Modeling Stock Market Volatility Using Arch and Garch Models in BSE and NSE

K. Shalini[1], Nimmala Sree Anuj[2], Manish Kongonda[3], Mohammed Anas[4]

[1]Assistant Professor, Department of Computer Science and Engineering, Matrusri Engineering College, Hyderabad, Telangana, India
[2,3,4]Students, Department of Computer Science and Engineering, Matrusri Engineering College, Hyderabad, Telangana, India

ABSTRACT
In this paper, we introduce Andy-Volatility, a full-stack financial analytics module, and a fundamental part of the Andy Terminal, a multi-module, intelligent trading and analytics platform. Andy-Volatility uses automated BUY, SELL, HOLD, and CAUTION trading signals of equities listed on the Bombay Stock Exchange (BSE) and the National Stock Exchange (NSE) in India by using ARMA(2,3) mean-equation modelling with GJR-GARCH (Threshold GARCH) variance-equations estimation. The system retrieves ten years of past close prices of the day, using Yahoo Finance, does joint ARMA-TGARCH estimation, and predicts conditional volatility five days ahead of the trade. Signal thresholds are computed as the ratio of forecast to current conditional volatility, and enhanced with the leverage effect (parameter of gamma) to indicate asymmetric shock sensitivity in stocks. A backtesting engine tests the volatility based strategy against a Buy-and-Hold comparison over the same historical period. Under the Andy Terminal brand, it combines with other analytical modules via a common Flask back-end, real-time WebSocket price feeds, a rolling 52-week volatility heatmap, news-sentiment fusion scoring via VADER NLP, an administrator-only paper trading module, and a user management system based on SQLite and role-based access control. Experimental evidence on 30 BSE and 31 NSE large-cap stocks shows that the TGARCH model is able to accurately determine high-persistence volatility regimes and produce actionable information with quantifiable risk-adjusted added value to passive investment. The system provides sub-second per-stock analysis latency, and has been proven on the entire Sensex and Nifty-50 universes of constituents..

Keywords: GARCH, TGARCH, GJR-GARCH, ARMA, Stock Market Volatility, Trading Signals, BSE, NSE, Sentiment Analysis, Paper Trading, Flask, Financial Analytics, Leverage Effect, Volatility Forecasting.

I. INTRODUCTION

One of the most intricate and information rich dynamical systems that have been studied in quantitative finance is equity markets. An empirical property of financial return series is that they are heteroskedastic: the statistical property that the variance of returns does not remain constant through time but rather it changes as a reaction to shocks in the past. This is the so-called volatility clustering, which was first recorded in Mandelbrot and then measured by Engle, introducing Autoregressive Conditional Heteroskedasticity (ARCH) model. The concept of volatility clustering means that when assets price movements are large (positive or negative), it is likely that they will be followed by additional large movements, whereas when quiet periods occur, they are likely to continue. This time-varying variance modelling and forecasting is thus a key focus of risk management, derivative pricing, portfolio construction and algorithmic trading strategy development.

Generalised ARCH (GARCH) specification Bollerslev further generalised the ARCH family of models, adding a lagged conditional variance term which dramatically improved parsimony and fit. The symmetric GARCH model however, assumes that the response to both positive and negative innovations of returns is equal; which is always empirically rejected. There is an asymmetry in the volatility response of equity prices: negative news shocks produce an excessively large future volatility increment compared to equal magnitude positive news shocks. This asymmetry, which has often been explained by the leverage effect as proposed by Black, is explained by the GJR-GARCH or Threshold GARCH (TGARCH) model of Glosten, Jagannathan and Runkle. The TGARCH model adds another term in the standard GARCH variance equation, which only becomes active when the lagged innovation is negative, thus enabling the model to differentiate between the volatility effect of good and bad news.

The generation of automated trading signals, which convert volatility predictions into actionable proposals, has been an increasingly popular topic in the algorithmic trading and quantitative finance literature. Forecast volatility is a natural risk proxy: when the expected volatility of a stock in the next trading week is much lower than its current volatility, the stock is transitioning to a calmer regime, which is a more favourable entry point to long positions. On the other hand, when forecast volatility rises, it means that the risk is on the increase and it should be cautioned or exited. Andy-Volatility supplies more information to a decision-maker than a naive buy-or-sell signal by mapping the ratio of forecast-to-current conditional volatility to a discrete signal space (BUY, SELL, HOLD, CAUTION) and by further augmenting the signal with a real-time news sentiment overlay.

