# A Time Series Modelling and Inferential Analysis of The National Stock Exchange (NSE) of Energy Sector in India: A Contemporary Perspective

Dr. Jitendra Kumar
*Deparment of Mathematics*
*Vellore Institute of Technology*
Vellore, India
jitendra.kumar@vit.ac.in

Febin V Shiju
*Department of Mathematics*
*Vellore Institute of Technology*
Vellore, India
febinvs07@gmail.com

Abhinav V Sunil
*Department of Mathematics*
*Vellore Institute of Technology*
Vellore, India
abhinavvsunil07@gmail.com

Bela Sreekesh
*Department of Mathematics*
*Vellore Institute of Technology*
Vellore, India
belashenoi@gmail.com

***Abstract*—This study analyzes stock price dynamics of selected NSE Energy Sector companies using time series techniques. Stability and volatility are assessed through the coefficient of variation and confidence intervals, while clustering highlights firm-level differences. Forecasting is performed using ARIMA, XGBoost and LSTM, with LSTM showing superior performance in capturing nonlinear trends. The findings offer practical insights for investment strategies and policy decisions in volatile market conditions.**

***Index Terms*—NSE, Volatility, LSTM, ARIMA**

## I. Introduction

Stock market analysis has long been a cornerstone of financial research, offering critical insights into market dynamics, volatility, and investor behaviour [8], [13]. In the context of the Indian stock market, particularly the National Stock Exchange (NSE), the energy sector holds a pivotal position, influencing both industrial productivity and national economic stability [12], [15].

In the recent years integration of statistical methods with machine learning models has significantly helped financial forecasting [4], [17]. Traditional approaches like ARIMA remain widely used due to their interpretability and robustness in capturing linear temporal dependencies [5], [6]. However, the emergence of deep learning and machine learning architectures—particularly the Long Short-Term Memory (LSTM) network—has enabled researchers to model complex nonlinearities and long-range dependencies in financial time series with superior predictive accuracy [9], [16].

This research work develops an integrated framework for the analysis of the NSE energy sector using statistical and machine learning methods. Trend and momentum signals are captured by indicators derived from stock prices, while returns describe performance and volatility. Coefficient of variation and confidence interval analysis are used to rank companies based on price stability and uncertainty. The K-Means clustering is then used to segment stocks into three movement profiles, namely Strong Movers, Neutral Movers, and Quiet Movers, giving a data-driven segmentation of market behavior. Furthermore, ARIMA, XGBoost and LSTM models are used for predictive modelling, allowing short and long-term forecasting [18], [24].

This integrated approach bridges the gap between classical statistical inference and modern machine learning techniques, offering a general view of the dynamics of stock. [10], [11].

## II. Methodology

### A. Data Collection and Pre-processing

The study analyzed daily stock prices of the following mentioned energy sector companies Adani Power, Bharat Petroleum, Castrol, CESC, Chennai Petroleum, Coal India, GAIL, Gujarat Gas, Gujarat Industries Power, Gulf Oil Lubricants, Hindustan Oil, Hindustan Petroleum, Indian Oil, JSW Energy, Linde India, MRPL, NHPC, NTPC, Oil India, ONGC, Orient Green Power, Petronet LNG, Power Grid, RattanIndia Power, Reliance Industries, SJVN, Suzlon Energy, Tata Power, and Veedol Corporation. The dataset consisted a 10-year period from 2014 to 2024.

The data included closing price, trading volume, and date, among others, that were observed for the set period. It then identifies missing values or outliers and cleans them using forward-fill techniques to ensure intact data. To eliminate differences among companies some features were normalized.

### B. Feature Derivation and Technical Indicator Computation

To capture the behavioural and directional characteristics of price movements, several derived indicators [1]were computed:

- Price Return (R) – representing daily percentage change in stock prices.
- Relative Strength Index (RSI 14-day) – captures short-term momentum and give idea on overbought/oversold zones. [2]
- Moving Average Convergence Divergence (MACD) – indicating long-term trend shifts through signal crossovers. [2]
- Normalized Volume (V_norm) – standardizing volume to highlight relative liquidity changes. [20]

These derived features were employed to quantify the temporal and volatility characteristics of each company's stock movement.

