures fairness by giving all models the same historical data for calibration, starting them from identical initial conditions, and using the same number of days, runs, and random seeds. Evaluation metrics are applied in exactly the same way to each model, so any performance differences reflect the models' structures rather than variations in data or simulation setup.

To ensure reproducibility, a global random seed was fixed, with sub-seeds generated for each model and each simulation path. All models were standardised to start from the same initial price (the first day of the test window), and were simulated over the same number of trading days. Indices were aligned explicitly before computing RMSE to guarantee that differences in evaluation arise only from model dynamics, not from misaligned starting points or indexing inconsistencies.

3.7 Model Evaluation Metrics

After simulating price and return series for each model, we need an objective way to decide which model performs best. No single metric can fully capture model quality, so we use a set of complementary evaluation tools. These metrics look at accuracy, statistical fit, complexity, and how well each model reproduces the shape of the real return distribution.

3.7.1 RMSE – Root Mean Square Error

RMSE measures the average size of the differences between the simulated prices and the actual historical prices, with larger errors penalised more heavily:

$$\text{RMSE} = \sqrt{\frac{1}{T} \sum_{t=1}^T (P_t^{\text{sim}} - P_t^{\text{real}})^2}$$

Where:

- P_t^{sim} is the simulated price at time t ,
- P_t^{real} is the actual historical price,
- T is the total number of days.

In finance, a lower RMSE means the simulated prices track the real price path more closely, with the metric being especially sensitive to large deviations — important when big errors, such as during market crashes, are costly. It is included here because it directly measures price path accuracy and is widely understood, making the results easy to communicate to both technical and non-technical audiences.

3.7.2 AIC & BIC – Balancing Fit and Complexity

A model can achieve a low RMSE by adding more parameters, but extra parameters can lead to overfitting — performing well on historical data but poorly on new data.

AIC (Akaike Information Criterion) and BIC (Bayesian Information Criterion) counteract this by penalising complexity.

If k is the number of model parameters, L is the maximised likelihood value, and n is the number of data points:

$$\text{AIC} = 2k - 2 \ln(L)$$

$$\text{BIC} = k \ln(n) - 2 \ln(L)$$

Lower AIC or BIC indicates a better balance between model fit and complexity, with AIC applying a lighter penalty for complexity and BIC being more stringent. For GBM and GARCH, likelihood-based AIC and BIC were obtained directly (GBM via Gaussian likelihood, GARCH via the `arch` package). For Langevin and OU, no closed-form likelihood was available, so an AR(1) surrogate was fitted (on returns for Langevin, on log-prices for OU) and the resulting AIC/BIC values were used as proxies. For the Fokker–Planck (KM-inspired) model, which is non-parametric, no likelihood can be defined in a straightforward way, and AIC/BIC values are therefore left undefined.

3.7.3 Kolmogorov–Smirnov (KS) Test

Accuracy in the price path is important, but so is accuracy in the distribution of returns. Even if two models produce similar average prices, one may generate unrealistic return patterns. The Kolmogorov–Smirnov (KS) test compares the cumulative distribution functions (CDFs) of simulated and real returns. The statistic is

$$D = \sup_x |F_{\text{sim}}(x) - F_{\text{real}}(x)|,$$

where F is the CDF. A small D means the simulated distribution is very close to the real one, while a large D indicates significant differences in shape.

3.7.4 Wasserstein Distance

The Wasserstein distance (also called Earth Mover’s Distance) measures the “minimum work” needed to reshape one probability distribution into another. The intuition is simple: imagine two histograms as piles of sand; the Wasserstein distance is how much sand you would need to move, times how far you would need to move it, to make them match. A lower value indicates the distributions are more similar, while a higher value means more reshaping is needed for them to align.

3.7.5 Jarque–Bera (JB) Test for Normality

The Jarque–Bera (JB) test checks whether a dataset follows a normal (Gaussian) distribution by comparing the sample skewness and kurtosis with those expected under normality. The test statistic is

$$JB = \frac{n}{6} \left(S^2 + \frac{(K - 3)^2}{4} \right),$$

where:

- n is the sample size,
- S is the sample skewness,
- K is the sample kurtosis.

The null hypothesis is that the data are normally distributed. Large values of JB with small p -values lead to rejection of the null, indicating non-normality. In financial markets, this is crucial because returns often show negative skewness and high kurtosis (“fat tails”), deviating from Gaussian assumptions. Confirming non-normality motivates the use of advanced models such as GARCH, Langevin, Ornstein–Uhlenbeck, and Fokker–Planck, which explicitly account for volatility clustering and heavy-tailed behaviour.

3.7.6 Augmented Dickey–Fuller (ADF) Test for Stationarity

The Augmented Dickey–Fuller (ADF) test examines whether a time series is *stationary* or has a unit root (non-stationary). Stationary series have constant statistical properties, whereas non-stationary series can drift or wander. The ADF regression takes the form:

$$\Delta y_t = \alpha + \beta t + \gamma y_{t-1} + \sum_{i=1}^p \delta_i \Delta y_{t-i} + \varepsilon_t,$$

and tests the null hypothesis $H_0 : \gamma = 0$ (unit root) against the alternative $H_1 : \gamma < 0$ (stationary).

In finance, prices often behave like near-unit-root processes, but returns are generally stationary. Confirming this with the ADF test justifies analysing returns rather than raw prices. Other tests exist (e.g., KPSS, Phillips–Perron, DF–GLS), but ADF is standard and well suited to daily data.

3.7.7 ARCH Test for Volatility Clustering

Financial time series often show that large price moves are followed by other large moves, and calm periods by more calm periods. This is called *volatility clustering*. The Autoregressive Conditional Heteroskedasticity (ARCH) test detects it by checking whether squared residuals from a fitted model are predictable. If past squared errors explain current squared errors, then the series has ARCH effects, meaning volatility is time-dependent. The null hypothesis is no ARCH effects; the test statistic follows a chi-squared distribution.

Detecting volatility clustering is important because it justifies using models like GARCH, which explicitly model changing volatility. A significant ARCH test result confirms that financial returns do not have constant variance, strengthening the case for volatility models.

3.7.8 Skewness and Kurtosis (Distributional Moments)

Beyond mean and variance, two important features of return distributions are *skewness* and *kurtosis*. Skewness measures asymmetry: a negative skew means large negative returns (crashes) are more likely than equally large positive gains, while positive skew means the opposite. Kurtosis measures the “tailedness” of a distribution: the normal distribution has kurtosis of 3, while higher values indicate *fat tails*, i.e. more frequent extreme events.

These metrics provide a clearer picture of risk than variance alone. They also feed into formal tests like the Jarque–Bera (Section 3.7.5), which combines them to test normality. In markets, negative skew and high kurtosis are commonly observed and motivate models that go beyond Gaussian assumptions.

3.8 Comparative Analysis Procedure

Once each model has been calibrated, simulated, and evaluated using the metrics in section 3.7, the final step is to compare them systematically.

1. **Aligning Model Outputs** – All models are simulated over the same time horizon and matched to historical dates for point-by-point comparisons.

2. **Evaluating with Multiple Metrics** – RMSE, AIC, BIC, KS, Wasserstein, and entropy tracking are all applied.
3. **Aggregating Over Multiple Runs** – 500 runs per model per asset, results averaged, standard deviations recorded.
4. **Ranking Models** – Best performer in each metric gets rank one; ranks aggregated for composite score.
5. **Interpreting in Context** – Trade-offs considered between metrics, interpretability, and feasibility.
6. **Cross-Checking with Entropy** – Models replicating entropy build-up before volatility spikes get extra credit.

Outcome: This structured framework delivers a clear and defensible answer to whether physics-inspired models offer measurable advantages over traditional models.

3.9 Tools and Libraries

All analysis was performed in Python for its numerical libraries, open-source nature, and reproducibility.

Core libraries used:

- `yfinance` – Downloading historical market data.
- `pandas` – Data handling and cleaning.
- `numpy` – Numerical operations and RNG.
- `scipy` – Statistics, integration, optimisation.
- `matplotlib & seaborn` – Plotting.
- `arch` – GARCH estimation.
- `statsmodels` – Regression, KS tests.
- `wasserstein_distance` – Earth Mover's Distance.

Reproducibility Measures: Version control, fixed seeds, parameter logging, and environment isolation ensure results are fully replicable.

Results & Discussion

This chapter presents the findings from the analysis of five major markets: the S&P 500, QQQ (NASDAQ ETF), FTSE 100, Nikkei 225, and Bitcoin (BTC-USD), covering the period 1992–2025. To find how well different models capture the key features of real financial data, the results are organised into eight subsections:

1. **Descriptive statistics and correlation:** A first look at the data, reporting averages, variability, skewness, heavy tails, normality tests, correlations across markets, and rolling volatility.
2. **Volatility patterns and return distributions:** Time-series plots of prices and returns, rolling volatility charts, and histograms that show volatility clustering and fat tails.
3. **Entropy analysis:** Rolling Shannon and Tsallis entropy compared with volatility, with lead–lag analysis showing whether entropy can act as an early warning signal of turbulence.
4. **Model parameter calibration:** Estimated parameters for all models (GBM, GARCH, Langevin, OU, FP-KM), with comparisons across the five markets.
5. **Empirical cumulative distribution functions (ECDF):** Plots of actual versus simulated returns, showing which models best capture the shape of distributions and the behaviour of tails.

6. **Model evaluation metrics:** Quantitative measures such as RMSE (price accuracy), KS test, and Wasserstein distance, comparing model performance across 500 simulation runs.
7. **Entropy-based model evaluation:** How well each model replicates actual entropy dynamics, using Shannon and Tsallis entropy as benchmarks.
8. **Summary of findings:** A synthesis of which models perform best at different tasks, highlighting complementarities between traditional finance models and physics-inspired approaches.

We will first show how markets behave in reality, and then test how different models succeed—or fail—in capturing those behaviours.

4.1 Descriptive Statistics and Correlation

The analysis begins with the computation of daily log returns for all five assets, combined into both a union of dates (maximizing the available history for each series) and an aligned intersection (restricting to common trading days). The aligned dataset spans 2,388 observations and is used for direct cross-market comparisons. This section establishes the baseline stylized facts that motivate the modelling choices in subsequent chapters.

4.1.1 Core Statistics

Table 4.1 gives a first overview of the data. For the stock indices, the average daily return is very small: about 0.045% for the S&P 500, 0.07% for QQQ, 0.008% for the FTSE 100, and 0.028% for the Nikkei. These values reflect the slow but steady growth we typically see in developed markets. Their typical ups and downs, measured by annualized volatility, range between 16–22%. The FTSE 100 is at the lower end (about 16%), while the S&P 500 and Nikkei are closer to 18–21%. QQQ, which follows U.S. technology stocks, shows slightly bigger swings at around 22%, but also the best risk–return balance (Sharpe ratio of 0.81), consistent with the strong performance of tech companies over the sample period.

Bitcoin, however, looks very different. Its average daily return is much higher, at 0.11%, but this comes with very large fluctuations: a daily standard deviation of nearly 4%, which translates into an annualized volatility of more than 63%. Its Sharpe ratio (0.45) is lower than QQQ's, meaning the extra return comes with much more risk. These results underline Bitcoin's role as a high-risk, high-reward asset that behaves very differently from traditional equity benchmarks. In this dissertation, Bitcoin therefore acts as a stress test for the models, showing whether they can capture such extreme behaviour.

Table 4.1: Core descriptive statistics of daily log returns (aligned intersection)

Asset	n	mean_daily	std_daily	ann_vol	ann_sharpe
S&P 500	2388	0.000455	0.011273	0.178958	0.641016
QQQ	2388	0.000702	0.013719	0.217778	0.812485
FTSE 100	2388	0.000082	0.010028	0.159184	0.130132
Nikkei	2388	0.000282	0.013105	0.208037	0.341829
Bitcoin	2388	0.001132	0.039699	0.630202	0.452854

4.1.2 Distributional Moments

Table 4.2 shows how returns are distributed across the five assets. Two features stand out: *skewness* (which tells us about asymmetry) and *kurtosis* (which tells us about the thickness of the tails, or how extreme the ups and downs can be).

For all assets, skewness is negative, meaning large downward moves are more common than equally large upward ones. The FTSE 100 has the strongest negative skew (-0.90), followed closely by Bitcoin (-0.88) and the S&P 500 (-0.82). This suggests that these markets are more prone to sudden crashes than to equally large rallies. QQQ and the Nikkei also lean negative, but less strongly (-0.56 and -0.45 respectively).

Kurtosis values are also very high, confirming the presence of “fat tails,” or extreme events that occur much more often than they would in a normal bell-shaped distribution. The S&P 500 shows the highest excess kurtosis (16.3), indicating a high frequency of unusually large price changes. The FTSE 100 (13.3) and Bitcoin (11.9) also display

heavy-tailed behaviour, while the Nikkei (8.85) and QQQ (7.10) are still far above what would be expected under normality (which has kurtosis of zero).

Together, these results show that none of the assets behave like a simple Gaussian (normal) distribution. Instead, they all show a tilt toward downside risk and a tendency for extreme moves. This underlines the need for models that can handle asymmetry and fat tails — such as GARCH, Langevin, OU, and Fokker–Planck — rather than relying only on Gaussian-based models like GBM.

Table 4.2: Distributional moments of daily log returns (aligned intersection)

Asset	Skew	Excess Kurtosis
S&P 500	-0.822451	16.295155
QQQ	-0.563156	7.103047
FTSE 100	-0.899390	13.348019
Nikkei	-0.447216	8.847932
Bitcoin	-0.878689	11.857545

4.1.3 Tails and Statistical Tests

Table 4.3 looks at how extreme returns can get (the tails) and runs three diagnostic tests.

First, considering the extremes: The largest single-day losses range from about -13.2% for the Nikkei, -12.7% for the S&P 500, and -12.8% for QQQ. The FTSE 100 is only slightly smaller at -11.5%. Bitcoin is much more extreme, with a worst daily drop of -46.5%. On the upside, equities show maximum daily gains between 8–9% (S&P 500: +8.97%, Nikkei: +9.74%), whereas Bitcoin again stands out with a massive single-day jump of +22.5%. These figures underline how much more violent Bitcoin’s swings are compared to traditional stock markets.

Second, the **Jarque–Bera (JB) test** checks whether returns follow a normal bell-shaped distribution. For all assets, the JB values are extremely large (e.g., 26,570 for the

S&P 500, 5,121 for QQQ, 17,968 for FTSE, 7,831 for Nikkei, and 14,232 for Bitcoin), with p -values of 0.000. This means we can strongly reject normality, confirming that returns are not Gaussian and instead have fat tails and asymmetry.

Third, the **Augmented Dickey–Fuller (ADF) test** checks if returns are “stationary” — meaning their statistical properties remain stable over time. For all series, the ADF statistics are strongly negative (e.g., -11.3 for the S&P 500, -11.9 for QQQ, -49.8 for FTSE, -22.6 for Nikkei, and -14.6 for Bitcoin), with p -values of 0.000. This shows the returns are stationary, making them suitable for time-series modelling.

Finally, the **ARCH test** checks whether volatility clusters — that is, whether periods of calm are followed by periods of turbulence. The test is highly significant for all markets: for example, the ARCH statistic is 967.5 for the S&P 500, 668.7 for QQQ, 440.9 for FTSE, 436.2 for Nikkei, and 63.2 for Bitcoin, all with p -values of 0.000. This confirms volatility clustering as a universal feature.

Taken together, these results show that returns across all assets are non-normal, stationary, and exhibit volatility clustering. Equity markets (S&P 500, QQQ, FTSE 100, Nikkei 225) show more moderate extremes, while Bitcoin experiences much larger shocks in both directions. These patterns justify the use of advanced models such as GARCH and physics-inspired approaches, which are specifically designed to capture fat tails and clustered volatility.

Table 4.3: Distributional tails and diagnostic tests (aligned intersection)

Asset	min	p1	p5	p95	p99	max	jb_stat	jb_p	adf_stat	adf_p	arch_stat	arch_p
S&P 500	-0.127652	-0.032394	-0.016632	0.015726	0.025658	0.089683	26570.737150	0.000	-11.289009	0.000	967.509	0.000
QQQ	-0.127592	-0.039128	-0.022316	0.021050	0.033731	0.081309	5121.651203	0.000	-11.894536	0.000	668.680	0.000
FTSE 100	-0.115117	-0.030885	-0.015259	0.014532	0.025314	0.086664	17968.641736	0.000	-49.754598	0.000	440.942	0.000
Nikkei	-0.132341	-0.036860	-0.020533	0.019862	0.031054	0.097366	7831.969984	0.000	-22.639797	0.000	436.203	0.000
Bitcoin	-0.464730	-0.117078	-0.061495	0.060162	0.103923	0.225119	14232.342888	0.000	-14.588170	0.000	63.242	0.000

4.1.4 Correlation Matrices

Table 4.4 shows how closely the markets move together, using both Pearson and Spearman correlations. The U.S. indices — the S&P 500 and QQQ — are very tightly linked, with a correlation of about 0.93. This makes sense, since QQQ is a tech-heavy U.S. ETF and shares many drivers with the broader U.S. market.

The FTSE 100 (UK) shows only moderate links with U.S. markets: about 0.55 with the

S&P 500 and 0.42 with QQQ. The Nikkei (Japan) is even weaker, with correlations of 0.21 (S&P 500) and 0.18 (QQQ). This suggests that while global equity markets sometimes move together, Japan has more independent dynamics.

Bitcoin stands out as almost independent. Its correlation with the S&P 500 is only 0.23, and with the Nikkei it is basically zero (0.005). This means Bitcoin often moves on its own, separate from traditional equity markets.

The Spearman correlations (which look at the rank order of movements rather than exact values) tell a similar story: strong correlation between the U.S. indices (0.91), weaker ties for the FTSE 100 (0.44 with S&P 500), and almost no connection for Bitcoin (0.17 with S&P 500, -0.05 with Nikkei).

These heterogeneous dependence structures highlight the need for flexible modelling approaches capable of capturing both integration and segmentation effects.

Table 4.4: Pearson and Spearman correlations of daily log returns (aligned intersection)

	S&P 500	QQQ	FTSE 100	Nikkei	Bitcoin
Pearson Correlation					
S&P 500	1.000	0.929	0.545	0.213	0.228
QQQ	0.929	1.000	0.418	0.182	0.234
FTSE 100	0.545	0.418	1.000	0.319	0.130
Nikkei	0.213	0.182	0.319	1.000	0.005
Bitcoin	0.228	0.234	0.130	0.005	1.000
Spearman Correlation					
S&P 500	1.000	0.907	0.438	0.156	0.167
QQQ	0.907	1.000	0.339	0.142	0.175
FTSE 100	0.438	0.339	1.000	0.255	0.059
Nikkei	0.156	0.142	0.255	1.000	-0.052
Bitcoin	0.167	0.175	0.059	-0.052	1.000

Looking across all the statistics, three big lessons stand out:

1. Returns don't follow a neat bell curve. Instead, they have *fat tails* and a *downward tilt*, meaning extreme moves — especially big drops — happen more often than simple models expect.
2. Volatility doesn't come randomly. Calm and turbulent periods tend to cluster together — quiet days are followed by more quiet days, while stormy markets often stay stormy.
3. Markets don't all move in the same way. U.S. markets (S&P 500 and QQQ) move closely together, the UK and Japan are looser, and Bitcoin often moves on its own.

