FINANCIAL DYNAMICS UNDER SHOCKS: A COMPARATIVE ANALYSIS OF DEEP LEARNING, GARCH MODELS, AND BLACK-SCHOLES ASSUMPTION FAILURES

KONRAD BARANEK AND MAKSYMILIAN DĘBOWSKI

Abstract. Classical Black-Scholes assumptions, like constant volatility, often fail during market shocks. This study investigates these limitations using a case study of a company experiencing a specific external shock. We employ GARCH(1,1) and asymmetric GJR-GARCH models to analyze pre-shock volatility dynamics. We generate and compare post-shock price path simulations using: (1) fixed pre-shock GARCH/GJR parameters (static forecast), (2) adaptive expanding-window GARCH/GJR parameters (dynamic forecast), (3) a fixed naive approach based on pre-shock averages (static forecast), (4) a dynamic naive approach using lagged historical data (dynamic forecast), and (5) Deep Learning (DL) models trained pre-shock using the AutoGluon-TimeSeries framework (static forecast). Performance is assessed against the actual price path and realized volatility. Results highlight the failure of static forecasts (Fixed GARCH/Naive, AutoGluon). Adaptive GARCH models excelled at volatility tracking post-shock, while a dynamic naive approach surprisingly provided the best median price path simulation, outperforming the non-adaptive AutoML baseline in this challenging scenario. Findings underscore the critical role of model adaptability post-shock.

1. Introduction

Accurate modeling and forecasting of financial asset volatility are crucial for effective risk management, derivative pricing, and portfolio allocation [5, 3]. The Black-Scholes model [2], while foundational, describes the asset price dynamics S_t via a GBM (geometric Brownian motion).

$$dS_t = \mu S_t dt + \sigma S_t dW_t \tag{1}$$

where μ is the "percentage drift" (or expected return), W_t is a standard Wiener process, and crucially, σ is the volatility, which is assumed to be constant.

This assumption is recognized as empirically unrealistic, particularly during periods of market stress or some idiosyncratic shocks affecting specific assets [4, 1]. Such shocks often induce significant, abrupt changes in volatility dynamics, making standard B-S pricing and hedging formulas (σ = constant) unreliable just when accurate risk assessment is most critical. This project addresses the challenge of modeling and forecasting volatility under market shocks, focusing on addressing the limitations of the B-S framework and comparing the performance of traditional econometric models with modern machine learning techniques. Specifically, we examine the impact of a documented external shock on the volatility dynamics of a selected company's stock price.

Standard approaches to capture time-varying volatility include the family of GARCH models [3]. The canonical GARCH(1,1) model, for instance, describes the conditional variance $\sigma_t^2 = \mathrm{Var}(r_t|\mathcal{F}_{t-1})$ of returns r_t (given the past information $\mathcal{F}_{t-1}$) as:

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

where $\epsilon_t = r_t - \mathbb{E}[r_t|\mathcal{F}_{t-1}]$ are the residuals, and $\omega > 0$, $\alpha \geq 0$, $\beta \geq 0$ are parameters usually constrained to ensure stationarity (e.g., $\alpha + \beta < 1$).

While GARCH models effectively capture "groups" of different volatilities, asymmetric variants such as GJR-GARCH models [7] also take into account leverage (negative shocks increase volatility more than positive ones):

$$\sigma_t^2 = \omega + (\alpha + \gamma\mathbb{1}_{\epsilon_{t-1}<0})\epsilon_{t-1}^2 + \beta\sigma_{t-1}^2, \tag{3}$$

where $\mathbb{1}_{\epsilon_{t-1}<0}$ is equal to 1 if the previous shock was negative, and $\gamma \geq 0$ captures the asymmetry.

Despite their broad utility, the responsiveness and predictive accuracy of GARCH models during sudden, high-severity shocks require careful investigation, particularly concerning parameter stability and the need for adaptive estimation versus relying on pre-shock parameters.

In recent years, Deep Learning (DL) architectures have become powerful tools for modeling complex time series data [10, 8]. Automated Machine Learning (AutoML) frameworks like AutoGluon [6, 12] simplify their application. However, their effectiveness in forecasting through major structural breaks when trained only on pre-break data remains an open question compared to models that adapt using post-break information.

1.1. **Main Contributions.** The main contribution of this paper is divided into three parts:

(1) We demonstrate the failure of Black-Scholes assumptions during a specific shock event through empirical analysis. We quantify the shock's impact by comparing actual post-shock dynamics to counterfactual simulations based on fixed pre-shock GARCH/Naive parameters (though the primary focus here is on the B-S assumptions themselves).

(2) We conduct a comparative analysis of post-shock simulation performance between models using fixed pre-shock parameters (Fixed GARCH, Fixed Naive, AutoGluon trained pre-shock) and models incorporating post-shock adaptation (Adaptive GARCH, Dynamic Naive).