The equity markets of India provide a very attractive target of this methodology. The BSE Sensex and NSE Nifty-50 are the largest and most liquid emerging-market indices in the world, but have more volatility persistence and leverage effects than developed-market indices. The majority of the online algorithmic trading systems that are used to trade Indian equities are closed-source and not accessible to individual investors and academics. Andy-Volatility fills this gap by offering an open source and transparent end-to-end pipeline that is validated on the full constituent universes of both indices.

Andy-Volatility is created as a native component of Andy Terminal a more comprehensive, unified intelligent trading platform which brings together various analytical applications such as fundamental screeners, options analytics, portfolio optimisers and macroeconomic dashboards into a single, unified interface. Andy Terminal has been designed with a consistent Flask backend, universal authentication and role-based access control layer, and single real-time data infrastructure. All modules such as Andy-Volatility are connected with this common infrastructure but have independent analytical logic. The volatility indicators produced by Andy-Volatility are displayed in the master dashboard of Andy terminal with the output of the other modules so that users can make highly integrated, multi-dimensional investment decisions. Particular attention in this paper is given to the design, implementation and empirical validation of the Andy- Volatility module.

The main contributions of this work are as follows:

- An augmented ARMA(2,3)TGARCH(1,1,1)-ARCHE estimation pipeline with ARCH-LM diagnostic gating to test the existence of heteroskedasticity before GARCH estimation.

- A five-day ahead analytic conditional variance prediction to obtain a four-class trading signal with asymmetry-based CAUTION overlay.
- A sentiment-GARCH fusion layer with model-driven signals (60-percent) and Vader NLP sentiment scores based on the Google News RSS headlines (40-percent).
- An application of Flask with production quality including real-time WebSocket feeds, a rolling 52 week heatmap, PDF export, and a Bloomberg terminal-like paper trading module.
- Empirical confirmation on 61 Indian large cap equities of the Sensex-30 and Nifty-50 universes, and full backtesting against passive Buy-and-Hold benchmark.

The rest of this paper is structured in the following way. Section II gives the mathematical model of the ARMA-TGARCH pipeline. The logic of signal generation and sentiment fusion is explained in Section III. IV describes the system architecture. Empirical results are reported in section V. Section VI reviews pertinent literature. Part VII summarizes and provides future research directions.

II. MATHEMATICAL FRAMEWORK

The Andy-Volatility modelling pipeline is based on the classical time-series econometrics and is executed in a chain of estimation: mean-equation specification through ARMA, test of heteroskedasticity diagnostic through ARCH-LM, and estimation of variance through GJR-GARCH. This section documents each part.

1. Log Returns

Let P(t) denote the split- and dividend-adjusted closing price of a stock on trading day t, where t = 1, 2, ..., T. The daily log-return series is defined as:

$$r(t) = \ln(P(t)) - \ln(P(t-1)) = \ln(P(t) / P(t-1))$$

Log returns are preferred over arithmetic simple returns $r\_s(t) = (P(t) - P(t-1)) / P(t-1)$ for several reasons. First, multi-period log returns are time-additive: the multi-period log return of k days is the sum of the daily log returns. Second, they have a natural lower limit of -1 (i.e. -100%), and thus are better suited to the normal distribution assumptions. Third, for small returns, log returns and simple returns are approximately equal, ensuring negligible practical difference in high-frequency settings. The stationarity of r(t) is verified using the Augmented Dickey-Fuller (ADF) test prior to ARMA estimation.

2. Mean Equation — ARMA(2,3)

The conditional mean of the return series is modelled as an Autoregressive Moving-Average (ARMA) process. The order (p, q) is selected by minimising the Akaike Information Criterion (AIC) over a grid of $p \in \{0,1,2,3\}$ and $q \in \{0,1,2,3\}$. Across the majority of constituent equities in both the BSE and NSE universes, ARMA(2,3) is selected as the optimal specification:

$$r(t) = \mu + \varphi_1 r(t-1) + \varphi_2 r(t-2) + \theta_1\varepsilon(t-1) + \theta_2\varepsilon(t-2) + \theta_3\varepsilon(t-3) + \varepsilon(t)$$

where μ is the unconditional mean return, φ_1 and φ_2 are the autoregressive coefficients capturing the linear dependence of the current return on its two most recent lags, θ_1, θ_2, and θ_3 are the moving-average coefficients capturing the dependence on past innovations, and ε(t) is the innovation process. The innovation is assumed to satisfy E[ε(t)] = 0 and to exhibit conditional heteroskedasticity, meaning that while innovations are uncorrelated, their squared values are not. This serial dependence in squared innovations motivates the GARCH variance equation.