### C. Statistical Measures and Composite Index Construction

To evaluate the variability and uncertainty within indicators, two major statistical measures were computed:

- Coefficient of Variation (CV) — quantifying relative variability across features to identify companies with higher volatility. [22]

$$CV = \frac{\sigma}{\mu} \times 100 \tag{1}$$

where:
$\sigma$ = standard deviation
$\mu$ = mean

- Confidence Interval (CI) — estimating the precision of the mean indicator values to represent statistical stability. [22]

$$CI = \overline{x} \pm t_{\alpha/2,n-1} \times \frac{s}{\sqrt{n}} \tag{2}$$

where:
$\overline{x}$ = mean
s = standard deviation
n = number of observation

- Spearman Rank Correlation - was applied between CV and CI to assess the monotonic association between volatility and confidence-based stability measures.

$$\rho = 1 - \frac{6\sum d^2}{n(n^2-1)} \tag{3}$$

where $d_i$ represents the difference between paired ranks and $n$ is the number of observations. [19]

### D. Clustering Analysis

An unsupervised clustering approach (K-Means) [23] was used to categorize companies regarding on their technical indicators (Volume_normalized, Return, RSI_14, and MACD). The suitable number of clusters was determined using the Elbow Method which allowed grouping of companies with similar volatility and behavioural patterns.

### E. Forecasting Models

To evaluate and compare predictive performance, three forecasting models were implemented:

- ARIMA (1,1,1) – a statistical model designed for linear trend-based forecasting after differencing to ensure stationarity. [3]
- LSTM (Long Short-Term Memory) – a deep learning model capable of capturing temporal dependencies and nonlinear price dynamics using differenced price sequences. [9]
- XGBoost Regressor – a gradient boosting model trained on structured tabular data (Price, RSI, MACD, and Volume_normalized) to forecast future prices. [7]

### F. Evaluation Metrics

Model performance was evaluated using RMSE, and comparative analysis was performed across ARIMA, LSTM, and XGBoost to determine forecasting strength. In addition to that statistical correlation between CV and CI was used to understand the relationship between volatility and confidence levels across stocks.

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(\hat{y}_i - y_i)^2}{n}} \tag{4}$$

where $\hat{y}_i$ are predicted values, $y_i$ are observed values and n is the number of observations. [21]

## III. Results and Discussion

### A. Derived Technical Indicators and Market Behaviour

Based on the directional movement and indicator trends, the companies were classified as bullish or bearish, reflecting their overall momentum consistency across the observation window.

Bullish companies: Coal India, RattanIndia Power, Suzlon Energy, ONGC, and Orient Green Power. These stocks exhibited positive RSI trends, frequent MACD crossovers above the signal line, and price movements clustering near the upper Bollinger Band, suggesting steady upward momentum and buying strength.

Bearish companies: Gujarat Gas, Bharat Petroleum, Petronet LNG, HP, and Linde India. These firms showed persistent RSI values below the neutral 50 level, negative MACD values, and price movement near or below the lower Bollinger Band, indicating sustained selling pressure and downward bias.

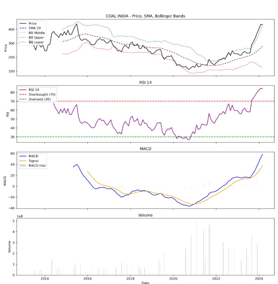

Fig. 1. Coal India

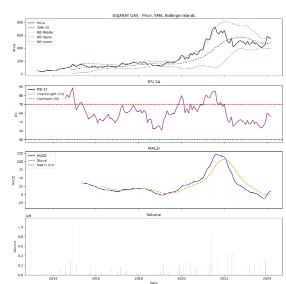

Fig. 2. Gujarat Gas

### B. Descriptive Statistics

*1) Coefficient of Variance:* The Coefficient of Variation (CV) was computed to assess relative variability. The feature Normalized volume had the highest CV which indicates strong trading fluctuations, while RSI, MACD, and price exhibited moderate dispersion. The composite CV highlighted firms with higher volatility and reactive market behaviour. [14]

