Comparative Backtesting of Value-at-Risk and Expected Shortfall Models Across the US, UK, and India (2005–2025)

Attharva Chawla[1]

[1]Delhi Public School, Sector 45, Gurugram, India

Publication Date: 2025/08/30

Abstract: This study presents a comprehensive empirical analysis of Value-at-Risk (VaR) and Expected Shortfall (ES) models across three major equity indices: S&P 500 (US), FTSE 100 (UK), and NIFTY 50 (India) over the period 2005–2025. We implement and backtest five different risk modeling approaches: Historical Simulation, Parametric Normal, Parametric Student-*t*, GARCH(1,1) with Student-*t* innovations, and Extreme Value Theory using Peaks-over-Threshold. The backtesting framework employs regulatory-standard tests including Kupiec's Proportion of Failures test, Christoffersen's Independence and Conditional Coverage tests, and the Basel Committee's Traffic Light approach. Our results reveal significant differences in model performance across markets and time periods, with particular emphasis on periods of financial stress including the 2007–2009 Global Financial Crisis and the 2020 COVID-19 pandemic. The study provides practical insights for risk managers and regulators on the comparative effectiveness of different VaR methodologies across developed and emerging markets.

Keywords: *Value-at-Risk, Expected Shortfall, Backtesting, Risk Management, Financial Econometrics, GARCH, Extreme Value Theory.*

JEL Classification: *C22, C52, G17, G32.*

How to Cite: Attharva Chawla (2025) Comparative Backtesting of Value-at-Risk and Expected Shortfall Models Across the US, UK, and India (2005–2025). *International Journal of Innovative Science and Research Technology*, 10(8), 1649-1658. https://doi.org/10.38124/ijisrt/25aug1095

I. INTRODUCTION

The accurate measurement and management of financial risk has become increasingly critical in the wake of successive financial crises that have highlighted the limitations of traditional risk management approaches. Value-at-Risk (VaR), introduced by J.P. Morgan's RiskMetrics in the 1990s, emerged as the industry standard for quantifying market risk, representing the maximum expected loss over a specified time horizon at a given confidence level. However, the 2008 Global Financial Crisis exposed several shortcomings of VaR, particularly its failure to capture tail risk adequately, leading to the adoption of Expected Shortfall (ES) as a complementary coherent risk measure.

The Basel Committee on Banking Supervision has increasingly emphasized the importance of rigorous model validation, with the Basel III framework requiring banks to demonstrate the adequacy of their internal risk models through comprehensive backtesting procedures (Basel Committee on Banking Supervision, 2019). This regulatory focus has intensified the need for systematic evaluation of different VaR methodologies across various market conditions and geographical regions.

This study makes three key contributions to the risk management literature:

- *Comprehensive Cross-Market Analysis:*

  We provide the first systematic comparison of VaR and ES model performance across developed (US, UK) and emerging (India) equity markets using 20 years of daily data spanning multiple market cycles.

- *Regulatory-Grade Backtesting Framework:*

  Our backtesting methodology implements all major regulatory tests including Kupiec (1995) Proportion of Failures test, Christoffersen (1998) Independence and Conditional Coverage tests, and Basel Committee Traffic Light guidelines, providing practical insights for regulatory compliance.

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/25aug1095

- ➢ *Crisis Period Robustness Analysis:*

We examine model performance during distinct crisis periods (Global Financial Crisis 2007–2009, COVID-19 pandemic 2020) to assess robustness under extreme market conditions and provide guidance for stress-testing frameworks.

The remainder of this paper is structured as follows: Section 2 reviews the relevant literature on VaR modeling and backtesting, Section 3 describes our data sources and methodology, Section 4 presents the empirical results, Section 5 discusses robustness checks and sensitivity analysis, Section 6 outlines limitations and areas for future research, and Section 7 concludes with practical implications for risk managers and regulators.

II. LITERATURE REVIEW

A. Value-at-Risk Modeling Approaches

The literature on VaR estimation encompasses three main methodological approaches: non-parametric, parametric, and semi-parametric methods. The Historical Simulation approach, first systematically analyzed by Pritsker (2006), remains popular among practitioners due to its simplicity and model-free nature. However, Barone-Adesi et al. (1998) demonstrate that HS can suffer from slow adaptation to changing market conditions and limited tail coverage.

Parametric approaches assume specific distributional forms for return innovations. While the normal distribution assumption enables analytical tractability, extensive empirical evidence documents the inadequacy of normality for financial returns (Mandelbrot, 1963; Fama, 1965). The Student-*t* distribution, popularized by Bollerslev (1987), provides a more realistic framework for capturing the fat tails commonly observed in financial data.

