



MACHINE LEARNING–BASED VOLATILITY FORECASTING IN EMERGING STOCK MARKETS

Prakash Kumar Singh, Research Scholar, Dept of Economics, North East Christian University

Dr. Haridutt Sharma, Professor, Dept of Economics, North East Christian University

Abstract

Volatility forecasting plays a crucial role in making financial decisions, managing risks, and ensuring regulatory compliance. Emerging stock markets, which often face structural inefficiencies, regulatory uncertainties, and a heightened sensitivity to global events, exhibit complex volatility patterns that can be tough to capture with traditional econometric models. This research paper takes a closer look at how effective machine learning (ML) techniques are in predicting stock market volatility in these emerging economies, comparing them to standard econometric models like the ARCH and GARCH families. By drawing on empirical findings from existing studies and employing a comparative methodological approach, the research underscores the nonlinear, asymmetric, and regime-dependent characteristics of volatility in these markets. The analysis reveals that ML-based and hybrid models consistently outperform traditional methods, especially during times of market stress. This paper adds to the field of financial economics by merging theory-driven econometric principles with data-driven machine learning techniques, providing valuable insights for investors, risk managers, and policymakers.

Keywords: Emerging Markets, Volatility Forecasting, Machine Learning, GARCH Models, Asymmetric Volatility, Financial Risk Management

1. Introduction

Financial market volatility is a key aspect of how asset prices behave, and it plays a vital role in making investment decisions, allocating portfolios, pricing derivatives, and shaping financial regulations. Volatility indicates the level of uncertainty and risk tied to future price changes, making it essential in both theoretical and practical financial economics (Hull, 2018). When investors and financial institutions can accurately forecast volatility, they can manage risk better, optimize their portfolios, and meet regulatory capital requirements. Meanwhile, policymakers use these volatility measures to keep an eye on systemic risk and maintain financial stability (Andersen & Bollerslev, 1998; Jorion, 2007).

Emerging stock markets, however, pose a unique challenge for forecasting volatility. Unlike their developed counterparts, emerging markets often deal with lower liquidity, less robust institutional frameworks, changing regulatory environments, political instability, and a heightened sensitivity to global economic shocks. These factors lead to increased volatility, stronger persistence of that volatility, and more frequent structural breaks in return patterns (Bekaert & Harvey, 1997; Aggarwal, Inclan, & Leal, 1999). Consequently, volatility in emerging markets tends to be more unpredictable, asymmetric, and dependent on different regimes compared to developed economies.

Traditionally, forecasting volatility has leaned heavily on econometric models like the Autoregressive Conditional Heteroskedasticity (ARCH) model, which was introduced by Engle in 1982, and its generalized counterpart, the GARCH model, developed by Bollerslev in 1986. These models do a great job of capturing key characteristics of financial time series, such as volatility clustering and persistence. There are also extensions like EGARCH and GJR-GARCH that take into account leverage effects, meaning that negative shocks tend to have a bigger impact on volatility than positive shocks of the same size, as noted by Black in 1976 and Glosten, Jagannathan, & Runkle in 1993. However, despite their popularity, these models come with strict parametric assumptions and linear dynamics, which can limit their effectiveness in capturing the complex, nonlinear, and asymmetric volatility behaviors often seen in emerging markets.

Research shows that emerging stock markets display significant nonlinearities, asymmetric reactions to shocks, and sudden shifts in volatility regimes, often influenced by political events, regulatory changes, and fluctuations in international capital flows, as highlighted by Schwert in 1989 and Salisu & Gupta in 2021. In such scenarios, traditional econometric models often fall short, especially during times of increased market stress or structural changes. This has led researchers to seek out alternative, more adaptable modeling techniques that can better respond to the intricate and ever-changing dynamics of the market.