*2) Confidence Interval:* In this study, A combination of several technical indicators (e.g. Price, RSI, MACD, Volume normalized) transformed into a Composite CI (Complexity Index) that quantifies the overall behavioural variability or complexity of a company's stock movement. [14]

TABLE I
RANKING BASED ON CV

| Company Name | Mean | Std | CV_percent | Rank |
|---|---|---|---|---|
| ONGC | 0.35 | 0.14 | 38.88 | 1 |
| COAL INDIA | 0.36 | 0.15 | 41.04 | 2 |
| PETRONET LNG | 0.42 | 0.18 | 43.04 | 3 |
| CESC | 0.34 | 0.15 | 45.52 | 4 |
| BP | 0.38 | 0.18 | 46.73 | 5 |

TABLE II
RANKING BASED ON CI

| Name | Price | RSI | MACD | CI | Rank |
|---|---|---|---|---|---|
| NHPC | 4.42 | 4.00 | 0.99 | 0.09 | 1 |
| CESC | 7.30 | 3.70 | 1.45 | 0.10 | 2 |
| POWER GRID | 15.6 | 3.72 | 2.51 | 0.11 | 3 |
| CASTROL | 13.8 | 3.83 | 4.51 | 0.11 | 4 |
| RATTANINDIA | 1.09 | 3.95 | 0.28 | 0.14 | 5 |

*3) Spearman Rank Correlation:* Spearman rank correlation between Composite CV and Composite CI (Confidence Interval) was computed and resultant $\rho$ = 0.07. This reflects that the relationship is negligible which means that the relative variability among companies does not directly relate to the confidence width of their statistical estimates. In other words, even for those companies with higher internal dispersion, it can be observed that the confidence bounds were stable, reflecting robustness in their indicator estimation.

## C. Clustering

From the Elbow plot, a clear bend was observed near three clusters, indicating that K = 3 provides the most meaningful separation among the companies.

TABLE III
CLUSTERING OF COMPANIES

| **Cluster** | **Company** | **Interpretation** |
|---|---|---|
| **Cluster 0** | Adani Power, Bharat Petroleum, CESC | Steady momentum and moderate investor activity,stocks show balanced technicals, neither strongly overbought nor oversold. |
| **Cluster 1** | Gujarat Gas, Gujarat Industries Power ,Gulf Oil Lubricants | This group shows bullish tendencies — moderate returns with strong RSI suggest ongoing buying momentum. These companies may represent technically strong or strengthening stocks. |
| **Cluster 2** | Coal India, GAIL, Linde India | Mixed or corrective behavior — some show oversold signs (e.g., Coal India, ONGC) while others (e.g., Linde India) are spiking in RSI,represents volatile or transitional stocks, where traders' sentiment is shifting. |

## D. Forecasting of Stock Prices

It is observed from the comparative results of ARIMA, LSTM, and XGBoost that there is a clear difference in the forecasting accuracy. Overall, LSTM produced the lowest RMSE for most firms by capturing nonlinear and temporal dependencies. ARIMA works well on stable, stationary stocks like GAIL and Indian Oil. XGBoost showed better performance for moderately volatile companies like Hindustan Petroleum and Bharat Petroleum. Bullish firms such as Adani Power and Coal India have shown improvement in trend capture under LSTM, while XGBoost's feature-based learning has been beneficial to bearish stocks like Gujarat Gas and Petronet LNG. These observations have verified that LSTM has an advantage in modeling dynamic, trend-sensitive series, along with the complementary stability and interpretability provided by ARIMA and XGBoost.