These patterns the necessity of using different models. The basic GBM model cannot explain the heavy tails. GARCH is designed to capture the clustering of volatility. Physics-inspired models (Langevin, Ornstein–Uhlenbeck, Fokker–Planck) are good candidates to capture the heavy-tailed, irregular behaviour we actually see. Bitcoin, with its extreme swings and weak connection to equities, provides the toughest test of how well these models work.

4.2 Volatility Patterns and Return Distributions

Here we examine how prices evolve over time, how daily returns behave, and how volatility and distributions reveal deeper regularities in financial markets. Looking across both equity indices (S&P 500, QQQ, FTSE 100, Nikkei 225) and Bitcoin allows us to compare traditional markets with a high-volatility digital asset.

4.2.1 Price and Return Patterns

Figure 4.1 shows how the prices of the five assets have changed over time. The S&P 500 and QQQ generally rise in the long run, but both experienced sharp drops during major crises such as the 2008 financial crash and the 2020 pandemic. The FTSE 100 grows at a slower pace, while the Nikkei 225 reflects Japan’s weaker stock market performance over the same period. Bitcoin looks very different: its price swings between dramatic booms and deep crashes, showing its highly speculative and risky nature compared to traditional equity markets.

Figure 4.2 illustrates the day-to-day changes in these markets. For the equity indices, most of the daily moves are small and remain close to zero. Bitcoin, in contrast, often shows very large jumps and drops. This difference is confirmed by the numbers in Table 4.1: Bitcoin’s daily ups and downs average around 3.97%, compared with only about 1.13% for the S&P 500 and 1.00% for the FTSE 100. Over a full year, this translates into Bitcoin having an annual volatility of 63%, while the stock indices remain in the much lower range of 18–22%.

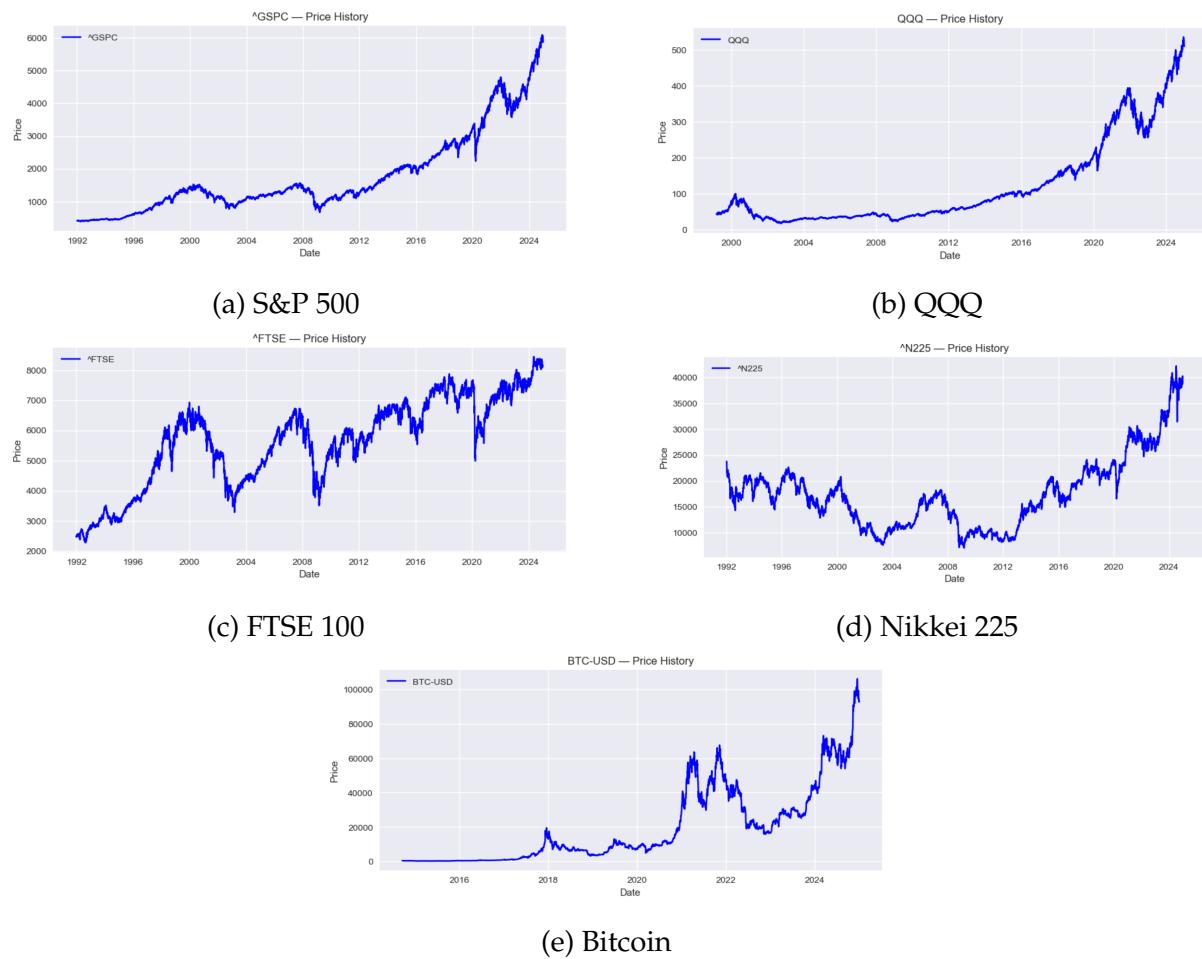


Figure 4.1: Price series (1992–2025). All panels share the same vertical scale within each subplot.

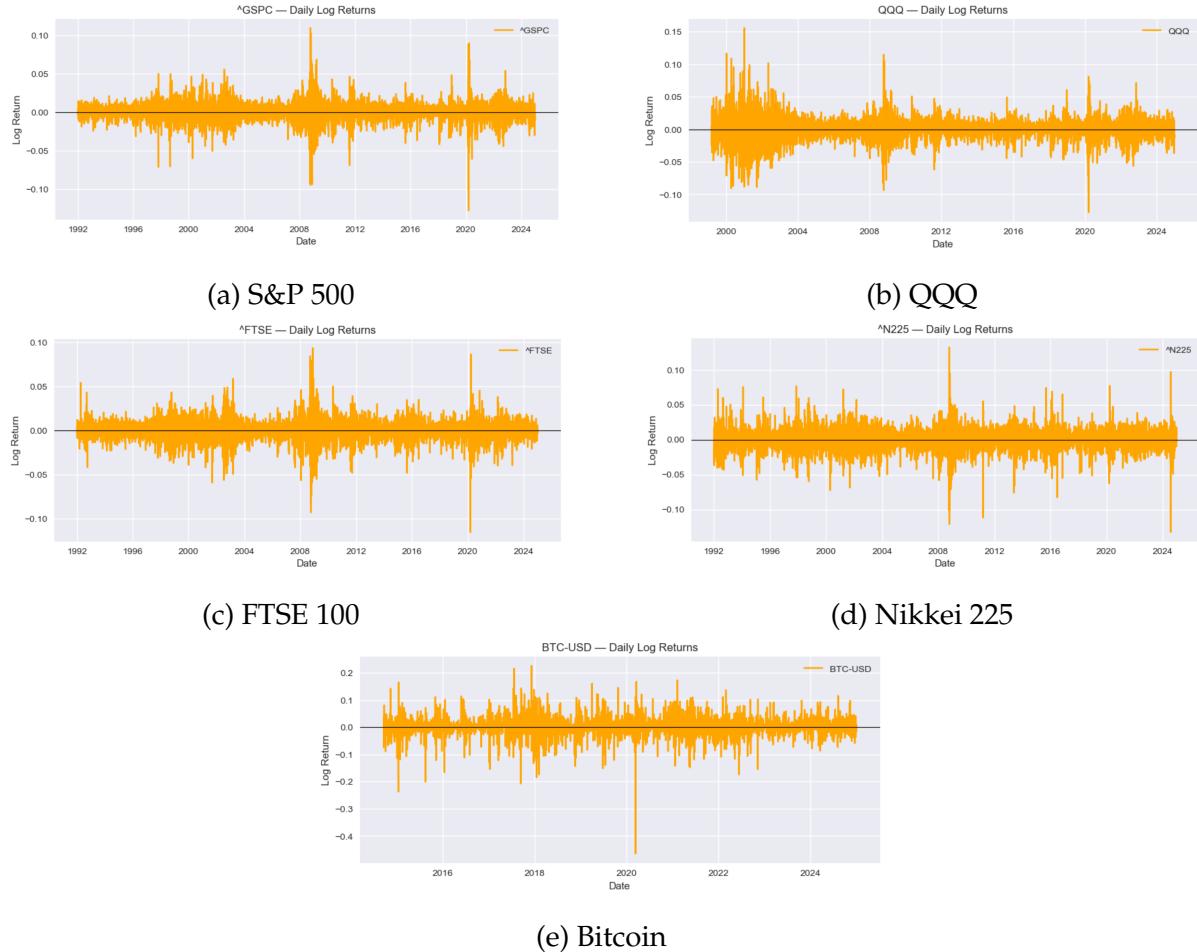


Figure 4.2: Daily log returns (1992–2025). Equity returns cluster tightly around zero; Bitcoin shows frequent large spikes.

4.2.2 Rolling Volatility

Figure 4.3 shows how market volatility — the size of price swings — changes over time when measured over a rolling 21-day window. In all markets, quiet periods and stormy periods tend to come in clusters. This means calm days often follow other calm days, and turbulent times often arrive in bursts rather than being spread evenly. For the stock indices, we see sharp jumps in volatility during big global crises, after which conditions usually settle back down.

Bitcoin, however, behaves very differently. Its volatility spikes happen not only during crises but also much more often and at much higher levels than equities. According to the figures in Table 4.1, Bitcoin's typical annual volatility reaches about 63%, compared with only 18–22% for the stock indices.

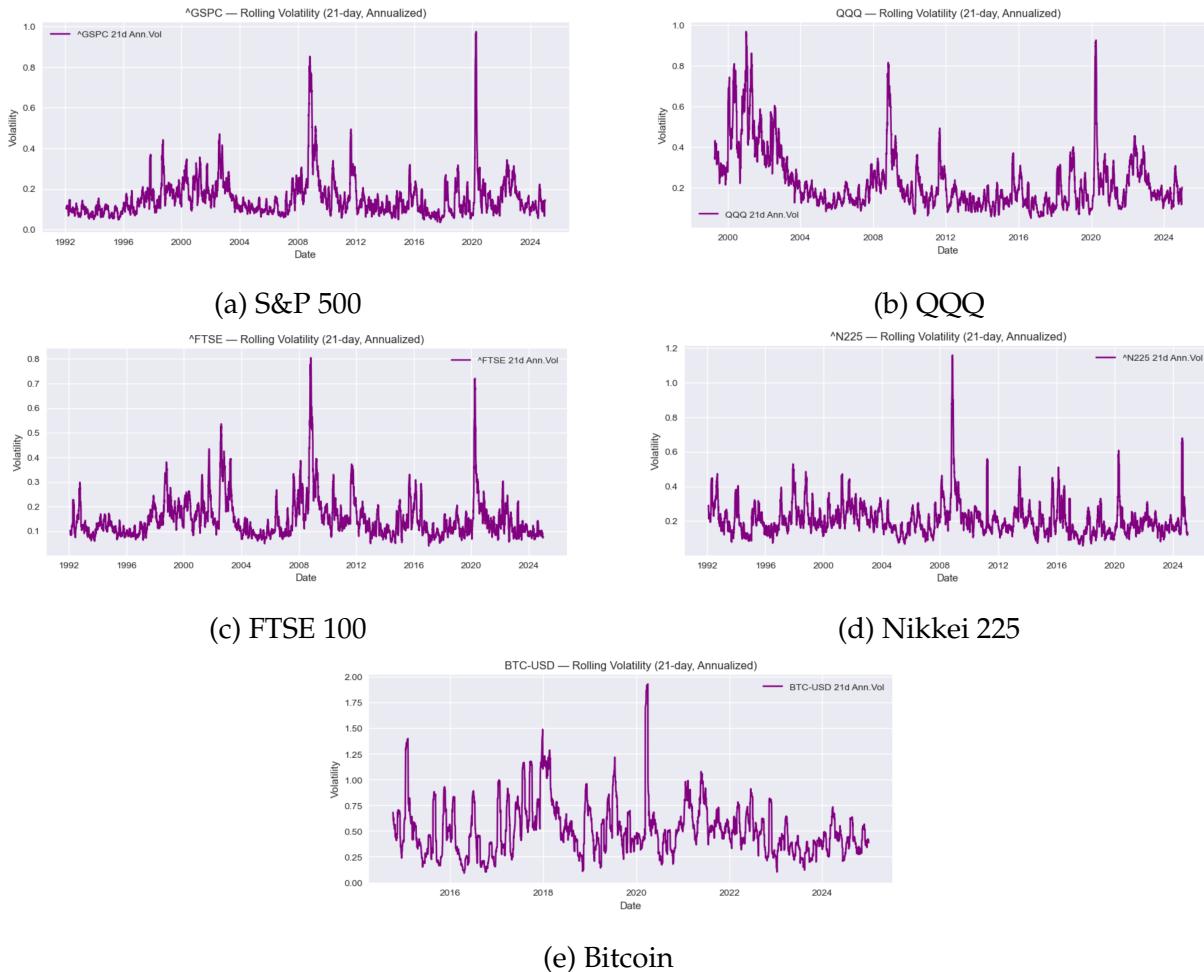


Figure 4.3: 21-day rolling annualized volatility. Clustering is visible for all assets; Bitcoin's spikes are larger and more frequent.

4.2.3 Return Distributions

Figure 4.4 looks at the shape of the daily return distributions for each asset. The main finding is that extreme price changes, especially sharp losses, happen far more often than what we would expect if markets followed a simple bell-curve pattern. In other words, all the assets show “fat tails” (many more extreme events than normal) and “negative skew” (downward moves tend to be larger than upward ones).

The numbers in Table 4.2 confirm this. For example, the S&P 500 has a skew of -0.82 and a kurtosis of 16.3 . The high kurtosis indicates that extreme returns, in either direction, are much more likely than under a normal distribution, while the negative skew suggests that these extremes are tilted toward the downside. Taken together, these measures point to a heightened risk of large negative shocks. The FTSE 100 and Bitcoin

are similar, with skew close to -0.9 and kurtosis around 12 to 13. Table 4.3 shows the real impact of these extremes: while equity indices have single-day gains of about +9% and losses of around -12% , Bitcoin is far more volatile, with jumps as high as +22.5% and drops as large as -46.5% . In simple terms, stock markets sometimes make big moves, but Bitcoin's swings are much more dramatic.

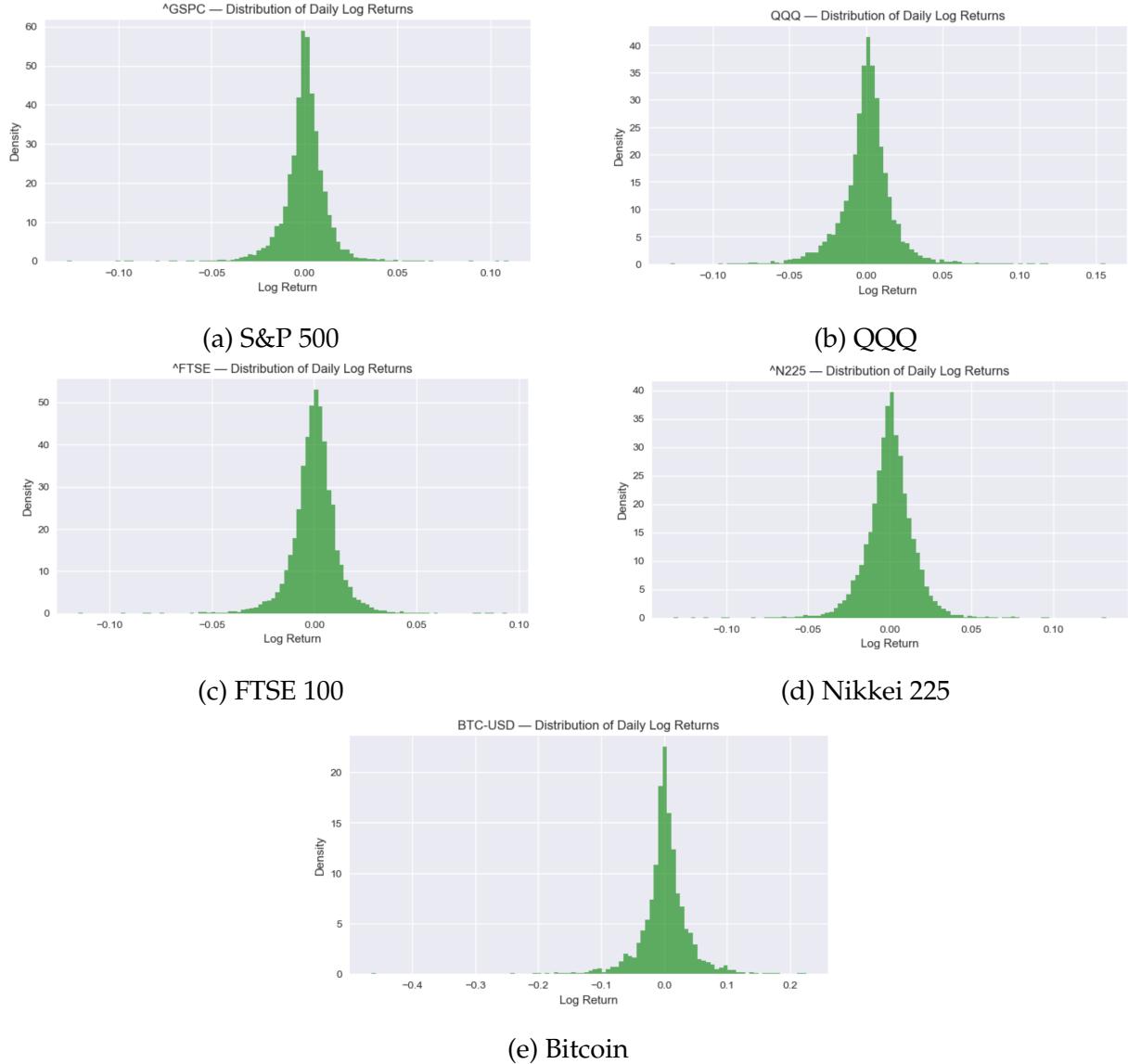


Figure 4.4: Histograms of daily log returns. Heavy tails and negative skew are present in all assets, most pronounced for Bitcoin.