(3) We assess the impact of model adaptability on simulation accuracy in the post-shock period, highlighting the benefits of adaptation (Adaptive GARCH/Dynamic Naive) versus static forecasts (Fixed GARCH/Naive, AutoGluon) for both volatility tracking and price path prediction in this specific shock scenario.

Our methodology uses data from Rheinmetall AG. We simulate/forecast post-shock price paths and volatility using the different modeling approaches. The results aim to shed light on the relative strengths and weaknesses, particularly regarding adaptability, in volatile, shock-prone conditions.

The remainder of this article is organized as follows. Section 2 describes the data, models, simulation/forecasting strategies, and evaluation metrics. Section 3 presents empirical analysis and comparative performance. Section 4 discusses the findings. Section 5 provides information about the code repository.

2. Methodology

For the analysis, we chose the German joint-stock company Rheinmetall AG, which is a leading international manufacturer of systems in the defense industry.

2.1. **Data and Preparation.** Data was acquired using the `yfinance` Python library for the period starting from January 1, 2019 up to April 4, 2025 (though analysis focuses on the period around the shock). Data was retrieved at a daily frequency. We utilize the daily Adjusted Close price (P_t), which accounts for dividends and stock splits.

Logarithmic returns (r_t) were calculated using the standard formula:

$$r_t = \ln\left(\frac{P_t}{P_{t-1}}\right). \tag{4}$$

The first observation in the returns series was subsequently removed. We also calculate rolling mean log returns over a window $W = 21$:

$$\bar{r}_{roll,t} = \frac{1}{W}\sum_{i=0}^{W-1} r_{t-i}. \tag{5}$$

2.2. **Realized Volatility Calculation.** The target variable for volatility forecasting is the realized annualized volatility ($\sigma_{RV,t}$). We estimate it using a rolling window ($W = 21$ trading days) of daily logarithmic returns (r_t). We also define the daily realized volatility $\sigma_{RV,daily,t}$. The calculation, based on [9], is:

$$\sigma_{RV,t} = \sqrt{N} \times \sigma_{RV,daily,t} = \sqrt{N} \times \sqrt{\frac{1}{W-1}\sum_{i=0}^{W-1}(r_{t-i} - \bar{r}_{roll,t})^2}, \tag{6}$$

where $N = 260$ is the annualization factor and $\bar{r}_{roll,t}$ is the rolling average return (Eq. 5). This $\sigma_{RV,t}$ serves as the ground truth proxy for actual annualized volatility. $\sigma_{RV,daily,t}$ is used as the target for AutoGluon volatility prediction.

2.3. **Shock Event.** The chosen shock date is February 24, 2022, marking the start of the full-scale Russian invasion of Ukraine. This event significantly impacted the defense sector, including Rheinmetall AG. Figure 1 visually confirms a structural change around this date, with sharp increases in both price and realized volatility.

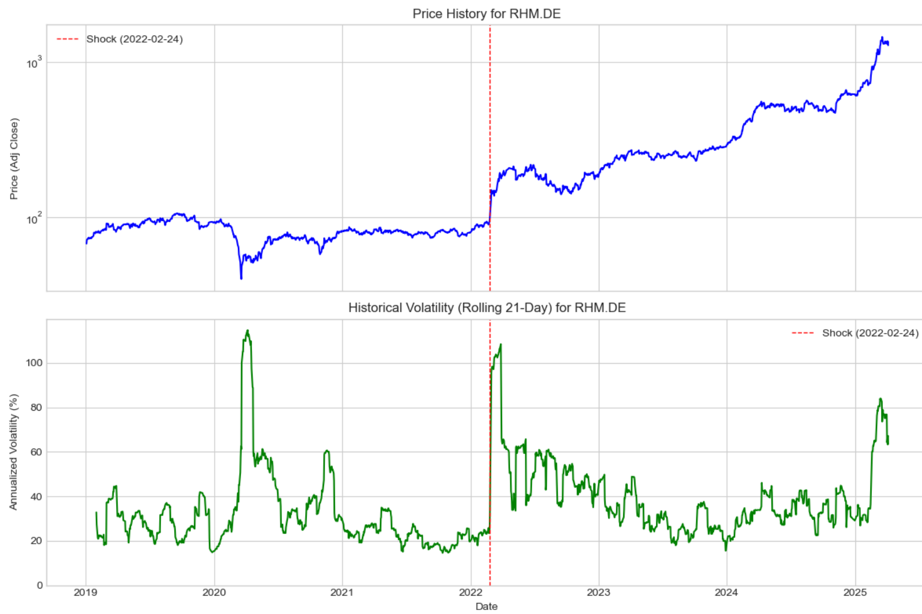

FIGURE 1. Adjusted Closing Price and Annualized 21-Day Realized Volatility for RHM.DE. The vertical dashed red line indicates the identified shock date (2022-02-24).