TABLE IV
RMSE FORECASTING RESULTS

| RMSE | | | |
|---|---|---|---|
| Name | ARIMA | XGBoost | LSTM |
| ADANI POWER | 37.83 | 140.59 | 62.38 |
| BHARAT PETROLEUM | 23.36 | 20.87 | 16.25 |
| CASTROL | 18.63 | 12.40 | 12.66 |
| CESC | 14.09 | 21.98 | 9.45 |
| CHENNAI PETROLEUM | 51.32 | 190.73 | 44.88 |
| COAL INDIA | 27.84 | 51.87 | 24.68 |
| GAIL | 11.89 | 25.04 | 8.24 |
| GUJARAT GAS | 45.14 | 51.91 | 34.89 |
| GUJARAT INDUSTRIES POWER | 23.33 | 28.00 | 16.12 |
| GULF OIL LUBRICANTS | 85.27 | 101.64 | 71.50 |
| HINDUSTAN OIL | 23.29 | 18.23 | 20.05 |
| HINDUSTAN PETROLEUM | 29.57 | 28.52 | 19.59 |
| INDIAN OIL | 10.65 | 11.95 | 8.29 |
| JSW ENERGY | 45.23 | 110.20 | 45.68 |
| LINDE INDIA | 552.47 | 1841.85 | 450.11 |
| MRPL | 20.25 | 29.88 | 14.94 |
| NHPC | 8.91 | 25.63 | 5.97 |
| NTPC | 19.56 | 91.24 | 17.15 |
| OIL INDIA | 26.31 | 86.06 | 23.71 |
| ONGC | 17.05 | 32.08 | 18.51 |
| ORIENT GREEN POWER | 3.27 | 5.02 | 1.85 |
| PETRONET LNG | 22.00 | 14.74 | 5.36 |
| POWER GRID | 15.85 | 47.75 | 5.02 |
| RATTANINDIA POWER | 1.37 | 0.78 | 1.92 |
| RELIANCE INDUSTRIES | 66.23 | 134.31 | 24.42 |
| SJVN | 14.23 | 41.98 | 1.804 |
| SUZLON ENERGY | 4.95 | 13.67 | 3.841 |
| TATA POWER | 30.03 | 71.95 | 8.916 |
| VEEDOL CORPORATION | 124.59 | 113.11 | 180.227 |

## E. Conclusion

This study provided an analytical framework to model and forecast the stock price behavior of companies listed as part of the energy sector of the National Stock Exchange (NSE). By integrating components of descriptive, inferential, and machine learning techniques, the work highlighted shared volatility and trend patterns across companies. The indicators derived from the analysis successfully captured features of behavior that distinguish bullish and bearish stocks.

The analysis of Coefficient of Variation (CV) and Confidence Interval (CI) of the different investments provided variation in price stability and investor activity. The weak rank correlation ($\rho$ = 0.07) between them suggested that statistical precision and volatility were relatively independent. K-means clustering, grouping firms into three behaviourally consistent clusters, offered meaningful classification of market dynamics.

During the forecasting analysis, the LSTM model appeared to provide better prediction accuracy due to its nonlinear temporal dependencies. ARIMA remained reliable for stable, stationary data, while XGBoost was effective when feature-based contextual information was incorporated. On the whole, this hybrid comparison highlights the relevance of deep learning models in financial time series prediction.

### F. Future Work

Future research may explore the following directions:

- Integration of Sentiment and Macro Variables – Including textual data such as news sentiment or macroeconomic indicators could improve contextual accuracy.
- Hybrid and Ensemble Forecasting Models – Combining ARIMA, LSTM, and boosting-based learners may enhance performance stability across market conditions.
- Regime-Switching and Volatility Modelling – Incorporating models such as GARCH or Markov-switching LSTM to explicitly capture volatility clustering.
- Long-Horizon and Cross-Sectoral Analysis – Extending the study beyond the energy sector to compare sectoral sensitivities and diversification effects.
- Explainable AI in Finance – Applying SHAP or attention-based interpretations to make neural forecasts more transparent to investors and regulators.

## References


[1] Steven B Achelis. *Technical Analysis from A to Z*. McGraw-Hill, 2001.

[2] Erman Arif, Suherman Suherman, and Aris Puji Widodo. Integration of technical analysis and machine learning to improve stock price prediction accuracy. *Mathematical Modelling of Engineering Problems*, 11(11), 2024.