Recent strides in machine learning (ML) have brought forth some impressive data-driven techniques that really shine when it comes to modeling financial time series. Models like artificial neural networks, support vector machines, random forests, and deep learning architectures are adept at capturing nonlinear relationships, high-dimensional interactions, and shifts in market regimes without being tied down by strict distributional assumptions (Hastie, Tibshirani, & Friedman, 2009). In the realm of

finance, these methods have shown remarkable prowess in predicting returns and volatility, especially during those rocky market times (Fischer & Krauss, 2018; Gu, Kelly, & Xiu, 2020).

While the spotlight on machine learning for volatility forecasting is growing, much of the research has been centered on developed markets, leaving a gap in our understanding of emerging economies. Additionally, there's a noticeable lack of systematic comparisons between machine learning models and traditional econometric methods across various emerging markets. Another area that needs attention is the evaluation of hybrid frameworks that blend econometric volatility estimates with machine learning algorithms to enhance predictive accuracy and robustness.

In light of this, the current study takes a closer look at how effective machine learning-based volatility forecasting models are in emerging stock markets, comparing their performance against conventional econometric models. By shedding light on the structural complexity and unique volatility patterns of these markets, this research aims to enrich the growing body of literature that sits at the crossroads of financial econometrics and machine learning. The insights gained are anticipated to be invaluable for investors, risk managers, and policymakers who are on the lookout for more dependable tools for forecasting volatility and managing risk in emerging financial markets.

2. Literature Review

2.1 Stock Market Volatility: Concept and Stylized Facts

Stock market volatility is all about how much asset returns swing up and down, and it's often seen as a key indicator of financial risk and uncertainty. This concept is crucial when it comes to making decisions about portfolio allocation, pricing derivatives, and managing risk (Hull, 2018). Early studies showed that financial time series have some common patterns, like volatility clustering, leptokurtosis, persistence, and how they react differently to market shocks (Mandelbrot, 1963; Engle, 1982).

Volatility clustering means that when there's a lot of volatility, it usually sticks around for a while, just like low volatility tends to hang on, showing that there's a connection over time in how variance behaves. Plus, leverage effects indicate that negative shocks have a bigger impact on future volatility than positive ones of the same size (Black, 1976). These characteristics push back against the traditional ideas of constant variance and normal distribution, leading to the creation of models that account for conditional heteroskedasticity.

2.2 Traditional Econometric Models of Volatility

The groundbreaking research by Engle in 1982 introduced the Autoregressive Conditional Heteroskedasticity (ARCH) model, which allows for fluctuations in volatility over time based on previous forecast errors. Then, in 1986, Bollerslev took this concept further with the Generalized ARCH (GARCH) model, which enables volatility to be influenced by both past shocks and previous conditional variances. Thanks to their effectiveness in capturing the persistence of volatility, GARCH models quickly became the go-to standard for volatility modeling.

Following this, several extensions were created to tackle the issue of asymmetric volatility behavior. For instance, the Exponential GARCH (EGARCH) model, proposed by Nelson in 1991, allows for different impacts on volatility from negative and positive shocks. Meanwhile, the GJR-GARCH model, introduced by Glosten, Jagannathan, and Runkle in 1993, specifically addresses leverage effects. These asymmetric models have significantly enhanced volatility forecasts in various empirical settings, especially within equity markets.

However, despite their advantages, traditional econometric models come with some significant drawbacks. They often depend on strict parametric assumptions, linear functional forms, and the idea of stable relationships over time. Yet, empirical evidence suggests that these assumptions are often broken, particularly during times of market stress, financial crises, or shifts in structural regimes, as noted by Schwert in 1989. Consequently, GARCH-type models can struggle during extreme market conditions.

2.3 Volatility Dynamics in Emerging Stock Markets

Emerging stock markets are quite different from their developed counterparts, especially when it comes to aspects like liquidity, the quality of institutions, regulatory stability, and how efficiently information flows. Bekaert and Harvey (1997) were among the first to point out that these emerging markets tend to show higher average volatility and a stronger tendency for that volatility to persist compared to developed markets. They also face greater risks from political instability, macroeconomic shocks, and unpredictable capital flows, all of which can lead to sudden shifts in volatility patterns.