Three key facts emerge from the analysis. Firstly, volatility clearly clusters in time across all markets, consistent with the ARCH results reported in Table 4.3. Secondly, all assets also display fat tails and a downward tilt in their return distributions (Table 4.2), which shows that purely Gaussian models are not sufficient to describe market be-

haviour. Finally, Bitcoin stands out as an outlier: its much higher day-to-day dispersion (Table 4.1) and extreme tail magnitudes (Table 4.3) far exceed those of the equity indices, making it a particularly demanding stress test for the models assessed in the following sections.

4.3 Entropy Analysis

Entropy provides a complementary perspective on market dynamics by capturing the degree of unpredictability in return distributions. Unlike volatility, which only measures the size of price movements, entropy reflects both the spread and the shape of return distributions. This section uses two measures: Shannon entropy and Tsallis entropy (with $q = 1.5$). The results compute rolling entropies alongside volatility and perform a lead–lag correlation analysis.

4.3.1 Entropy and Volatility

Figure 4.5 brings together the rolling Shannon and Tsallis entropy series alongside 21-day annualized volatility for all five assets. Across equity markets, entropy and volatility both rise sharply during global crises such as the 2008 financial collapse and the COVID-19 shock in 2020. Importantly, entropy often reacts earlier than volatility, spiking in advance of turbulent episodes and settling down once markets stabilize.

The contrast with Bitcoin is stark. As shown in Table 4.1, Bitcoin’s daily standard deviation (3.97%) and annualized volatility (63%) far exceed those of the equity indices (1.0–1.3% daily; 18–22% annualized). These numbers align with the entropy plots: Bitcoin shows much more frequent and much larger spikes in entropy, reinforcing its status as the most unstable of the five markets.

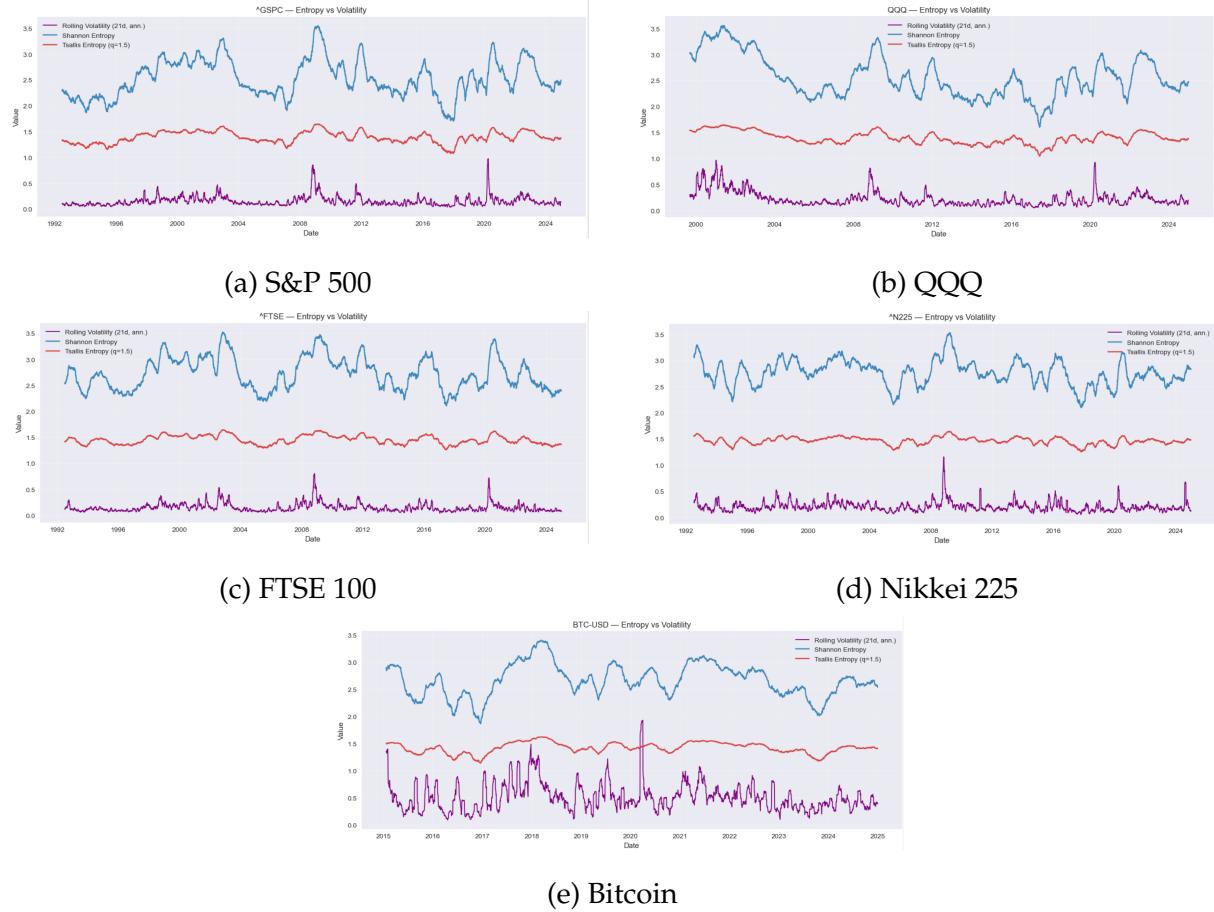


Figure 4.5: Rolling Shannon and Tsallis entropy compared with 21-day annualized volatility across all five assets. Entropy often increases before volatility spikes, suggesting early-warning potential.

4.3.2 Lead–Lag Analysis

Figure 4.6 explores how entropy and volatility move relative to one another by plotting lead–lag correlations for all markets. A positive lag means that changes in entropy tend to appear first, followed by volatility a few days later. A negative lag would imply the reverse relationship.

For the major equity indices, the strongest correlations occur at positive lags of around 5–10 days. In practical terms, this suggests that increases in entropy often provide a short early warning of upcoming turbulence. The effect is clearest for the S&P 500 and QQQ, which are highly integrated within U.S. markets and therefore act as leaders. The FTSE 100 and Nikkei also show positive lead effects, though the signals are weaker, reflecting their less intense volatility patterns (see Table 4.1).

Bitcoin behaves very differently. Its lead–lag profile exhibits both stronger correlations and much larger swings, showing that entropy is especially sensitive to the erratic and explosive nature of cryptocurrency markets. This highlights entropy’s potential as a diagnostic tool: it can capture signs of instability earlier than volatility measures alone, particularly in markets prone to extreme and irregular movements.

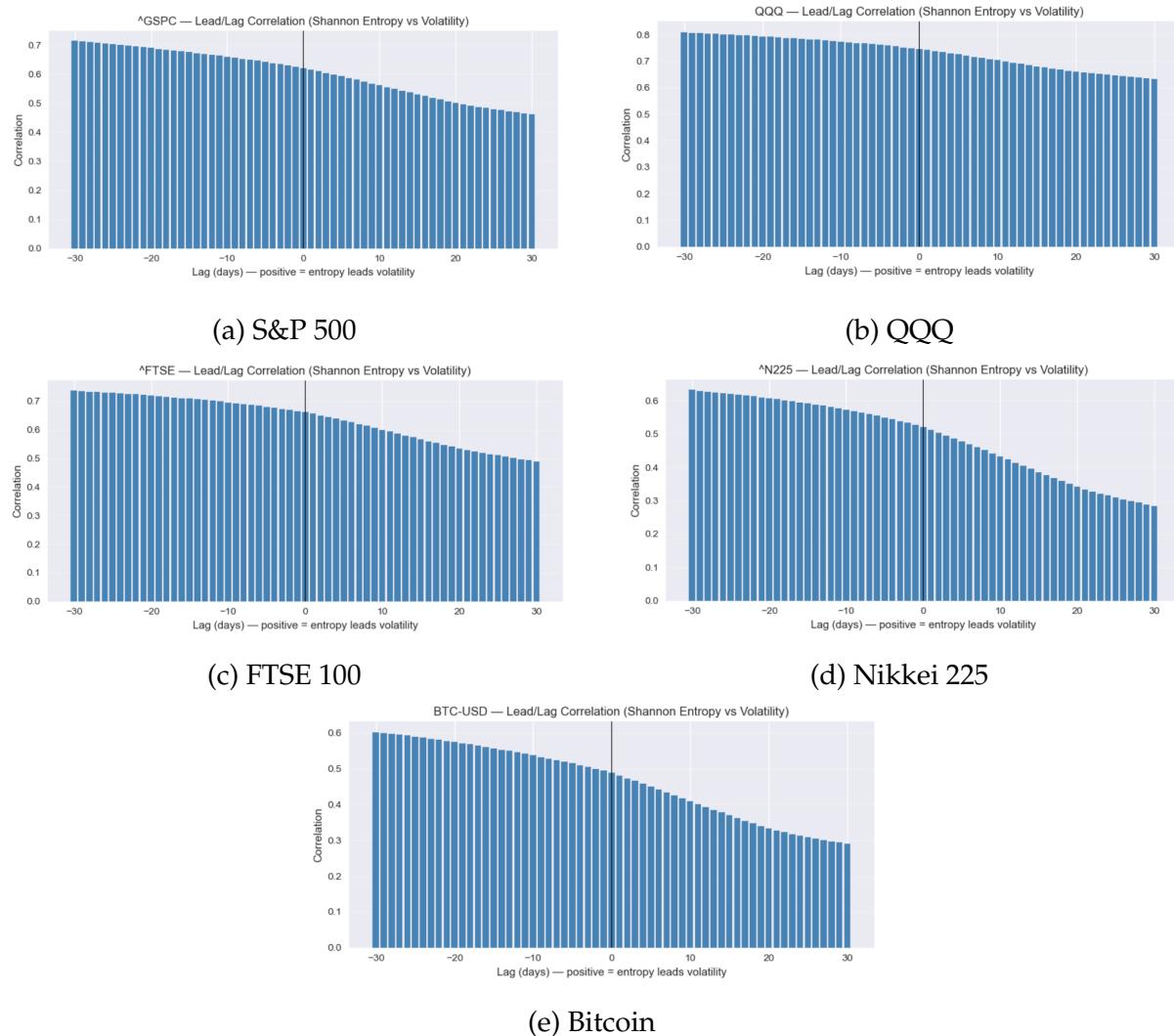


Figure 4.6: Lead–lag correlation between Shannon entropy and volatility across all five assets. Positive lags indicate entropy leads volatility.

4.3.3 Cross-Market Comparisons

Looking across all markets, three main points stand out.

First, entropy and volatility usually move together, but entropy often reacts faster. This means entropy can give an earlier signal when market conditions are changing.

Second, the major stock indices behave in broadly similar ways: entropy rises sharply during global crises and then falls back as markets calm down.

Third, Bitcoin is very different. Its entropy shows much bigger and more frequent swings, confirming that it is the most unpredictable and unstable asset in the sample.

Overall, these results suggest that entropy is a useful diagnostic tool. It can reveal signs of instability that volatility on its own might miss.

4.4 Model Parameter Calibration

Before we can run simulations and judge how well the models work, we first need to estimate their key parameters from past market data. These parameters act like the “settings” of each model, making sure the simulated behaviour reflects the patterns we actually observe in history.

One exception is the Fokker–Planck (FP–KM) approach. This is a non-parametric model, which means it does not rely on a fixed set of parameters to describe the market. Instead, it reconstructs the full probability distribution of returns directly from the data, capturing how that distribution changes over time. Because of this, the FP–KM model cannot be calibrated in the same way as the other models. Its performance will be evaluated later, based on how well it matches real return distributions and entropy patterns.

Tables 4.5 and 4.6 report the calibrated values for all five markets across the parametric model families.

Table 4.5: Calibrated parameters for GBM and GARCH models across five markets (daily frequency).

Ticker	GBM_mu	GBM_sigma	GARCH0_omega	GARCH0_alpha	GARCH0_beta	GARCHd_mu	GARCHd_omega	GARCHd_alpha	GARCHd_beta
S&P 500	0.000318	0.011479	0.017500	0.102537	0.883160	0.062064	0.018476	0.108193	0.877098
QQQ	0.000380	0.017097	0.023523	0.094041	0.896936	0.090798	0.025037	0.098519	0.892070
FTSE 100	0.000142	0.010961	0.018771	0.096466	0.886471	0.036197	0.019258	0.098327	0.884220
Nikkei	0.000064	0.014617	0.054389	0.104709	0.870711	0.047462	0.054039	0.106251	0.869570
Bitcoin	0.001416	0.036457	0.623335	0.124842	0.839083	0.174120	0.622684	0.125854	0.838222

Table 4.6: Calibrated parameters for Langevin and Ornstein–Uhlenbeck (OU) models across five markets (daily frequency).

Ticker	Langevin_gamma	Langevin_sigma	Langevin_a	OU_kappa	OU_mu	OU_sigma
S&P 500	1.085276	0.011438	0.000345	0.000083	11.101043	0.011480
QQQ	1.054462	0.017073	0.000400	0.000001	-151.002858	0.017098
FTSE 100	1.016887	0.010960	0.000144	0.000994	8.742822	0.010963
Nikkei	1.034826	0.014609	0.000067	0.000470	9.836992	0.014621
Bitcoin	1.020122	0.036433	0.001464	0.000213	15.458068	0.036464

The estimates reveal several important patterns:

GBM. Drift terms (μ) are small but positive for all equity indices, consistent with the modest upward trend observed in long-run prices. Volatility (σ) lies between 1.1% and 1.7% per day for equities, translating to annualized levels of 18–22%. Bitcoin’s daily volatility (3.6%) is more than double, consistent with the descriptive statistics in Table 4.1.

GARCH. The persistence parameters ($\alpha + \beta$) are close to one for all markets, indicating strong volatility clustering. For equities, β dominates, showing slow decay of volatility shocks. Bitcoin has larger ω and α values, pointing to higher baseline variance and stronger immediate reactions to shocks.

Langevin. The γ coefficients are close to one for all assets, suggesting that the force term acts as a stabilizer but with different strengths. The noise term σ is again much higher for Bitcoin, reflecting its greater randomness.

Ornstein–Uhlenbeck. The mean-reversion speed (κ) varies widely across markets. For the S&P 500 and FTSE 100, κ is moderate (0.00008–0.0010), meaning that prices slowly revert to their long-term mean. For QQQ, the estimate of κ is very close to zero, while the long-term mean (μ) is strongly negative, which may reflect instability or overfitting in the calibration. Bitcoin shows the highest κ (0.00021) and a large long-term mean ($\mu \approx 15.5$), again highlighting its distinct behaviour relative to equities.

4.4.1 Cross-Market Comparison

Three broad conclusions emerge. First, equity indices share broadly similar parameter values, with low drifts, moderate daily volatilities, and strong GARCH persistence. Sec-

ond, the Nikkei shows slightly higher volatility and weaker mean reversion, reflecting its historically lower returns. Third, Bitcoin consistently stands out: it has far larger volatility, stronger ARCH effects, and higher mean-reversion intensity. These features underline its role as an outlier and a useful stress test for the models examined in the following sections.

4.5 Empirical Cumulative Distribution Function (ECDF) Analysis

A key test of model realism is how well simulated returns reproduce the shape of the empirical distribution. This is assessed using empirical cumulative distribution functions (ECDFs). The ECDF shows, for each return size, the proportion of days on which the return is less than or equal to that value. Matching the ECDF means that the model reproduces both the center of the distribution and its extreme tails.

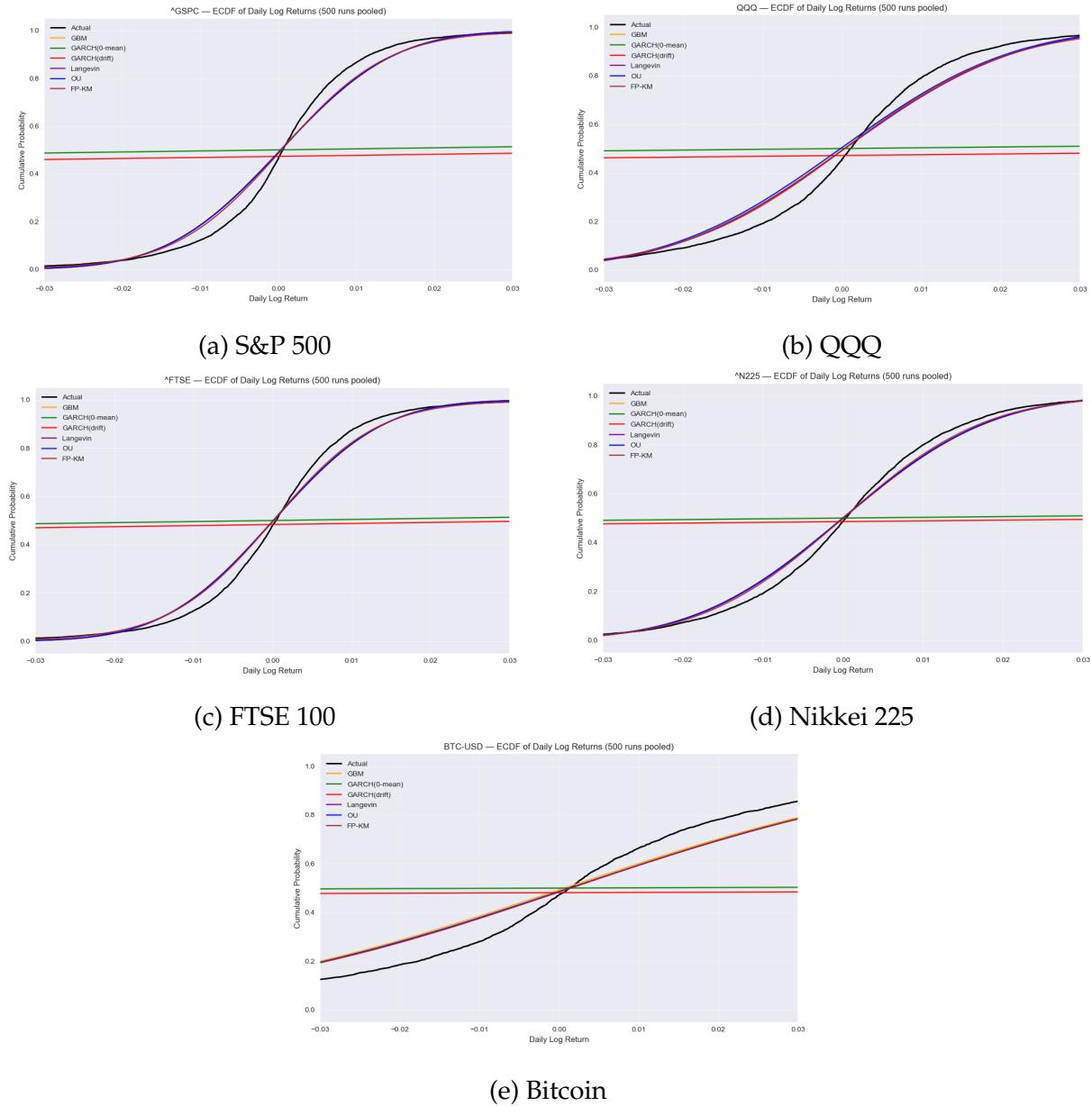


Figure 4.7: ECDF of daily log returns: actual data (black) vs pooled simulations from 500 runs of GBM, GARCH, Langevin, OU, and FP–KM models. In some cases (e.g., S&P 500), the Langevin, OU, and FP–KM curves overlap because their fits are very similar.