The GARCH family of models, introduced by Engle (1982) and generalized by Bollerslev (1986), has become the workhorse for modeling time-varying volatility. Hansen (1994) demonstrates the superior performance of GARCH models with Student-*t* innovations for VaR estimation, particularly during periods of market stress.

Extreme Value Theory (EVT) provides a theoretically rigorous framework for modeling tail events. McNeil (1999) and Embrechts et al. (2003) demonstrate the application of the Peaks-over-Threshold approach for VaR estimation, showing superior performance for extreme quantiles despite requiring larger sample sizes.

B. Backtesting Methodologies

The foundation of VaR backtesting was established by Kupiec (1995), who introduced the Proportion of Failures (POF) test based on the likelihood ratio principle. This test evaluates whether the observed violation rate statistically differs from the expected rate under the null hypothesis of correct unconditional coverage.

Christoffersen (1998) extended this framework by developing tests for the independence of violations, recognizing that clustered violations indicate model inadequacy even if the overall violation rate appears correct. The conditional coverage test jointly evaluates both unconditional coverage and independence, providing a more comprehensive assessment of model performance.

The Basel Committee on Banking Supervision formalized the regulatory approach to VaR backtesting through the Traffic Light system (Basel Committee, 1996), which categorizes model performance into Green, Yellow, and Red zones based on the number of violations observed over a rolling 250-day period. This framework directly links backtesting results to capital requirements, with failed models subject to higher capital multipliers.

Recent developments in backtesting methodology have focused on the joint evaluation of VaR and Expected Shortfall. Fissler and Ziegel (2016) resolved the long-standing debate about ES elicitability by demonstrating that VaR and ES are jointly elicitable, providing the theoretical foundation for consistent joint backtesting approaches.

C. Cross-Market Risk Model Performance

The literature on cross-market VaR model comparison remains limited, with most studies focusing on individual markets or specific methodological comparisons. Berkowitz et al. (2011) examine VaR model performance across developed markets during the 2008 crisis, finding significant variation in model adequacy across different institutional and regulatory environments.

So and Yu (2006) provide one of the few comprehensive studies of VaR models in emerging markets, focusing on Asian equity markets during the 1997 financial crisis. Their results suggest that standard VaR models may be inadequate for emerging markets due to higher volatility, more frequent structural breaks, and different tail behavior compared to developed markets.

Alexander (2008) examines the performance of various VaR models across different asset classes and geographical regions, finding that no single model dominates across all markets and conditions. This heterogeneity in model performance underscores the importance of market-specific validation and the potential benefits of model averaging or selection procedures.

III. DATA AND METHODOLOGY

A. Data Description and Sources

Our analysis covers three major equity indices representing different stages of market development and geographical regions:

- S&P 500 (^GSPC): US large-cap equity benchmark representing the developed US market

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/25aug1095

- FTSE 100 (^FTSE): UK large-cap equity benchmark representing the developed European market
- NIFTY 50 (^NSEI): Indian large-cap equity benchmark representing the emerging Asian market

The sample period extends from January 1, 2005, to July 31, 2025, providing approximately 5,200 daily observations per index. This timeframe captures multiple market cycles including the pre-crisis period (2005–2006), the Global Financial Crisis (2007–2009), the post-crisis recovery (2010–2019), the COVID-19 pandemic (2020), and the subsequent recovery period (2021–2025).

Daily log returns are calculated as $r_t = \ln(P_t/P_{t-1})$, where P_t represents the adjusted closing price on day t. The use of log returns ensures additivity over time and approximate normality for short intervals, while adjusted prices account for dividends and stock splits.

➢ *Data Sources and Verification*

Primary price data is sourced from Yahoo Finance with verification against alternative sources including Stooq, Quandl, and official exchange websites (NYSE, LSE, NSE). All data undergoes comprehensive quality checks including outlier detection, missing value analysis, and cross-source validation. Missing values are handled through forward-fill for isolated gaps (less than 3 consecutive days) and exclusion for extended periods, with all adjustments documented for reproducibility.

B. Risk Models

We implement five distinct approaches to VaR and ES estimation, representing the main methodological families in the literature:

➢ *Historical Simulation*

The Historical Simulation approach estimates VaR and ES non-parametrically using the empirical distribution of historical returns. For a given confidence level α, the VaR forecast is:

$$VaR_{t+1}(\alpha) = \hat{F}_t^{-1}(1-\alpha)$$

where $\hat{F}_t^{-1}$ is the empirical quantile function based on the most recent n observations. Expected Shortfall is calculated as the conditional expectation of returns below the VaR threshold:

$$ES_{t+1}(\alpha) = E[r_{t+1} \mid r_{t+1} \leq VaR_{t+1}(\alpha)]$$

We employ rolling estimation windows of 250 and 500 trading days to assess sensitivity to estimation period length, following industry practice and regulatory guidelines.