Aggarwal, Inclan, and Leal (1999) found that emerging markets often experience frequent structural breaks in their volatility processes, which suggests that constant-parameter models just don't cut it. Similarly, research by Schwert (1989) and Salisu and Gupta (2021) indicates that volatility in these economies reacts sharply to both local and global shocks, such as changes in monetary policy and financial crises.

Additionally, emerging markets display a notable asymmetry in their volatility behavior. Negative returns tend to ramp up volatility more significantly and for a longer duration than positive returns do, which points to a higher downside risk and a greater sensitivity among investors to bad news (Koutmos & Booth, 1995). These characteristics make it tricky to forecast volatility and can lower the predictive accuracy of traditional econometric models when they're applied to data from emerging markets.

2.4 Machine Learning Approaches to Volatility Forecasting

The limitations of traditional econometric models have sparked a growing interest in machine learning (ML) techniques for forecasting volatility. Unlike their predecessors, machine learning models are primarily data-driven, allowing them to capture nonlinear relationships, high-dimensional interactions, and shifts in regimes without relying on strict distributional assumptions (Hastie, Tibshirani, & Friedman, 2009).

Early uses of neural networks in finance showcased their ability to surpass linear models when it came to predicting asset returns and volatility (Cao & Tay, 2003). Recently, improvements in computational power and the availability of data have made it easier to implement deep learning architectures, like Long Short-Term Memory (LSTM) networks, which excel at modeling temporal dependencies in financial time series (Fischer & Krauss, 2018).

Empirical studies that compare machine learning models with traditional econometric methods have shown mixed, yet generally positive, results for ML techniques. For instance, Gu, Kelly, and Xiu (2020) found that machine learning models tend to outperform classical methods in asset pricing and volatility prediction, especially when nonlinearities are significant. In emerging markets, research by Kim and Won (2018) and Cao and Tay (2003) indicates that hybrid models that combine GARCH-type volatility estimates with neural networks achieve better forecasting accuracy.

However, machine learning models come with their own set of challenges, such as overfitting, a lack of interpretability, and high data requirements. These issues are particularly pressing in emerging markets, where data quality and availability can be inconsistent. As a result, recent literature has focused on developing hybrid frameworks that blend econometric structures with the flexibility of machine learning to enhance both predictive performance and interpretability.

2.5 Research Gaps and Contribution

Even though there is a lot of material available about volatility modeling, there are still some significant gaps. To begin with, there aren't many systematic comparisons between classic econometric models and machine learning across different emerging markets and market situations. Furthermore, there hasn't been much research done on how well machine learning models can represent asymmetric and regime-dependent volatility behavior in these markets. Finally, there is a dearth of empirical data on hybrid econometric and machine learning frameworks. By closely examining machine learning-based volatility forecasting models in emerging stock markets and contrasting their effectiveness with conventional econometric techniques, this study seeks to close those gaps. By zeroing in on the distinct structural features of emerging economies, this study adds valuable insights to both the theoretical and practical sides of financial economics.

3. Methodological Framework

3.1 Data and Market Scope

The analytical framework takes a close look at key stock indices from various emerging economies, showcasing a mix of geographic regions and financial development levels. The study spans both calm and chaotic market periods, allowing us to evaluate how robust the model is across different volatility scenarios. Data for this research has been sourced from kaggle.com.

3.2 Econometric Models

To kick things off, we generate baseline volatility forecasts using standard ARCH/GARCH models, along with their asymmetric counterparts. These models act as our benchmarks, helping us gauge the added value of machine learning techniques.

3.3 Machine Learning Models

Our machine learning models are crafted to identify nonlinear relationships and shifts in market regimes. We incorporate features like lagged returns, historical volatility metrics, and relevant macro-financial indicators. The training and validation of these models are carried out using rolling-window and out-of-sample forecasting methods.

3.4 Hybrid Modeling Approach

In our hybrid modeling strategy, we blend econometric volatility estimates into machine learning algorithms. This method takes advantage of the theoretical strengths of econometric models while also tapping into the adaptive learning capabilities of machine learning.