2.4. **GARCH Model Estimation.** We utilize GARCH(1,1) (Eq. 2) and GJR-GARCH(1,1) (Eq. 3) models with Student's t-distributed innovations. Parameters are estimated via MLE using the `arch` library. Model orders (1,1) were selected based on BIC using pre-shock data (Tables 4, 5).

2.5. **Naive Baseline Approaches.** Two simple baselines are used:

2.5.1. *Dynamic Naive Approach (Rolling History T-1 Year).* Uses rolling historical mean log-return $\hat{\mu}_{log,hist}$ and historical daily realized volatility $\hat{\sigma}_{daily,hist}$ calculated over the period $[t-\Delta_{1yr}, t-1]$ to simulate $r_t^{(j)} \sim N(\hat{\mu}_{log,hist}, \hat{\sigma}^2_{daily,hist})$ for each post-shock day t. This approach adapts based on the most recent year's data.

2.5.2. *Fixed Naive Approach (Average Pre-Shock Parameters).* Uses fixed average mean log-return $\bar{\mu}_{log,fixed}$ and average daily realized volatility $\hat{\sigma}_{daily,fixed}$ calculated over the year `before` the shock ($[t_{shock}-\Delta_{1yr}, t_{shock}-1]$) to simulate $r_t^{(j)} \sim N(\bar{\mu}_{log,fixed}, \hat{\sigma}^2_{daily,fixed})$ for all post-shock days t. This is a static approach.

2.6. **AutoGluon-TimeSeries Setup.** We use AutoGluon-TimeSeries [12] version 1.2 to generate DL-based forecasts. Crucially, AutoGluon was trained `only` on data available `before` the shock date (up to Feb 23, 2022). Two separate predictors were trained:

(1) **Volatility Predictor:** Trained to forecast `Realized Volatility Daily`($\sigma_{RV,daily,t}$) over the entire post-shock horizon.
(2) **Price Predictor:** Trained to forecast `Adj Close` over the same horizon.

Both used the `best_quality` preset, RMSE evaluation, a 3600s time limit, GPU acceleration, and available past covariates (e.g., price, volume, returns for volatility predictor). AutoGluon selected a `WeightedEnsemble` as the best model based on validation performance (using validation folds within the pre-shock data). This pre-trained ensemble was then used to generate a **single, static forecast** for the entire post-shock period (median quantile "0.5" for price, "mean" for volatility) without any updates using post-shock data.

2.7. **Simulation and Forecasting Strategies.** We compare five strategies post-shock, starting from $S_{shock-1}$:

(1) **Fixed GARCH/GJR:** Models estimated once on pre-shock data. Used for static simulation ($n_{sims} = 20000$) and volatility forecast over the entire post-shock period.
(2) **Adaptive GARCH/GJR:** Parameters re-estimated daily using an expanding window including post-shock data. Used for dynamic one-step-ahead simulation ($n_{sims} = 20000$) and volatility forecast ($\hat{\sigma}_t$).
(3) **Fixed Naive:** Uses fixed pre-shock averages for static simulation ($n_{sims} = 20000$) and volatility forecast.
(4) **Dynamic Naive:** Uses rolling T-1 year history for dynamic simulation ($n_{sims} = 20000$) and volatility forecast (lagged rolling volatility).
(5) **AutoGluon Forecast:** Uses the pre-trained AutoGluon ensembles (Sec 2.6) to generate a single, static forecast path for price and volatility over the entire post-shock horizon.

Strategies (1), (3), and (5) represent forecasts made *at the time of the shock* using only past information. Strategies (2) and (4) represent forecasts that **adapt** using information revealed after the shock.

2.8. **Evaluation Metrics.** Performance is evaluated over the post-shock period T.

2.8.1. *Price Path Accuracy Metrics.* Median simulated price ($\hat{S}_{t,median}$ for GARCH/Naive) or AutoGluon median price forecast ($\hat{S}_{t,AG,0.5}$) compared against actual price (S_t). Lower is better.

(1) **RMSE**: $RMSE_{price} = \sqrt{\frac{1}{T}\sum_{t=t_0}^{t_0+T-1}(S_t - \hat{S}_t)^2}$
(2) **MAE**: $MAE_{price} = \frac{1}{T}\sum_{t=t_0}^{t_0+T-1} |S_t - \hat{S}_t|$

where $\hat{S}_t$ is $\hat{S}_{t,median}$ or $\hat{S}_{t,AG,0.5}$.

2.8.2. *Volatility Forecast Accuracy Metrics.* Model's annualized forecast volatility ($\hat{\sigma}_{model,t}$) compared against actual annualized realized volatility ($\sigma_{RV,t}$) over T_{vol} days (from $t_0 + W$ to $t_0 + T - 1$). AutoGluon's daily mean forecast $\hat{\sigma}_{AG,daily,mean}$ is annualized: $\hat{\sigma}_{AG,t} = \sqrt{N} \times \hat{\sigma}_{AG,daily,mean}$. GARCH/Naive daily σ_t are similarly annualized. Lower is better.