3. ARCH-LM Diagnostic Test

Before fitting the variance equation, the presence of ARCH effects in the ARMA residuals must be confirmed. Engle's Lagrange Multiplier (LM) test regresses the squared residuals $\varepsilon^2(t)$ on a constant and q lagged values:

$\varepsilon^2(t) = \alpha_0 + \alpha_1\varepsilon^2(t-1) + \alpha_2\varepsilon^2(t-2) + ... + \alpha q\varepsilon^2(t-q) + u(t)$

The null hypothesis H_0: $\alpha_1 = \alpha_2 = ... = \alpha q = 0$ states that there are no ARCH effects. The test statistic LM = $T \cdot R^2$ follows a chi-squared distribution with q degrees of freedom under H_0:

$LM = T \cdot R^2(q) \sim \chi^2(q)$

where T is the effective sample size and $R^2(q)$ is the coefficient of determination from the auxiliary regression. The ARCH-LM test is used in Andy-Volatility, with q = 5 lags. Stocks with a non-rejected null at the 5% level of significance are highlighted in the UI with a warning, and no GARCH estimation is done on those assets. In the validation set of 61 equities, the ARCH-LM test rejected H0 in 92% of the samples, which validates that GARCH-family modelling is suitable in the vast majority of large-cap Indian equities.

4. Variance Equation — GJR-GARCH(1,1,1)

The conditional variance 2(t) = Var[8(t) | I(t−1)], with I(t−1) being the information set at time t, is of GJR-GARCH(1,1,1) type introduced by Glosten, Jagannathan and Runkle:

$\sigma^2(t) = \omega + \alpha\varepsilon^2(t-1) + \gamma\varepsilon^2(t-1)\cdot I[\varepsilon(t-1) < 0] + \beta\sigma^2(t-1)$

The model parameters have the following interpretations:

- **$\omega > 0$:** the constant term contributing to the long-run unconditional variance. It makes sure that the conditional variance is positive.
- **$\alpha \geq 0$:** the ARCH coefficient, which is the sensitivity of current conditional variance to the squared previous period innovation. A big α means that new shocks have a strong impact on increasing short-term volatility.
- **$\gamma \geq 0$:** the asymmetry/ leverage coefficient. In case of γ greater than 0, a negative innovation at time t-1, will add another γe 2(t-1) to the existing variance, on top of the similar ARCH term. This formalises the leverage effect.
- **$\beta \geq 0$:** the coefficient in GARCH, which quantifies the extent to which the current conditional variance is related to the lagged value of its own conditional variance. A big β means the volatility shocks are persistent and slow to dissipate with time.

The positivity constraints $\omega > 0$, $\alpha \geq 0$, $\gamma \geq 0$, $\beta \geq 0$ and the second-moment stationarity condition $\alpha + \gamma/2 + \beta < 1$ are imposed during maximum likelihood estimation using the quasi-maximum likelihood (QML) estimator with a Student-t distributed error term to accommodate fat tails.

5. Volatility Persistence

The scalar measures the level of volatility persistence:

$\rho = \alpha + \gamma/2 + \beta$

In the case of ρ <1, the conditional variance is covariance-stationary and shocks eventually fade away. The volatility shock half-life is h = ln (0.5)/ln ( ), (trading days). In cases where rho is near 1, the half-life is extremely large meaning that the effects of a big shock on variance last weeks or months. In case ρ is greater than 1, the process is known as integrated GARCH (I-GARCH), i.e. the shocks are permanent and the unconditional variance is non-existent. In Andy-Volatility, stocks with 0.999 0.999 are labeled with a near-unit-root persistence warning in the dashboard to warn users that volatility forecasting of these assets is less precise.

6. Five-Day Ahead Conditional Variance Forecast

Assuming parameter estimates (ω hat, α hat, gamma hat, 8 hat, 8), and the current conditional variance 8 2 (t), the h-step ahead analytic prediction of conditional variance is calculated recursively. For a symmetric GARCH model the recursion is straightforward; for GJR-GARCH the conditional expectation of the asymmetric term requires integration over the error distribution:

$E[\sigma^2(t+h) \mid I(t)] = \omega + (\hat{\alpha} + \hat{\gamma}/2 + \hat{\beta}) \cdot E[\sigma^2(t+h-1) \mid I(t)]$ for $h \geq 2$

with $E[\sigma^2(t+1) \mid I(t)] = \sigma^2(t+1)$ computed directly from the observed $\varepsilon(t)$. The volatility of the reported forecasts to generate the signal is the annualised square root of the mean 5-day ahead forecast variance:

$\hat{\sigma}\_forecast = \sqrt{(1/5)\, \Sigma(h=1..5)\, E[\sigma^2(t+h) \mid I(t)]} \times \sqrt{252}$

where 365 is the number of trading days per year and 252 is the annualisation of the daily standard deviation. This amount is easily comparable to the realised volatility measures that are typically reported by data vendors and is an intuitive scale to practitioners.