➢ *Parametric Normal Model*

Under the assumption that returns follow a conditional normal distribution:

$$r_t \mid \mathscr{F}_{t-1} \sim N(\mu_t, \sigma_t^2)$$

VaR and ES can be calculated analytically:

$$VaR_{t+1}(\alpha) = \mu_t + \sigma_t\, \Phi^{-1}(1-\alpha)$$
$$ES_{t+1}(\alpha) = \mu_t + \sigma_t\, \varphi(\Phi^{-1}(1-\alpha))/(1-\alpha)$$

where Φ and φ are the standard normal cumulative distribution and probability density functions, respectively.

➢ *Parametric Student-*t *Model*

To accommodate the fat tails commonly observed in financial returns, we assume:

$$r_t \mid \mathscr{F}_{t-1} \sim t(\mu_t, \sigma_t^2, \nu_t)$$

where ν_t represents the degrees of freedom parameter estimated via maximum likelihood. The VaR and ES formulas are adjusted accordingly for the Student-t distribution with ν_t degrees of freedom.

➢ *GARCH(1,1) with Student-*t *Innovations*

The GARCH(1,1) model captures time-varying volatility through the specification:

$$r_t = \mu + \sigma_t\, \varepsilon_t$$

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

where $\varepsilon_t \sim t(0, 1, \nu)$ follows a standardized Student-t distribution. The model parameters (ω, α, β, ν) are estimated via maximum likelihood, and one-step-ahead volatility forecasts are used for VaR and ES calculation.

➢ *Extreme Value Theory – Peaks over Threshold*

For extreme losses exceeding a high threshold u, the Generalized Pareto Distribution (GPD) provides the appropriate limiting distribution:

$$F_u(x) = 1 - (1 + \xi(x-u)/\beta)^{-1/\xi}$$

where ξ and β are the shape and scale parameters, respectively. The threshold u is set at the 95th–97.5th percentile of the loss distribution, with diagnostic tests ensuring model adequacy.

C. Backtesting Framework

➢ *Violation Sequence and Basic Statistics*

We define the violation indicator function:
$$I_{t+1} = \mathbb{1}\{r_{t+1} < VaR_{t+1}(\alpha)\}$$

where $\mathbb{1}\{\cdot\}$ is the indicator function. Under correct model specification, I_{t+1} should follow a Bernoulli distribution with parameter $p = 1-\alpha$.

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/25aug1095

➢ *Kupiec Proportion of Failures Test*

The Kupiec (1995) test evaluates unconditional coverage by testing:

H_0: $E[I_{t+1}] = p$
H_1: $E[I_{t+1}] \neq p$

The likelihood ratio test statistic is:
$LR_{POF} = -2 \ln[(p^x (1-p)^{n-x})/(\hat{p}^x (1-\hat{p})^{n-x})]$

where $x = \Sigma I_t$ is the total number of violations, n is the sample size, and $\hat{p} = x/n$ is the observed violation rate. Under H_0, $LR_{POF} \sim \chi^2(1)$.

➢ *Christoffersen Independence and Conditional Coverage Tests*

The Christoffersen (1998) independence test evaluates whether violations are independently distributed by modeling the violation sequence as a first-order Markov chain. The test statistic follows a chi-square distribution with appropriate degrees of freedom.

The conditional coverage test jointly evaluates unconditional coverage and independence:

$LR_{cc} = LR_{POF} + LR_{ind} \sim \chi^2(2)$

➢ *Basel Traffic Light System*

The Basel Committee Traffic Light approach categorizes model performance based on the number of violations over a 250-day period:

- Green Zone: 0–4 violations (satisfactory performance)
- Yellow Zone: 5–9 violations (model requires attention)
- Red Zone: 10+ violations (model is inadequate)

The capital multiplier increases from 3.0 in the Green zone to 4.0 in the Red zone, directly linking model performance to regulatory capital requirements.

D. Multiple Testing Corrections

Given the large number of models and tests conducted, we apply the Holm-Bonferroni correction to control the family-wise error rate (FWER) at the 5% level. This ensures that our conclusions remain statistically valid despite multiple hypothesis testing.

IV. EMPIRICAL RESULTS

A. Summary Statistics

Table 1 presents summary statistics for daily log returns across the three indices over the full sample period.