(1) **RMSE**: $RMSE_{vol} = \sqrt{\frac{1}{T_{vol}}\sum(\sigma_{RV,t} - \hat{\sigma}_{model,t})^2}$
(2) **MAE**: $MAE_{vol} = \frac{1}{T_{vol}}\sum|\sigma_{RV,t} - \hat{\sigma}_{model,t}|$
(3) **QLIKE** [11]: $QLIKE_{vol} = \frac{1}{T_{vol}}\sum\left(\frac{\sigma^2_{RV,t}}{\hat{\sigma}^2_{model,t}} - \ln\left(\frac{\sigma^2_{RV,t}}{\hat{\sigma}^2_{model,t}}\right) - 1\right)$

3. Results and Comparisons

3.1. **Empirical Analysis of Black-Scholes Assumptions.** Before proceeding with volatility forecasting, we begin by empirically examining the key mathematical assumptions of the Black-Scholes model, namely constant volatility and normality of logarithmic returns. This is crucial to demonstrate the context of our study and the need to go beyond the B-S framework, especially in a market environment highly susceptible to shocks.

3.1.1. *Volatility Dynamics Pre- and Post-Shock.* Descriptive statistics in Table 1 and Figure 2 clearly demonstrate a significant change in volatility dynamics after the shock. The mean annualized realized volatility increased significantly, rising from 31.37% pre-shock to 40.11% post-shock. The standard deviation also rose substantially, indicating wider volatility swings.

Table 1. Descriptive Statistics of Annualized Realized Volatility Before and After Shock (2022-02-24)

Statistic	Before Shock	After Shock
Mean (% pa)	31.37	40.11
Std Dev (% pa)	26.15	35.42
Median (% pa)	16.60	16.63
Min (% pa)	14.53	15.53
Max (% pa)	114.81	108.48

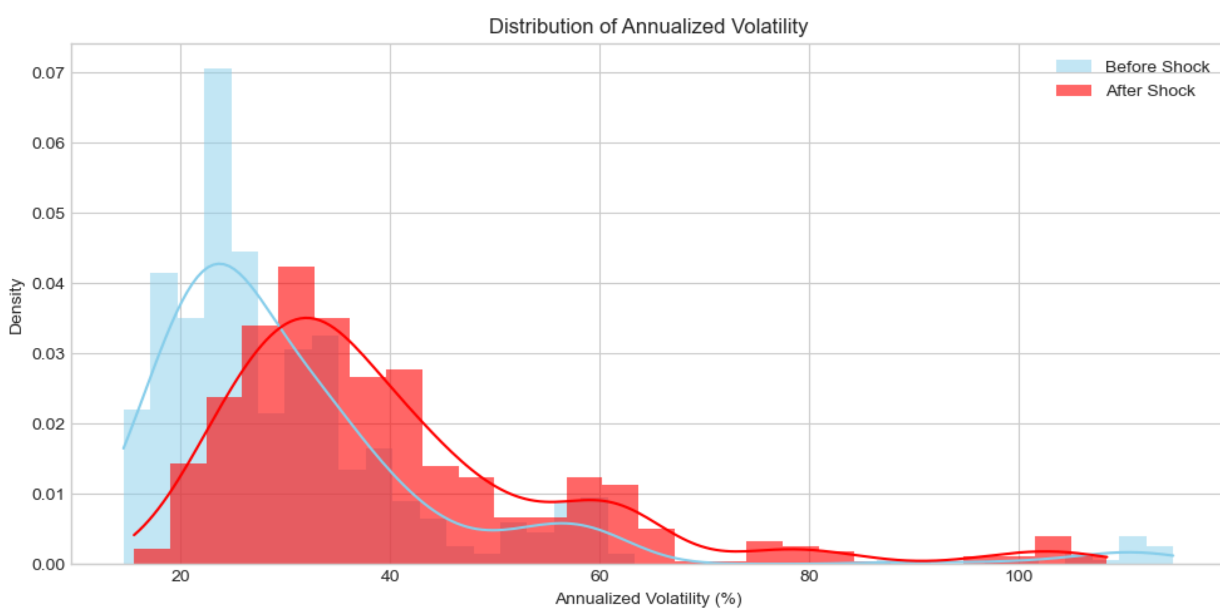

Figure 2. Distribution of Annualized Realized Volatility Before and After Shock (2022-02-24)

Figure 2 visually confirms this shift, showing the post-shock volatility distribution shifted towards higher levels and exhibiting greater dispersion compared to the pre-shock period. This evidence points to a clear structural break in volatility behavior coinciding with the shock, directly challenging the Black-Scholes assumption of constant volatility.