III. SIGNAL GENERATION LOGIC

1. Four-Class Signal Framework

The trading signal is determined as the ratio R = 5-day ahead forecast volatility (as above) / 5-day current conditional variance = 5-day ahead forecast volatility / 5-day current conditional variance = 5-day ahead forecast volatility / 5-day current conditional variance. R is a ratio of how the market will be likely to be calmer ($R < 1$) or turbulent ($R > 1$) in the upcoming trading week. The following are the rules of signal classification:

Signal	Primary Condition	Additional Condition
BUY	$R < 0.90$	Any γ
SELL	$R > 1.10$	Any γ
CAUTION	$0.90 \leq R \leq 1.10$	$\gamma > 0.10$
HOLD	$0.90 \leq R \leq 1.10$	$\gamma \leq 0.10$

Table 1: Four-class signal classification rules.

A BUY signal is that the forecast volatility is predicted to fall by over 10% compared to the current level, which means that the market is becoming more stable and the risk-adjusted entry point of long positions is more favourable. A SELL signal means that we are likely to experience a volatility increase of more than 10 percent, which implies an increasingly turbulent environment and high downside risk. The leverage parameter γ serves as a second-order discriminant in stocks with a forecast volatility nearly equal to its current value (= -10-+10% change) where the stock has shown a high level of asymmetric shock sensitivity in the past: a 10-14 incurred leverage 0.10 value indicates a CAUTION designation; a 10-14 value of 0.05

Signal thresholds (0.90 and 1.10 in the case of R, 0.10 in the case of γ) were optimized on a training window of 2015-2021 by maximizing Risk-adjusted returns over a Buy-and-Hold benchmark on a held-out 2022-2024 validation sample. Thresholds are configured in the application settings to enable analysts to change the sensitivity according to their risk appetite and assessment of the market regime.

2. News-Sentiment Fusion Signal

To complement the purely quantitative model-driven signal, Andy-Volatility incorporates a real-time news sentiment layer. For each tracked equity, up to eight of the most recent headlines are retrieved from Google News RSS feeds using the stock's name and ticker as search terms. Each headline is scored using the VADER (Valence Aware Dictionary and sEntiment Reasoner) lexicon, which assigns a compound sentiment score in the range [−1, +1], where −1 denotes maximally negative and +1 denotes maximally positive sentiment. VADER is particularly well-suited to short, informal financial news text due to its handling of punctuation, capitalisation, and domain-specific modifiers.

The per-headline VADER scores are averaged to produce a single normalised sentiment score $S\_NLP \in [-1, +1]$. The four GARCH signal classes (SELL, CAUTION, HOLD, BUY) are mapped to a numerical scale $S\_GARCH \in \{-1, -0.33, +0.33, +1\}$. The final fusion score is a weighted linear combination:

$$F = 0.60 \cdot S\_GARCH + 0.40 \cdot S\_NLP$$

The fusion score F is then discretised into five final signal classes using fixed thresholds: STRONG BUY ($F > 0.6$), LIKELY BUY ($0.2 < F \leq 0.6$), NEUTRAL ($-0.2 \leq F \leq 0.2$), LIKELY SELL ($-0.6 \leq F < -0.2$), and STRONG SELL ($F < -0.6$). The 60/40 weighting was chosen to preserve the primacy of the econometric model while allowing sentiment to modify signals near classification boundaries. An example of this would be a HOLD signal with a news sentiment uniformly negative, which would be re-classified as LIKELY SELL and could warn the user of an impending risk earlier than the entirely price-based model.

3. Backtesting Engine

Backtesting can be used to compare the historical performance of the volatility-based strategy with a passive Buy-and-Hold benchmark over a specified window. The strategy has a long position when the daily signal is BUY, a cash exit when the signal changes to SELL or CAUTION and the current position when the signal is not SELL, CAUTion or CASH. Transaction costs are assumed to be a fixed round-trip commission of 0.10% of trade value, which is in line with the average Indian retail brokerage. The performance is reported as a total return, annualised Sharpe ratio, maximum drawdown, and signal transitions. These indicators are given under a per-stock performance tab that can be viewed on the main dashboard.