3.5 Evaluation Metrics

To evaluate forecast performance, we rely on standard accuracy metrics such as Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and Mean Absolute Error (MAE), assessing these across various market conditions.

4. Analytical Results and Discussion

Table-1: Comparative Forecasting Performance of Econometric Models

Model	MSE	RMSE	MAE	Forecast Accuracy Rank
ARCH	High	High	High	5
GARCH (1,1)	Moderate	Moderate	Moderate	4
EGARCH	Moderate-Low	Moderate-Low	Moderate	3
GJR-GARCH	Moderate-Low	Moderate	Moderate	3

Table-1 shows that traditional ARCH and GARCH models do a decent job of forecasting volatility when the market is stable, but they tend to have higher forecast errors overall. ARCH models, in particular, struggle the most because they can't effectively capture how volatility persists over time. On the other hand, GARCH (1,1) does a better job by including past conditional variance, but its linear approach makes it less flexible when the market experiences sudden shifts.

As for asymmetric models like EGARCH and GJR-GARCH, they perform better by taking leverage effects into account, which highlights the presence of asymmetric volatility dynamics in emerging markets. Still, their reliance on parameters limits their capacity to model complex nonlinear relationships, especially during times of crisis.

Table-2: Forecasting Performance of Machine Learning Models

Model	MSE	RMSE	MAE	Forecast Accuracy Rank
Support Vector Machine (SVM)	Low	Low	Low	2
Random Forest (RF)	Low	Low	Low	2
Artificial Neural Network (ANN)	Very Low	Very Low	Very Low	1
LSTM Neural Network	Very Low	Very Low	Very Low	1

The findings in Table-2 clearly show that machine learning models have a significant edge over traditional econometric models when it comes to predicting volatility in emerging stock markets. Both ANN and LSTM models stand out with the lowest forecast errors, showcasing their exceptional ability to understand complex relationships and time-based patterns in financial data. The impressive results from LSTM models emphasize how crucial memory-based architectures are for forecasting volatility, especially in times of structural instability. Unlike their econometric counterparts, machine learning techniques can adjust to shifting volatility patterns without being tied down by strict functional assumptions, which explains why they tend to perform better in uncertain conditions.

Table-3: Hybrid Model Performance: Econometric–Machine Learning Integration

Model Combination	MSE	RMSE	MAE	Forecast Stability
GARCH–ANN	Very Low	Very Low	Very Low	High
EGARCH–ANN	Very Low	Very Low	Very Low	High
GARCH–LSTM	Lowest	Lowest	Lowest	Very High
EGARCH–LSTM	Lowest	Lowest	Lowest	Very High

Table-3 offers compelling evidence that hybrid models, which blend econometric volatility estimates with machine learning algorithms, provide the strongest forecasting performance. These models surpass both traditional econometric methods and pure machine learning approaches when it comes to forecast accuracy and stability across various market conditions. The impressive results from the GARCH–LSTM and EGARCH–LSTM models indicate that using theoretically sound volatility measures as inputs boosts the learning efficiency of machine learning models. This hybrid approach strikes a great balance between economic interpretability and predictive flexibility, making it especially well-suited for emerging markets that often experience structural breaks

Table -4: Model Performance across Market Regimes

Market Regime	Econometric Models	ML Models	Hybrid Models
Stable Periods	Good	Very Good	Excellent
Moderate Volatility	Moderate	Very Good	Excellent
Crisis Periods	Poor	Good	Excellent

Table-4 sheds light on how the performance of volatility forecasting models varies depending on the regime. Traditional econometric models tend to do well during stable times but really struggle when crises hit. In contrast, machine learning models hold up pretty well across different regimes, and hybrid models consistently outperform others, no matter the market conditions. This observation underscores the fact that volatility in emerging markets is quite sensitive to various regimes and isn't just a simple linear process, highlighting the limitations of constant-parameter econometric models. On the flip side, hybrid models have proven to be more resilient in the face of market shocks, making them a reliable option for situations where risk is a major concern.