3.1.2. *Normality of Log-Returns.* Analysis shows logarithmic returns fail to meet the Black-Scholes normality assumption. Statistics in Table 2 indicate positive skewness and high kurtosis (leptokurtosis) in both periods (pre and post-shock), demonstrating a right-skewed, heavy-tailed distribution.

TABLE 2. Descriptive Statistics of Daily Log-Returns Before and After Shock (2022-02-24)

Statistic	Before Shock	After Shock
Mean	0.000338	0.003321
Std Dev	0.022061	0.027382
Skewness	0.375239	0.738859
Kurtosis	9.976945	9.294488

Histograms (Figure 3) and Q-Q plots (Figure 4) confirm significant deviations from normality, especially in the tails. The Shapiro-Wilk test (Table 3) firmly rejects normality ($p \approx 0.0000$) for both periods. Increased volatility after the shock and persistent non-normality weaken the Black-Scholes model's applicability, particularly during shocks when extreme price movements are more likely.

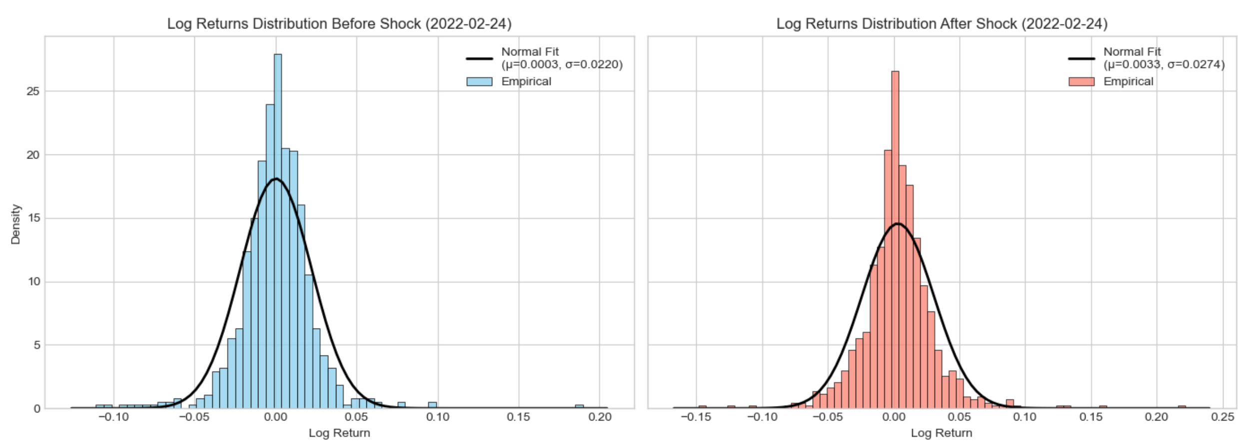

FIGURE 3. Histograms of Daily Log Returns Before and After Shock (2022-02-24) with Fitted Normal Distributions

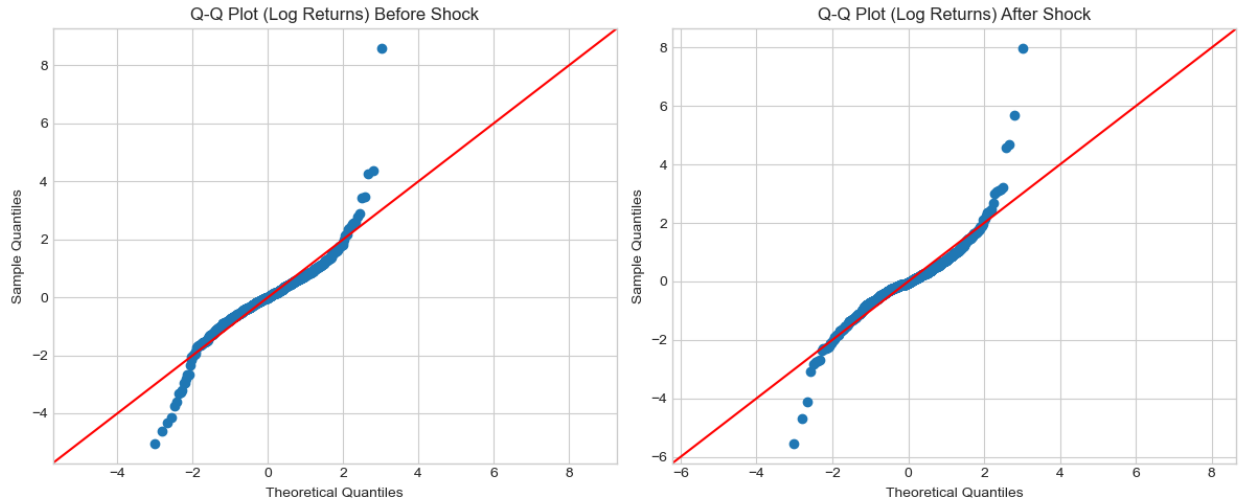

FIGURE 4. Q-Q Plots of Daily Log Returns Against Normal Distribution Before and After Shock (2022-02-24)