IV. SYSTEM ARCHITECTURE

Andy-Volatility platform is a layered architecture that is designed to be a modular architecture that separates data ingestion, analytical computation, persistence, and presentation. These key elements are explained below.

1. Data Ingestion and Caching Layer

Adjusted OHLCV (Open, High, Low, Close, Volume) data at a historical scale are requested through the yfinance Python package, which uses the public API of Yahoo Finance. The lookback window is 10 years long and offers adequate observations to estimate the stable GARCH parameters (around 2,500 trading days). The PyArrow backend serialises the data that is retrieved to Apache Parquet on disk, allowing read access via columnar access. A cache validity check is an operation to compare the modification time of the Parquet file to a configurable TTL (default 6 hours); files that are stale are automatically updated when next accessed. This caching plan saves Yahoo Finance API calls by about 80 with typical usage patterns and provides sub-second loading of data on previously analysed tickers.

Two of the curated ticker lists define the stock universe: 30 BSE equities (the complete Sensex-30 index), and 31 NSE equities (the Nifty-50 index). The ticker symbols are in the Yahoo Finance convention (e.g., HDFCBANK).BO of BSE, HDFCBANK.NS for NSE). The universe lists are held in a JSON configuration file, and are extendable without writing code.

2. Analytical Computation Layer

ARMA-TGARCH estimation pipeline is done in Python with the arch Python library of GARCH modelling and statsmodels library of order selection in ARMA. This calculation is parallelised over stocks with Python concurrent.futures.ThreadPoolExecutor that has an adjustable number of worker threads (4 by default). Full universe re-analysis (61 stocks) on a typical quad-core server takes about 45 seconds. Serialised to a results cache in JSON format with a TTL of 15 minutes, per-stock results (estimated parameters, signal, forecast volatility, persistence, and ARCH-LM p-value) are refreshed in the background when the next user request arrives.

The error handling is provided on several levels: stocks that fail ARCH-LM test, fail to converge in the estimation of GARCH or provide less than 500 history days with valid trading records are given a signal of N/A and shown with proper warning in the UI. This provides gracious degradation of new listed stocks or tickers having data quality concerns.

3. Backend: Flask Web Application

The web server is written in Python 3.12 and based on Flask and Flask-SocketIO and Eventlet to support non-blocking asynchronous WebSocket services. A Redis Pub/Sub channel is fed live intraday price data by a second process of price feeds, and pushed out to all connected WebSocket clients, allowing real-time updates to dashboard without page reload. Flask-APScheduler extension will re-analyze the entire universe after every 15 minutes during the trading hours of the Indian Stock Exchange (09:15-15:30 IST Monday-Friday). Cached results are also provided outside trading hours and do not re-compute.

- **Authentication:** Flask cookie session and Werkzeug bcrypt password hashing (bcrypt cost factor 12). No plaintext passwords are ever kept.
- **Role-Based Access Control:** There are two roles, Admin (full access, including paper trading and user management), and Analyst (access to dashboards, screener, and heatmap only).
- **PDF Export:** WeasyPrint will generate a per-stock Jinja2 HTML template to a downloadable PDF report containing parameter estimates, signal history, and volatility chart.
- **Rate Limiting:** Flask-Limiter will rate limit API endpoints to 60 requests per minute per IP to avoid abuse.

4. Database Schema

Application state is stored in a SQLite database, with Flask-SQLAlchemy. The schema has four major tables:

Table	Primary Key	Key Columns	Purpose
users	id (int)	username, password_hash, role, created_at	User credentials and role assignment
accounts	id (int)	user_id, cash_balance, last_updated	Virtual trading account balance (₹10,00,000 initial)
portfolios	id (int)	user_id, ticker, quantity, avg_buy_price	Open equity holdings per user
trades	id (int)	user_id, ticker, action, price, quantity, signal, timestamp	Complete immutable trade history

Table 2: SQLAlchemy database schema summary.

Portfolio P&L is calculated dynamically by combining portfolios table with most recent price on the ticker that is in the cache. The unrealised gains and losses are presented in real time on the paper trading dashboard. The GARCH signal at the time of execution captures all executions of the trade, and post-hoc analysis of the existence or absence of consistency of the trade with the recommendations of the model can be conducted.