[3] A Mohamed Ashik and K Senthamarai Kannan. Time series model for stock price forecasting in india. In *Logistics, Supply Chain and Financial Predictive Analytics: Theory and Practices*, pages 221–231. Springer, 2018.

[4] George S Atsalakis and Kimon P Valavanis. Surveying stock market forecasting techniques – part ii: Soft computing methods. *Expert Systems with Applications*, 36(3):5932–5941, 2009.

[5] George E P Box, Gwilym M Jenkins, Gregory C Reinsel, and Greta M Ljung. *Time Series Analysis: Forecasting and Control*. John Wiley & Sons, 2015.

[6] Chris Chatfield. *Time-series forecasting*. CRC Press, 2000.

[7] Tianqi Chen and Carlos Guestrin. Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*, pages 785–794, 2016.

[8] Eugene F Fama. Efficient capital markets: A review of theory and empirical work. *The Journal of Finance*, 25(2):383–417, 1970.

[9] Thomas Fischer and Christopher Krauss. Deep learning with long short-term memory networks for financial market predictions. *European journal of operational research*, 270(2):654–669, 2018.

[10] Ankur Gupta, Bhanu Dhingra, et al. Machine learning applications in stock market prediction: A survey. *Proceedings of the International Conference on Computational Intelligence and Data Science*, 2020.

[11] R Jain and S Kumar. Forecasting stock market trends using hybrid deep learning models: Evidence from emerging markets. *Applied Soft Computing*, 120:108651, 2022.

[12] Archana Kumari and Anil Sharma. Energy consumption and economic growth: Empirical evidence from india. *Energy Policy*, 118:304–312, 2018.

[13] Andrew W Lo. The adaptive markets hypothesis: Market efficiency from an evolutionary perspective. *Journal of Portfolio Management*, 30(5):15–29, 2004.

[14] Matteo Mazziotta and Adriano Pareto. Weighting in composite indices construction: The case of the mazziotta-pareto index. *Rivista Italiana di Economia Demografia e Statistica*, pages 17–26, 2022.

[15] Paresh Kumar Narayan and Susan Sharma. Oil price shocks and stock returns: Evidence from major oil-exporting and oil-importing countries. *Energy Economics*, 33(5):932–947, 2011.

[16] David M Q Nelson, Adriano C M Pereira, and Renato A de Oliveira. Stock market's price movement prediction with lstm neural networks. *2017 International Joint Conference on Neural Networks (IJCNN)*, pages 1419–1426, 2017.

[17] Jigar Patel, S Shah, P Thakkar, and K Kotecha. Predicting stock and stock price index movement using trend deterministic data preparation and machine learning techniques. *Expert Systems with Applications*, 42(1):259–268, 2015.

[18] Ming Qiu and Yu Song. Application of artificial neural network for the prediction of stock market returns: The case of the japanese stock market. *Chaos, Solitons & Fractals*, 85:1–7, 2016.

[19] Luisa Trombetta Burić. Correlation coefficients in capital market analyses. *International journal of innovative science and research technology*, 8(12):2269–2274, 2023.

[20] Panagiotis Tsinaslanidis and Dimitris Kugiumtzis. A comprehensive evaluation of trading strategies based on technical indicators. *Applied Economics*, 51(14):1494–1507, 2019.

[21] G Venkata Manish Reddy, Iswarya, Jitendra Kumar, and Dilip Kumar Choubey. Time series analysis of national stock exchange: A multivariate data science approach. In *Soft Computing for Problem Solving: Proceedings of the SocProS 2022*, pages 691–707. Springer, 2023.

[22] Dongliang Wang, Margaret K Formica, and Song Liu. Nonparametric interval estimators for the coefficient of variation. *The International Journal of Biostatistics*, 14(1):20170041, 2018.

[23] Junjie Wu. Cluster analysis and k-means clustering: an introduction. In *Advances in K-Means clustering: A data mining thinking*, pages 1–16. Springer, 2012.

[24] G Peter Zhang. Time series forecasting using a hybrid arima and neural network model. *Neurocomputing*, 50:159–175, 2003.