TABLE 3. Shapiro-Wilk Normality Test for Daily Log Returns

Test Period	Statistic	p-value
Before Shock	0.9062	< 0.0001
After Shock	0.9055	< 0.0001

3.1.3. *Implications for Modeling.* Empirical evidence confirms Black-Scholes assumptions fail for the analyzed asset, especially during the market shock. Volatility shows a structural shift post-shock, while log-returns display non-normality and heavy tails throughout. These findings highlight the standard model's limitations for pricing and risk management during market stress, necessitating more sophisticated approaches like GARCH and potentially Deep Learning to capture time-varying volatility and non-normality.

3.2. **Post-Shock Simulation with Fixed Pre-Shock Parameters.** We first examine simulations based on models and parameters derived solely from pre-shock data to represent a counterfactual "no-shock adaptation" scenario. This includes GARCH(1,1) and GJR-GARCH(1,1,1) models estimated on data up to Feb 23, 2022 (BIC selection in Tables 4, 5), and the Fixed Naive approach using average parameters from the year before the shock. These models generate static forecasts for the entire post-shock period.

Visual inspection of the simulation paths (not shown here) and quantitative metrics (Table 6) confirm the failure of these static approaches. Actual prices diverge sharply upwards from the simulated paths, and actual volatility remains significantly elevated compared to the low, decaying, or constant volatility forecasts generated by these models using only pre-shock parameters. This demonstrates the inadequacy of relying solely on pre-shock information after a major structural break has occurred.

Table 4. GARCH(p,q) Model Comparison (Pre-Shock Data, Sorted by BIC)

p	q	AIC	BIC	LogLikelihood	$\alpha + \beta$
1	1	3252.34	3275.75	−1621.17	0.9676
1	2	3254.26	3282.35	−1621.13	0.9509
2	1	3254.28	3282.36	−1621.14	0.9662
2	2	3255.79	3288.55	−1620.89	0.9450
1	3	3256.26	3289.03	−1621.13	0.9509
3	1	3256.28	3289.04	−1621.14	0.9662
3	2	3257.26	3294.71	−1620.63	0.9415
2	3	3257.64	3295.09	−1620.82	0.9425
3	3	3258.71	3300.84	−1620.35	0.9186

Models fitted using pre-shock data ($N = 797$) for RHM.DE returns, constant mean, Student's t-distribution.

Table 5. GJR-GARCH(p,1,q) Model Comparison (Pre-Shock Data, Sorted by BIC)

p	o	q	AIC	BIC	γ
1	1	1	3244.49	3272.57	0.0793
1	1	2	3246.20	3278.97	0.1213
2	1	1	3246.25	3279.02	0.0801
2	1	2	3247.72	3285.17	0.1253
3	1	1	3248.11	3285.56	0.0829
1	1	3	3248.20	3285.65	0.1213
3	1	2	3249.67	3291.79	0.1238
2	1	3	3249.72	3291.85	0.1252
3	1	3	3251.67	3298.47	0.1238

Models fitted using pre-shock data ($N = 797$) for RHM.DE returns, constant mean, GJR-GARCH($o = 1$), Student's t-distribution.

Table 6. Evaluation Metrics for Fixed Pre-Shock Simulations

Model (Fixed Pre-Shock Params)	Price Path		Volatility		
	RMSE	MAE	RMSE	MAE	QLIKE
GARCH(1,1) Fixed	320.8753	235.1205	20.3088	13.3936	0.7943
GJR-GARCH(1,1,1) Fixed	336.6622	248.7592	21.4038	14.4542	0.9942
Fixed Naive (Avg Pre-Shock)	343.7507	255.4850	24.3017	17.9503	1.7356

Lower values indicate better performance. Price metrics compare median simulated price to actual price. Volatility metrics compare model's implied/used volatility to actual realized volatility (proxy). QLIKE is for variance/volatility forecast quality.

3.3. **Post-Shock Simulation with Adaptive/Dynamic Parameters.** Next, we examine models that adapt using post-shock data. This includes Adaptive GARCH/GJR (re-estimated daily) and Dynamic Naive (using rolling T-1 year data). We analyze the simulation paths and volatility forecasts generated by these adaptive approaches against the actual data.

Analysis of the underlying data (plots not shown) reveals improved volatility tracking by adaptive GARCH/GJR compared to fixed models. Their conditional volatility follows the actual realized volatility more closely after incorporating post-shock information. The Dynamic Naive volatility also adapts but might appear lagged compared to the GARCH models. Price path simulations generated by these adaptive models, while still underestimating the strong upward trend observed in the actual data, tend to be closer to the reality than those from the fixed models. Quantitative performance metrics are provided in Table 7.