4.5.1 Comparison of Model Fit

The ECDF plots reveal several systematic patterns. The GBM model consistently underestimates the probability of extreme events, producing overly thin tails. GARCH models improve tail behaviour, but tend to concentrate too much probability near the center, failing to fully match the heavy tails. Langevin and OU processes better capture asymmetry and fat tails, bringing the simulated ECDFs closer to the empirical

curves. The FP–KM approach performs best in reproducing the overall shape, since it is directly estimated from empirical drift and diffusion terms. In some markets, the fits of Langevin, OU, and FP–KM overlap so closely that their curves are visually indistinguishable.

4.5.2 Strengths and Weaknesses

Three key findings emerge. First, at the center of the distribution, GBM aligns reasonably well for small, everyday returns but breaks down for larger moves. GARCH tends to overweight the center, giving too few large moves. Second, in the tails, Langevin, OU, and FP–KM all produce fatter tails that are much closer to the empirical data, especially for equities. Bitcoin remains the hardest to fit, with none of the models fully capturing its extreme +22.5% and -46.5% tail events (see Table 4.3). Third, across markets, parametric models perform adequately for equities, but only FP–KM comes close for Bitcoin, and even then the fit remains imperfect.

Now, the ECDF analysis confirms that Gaussian-based GBM is inadequate for real markets, while physics-inspired models and non-parametric FP–KM offer clear improvements in capturing the distributional features observed in the data.

4.6 Model Evaluation Metrics

Beyond visual comparison of distributions, quantitative metrics provide a systematic way to rank model performance. Three measures are used:

- **Root Mean Squared Error (RMSE):** evaluates how closely simulated price paths match the actual price series. Lower RMSE indicates better pathwise fit.
- **Kolmogorov–Smirnov (KS) statistic:** compares the distribution of simulated and actual returns. Smaller values mean the simulated and empirical distributions are more similar.
- **Wasserstein distance:** measures the average distance between simulated and actual return distributions. Again, smaller is better.

Table 4.7 reports the mean and standard deviation of these metrics across 500 simulation runs for each model and market. The results show clear differences in strengths and weaknesses.

Table 4.7: Aggregated model evaluation metrics (500 runs). Reported as mean \pm standard deviation.

Ticker	Model	RMSE_mean	RMSE_sd	KS_mean	KS_sd	Wasserstein_mean	Wasserstein_sd
S&P 500	GBM	1760.82	1900.72	0.095	0.005	0.0023	0.0001
	OU	2103.67	2004.62	0.093	0.004	0.0023	0.0001
	Langevin	2353.51	2888.90	0.092	0.005	0.0023	0.0001
	FP-KM	8327.78	14296.99	0.082	0.005	0.0020	0.0001
	GARCH(0-mean)	∞	NaN	0.482	0.003	0.807	0.052
	GARCH(drift)	∞	NaN	0.505	0.005	0.813	0.052
QQQ	GBM	230.85	351.23	0.097	0.006	0.0034	0.0001
	OU	152.22	23.61	0.106	0.006	0.0035	0.0001
	Langevin	447.97	954.52	0.095	0.005	0.0034	0.0001
	FP-KM	437.91	906.62	0.090	0.006	0.0031	0.0001
	GARCH(0-mean)	∞	NaN	0.481	0.004	1.143	0.102
	GARCH(drift)	∞	NaN	0.506	0.006	1.156	0.119
FTSE 100	GBM	3662.84	2796.29	0.075	0.004	0.0019	0.0001
	OU	1697.37	448.02	0.073	0.003	0.0018	0.0001
	Langevin	5272.37	6938.15	0.073	0.004	0.0018	0.0001
	FP-KM	4776.32	5697.46	0.068	0.005	0.0016	0.0001
	GARCH(0-mean)	2.90e46	6.35e43	0.484	0.004	0.779	0.041
	GARCH(drift)	∞	NaN	0.497	0.005	0.778	0.039
Nikkei	GBM	26821.73	35088.07	0.062	0.005	0.0019	0.0001
	OU	11998.91	5271.07	0.061	0.004	0.0019	0.0001
	Langevin	45633.14	74472.15	0.060	0.004	0.0019	0.0001
	FP-KM	47079.69	88492.81	0.054	0.004	0.0016	0.0001
	GARCH(0-mean)	∞	NaN	0.485	0.004	1.116	0.046
	GARCH(drift)	∞	NaN	0.496	0.005	1.119	0.044
Bitcoin	GBM	38867.66	74016.62	0.111	0.008	0.0079	0.0003
	OU	124579.98	391894.48	0.105	0.006	0.0079	0.0003
	Langevin	262130.42	1462294.39	0.105	0.006	0.0079	0.0003
	FP-KM	431845.76	2006502.44	0.103	0.007	0.0076	0.0003
	GARCH(0-mean)	∞	NaN	0.488	0.005	3.125	0.159
	GARCH(drift)	∞	NaN	0.502	0.008	3.141	0.154

4.6.1 Comparative Ranking of Models

The evaluation metrics reveal consistent patterns across markets. GBM provides a simple baseline but systematically underestimates large moves. For instance, its KS statistic for the S&P 500 is 0.095 and the Wasserstein distance 0.0023, whereas FP-KM achieves a lower KS (0.082) and Wasserstein (0.0020). On Bitcoin, GBM struggles even more, with RMSE around 38,868 and KS 0.111, confirming its inability to capture the cryptocurrency's extreme volatility.

GARCH models demonstrate their strength in volatility modelling but break down when evaluated on price paths. Across all assets, both GARCH(0-mean) and GARCH(drift) diverge, with RMSE values blowing up to ∞ or returning NaN. Their KS values are much higher than physics-inspired alternatives (e.g., 0.482–0.506 for equities), and while Wasserstein distances for equities are moderate (around 0.78 for the FTSE and 1.14 for QQQ), Bitcoin again exposes their weakness, with Wasserstein distances above 3.1. These results highlight the instability of GARCH for long-horizon simulations despite its strong theoretical motivation.

Langevin and Ornstein–Uhlenbeck processes provide a more balanced compromise between realism and tractability. For equities, they achieve relatively low KS statistics (0.092–0.106) and close Wasserstein alignment (around 0.0018–0.0019), indicating that their simulated return distributions resemble the real ones. However, their price-path errors (RMSE) remain high. For the S&P 500, Langevin produces an RMSE of about 2,353, while OU is similar at just over 2,100. The Nikkei shows a sharper contrast: OU achieves an RMSE near 12,000, whereas Langevin diverges much further, above 45,000. On Bitcoin, both models deteriorate substantially, with RMSE values exceeding 120,000 and KS statistics around 0.105. These results suggest that Langevin and OU are effective for capturing distributional features of equities, but their ability to track long, volatile price paths—especially in highly unstable assets like Bitcoin—is limited.

FP-KM consistently performs best in distributional similarity, as expected from its non-parametric nature. It produces the lowest KS values across all markets (0.082 for the S&P 500, 0.054 for the Nikkei) and the smallest Wasserstein distances (0.0016 for FTSE and Nikkei). However, this comes at a cost: its price-path RMSE values are far larger than those of parametric models (e.g., 8,328 for the S&P 500 and 431,846 for Bitcoin), reflecting its lack of pathwise control.

The metrics show that no single model dominates. GBM offers a baseline but misses extremes, GARCH captures volatility clustering but collapses in simulation, Langevin and OU strike a balance between realism and tractability, and FP–KM excels in matching distributions while failing on price paths. These trade-offs reinforce the need for a multi-model perspective, where different approaches capture different aspects of market dynamics.

4.6.2 Comparison of Entropy Fit

The entropy results highlight sharp contrasts between model families. As shown in Table 4.8 and Table 4.9, GBM, Langevin, OU, and FP–KM all fail to track entropy dynamics closely, with RMSE values for Shannon entropy often in the tens of thousands (e.g., for the S&P 500, GBM produces an error of 88,698) and Tsallis errors reaching into the millions (GBM: 8.3 million). These models capture broad distributional features but not the temporal fluctuations in uncertainty.

By contrast, the GARCH family produces dramatically lower errors, with Shannon RMSE values typically in the range of 2,000–4,000 and Tsallis RMSE in the tens of thousands. For example, QQQ’s GARCH models achieve Shannon errors around 2,448 compared with nearly 58,000 for GBM. Even in the most volatile case of Bitcoin, GARCH keeps Shannon entropy errors as low as 557, while all other models exceed 20,000. This demonstrates the particular strength of GARCH in reproducing entropy dynamics, consistent with its focus on volatility clustering.

Taken together, these results show that while physics-inspired models and FP–KM provide valuable realism in capturing fat tails and distributions, they fall short in replicating entropy behaviour. The GARCH models stand out as the most reliable for entropy-based diagnostics, highlighting the importance of explicitly modelling volatility dynamics.

Table 4.8: RMSE of simulated entropy series vs. actual entropy (Shannon and Tsallis), Part I

Asset	Model	Shannon_RMSE	Tsallis_RMSE
S&P 500	GBM	88,698	8,292,222
	GARCH(0-mean)	3,728	62,137
	GARCH(drift)	3,743	62,175
	Langevin	90,756	8,576,244
	OU	92,585	8,822,214
	FP-KM	95,995	9,466,281
QQQ	GBM	57,735	3,851,234
	GARCH(0-mean)	2,448	30,631
	GARCH(drift)	2,447	30,627
	Langevin	56,996	3,849,714
	OU	57,900	3,895,245
	FP-KM	58,208	4,028,044
FTSE 100	GBM	94,425	9,251,818
	GARCH(0-mean)	3,676	55,126
	GARCH(drift)	3,668	55,075
	Langevin	98,399	9,646,255
	OU	97,764	9,544,095
	FP-KM	101,990	10,255,654

Table 4.9: RMSE of simulated entropy series vs. actual entropy (Shannon and Tsallis), Part II

Asset	Model	Shannon_RMSE	Tsallis_RMSE
Nikkei	GBM	69,607	5,266,870
	GARCH(0-mean)	2,170	24,445
	GARCH(drift)	2,170	24,442
	Langevin	68,205	5,197,766
	OU	70,366	5,401,364
	FP-KM	72,504	5,711,821
Bitcoin	GBM	23,352	832,817
	GARCH(0-mean)	557	4,013
	GARCH(drift)	557	4,011
	Langevin	23,939	870,572
	OU	24,129	870,906
	FP-KM	23,906	863,798

Conclusions

This dissertation explored how ideas borrowed from physics can provide richer insights into financial market dynamics than classical finance models alone. Traditional approaches, such as the Geometric Brownian Motion (GBM), assume that markets follow smooth, Gaussian processes with fixed risk. In reality, however, markets are far more irregular. Crashes happen more often than a bell-curve model would suggest, calm periods can suddenly give way to turbulence, and extreme events tend to cluster rather than occur in isolation.

By applying physics-inspired models, this study showed that markets can be described in more realistic terms. Processes such as the Langevin and Ornstein–Uhlenbeck captured how prices drift, revert, and fluctuate under the constant push and pull of random shocks. The Fokker–Planck approach added further depth by focusing on the evolution of entire return distributions rather than individual price paths. These methods do not replace traditional models, but they provide more convincing accounts of how instability builds and spreads, especially during stressed conditions.

Entropy analysis offered an additional diagnostic layer. Unlike volatility, which only measures the size of price changes, entropy captures how unpredictable or disordered returns become. This study found that entropy often rises before volatility, especially in equity indices such as the S&P 500 and QQQ, suggesting that entropy can act as a short-term early warning signal of turbulence. Bitcoin displayed even stronger and more erratic entropy signals, underlining the usefulness of entropy in highly volatile and less structured markets. These findings support the idea that entropy complements

traditional risk measures by detecting instability at an earlier stage.

The results also confirmed that return distributions across all markets exhibit negative skewness and high kurtosis. In practice, this means extreme events occur more frequently than a normal model predicts (fat tails), and large losses are more likely than equally large gains (negative skew). Taken together, these features point to a heightened probability of severe downside shocks, a pattern that classical Gaussian models fail to capture.

The findings suggest that markets are best understood as complex, adaptive systems rather than simple random walks. Physics-inspired models make this complexity visible, helping us appreciate features that traditional finance tends to overlook. While no model can predict markets with certainty, this dissertation has shown that borrowing ideas from physics provides more realistic insights into how shocks emerge, how volatility clusters, and why markets sometimes move in seemingly unpredictable ways.

5.1 Limitations and Market Structure Considerations

This study has several limitations that should be acknowledged.

First, it focused exclusively on index-level data. Indices can be dominated by a small number of very large firms. For example, in the S&P 500, mega-cap companies such as Apple and Microsoft carry disproportionate weight because of the market-capitalisation methodology. This concentration means that index behaviour may partly reflect the fortunes of a handful of firms rather than the broad market. Future research could apply physics-inspired tools at the sectoral or constituent level to examine whether results are driven by overall market complexity or by the influence of dominant firms.

Secondly, only a small set of indices plus Bitcoin were analysed. Other asset classes such as commodities, foreign exchange, or government bonds may display different dynamics that deserve investigation.

Finally, the models here were primarily evaluated in descriptive terms, focusing on their ability to capture stylised facts such as fat tails, skewness, and clustering. They were not designed or tested as predictive forecasting tools, which limits their direct practical application. Future work could bridge this gap by assessing out-of-sample forecasting performance and comparing against machine learning approaches.

5.2 Future Work

There are several promising directions for extending this research, which can be grouped into three main areas.

5.2.1 Market Regimes and Time Horizons

A first way to extend this study would be to see how the models perform under different market conditions instead of looking at the whole period as one block. For example, we could separate the data into bull markets (rising prices), bear markets (falling prices), and crisis periods. This might show that some models work better in calm times, while others are more useful during turbulence. In the same way, comparing the periods just before and after a crash could reveal whether tools like entropy and distribution tests can act as early-warning signals of trouble.

Another direction would be to test the models on intraday data, such as hourly or minute-by-minute prices. These shorter time frames often show more sudden bursts of activity and stronger volatility patterns. Studying them would help us see whether the models remain reliable across different time scales, or whether their strengths depend on the frequency of the data.

5.2.2 Hybrid and Extended Models

Another way forward would be to make the models themselves more flexible by combining their different strengths. The results of this dissertation suggest some natural pairings. For example, GBM gives a simple baseline but cannot capture volatility clustering, while GARCH does capture clustering but struggles with distributional accuracy. A combined model could take GBM's clear price structure and add GARCH's volatility patterns, filling in each other's gaps. In the same way, the Ornstein–Uhlenbeck (OU) process is good at showing mean reversion but assumes volatility is constant; joining it with GARCH could better reflect both the trend and the changing risk levels. The Fokker–Planck method matched real return distributions well but lacked predictive power; linking it with entropy-based measures could add an early-warning ability. Exploring these kinds of hybrid or extended models would be a natural next step, building on the strengths and weaknesses identified in this study.

5.2.3 Broader Scope

Future research could also widen the scope of analysis beyond the stock indices and Bitcoin examined in this dissertation. Other markets, such as commodities, foreign exchange, and government bonds, may behave very differently. Testing physics-inspired models on these assets would help us see whether the approaches used here are general tools or whether they only work well in certain markets.

Overall, extending the work along these three dimensions—market regimes and time horizons, hybrid models, and broader scope—would deepen our understanding of the conditions under which physics-inspired approaches provide the greatest value, and how they can be integrated into the wider toolkit of financial modelling.

Appendix



Supplementary Figures

A.1 Extended Result Visualisations

This appendix presents full-size versions of the key visualisations used in the main text. They are grouped by market and mirror the analyses in Chapter 4 (Results) and the corresponding methods in Chapter 3 (Methodology). Specifically:

- *Price and log-price plots* provide context for market regimes (bull, bear, and crisis periods).
- *Rolling volatility* illustrates time-varying risk, highlighting clustering behaviour.
- *Return distributions and ECDFs* support the distributional comparisons discussed in Section 4.5.
- *Entropy versus volatility* and *lead-lag correlations* connect directly to the entropy diagnostics discussed in Section 3.5.

For readability, each figure is shown at full text width with a descriptive caption reminding the reader how it was used in the analysis.

A.1.1 Bitcoin (BTC)

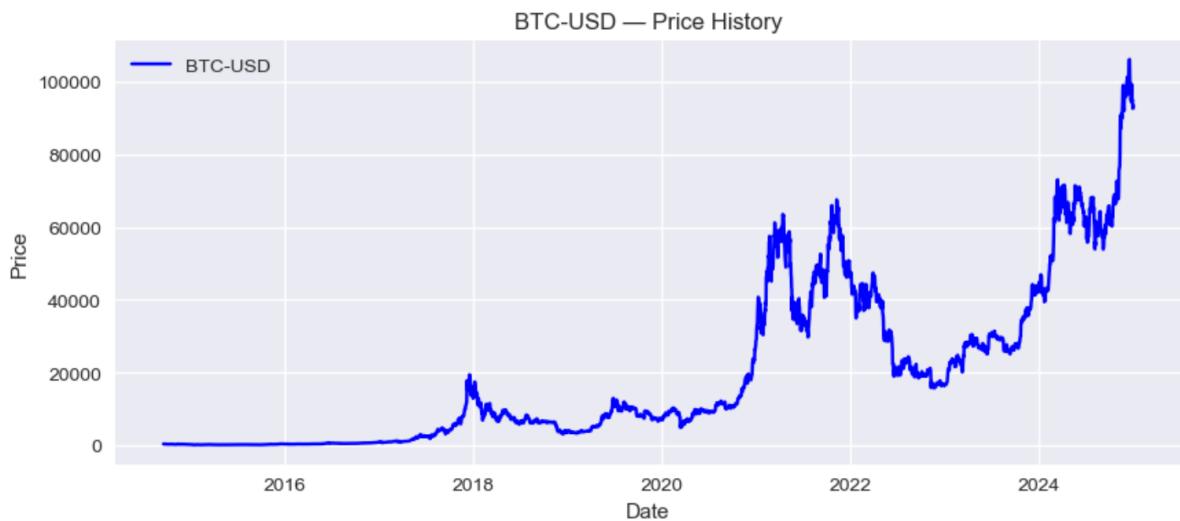


Figure A.1: BTC closing price (1992–2025). Used in Chapter 4 for regime context and entropy/volatility interpretation.