5. Paper Trading Module

Admin users are able to access a Bloomberg terminal-like three-paneling paper trading interface:

- **Left Panel:** Portfolio summary with semicircular gauge of capital utilisation (drawn on a canvas), P&L summary cards with unrealised gain/loss and total portfolio value, and a scrollable list of holdings with per-position weight.
- **Centre Panel:** Live signal table of all tracked equities with pill buttons (BUY / SELL / HOLD / CAUTION) which are filterable. The ticker, current price, signal, forecast volatility and persistence appear in every row. By clicking on a row, the right-panel trade form is pre-filled.
- **Right Panel:** Trade execution form where BUY/SELL is shown with a toggle, quantity is shown with a preview of the calculated order value, and a toast on successful execution. When there is not enough cash to buy a buy order or when the user does not own enough shares to sell them, an informative error message is given, causing the orders to be rejected..

All trades of paper products are marked with the live GARCH signal when they are executed, creating a history of verifiable connections between all decisions and the model output. This design is able to support retrospective performance attribution.

6. Frontend and Visualisation Layer

The frontend uses vanilla JavaScript and Bootstrap 5, the minimal dependency on heavy frameworks is used to minimise load times. Chart.js is used to render interactive charts; the 52-week rolling volatility heatmap is a bespoke Canvas-based element that colour-codes each cell of the stock-week grid by quantile rank of annualised volatility, in a green-yellowred diverging colour scheme. The heatmap is an effective visual summary of volatility regimes in the entire universe, over the last 12 months and is especially useful in identifying sector-wide volatility clustering events, like budget announcements, RBI policy decisions, or global risk-off events.

V. EMPIRICAL RESULTS AND ANALYSIS

1. Dataset and Experimental Setup

The empirical analysis is based on the daily adjusted close prices of 30 BSE equities (Sensex-30 constituents) and 31 NSE equities (Nifty-50 constituents) between 1 January 2015 and 14 April 2026, and each stock has around 2,750 observations. All prices are obtained at Yahoo Finance and are re-adjusted to splits and dividends. The ARMA order selection grid covers $p, q \in \{0, 1, 2, 3\}$. GARCH estimation uses QML with a Student-t error distribution; the degrees-of-freedom parameter ν is jointly estimated. The ARCH-LM test is applied with $q = 5$ lags at the 5% significance level. The backtesting window covers 1 January 2022 to 14 April 2026, with the preceding seven years used as the in-sample estimation period. All computations are performed on a single server with a 2.4 GHz quad-core CPU and 16 GB RAM.

2. ARCH-LM Test Results

The ARCH-LM test rejected the null hypothesis of zero ARCH effects ($p = 0.05$) in 56 of 61 stocks (91.8%), which supports the conclusion that GARCH is the right model of large-cap Indian equities. The five stocks with the test not rejecting H_0 were given an N/A signal and were not included in any further GARCH analysis. These exceptions were clumped around low-liquidity mid-cap stocks on the edge of the

Nifty-50 that had just been added to the index, implying that their returns series were not yet sufficiently statistically regular to be modelled using ARCH.

3. Signal Distribution — BSE Universe (April 2026)

On 30-stock Sensex universe running the entire BSE pipeline on 14 April 2026 generated the following signal distribution:

Signal	Count	% of Universe	Avg. Persistence (ρ)
BUY	12	40.0%	0.988
SELL	3	10.0%	0.984
CAUTION	1	3.3%	0.991
HOLD	13	43.3%	0.992
N/A	1	3.3%	—

Table 3: BSE signal distribution, 14 April 2026.

This distribution indicates a fairly optimistic market with 40% of Sensex constituents showing decreasing forecast volatility. Among the individual ones, one can mention: ICICIBANK.There was a strong leverage effect (γ = 0.109), the largest in the universe, which made BO flagged as CAUTION.BO and BHARTIARTL.BO demonstrated close to unit-root persistence (0.000) and, therefore, any volatility shock would be highly persistent; ITC.BO and MARUTI.BO got SELL signals where the forecast volatility increased 18.4% and 20.1% above the current values respectively.

4. Estimated TGARCH Parameters — Selected Stocks

The estimates of the GJR-GARCH(1,1,1) parameters of a sample of BSE equities are obtained to report this as Table 4. The volatility persistence of all stocks (0.97) is in line with the overall results in the emerging-market GARCH literature. The leverage parameter γ is significant across all stocks presented and this proves that there are asymmetric volatility dynamics.

Ticker	ω (×10⁻⁵)	α	γ	β	ρ	Signal
HDFCBANK.BO	1.82	0.041	0.032	0.937	0.994	BUY
ICICIBANK.BO	3.10	0.053	0.109	0.884	0.991	CAUTION
TATASTEEL.BO	0.41	0.002	0.001	0.997	1.000	HOLD
MARUTI.BO	4.25	0.061	0.019	0.920	0.991	SELL
LT.BO	5.60	0.072	0.028	0.891	0.977	BUY
WIPRO.BO	2.93	0.048	0.044	0.912	0.982	HOLD
SUNPHARMA.BO	3.77	0.055	0.061	0.903	0.989	BUY

Table 4: Estimated GJR-GARCH(1,1,1) parameters, BSE sample.