Table 7. Evaluation Metrics for Adaptive/Dynamic Simulations

Model (Adaptive/Dynamic Params)	Price Path		Volatility		
	RMSE	MAE	RMSE	MAE	QLIKE
GARCH(1,1) Adaptive	306.4631	220.1101	7.1723	5.0485	0.0518
GJR-GARCH(1,1,1) Adaptive	301.9481	197.8874	8.0955	5.3340	0.0611
Dynamic Naive (T-1 Year)	221.7467	160.5942	27.7763	21.9810	1.5855
Dynamic Naive (T-1 Day)	14.8315	7.3015	3.6707	1.5967	0.0167

Lower values indicate better performance. Price metrics compare median simulated price to actual price. Volatility metrics compare model's conditional/used volatility to actual realized volatility (proxy). QLIKE is for variance/volatility forecast quality.

3.4. **Deep Learning Approach using AutoGluon.** We employed AutoGluon-TimeSeries (Sec 2.6), training models solely on pre-shock data to generate static forecasts for the entire post-shock period. The best validation model (WeightedEnsemble) was used for prediction.

3.4.1. *AutoGluon Volatility Forecasting.* The AutoGluon mean volatility forecast, evaluated against actual realized volatility, yielded the metrics presented later in Table 8. Textually, the forecast captured the generally elevated post-shock level but was smoother and less reactive than adaptive GARCH models. Its performance was better than the naive models based on QLIKE but worse than adaptive GARCH across all metrics. Notably, the AutoGluon leaderboard indicated **ChronosFineTuned** had a better test RMSE (14.42) than the chosen **WeightedEnsemble** (test RMSE 17.33 based on calculation, 58.03 on leaderboard), suggesting pre-shock validation was not optimal for selecting the best model for the post-shock regime.

3.4.2. *AutoGluon Price Forecasting.* The AutoGluon median price forecast significantly underestimated the post-shock trend. Performance metrics (Table 8) show higher errors than adaptive GARCH and substantially higher than Dynamic Naive. Similar to volatility, the test leaderboard suggested **ChronosFineTuned** (test RMSE 213.61) performed better on the post-shock data than the validation-selected **WeightedEnsemble** (test RMSE 328.92 calculated, 335.45 on leaderboard), again highlighting the challenge of forecasting through the regime shift using only pre-shock data for model selection and training.

3.5. **Comparative Analysis of Simulation Performance.** Table 8 consolidates the evaluation metrics for all strategies, grouping them by whether they used adaptation post-shock.

Table 8. Consolidated Evaluation Metrics

Model / Strategy	Price Path Accuracy		Volatility Forecast Accuracy		
	$RMSE_{price}$	MAE_{price}	$RMSE_{vol}$	MAE_{vol}	$QLIKE_{vol}$
Static Forecasts (Trained/Calibrated Pre-Shock Only)					
GARCH(1,1) Fixed	320.88	235.12	20.31	13.39	0.794
GJR-GARCH(1,1,1) Fixed	336.66	248.76	21.40	14.45	0.994
Fixed Naive (Avg Pre-Shock)	343.75	255.49	24.30	17.95	1.736
AutoGluon (WeightedEnsemble)	328.92	262.49	17.33	11.26	0.084
Adaptive / Dynamic Forecasts (Using Post-Shock Data)					
GARCH(1,1) Adaptive	306.46	220.11	7.17	5.05	0.052
GJR-GARCH(1,1,1) Adaptive	301.95	197.89	8.10	5.33	0.061
Dynamic Naive (T-1 Year)	221.75	160.59	27.78	21.98	1.586
Dynamic Naive (T-1 Day)	14.8315	7.301	3.6707	1.5967	0.0167
TimesFM Rolling (1d step)	15.749	7.949	—	—	—
TimesFM Rolling (5d step)	24.082	12.072	—	—	—
TimesFM Rolling (10d step)	31.985	16.195	—	—	—
TimesFM Rolling (15d step)	40.485	20.00	—	—	—
TimesFM Rolling (20d step)	48.097	24.559	—	—	—

Lower values better. Price/Volatility: Forecast vs Actual (best in **bold**). AutoGluon: static pre-shock `WeightedEnsemble`. TimesFM: rolling forecast (update steps noted; Vol. N/A).

Key observations from Table 8:

(1) **Failure of Static Forecasts:** All models relying solely on pre-shock information (Fixed GARCH/GJR, Fixed Naive, AutoGluon) performed poorly compared to adaptive alternatives, especially in price path simulation. Within this static group, AutoGluon provided the best volatility forecast based on QLIKE (0.084), significantly better than Fixed GARCH (0.794), while Fixed GARCH had slightly better RMSE/MAE volatility than AutoGluon. However, all static volatility forecasts were substantially worse than adaptive GARCH. For price path, Fixed GARCH was marginally the best among static models but still very inaccurate.