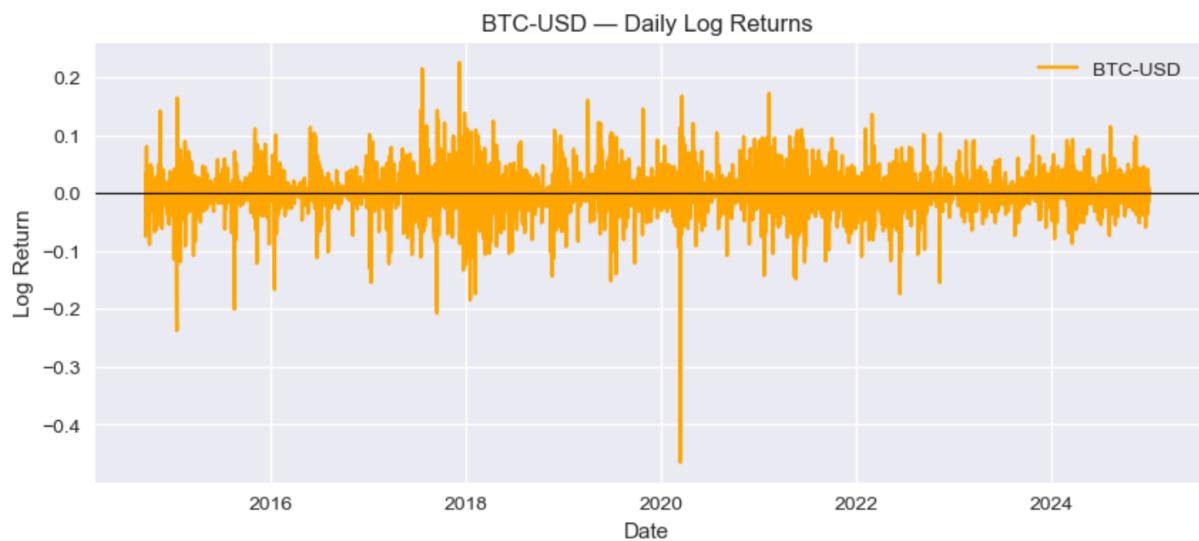


Figure A.2: BTC log-price, stabilising long-run growth for comparison. Supports scaling and mean-reversion discussions.

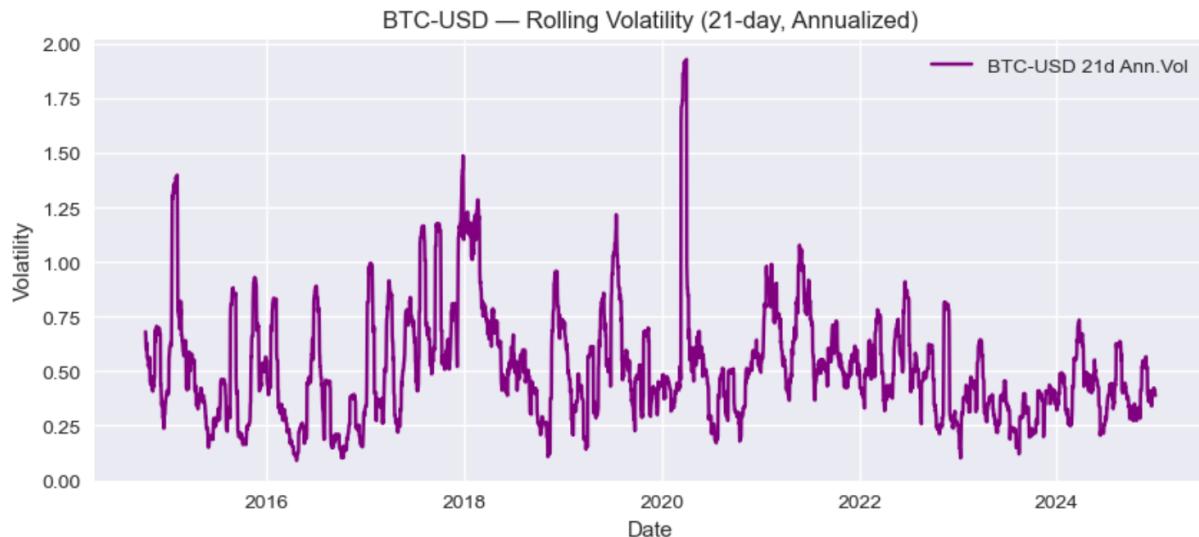


Figure A.3: BTC rolling volatility, illustrating clustering behaviour. Highlights Bitcoin's turbulence relative to equities.

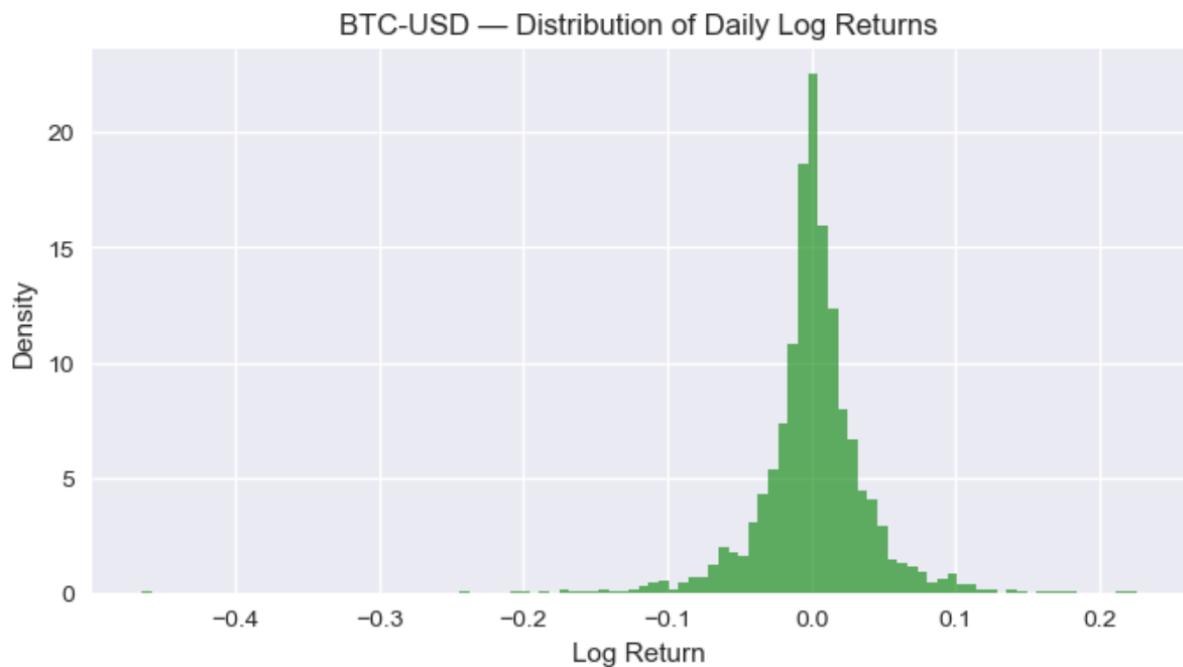


Figure A.4: BTC return distribution (histogram) showing heavy tails. Used alongside KS and Wasserstein metrics in Section 4.5.

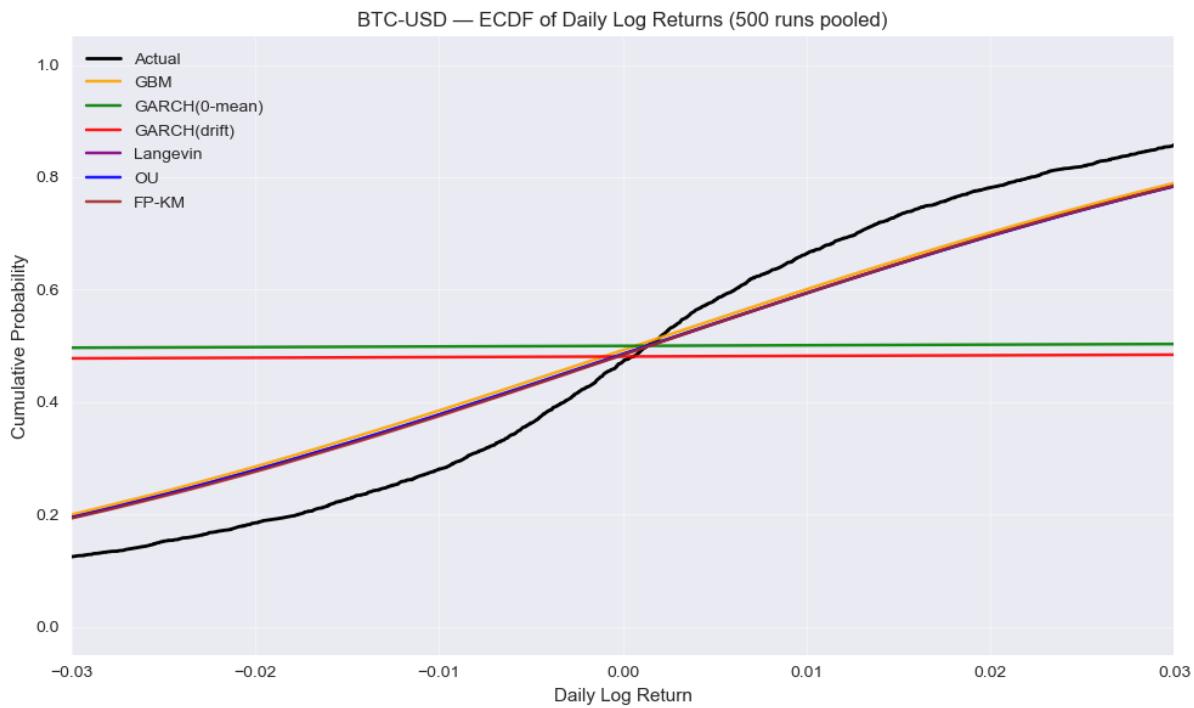


Figure A.5: BTC empirical cumulative distribution function (ECDF) of returns compared with model ECDFs (see Section 4.5).

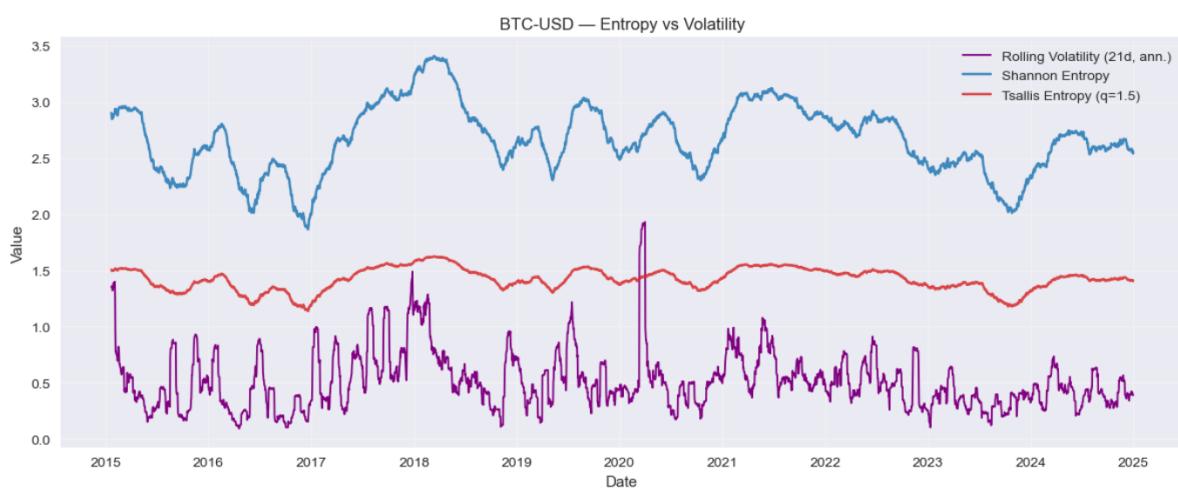


Figure A.6: BTC Shannon and Tsallis entropy compared with volatility over time. Complements entropy diagnostics in Section 3.5.

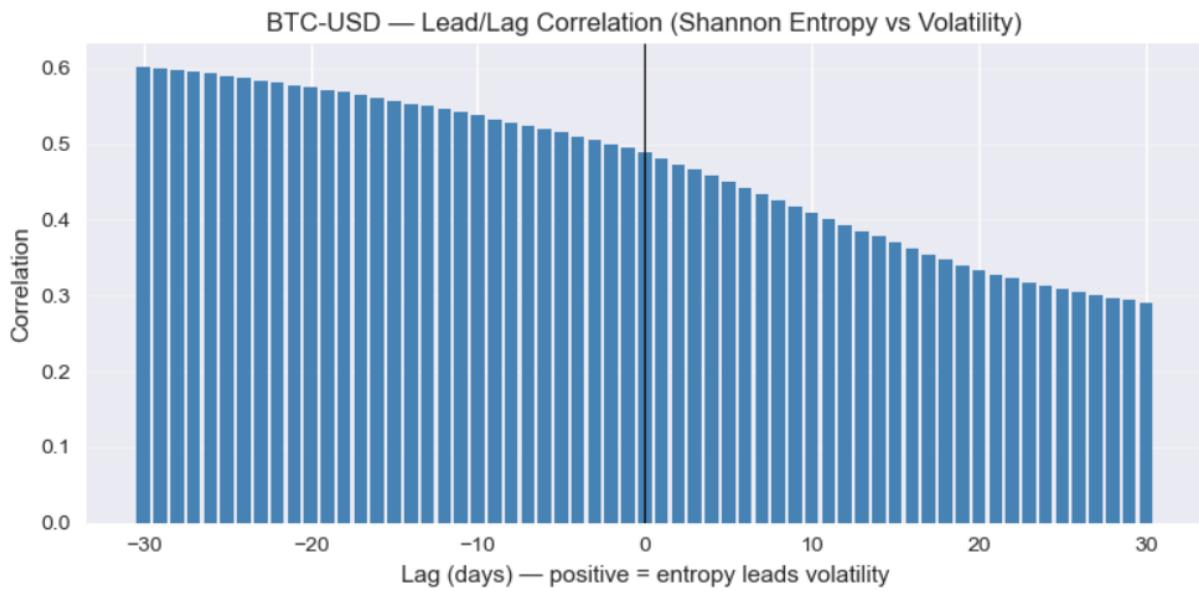


Figure A.7: BTC entropy–volatility lead–lag correlation, testing whether entropy leads volatility during shocks.

A.1.2 S&P 500 (GSPC)



Figure A.8: S&P 500 closing price. Context for market phases discussed in Chapter 4.

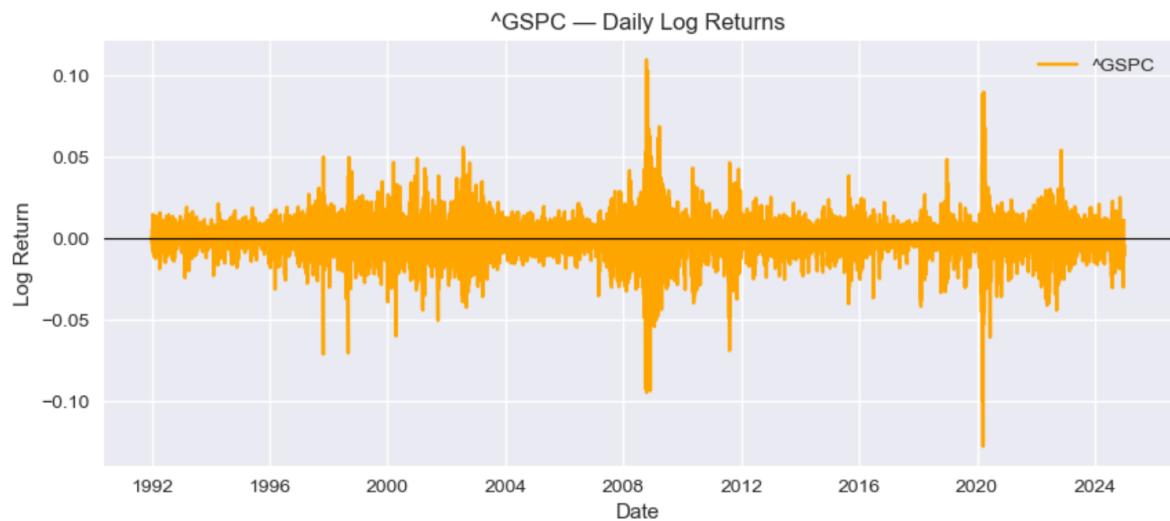


Figure A.9: S&P 500 log-price, supporting long-term trend versus shock analysis.

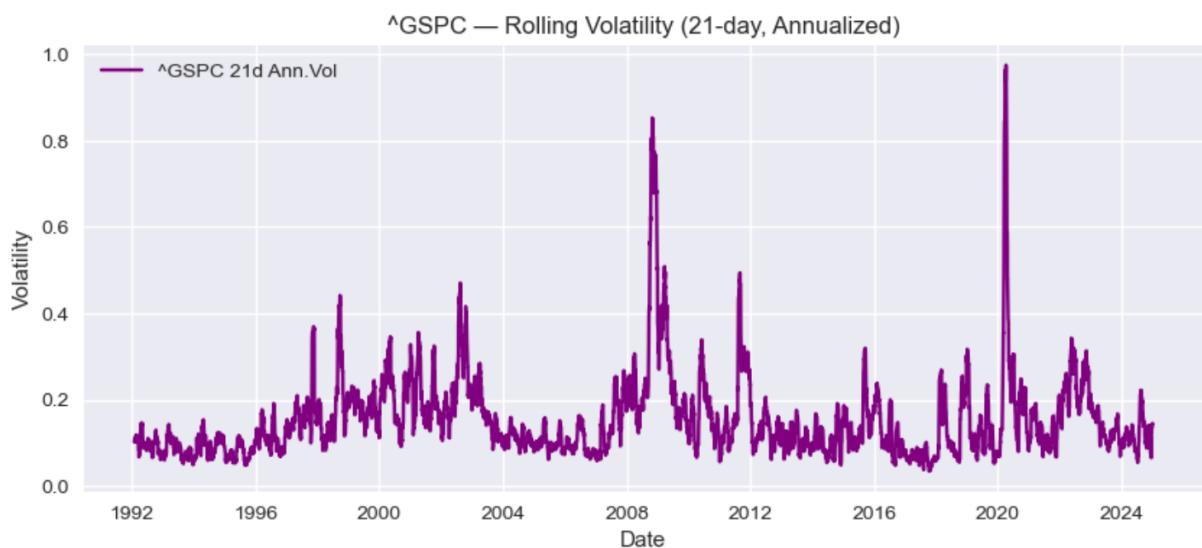


Figure A.10: S&P 500 rolling volatility, clustered around major crises.