5. Backtesting Performance

The volatility-signal strategy was backtested on the BSE universe over the 2022–2026 validation window.

Averaged across all 29 stocks for which signals were available (excluding the one N/A case), the strategy achieved an annualised Sharpe ratio of 0.84 compared to 0.61 for the equal-weighted Buy-and-Hold benchmark, representing a 38% improvement in risk-adjusted return. Maximum drawdown was reduced from 31.2% (Buy-and-Hold) to 22.8% (strategy), primarily because the SELL signal correctly triggered exits ahead of several high-volatility episodes including the October 2022 global equities correction and the March 2024 pre-election volatility spike.

Metric	Andy-Volatility Strategy	Buy & Hold Benchmark
Annualised Return	18.4%	16.1%
Annualised Volatility	14.2%	19.8%
Sharpe Ratio	0.84	0.61
Maximum Drawdown	−22.8%	−31.2%
Signal Transitions	147	—

Table 5: Backtesting performance summary, BSE universe 2022–2026.

VI. RELATED WORK

The theoretical foundations of Andy-Volatility rest on a well-established body of literature in financial econometrics and computational finance. Engle introduced the ARCH model in his seminal 1982 paper, demonstrating that the time-varying variance of UK inflation could be modelled as a function of past squared residuals. Bollerslev extended this to the GARCH specification, introducing a parsimonious lag structure that dramatically improved model fit while reducing parameter count. The leverage asymmetry in equity volatility was identified empirically by Black and theoretically rationalised through the debt-equity mechanism, whereby a fall in equity value increases financial leverage and therefore return variance. The GJR-GARCH formalisation of the leverage effect by Glosten, Jagannathan, and Runkle remains the most widely used asymmetric GARCH variant in applied work.

Poon and Granger provide a comprehensive survey of volatility forecasting methods, concluding that GARCH models consistently outperform historical volatility measures for short-horizon forecasting. Heston and Nandi demonstrated that closed-form option pricing is tractable under a GARCH variance process, providing a theoretical link between GARCH volatility forecasts and option market pricing. For the specific context of Indian equity markets, Karmakar documented statistically significant asymmetric volatility in BSE sectoral indices, with leverage parameters consistent with those estimated in this work.

On the natural language processing side, Hutto and Gilbert developed the VADER sentiment lexicon specifically for short social-media-style text, demonstrating superior performance relative to machine-learning alternatives in settings with limited training data. Subsequent work has applied VADER and similar lexicon-based approaches to financial news headline classification with promising results, motivating its adoption in the Andy-Volatility sentiment fusion layer.

Several prior systems have combined GARCH forecasts with machine-learning classifiers for trading signal generation. Atsalakis and Valavanis reviewed over 100 computational intelligence approaches to stock price forecasting, noting that hybrid systems combining time-series models with sentiment or technical indicators tend to outperform single-model approaches. Bollen, Mao, and Zeng demonstrated that Twitter mood states predict Dow Jones movements, establishing the conceptual basis for sentiment-price coupling. Andy-Volatility distinguishes itself from prior work by targeting the specific context of Indian large-cap equities, providing a complete production-quality open-source implementation, and delivering a transparent econometric modelling approach rather than a black-box machine-learning system.

VII. CONCLUSION

This paper presented Andy-Volatility, a complete end-to-end volatility modelling and automated trading signal generation platform designed for Indian equity markets. The core analytical engine is a joint ARMA(2,3)–GJR-GARCH(1,1,1) pipeline that models the time-varying conditional variance of daily log-return series for equities in the BSE Sensex-30 and NSE Nifty-50 universes. The pipeline incorporates ARCH-LM diagnostic gating to validate heteroskedasticity prior to GARCH estimation, maximum likelihood estimation with a Student-t error distribution to accommodate fat tails, and a five-day ahead analytic forecast recursion to generate forward-looking volatility estimates.

Trading signals are derived from the ratio of forecast-to-current conditional volatility and augmented by the GJR-GARCH asymmetry parameter to produce a four-class classification (BUY, SELL, HOLD, CAUTION). A news-sentiment fusion layer based on VADER NLP scoring of Google News headlines provides an additional real-time information signal, combined with the model-driven output in a 60/40 weighted fusion score. Backtesting results over the 2022–2026 validation window demonstrate that the strategy achieves a 38% improvement in Sharpe ratio relative to a passive Buy-and-Hold benchmark, with meaningfully reduced maximum drawdown.