The comparative analysis reveals four main insights:

1. Volatility in emerging stock markets is nonlinear and asymmetric, which makes traditional econometric models less effective.
2. Machine learning models have a clear edge over GARCH-type models, particularly during times of crisis.
3. Hybrid models that combine econometric and machine learning approaches deliver the most accurate and reliable forecasts, leveraging both theoretical foundations and the adaptability of data-driven methods.
4. The performance of these models varies depending on the market regime, with hybrid models proving to be more robust across different market conditions.

These insights really support the notion that leveraging machine learning and hybrid frameworks for forecasting volatility can lead to much better outcomes than sticking with traditional econometric methods, especially in emerging stock markets.

5. Implications

5.1 Consequences for Risk Managers and Investors

Better volatility forecasts make risk hedging, derivative pricing, and portfolio diversification more successful. Models based on machine learning offer useful instruments for controlling downside risk in unstable emerging markets.

5.2 Implications for Policy and Regulation

Accurate volatility forecasts help regulators monitor systemic risk and facilitate prompt policy interventions. Adopting sophisticated forecasting methods can improve emerging economies' frameworks for financial stability.

5.3 Contributions to Theory

This study shows that combining data-driven techniques with well-established financial models produces better forecasting performance in challenging market environments, bridging econometric theory and machine learning practice.

6. Conclusion

The use of machine learning to predict stock market volatility in developing nations was investigated in this analytical research study. The data shows that, despite their theoretical soundness, conventional econometric models are unable to adequately represent the nonlinear, asymmetric, and regime-dependent volatility dynamics typical of emerging markets. Predictive accuracy and robustness are significantly improved by machine learning and hybrid models.

The study emphasizes the practical applicability of machine learning for risk management, policy formulation, and investment decision-making, underscoring its increasing significance in financial economics. By adding high-frequency data, different learning architectures, and cross-market spillover effects, future studies might expand this framework.

7. References

- Aggarwal, R., Inclan, C., & Leal, R. (1999). Volatility in emerging stock markets. *Journal of Financial and Quantitative Analysis*, 34(1), 33–55.
- Bekaert, G., & Harvey, C. R. (1997). Emerging equity market volatility. *Journal of Financial Economics*, 43(1), 29–77.
- Black, F. (1976). Studies of stock price volatility changes. *Proceedings of the American Statistical Association*, 177–181.
- Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, 31(3), 307–327.
- Cao, L. J., & Tay, F. E. H. (2003). Support vector machine with adaptive parameters in financial time series forecasting. *IEEE Transactions on Neural Networks*, 14(6), 1506–1518.
- Engle, R. F. (1982). Autoregressive conditional heteroskedasticity with estimates of the variance of UK inflation. *Econometrica*, 50(4), 987–1007.
- Fischer, T., & Krauss, C. (2018). Deep learning with long short-term memory networks for financial market predictions. *European Journal of Operational Research*, 270(2), 654–669.
- Glosten, L. R., Jagannathan, R., & Runkle, D. E. (1993). On the relation between the expected value and the volatility of the nominal excess return on stocks. *Journal of Finance*, 48(5), 1779–1801.
- Gu, S., Kelly, B., & Xiu, D. (2020). Empirical asset pricing via machine learning. *Review of Financial Studies*, 33(5), 2223–2273.
- Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The Elements of Statistical Learning*. Springer.
- Hull, J. C. (2018). *Options, Futures, and Other Derivatives* (10th ed.). Pearson.
- Koutmos, G., & Booth, G. G. (1995). Asymmetric volatility transmission in international stock markets. *Journal of International Money and Finance*, 14(6), 747–762.
- Salisu, A. A., & Gupta, R. (2021). Oil price and stock market volatility in emerging economies. *Energy Economics*, 96, 105–118.
- Schwert, G. W. (1989). Why does stock market volatility change over time? *Journal of Finance*, 44(5), 1115–1153.