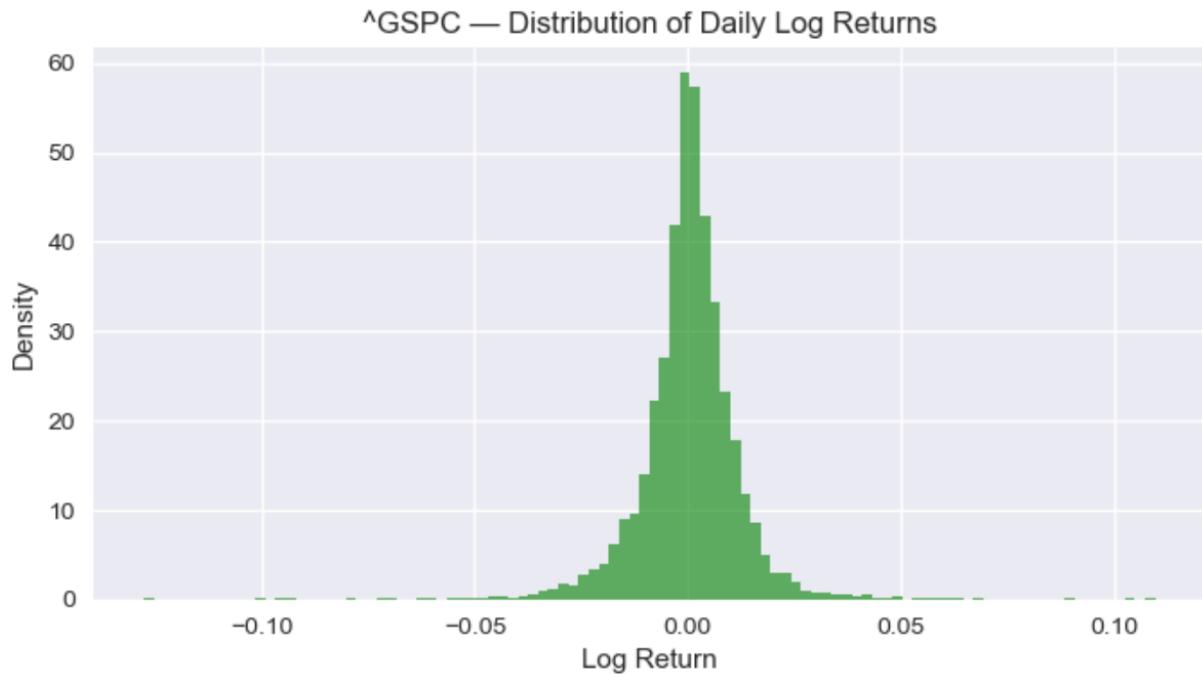


Figure A.11: S&P 500 return distribution used in distributional matching (KS and Wasserstein metrics, Section 4.5).

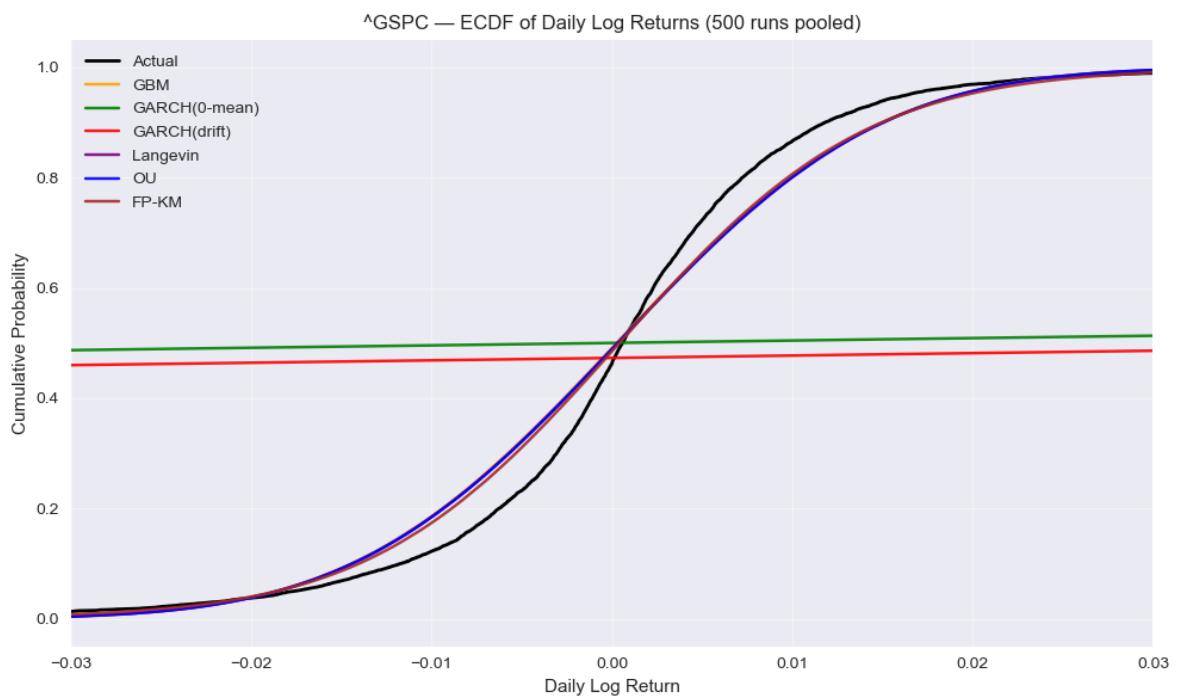


Figure A.12: S&P 500 ECDF of returns compared with simulated distributions (see Section 4.5).

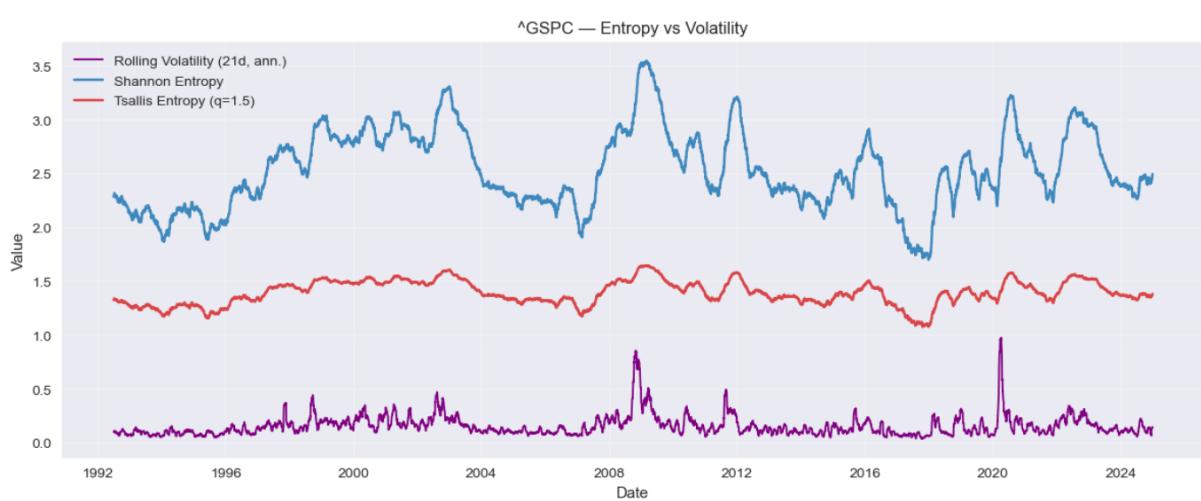


Figure A.13: S&P 500 entropy versus volatility. Compared across models in Tables 4.8–4.9.

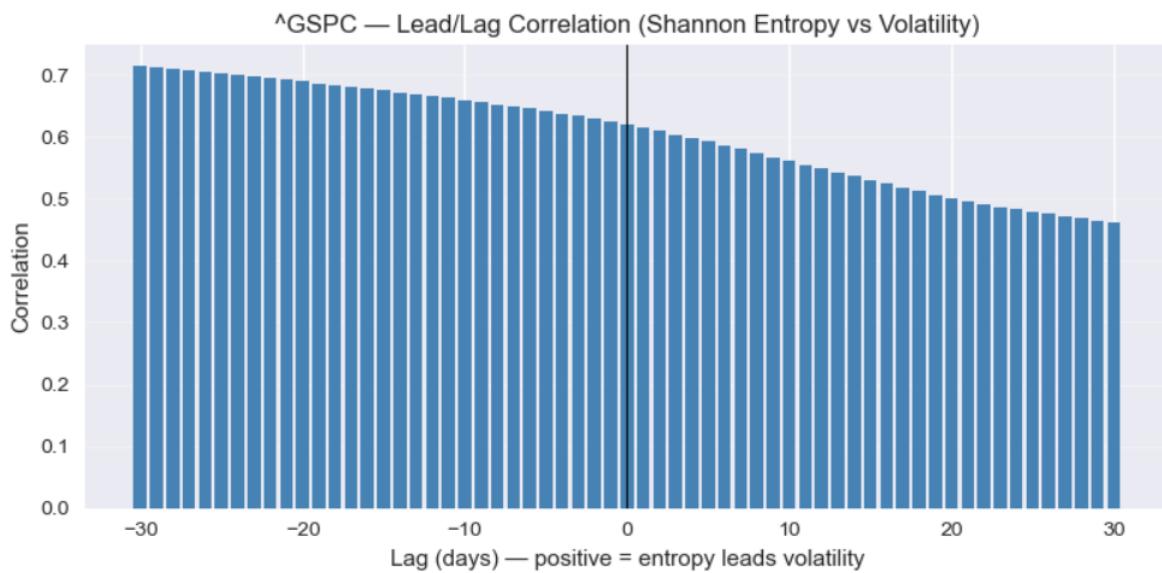


Figure A.14: S&P 500 entropy–volatility lead–lag analysis.

A.1.3 QQQ (NASDAQ-100)



Figure A.15: QQQ closing price, contextualising technology-led market cycles.

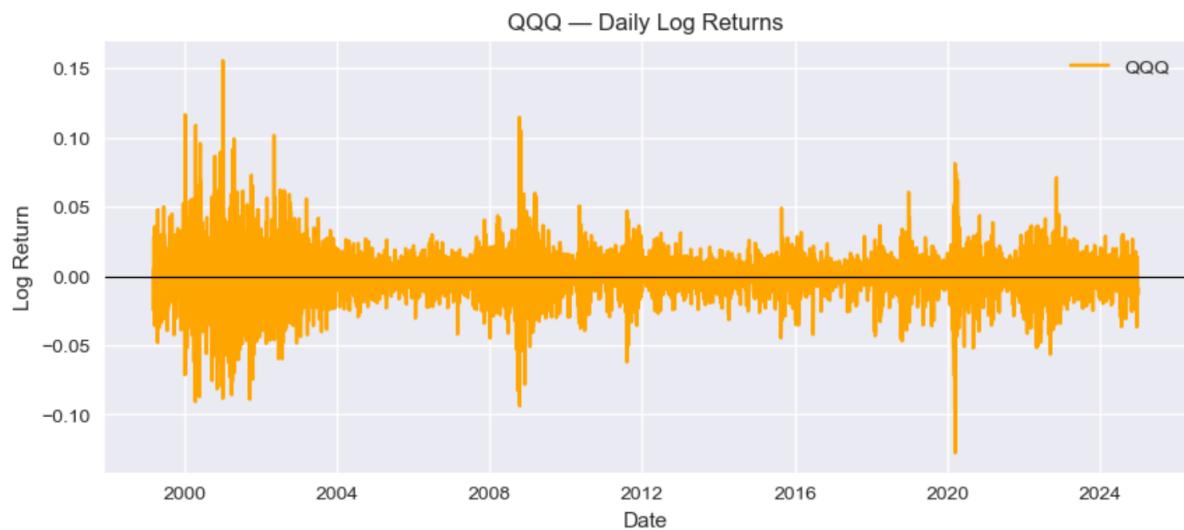


Figure A.16: QQQ log-price, aiding comparison of growth and drawdowns.

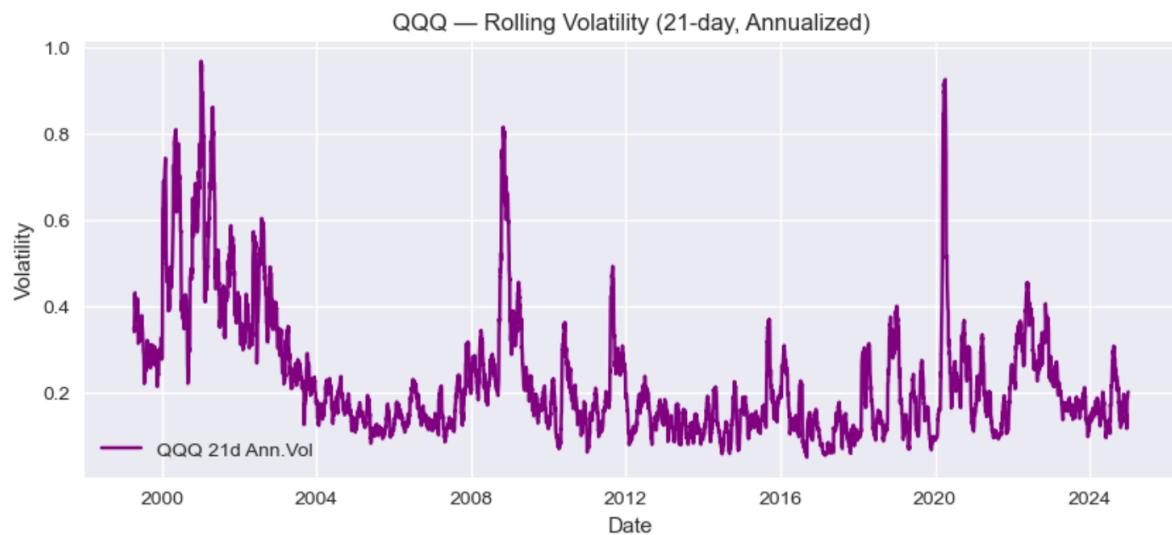


Figure A.17: QQQ rolling volatility, highlighting clustering during tech sell-offs.

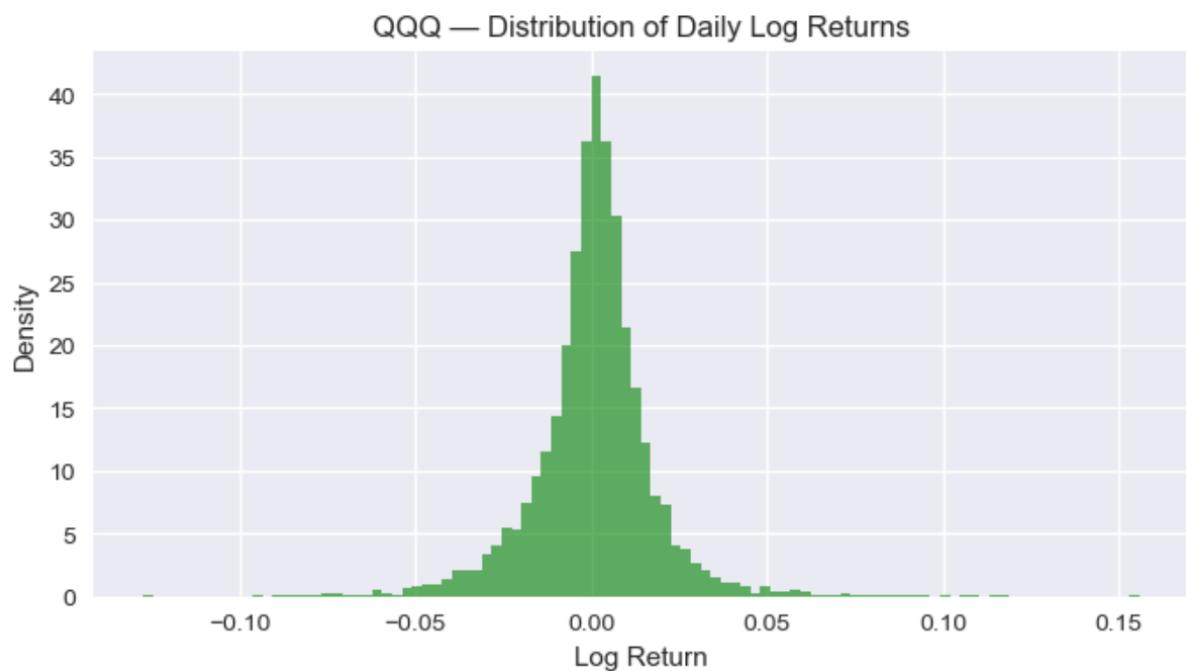


Figure A.18: QQQ return distribution for KS and Wasserstein comparisons (Section 4.5).

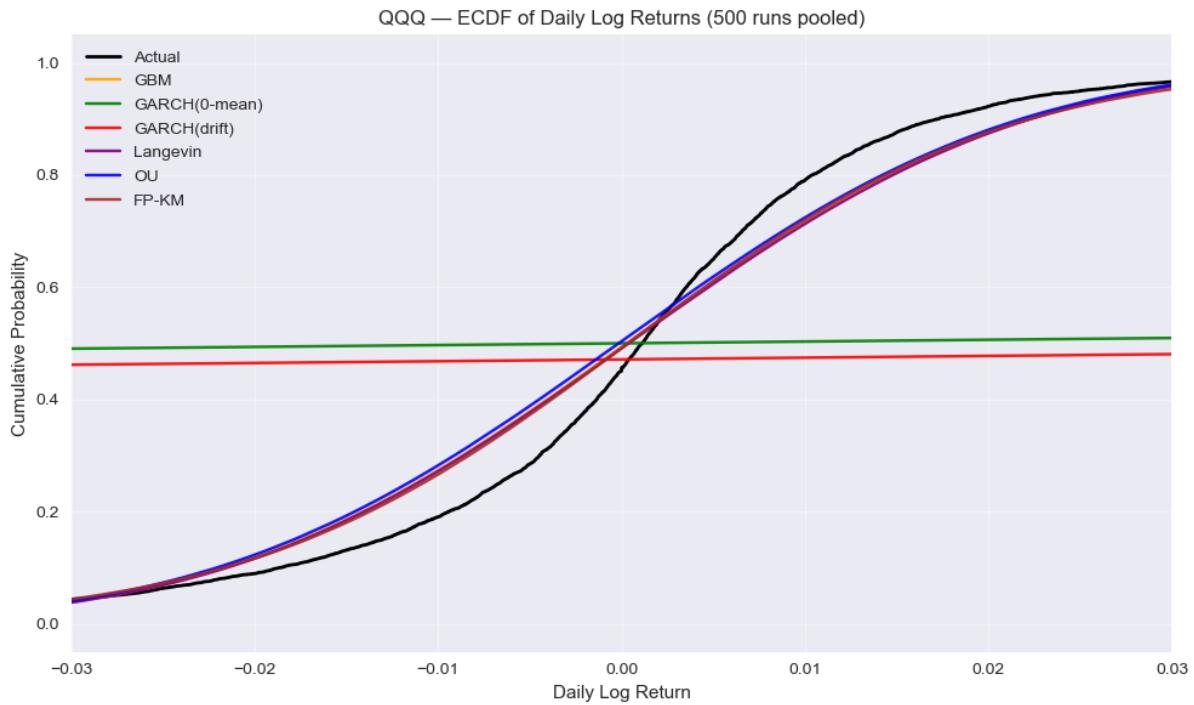


Figure A.19: QQQ ECDF of returns compared with model ECDFs in Section 4.5.