The site is served as a flask web app in production quality with live WebSocket price feeds, rolling 52-week volatility heatmap, PDF report export, a Bloomberg terminal-like paper trading system with role-based access control, and detailed SQLite-powered trade history logging. The re-analysis (61 stocks) of a full universe can be done in around 45 seconds using a typical quad-core server.

Some of the promising extensions will be discussed in future work. To begin with, the Dynamic Conditional Correlation (DCC-GARCH) models will be applied to include time-varying cross-asset correlations, which allows breaking down risk at the portfolio level. Second, it will incorporate India VIX, which is the implied volatility index of the NSE and will be used as a regime switching trigger to change signal thresholds when markets are stressed. The third one will be to reinforcement learners to learn adaptive trading rules by training on sequences of GARCH signals that capture persistence and regime shifts. Fourth, pricing of equity options will be investigated with the use of the Heston-GARCH model which will enhance the usability of the platform to derivatives analytics. Fifth, the sentiment layer will be replaced with a fine-tuned transformer model that is trained on text Indian financial news, which will have a higher accuracy in headline classification.

At platform level, it is intended that it be more integrated with the overall Andy Terminal ecosystem. Subsequent versions will open up Andy-Volatility signal API to other Andy Terminal modules so that the underlying screener and portfolio optimiser modules can take volatility forecasts as risk inputs. A single Andy Terminal mobile interface is in process, which will present condensed signal cards of Andy-Volatility among other module outputs, and allow on-the-go decision support to both retail and professional analysts.

VIII. ACKNOWLEDGMENT

The authors hereby take the opportunity to state that they are very much indebted to K. Shalini, Assistant

Professor, Computer Science and Engineering, Matrusri Engineering College, who guided us, motivated us and technically supported us during the course of this project. Her work in quantitative finance and system design greatly influenced the course of this study. The authors also acknowledge the Head of the Department of Computer science and engineering and faculty of Matrusri engineering college to have provided academic support as well as creating a research based learning environment. The management of the institution is also given special credit to ensure that they availed the infrastructure, as well as resources, required to undertake this work. Lastly, the authors would like to thank their families and peers who have been supportive and encouraging throughout this project in terms of morality.

IX. REFERENCES

1. R. F. Engle, "Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation," Econometrica, vol. 50, no. 4, pp. 987–1007, 1982.
2. K. Safi, S. Verma, and A. Kumar, "Financial Modelling of BSE-SENSEX Volatility using ARMA, ARCH and TGARCH Model," *Pacific Business Review (International)*, vol. 16, no. 8, pp. 103–111, Feb. 2024..
3. 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," Journal of Finance, vol. 48, no. 5, pp. 1779–1801, 1993.
4. F. Black, "Studies of stock price volatility changes," in Proc. 1976 Meetings of the American Statistical Association, Business and Economic Statistics Section, 1976, pp. 177–181.
5. S.-H. Poon and C. W. J. Granger, "Forecasting volatility in financial markets: A review," Journal of Economic Literature, vol. 41, no. 2, pp. 478–539, 2003.
6. S. L. Heston and S. Nandi, "A closed-form GARCH option valuation model," Review of Financial Studies, vol. 13, no. 3, pp. 585–625, 2000.
7. C. J. Hutto and E. E. Gilbert, "VADER: A parsimonious rule-based model for sentiment analysis of social media text," in Proc. 8th Int. AAAI Conf. Weblogs and Social Media (ICWSM), 2014.
8. M. Karmakar, "Asymmetric volatility and risk-return relationship in the Indian stock market," South Asia Economic Journal, vol. 8, no. 1, pp. 99–116, 2007.
9. R. Garvey and A. Murphy, "The profitability of active stock traders," Journal of Applied Finance, vol. 15, no. 2, pp. 93–100, 2005.
10. G. S. Atsalakis and K. P. Valavanis, "Surveying stock market forecasting techniques — Part II: Soft computing methods," Expert Systems with Applications, vol. 36, no. 3, pp. 5932–5941, 2009.
11. J. Bollen, H. Mao, and X. Zeng, "Twitter mood predicts the stock market," Journal of Computational Science, vol. 2, no. 1, pp. 1–8, 2011.
12. R. F. Engle and A. J. Patton, "What good is a volatility model?" Quantitative Finance, vol. 1, no. 2, pp. 237–245, 2001.
13. B. Mandelbrot, "The variation of certain speculative prices," Journal of Business, vol. 36, no. 4, pp. 394–419, 1963.