(2) **Benefit of Adaptability:** Introducing adaptability via daily updates (Adaptive GARCH/GJR) or rolling history (Dynamic Naive) yielded significant improvements.
  - **Volatility Forecasting:** Adaptive GARCH models dramatically outperformed all static forecasts and the Dynamic Naive approach. Adaptive GARCH(1,1) was the overall best volatility forecaster (lowest RMSE, MAE, QLIKE).
  - **Price Path Simulation:** The simple Dynamic Naive model (particularly the T-1 day variant) achieved the best price path accuracy (lowest RMSE, MAE), significantly outperforming the more complex adaptive GARCH models and all static models. The TimesFM rolling forecasts also showed strong price path performance, comparable to the Dynamic Naive (T-1 day).

(3) **AutoGluon Performance (Static Forecast):** The AutoGluon baseline, generating a static forecast based only on pre-shock training, did not match the performance of simpler adaptive models post-shock. Its volatility forecast was respectable relative to other static methods (best QLIKE) but inferior to adaptive GARCH. Its price forecast was poor, worse than adaptive GARCH and much worse than Dynamic Naive or rolling TimesFM. The discrepancy between validation-best and test-best models within AutoGluon further highlights the difficulty of forecasting through the shock using only pre-shock data.

(4) **Best Performing Model:** The best approach depends on the objective and whether adaptation is possible.
  - For **volatility forecasting post-shock**, `adaptive` GARCH(1,1) was superior.
  - For simulating the **median price path post-shock**, the **dynamic** Naive model (T-1 day) and rolling **TimesFM** were most accurate.
  - If only a **static** forecast from the shock date onwards is possible, AutoGluon offered the best volatility QLIKE, while Fixed GARCH had slightly better RMSE/MAE, but both were poor overall compared to adaptive methods.

  The price path result suggests that accurately capturing the post-shock drift was key, and the simple rolling average drift of the Dynamic Naive model (or the context update mechanism of TimesFM) happened to work well in this instance.

(5) **Relative Performance (Naive, GARCH):** Dynamic Naive strongly outperformed Fixed Naive. Adaptive GARCH/GJR strongly outperformed Fixed GARCH/GJR. Differences between adaptive GARCH and GJR were relatively small.

4. Conclusions

TBA

5. Code and Implementation

The analysis was implemented in Python (versions 3.12), using `pandas`, `numpy`, `arch`, `yfinance`, `autogluon.timeseries` (v1.2), `statsmodels`, and `matplotlib/seaborn`. The code, potentially including analyses beyond the scope of this paper, is available at:

https://github.com/maksdebowski/mathematical_foundations_of_ML_DL.git

in the **project** directory. Details might be found in the codebase and associated README files.

Exclusive footage of one of the authors deep into a late-night coding session.

References

1. J.-N. Barrot and J. Sauvagnat. Input specificity and the propagation of idiosyncratic shocks in production networks. *The Quarterly Journal of Economics*, 131(3):1543–1592, May 2016.
2. F. Black and M. Scholes. The pricing of options and corporate liabilities. *The Journal of Political Economy*, 81(3):637–654, 1973.
3. T. Bollerslev. Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, 31(3):307–327, 1986.
4. R. Cont. Empirical properties of asset returns: stylized facts and statistical issues. *Quantitative Finance*, 1(2):223–236, 2001.
5. R. F. Engle. Autoregressive conditional heteroscedasticity with estimates of the variance of united kingdom inflation. *Econometrica*, 50(4):987–1007, 1982.
6. N. Erickson, J. Mueller, A. Shirkov, H. Zhang, P. Larroy, M. Li, and A. Smola. AutoGluon-Tabular: Robust and accurate automl for structured data. In *Advances in Neural Information Processing Systems*, volume 33, pages 11700–11710, 2020.
7. L. R. Glosten, R. Jagannathan, and D. E. Runkle. On the relation between the expected value and the volatility of the nominal excess return on stocks. *The Journal of Finance*, 48(5):1779–1801, Dec 1993.
8. I. Goodfellow, Y. Bengio, and A. Courville. *Deep Learning*. MIT Press, 2016.
9. J. C. Hull. *Options, Futures, and Other Derivatives*. Pearson, 10th edition, 2018.
10. Y. LeCun, Y. Bengio, and G. Hinton. Deep learning. *Nature*, 521(7553):436–444, 2015.
11. A. J. Patton. Volatility forecast comparison using imperfect volatility proxies. *Journal of Econometrics*, 161(2):244–256, 2011.
12. O. Shchur, C. S. Aussignac, J. Gasthaus, and S. S. Rangapuram. AutoGluon-TimeSeries: Automl for probabilistic time series forecasting, 2023.