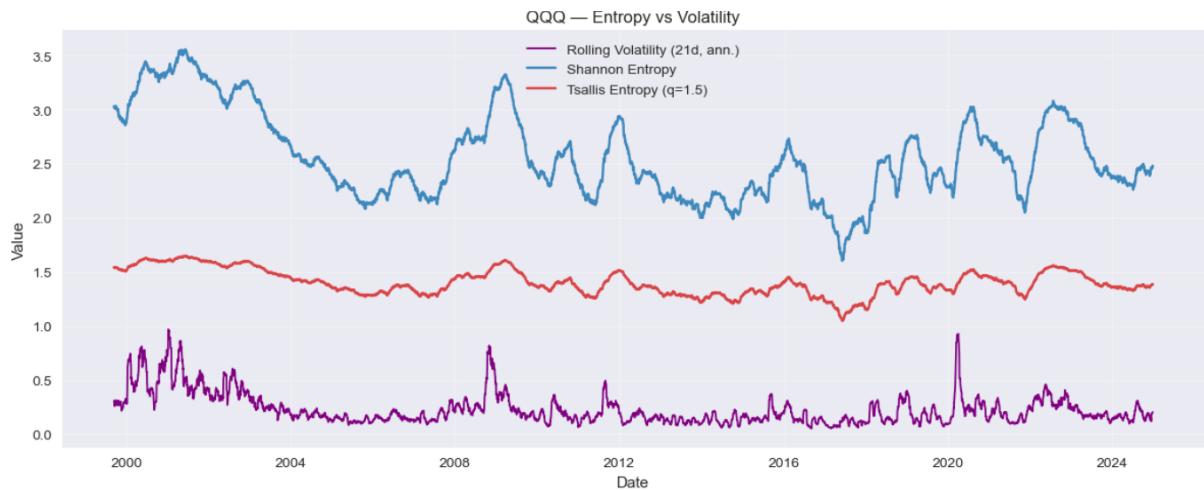


Figure A.20: QQQ entropy versus volatility. Benchmarks entropy tracking error across models.

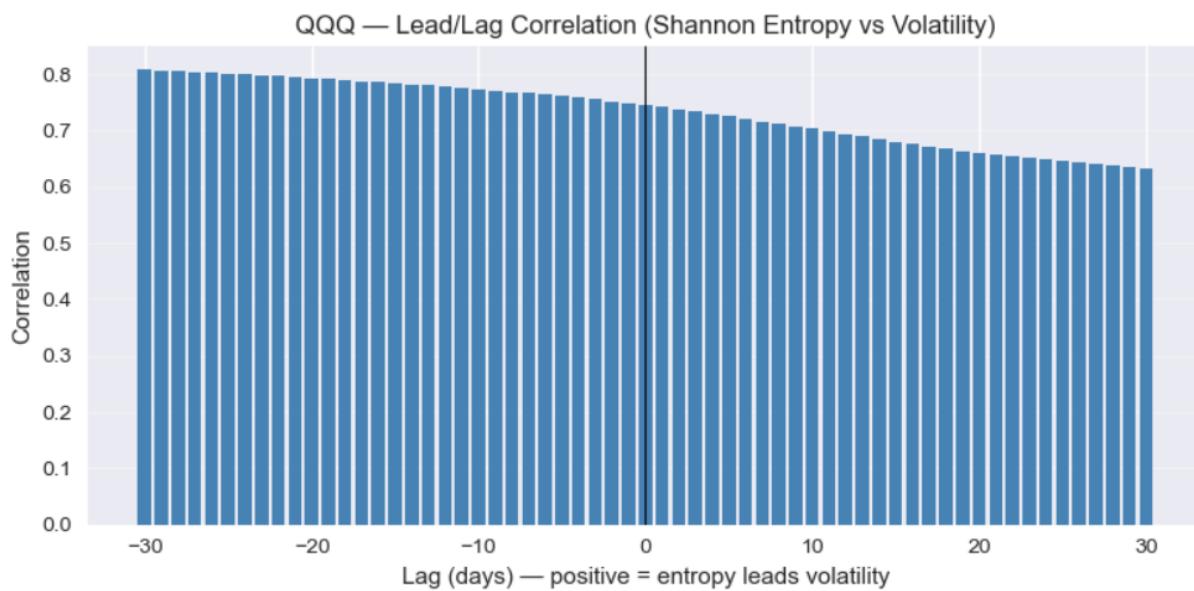


Figure A.21: QQQ entropy–volatility lead–lag correlation.

A.1.4 FTSE 100



Figure A.22: FTSE 100 closing price. Context for UK-specific dynamics in Chapter 4.

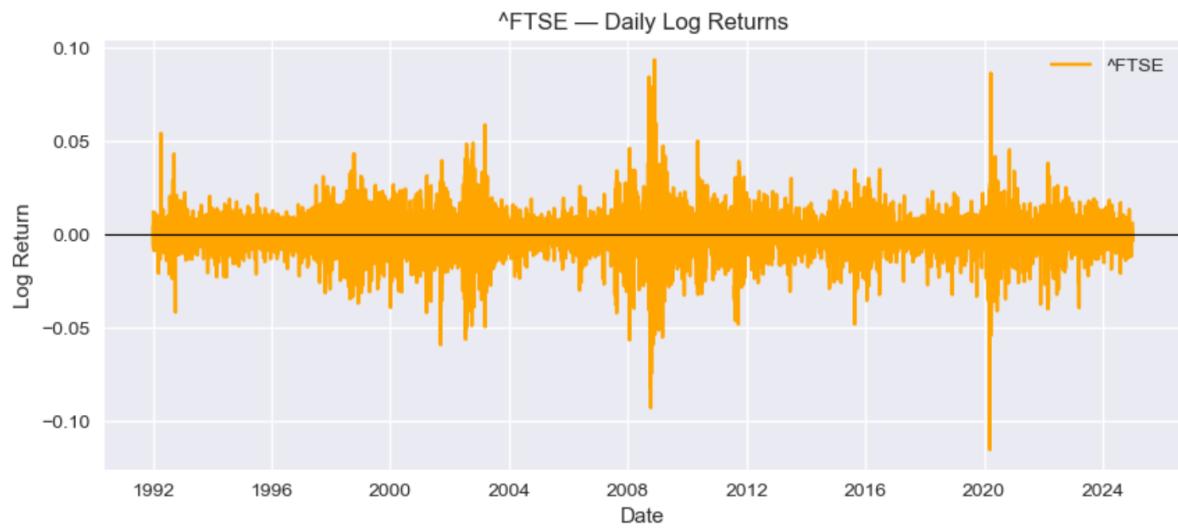


Figure A.23: FTSE 100 log-price over the sample period.

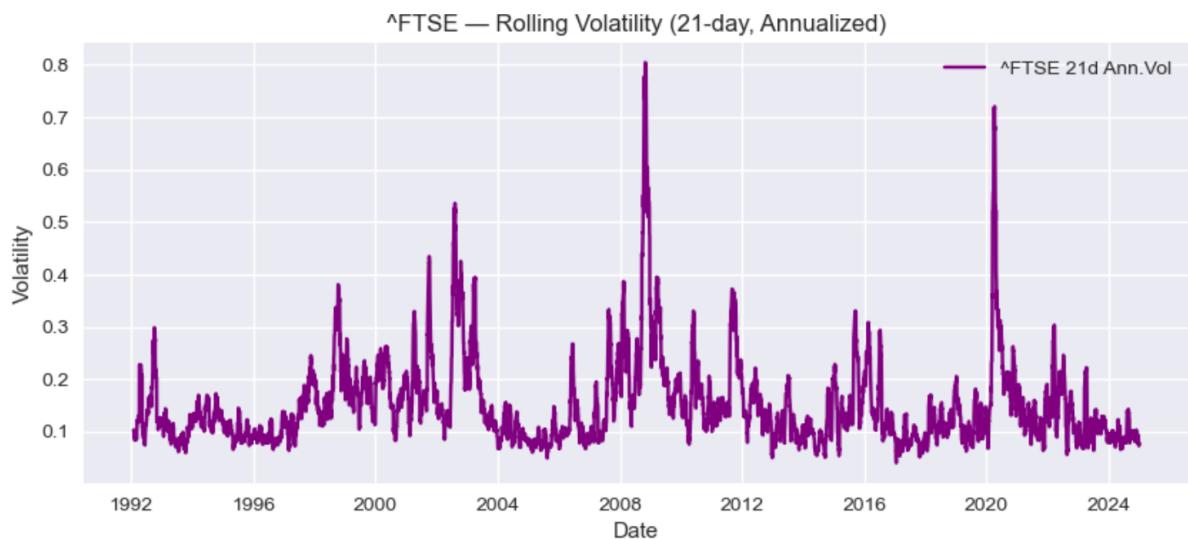


Figure A.24: FTSE 100 rolling volatility highlighting clustering around shocks.

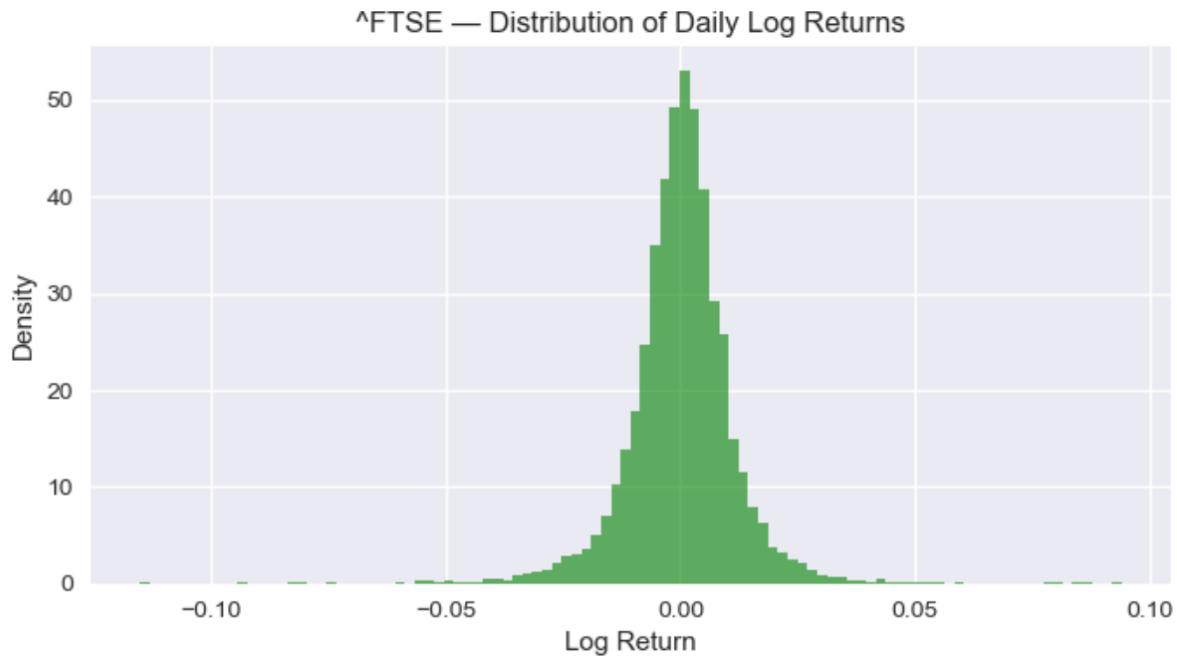


Figure A.25: FTSE 100 return distribution feeding KS/Wasserstein tests (Section 4.5).

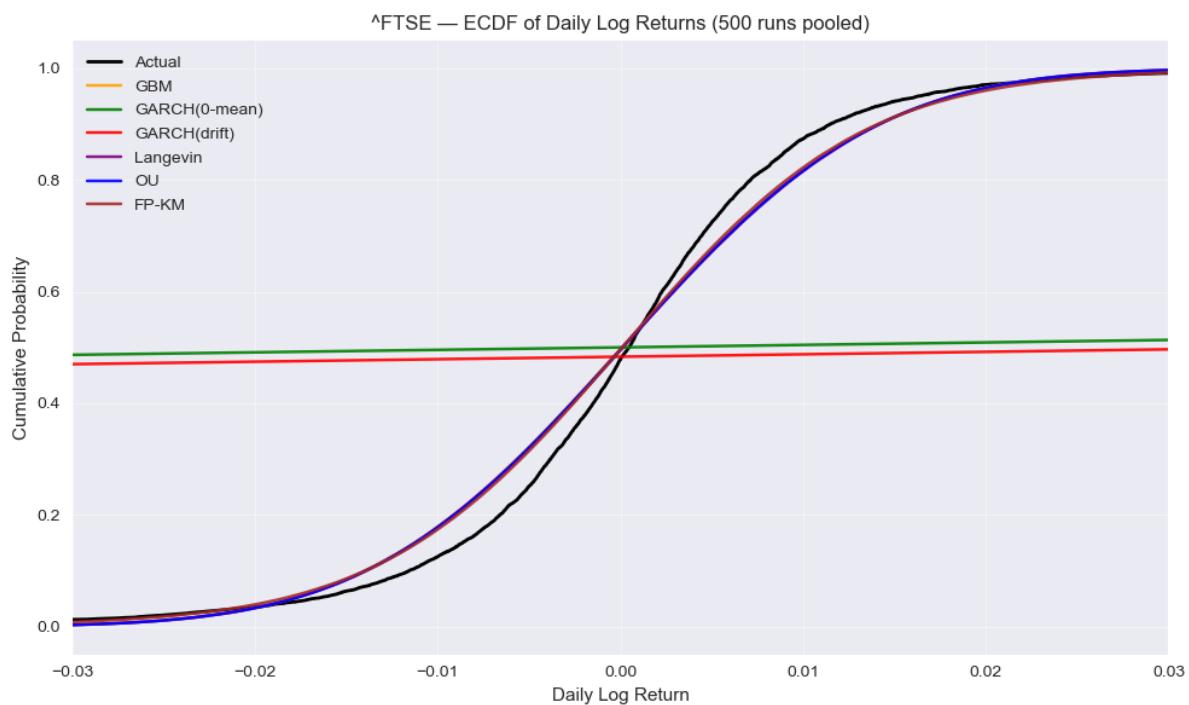


Figure A.26: FTSE 100 ECDF of returns compared with simulated distributions.

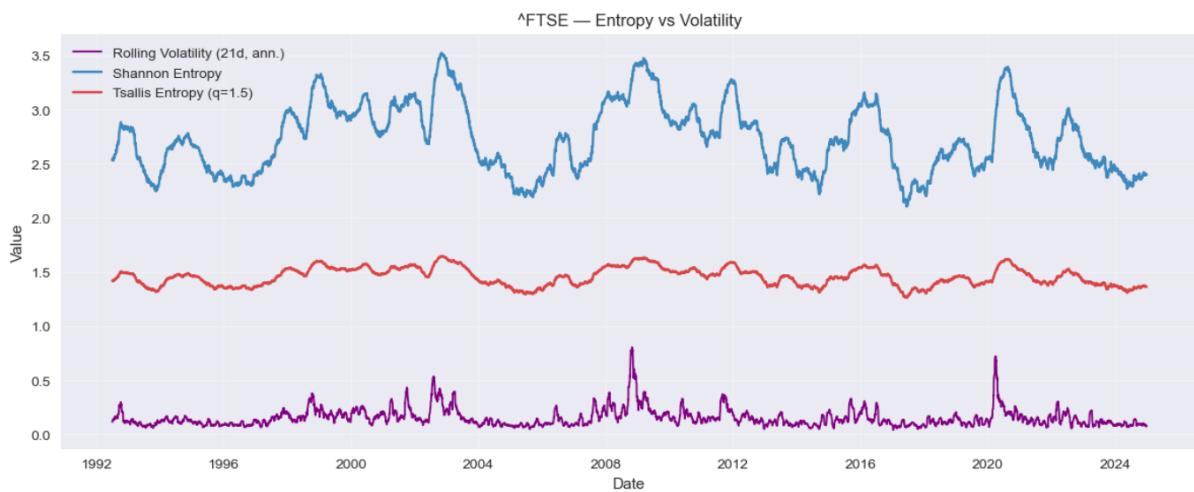


Figure A.27: FTSE 100 entropy versus volatility, used to assess entropy tracking (Tables 4.8–4.9).

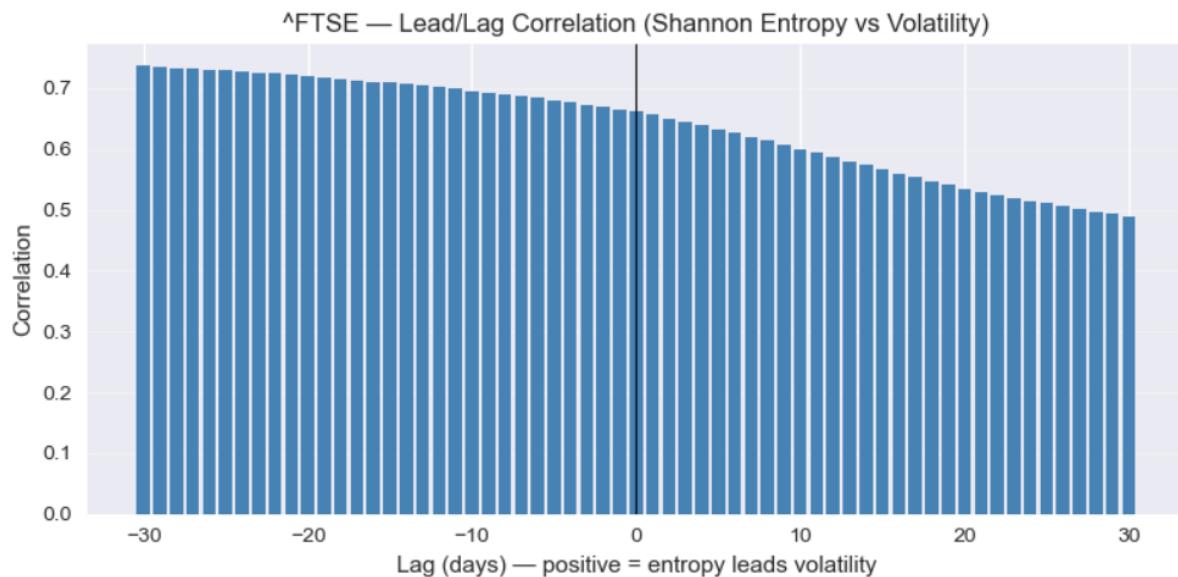


Figure A.28: FTSE 100 entropy–volatility lead–lag analysis.

A.1.5 Nikkei 225



Figure A.29: Nikkei 225 closing price. Context for Japan-specific market regimes in Chapter 4.

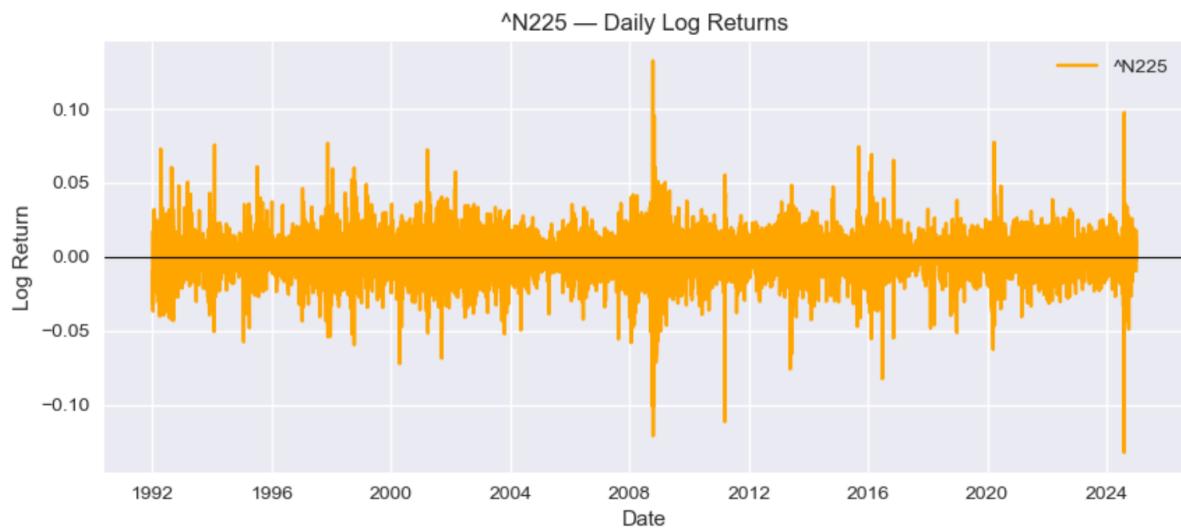


Figure A.30: Nikkei 225 log-price.

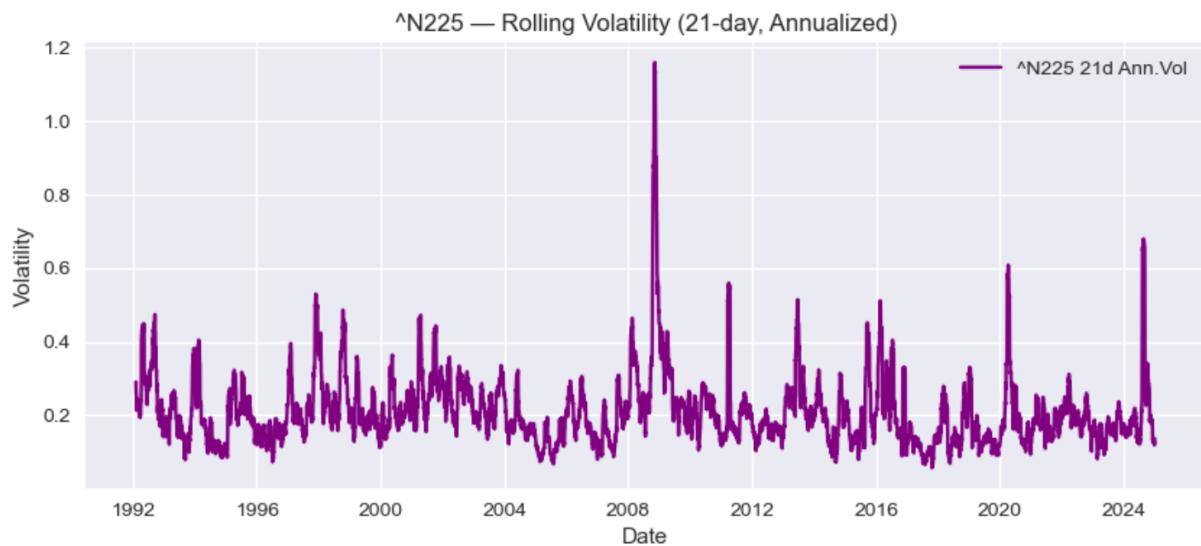


Figure A.31: Nikkei 225 rolling volatility showing persistent clustering.

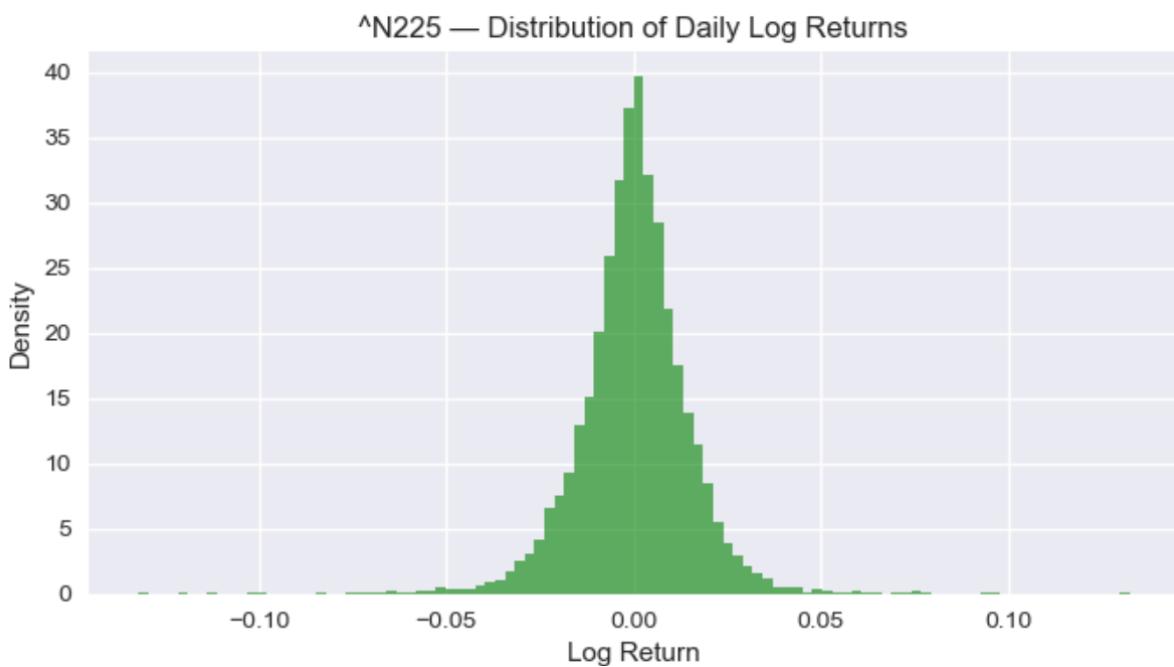


Figure A.32: Nikkei 225 return distribution for KS and Wasserstein comparisons (Section 4.5).

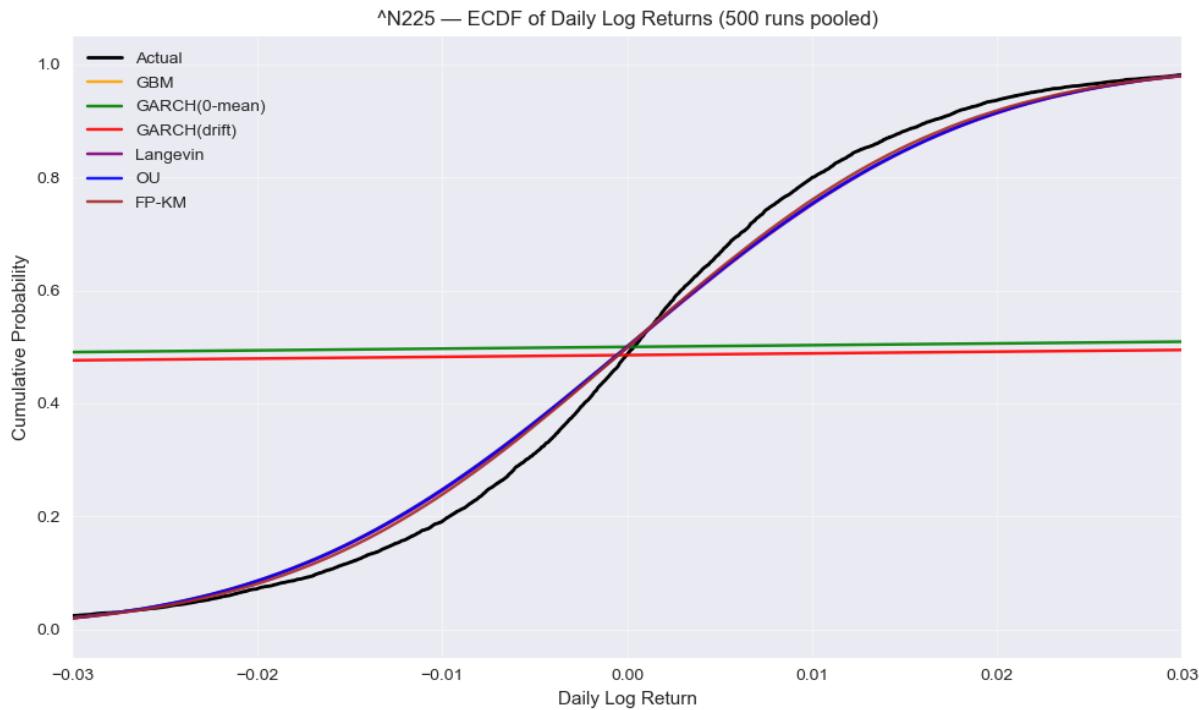


Figure A.33: Nikkei 225 ECDF of returns compared with model ECDFs.

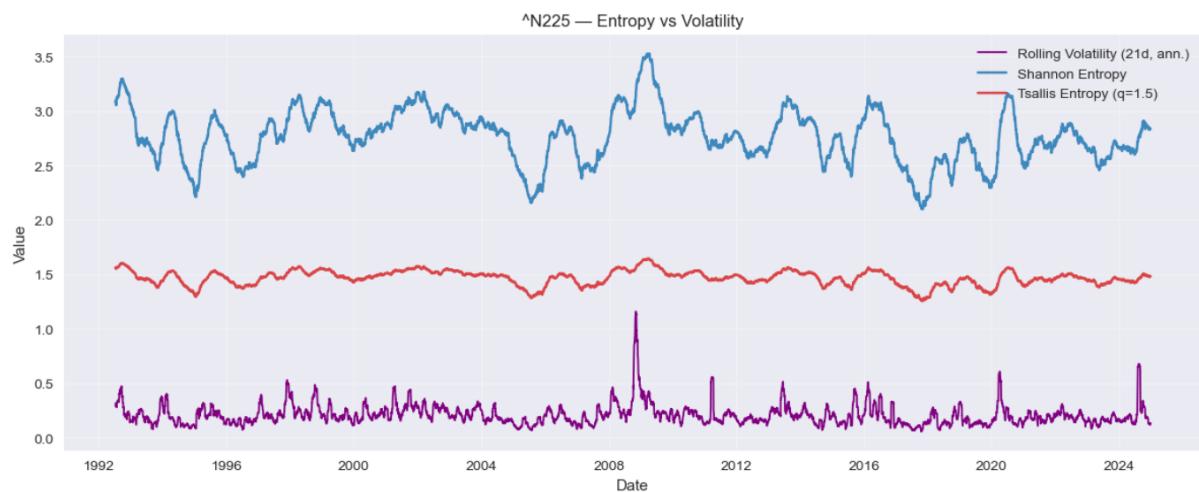


Figure A.34: Nikkei 225 entropy versus volatility, benchmarking model performance (Tables 4.8–4.9).

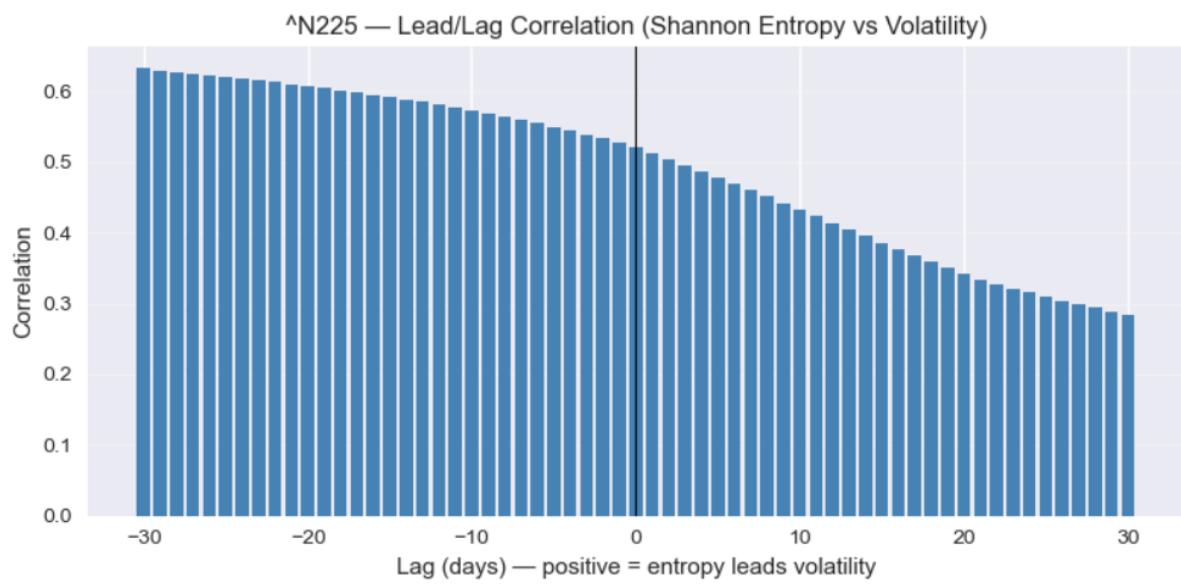


Figure A.35: Nikkei 225 entropy–volatility lead–lag analysis.



Methodology–Code Alignment Statement

To ensure transparency and reproducibility, every methodological step described in Chapter 3 has been directly implemented in Python. The Jupyter notebook accompanying this dissertation provides cell-by-cell execution, and the alignment is summarised as follows:

1. Data Collection and Preprocessing (Section 3.3)

Daily OHLCV data for the S&P 500, Nasdaq QQQ, FTSE 100, Nikkei 225, and Bitcoin (1992–2025) were downloaded via the `yfinance` API. Log returns were computed, datasets aligned, and descriptive statistics and correlations produced exactly as specified.

2. Baseline Financial Models (Section 3.3)

GBM: μ and σ from log returns; simulated via exponential random walk.

GARCH(1,1): Zero-mean and drifted variants estimated with `arch`; simulations via built-in and custom routines.

3. Physics-Inspired Models (Section 3.4)

Langevin: AR(1) on returns to obtain (γ, σ, a) .

OU: AR(1) on log-prices mapped to (κ, μ, σ) .

Fokker–Planck (FP–KM): Drift/diffusion via Kramers–Moyal; Euler–Maruyama with state-dependent coefficients.

4. Entropy-Based Diagnostics (Section 3.5)

Rolling Shannon and Tsallis ($q = 1.5$) entropy computed with fixed binning; compared with rolling volatility; lead–lag correlations calculated.

5. Simulation Framework (Section 3.6.4)

Full-horizon simulations (1992–2025), common P_0 , 500 paths per model/asset, fixed seeds.

6. Model Evaluation Metrics (Section 3.7)

Price-path RMSE; KS and Wasserstein for distributions; AIC/BIC for GBM/GARCH (AR(1) proxies for Langevin/OU; FP non-parametric).

7. Comparative Analysis (Section 4.5)

ECDF comparisons; entropy-series RMSE vs. observed entropy.

8. Tools and Libraries

Python 3.11 with `pandas`, `numpy`, `scipy`, `matplotlib`, `seaborn`, `arch`, `statsmodels`, `scikit-learn`, `yfinance`.

Bibliography

- [1] R. N. Mantegna and H. E. Stanley. *Introduction to Econophysics: Correlations and Complexity in Finance*. Cambridge University Press, 1999.
- [2] C. Tsallis. Possible generalization of Boltzmann–Gibbs statistics. *Journal of Statistical Physics*, 52(1–2):479–487, 1988.
- [3] T. Bollerslev. Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, 31(3):307–327, 1986.
- [4] T. Lux and M. Marchesi. Scaling and criticality in a stochastic multi-agent model of a financial market. *Nature*, 397(6719):498–500, 1999.
- [5] J. D. Farmer and S. Joshi. The price dynamics of common trading strategies. *Journal of Economic Behavior & Organization*, 49(2):149–171, 2000.
- [6] L. Borland. Option pricing formulas based on a non-Gaussian stock price model. *Physica A: Statistical Mechanics and its Applications*, 299(1–2):529–546, 2002.
- [7] J.-P. Bouchaud and R. Cont. A Langevin approach to stock market fluctuations and option pricing. *The European Physical Journal B*, 6(4):543–550, 1998.
- [8] P. Hänggi and H. Thomas. Stochastic processes: Time evolution, symmetries and linear response. *Physics Reports*, 88(4):207–319, 1982.
- [9] A. Das and C. Y. Fong. Efficient Fokker–Planck modeling of stochastic memory devices. *Journal of Applied Physics*, 129(23):234501, 2021.
- [10] L. Zunino, M. C. Soriano, and O. A. Rosso. Permutation entropy and its main biomedical and econophysics applications: A review. *Physics Reports*, 697:1–46, 2017.

- [11] V. Filimonov and D. Sornette. Quantifying reflexivity in financial markets: Toward a prediction of flash crashes. *Physical Review E*, 89(1):012801, 2014.
- [12] A. Johansen and D. Sornette. Critical crashes. *Risk*, 12(1):91–94, 1999.
- [13] M. Scheffer, J. Bascompte, W. Brock, V. Dakos, H. Held, E. van Nes, M. Rietkerk, and G. Sugihara. Early-warning signals for critical transitions. *Nature*, 461:53–59, 2009.
- [14] C. Diks, C. Hommes, and J. Wang. Critical slowing down as an early warning signal for financial crises. *Empirical Economics*, 57:1201–1228, 2019.
- [15] L. Lin, R. Ren, and D. Sornette. A consistent model of explosive financial bubbles with mean-reversing residuals. *Journal of Economic Behavior & Organization*, 74(3):297–311, 2010. arXiv:0905.0128.
- [16] J. Nielsen, M. Raissi, and D. Sornette. Deep LPPLS: Forecasting of temporal critical points in natural, engineering and financial systems. *arXiv preprint arXiv:2405.12803*, 2024.