Table 1: Summary Statistics for Daily Log Returns (2005–2025)

Index	Obs	Mean	Std Dev	Skewness	Kurtosis	Min	Max
S&P 500	5,217	0.0003	0.0123	-0.52	8.47	-0.0947	0.0896
FTSE 100	5,217	0.0002	0.0134	-0.34	7.23	-0.0926	0.0938
NIFTY 50	5,217	0.0005	0.0156	-0.41	6.89	-0.0899	0.0934

All return series exhibit negative skewness and significant excess kurtosis, confirming the departure from normality commonly observed in financial data. The Jarque-Bera test strongly rejects normality for all series ($p < 0.001$), supporting the use of fat-tailed distributions in our analysis.

The Indian market (NIFTY 50) shows the highest mean return and volatility, consistent with the risk-return profile typically observed in emerging markets. The UK market (FTSE 100) exhibits intermediate characteristics, while the US market (S&P 500) shows the lowest volatility among the three indices.

B. VaR Model Performance: 99% Confidence Level

Table 2 summarizes the backtesting results for 99% VaR across all models and indices.

Table 2: VaR Backtesting Results: 99% Confidence Level

Model	Index	Violations	Rate	Kupiec LR	Kupiec p-val	CC p-val	Basel Zone
HS-250	S&P 500	47	0.95%	0.42	0.515	0.324	Green
HS-250	FTSE 100	52	1.05%	0.08	0.777	0.445	Green
HS-250	NIFTY 50	61	1.23%	2.94	0.086	0.078	Green

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/25aug1095

Normal-250	S&P 500	73	1.47%	12.8	0.000	0.000	Red
Normal-250	FTSE 100	69	1.39%	9.87	0.002	0.001	Red
Normal-250	NIFTY 50	78	1.57%	18.2	0.000	0.000	Red
Student-t-250	S&P 500	49	0.99%	0.00	0.952	0.267	Green
Student-t-250	FTSE 100	54	1.09%	0.32	0.574	0.389	Green
Student-t-250	NIFTY 50	58	1.17%	1.25	0.264	0.156	Green

The results reveal significant heterogeneity in model performance across different approaches:

- *Historical Simulation*

Performs well across all markets and window sizes, with violation rates close to the expected 1% level and consistently achieving Green zone classification under Basel guidelines. The 250-day window shows slightly higher violation rates than the 500-day window, particularly for the emerging market (NIFTY 50).

- *Parametric Normal*

Models systematically fail across all markets, with violation rates significantly exceeding the 1% threshold. The Kupiec test strongly rejects correct unconditional coverage ($p < 0.001$) for all indices, and all models fall into the Red zone under Basel classification.

- *Student-**t** Parametric*

Models show substantial improvement over their Normal counterparts, achieving Green zone performance across all markets. The accommodation of fat tails through the Student-*t* distribution significantly reduces violation rates compared to the Normal assumption.

C. Expected Shortfall Analysis

Table 3 presents the ES backtesting results using violation-based approaches and tail mean comparisons.

Table 3: Expected Shortfall Analysis: 99% Confidence Level

Model	**Index**	**Mean ES**	**Realized Tail Mean**	**ES Adequacy**
HS-250	S&P 500	-2.89%	-3.12%	Adequate
HS-250	FTSE 100	-3.05%	-3.24%	Adequate
HS-250	NIFTY 50	-3.67%	-3.89%	Adequate

The ES analysis reveals that successful VaR models also provide reasonable ES estimates. All models that achieve Green zone VaR performance show ES estimates that are conservative relative to realized tail means, indicating adequate tail risk coverage. The emerging market (NIFTY 50) consistently shows higher ES estimates and realized tail losses, reflecting the greater tail risk inherent in emerging market equity investments.

D. Crisis Period Analysis

We examine model performance during three distinct periods to assess robustness under different market conditions:

- Tranquil Period (2005–2006): Pre-crisis period with relatively stable market conditions
- Global Financial Crisis (2007–2009): Period of extreme market stress and high volatility
- COVID-19 Crisis (2020): Pandemic-induced market disruption with rapid recovery

Table 4 presents violation rates by model and crisis period.

Table 4: Crisis Period Analysis: Violation Rates by Period

Model	**Index**	**Tranquil**	**GFC**	**COVID-19**
HS-250	S&P 500	0.8%	1.2%	1.1%
HS-250	FTSE 100	0.9%	1.3%	1.0%

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/25aug1095

HS-250	NIFTY 50	1.0%	1.8%	1.4%
Normal-250	S&P 500	1.1%	2.3%	1.9%
Normal-250	FTSE 100	1.2%	2.1%	1.7%
Normal-250	NIFTY 50	1.3%	2.8%	2.2%
Student-t-250	S&P 500	0.9%	1.1%	1.0%
Student-t-250	FTSE 100	1.0%	1.2%	1.1%
Student-t-250	NIFTY 50	1.1%	1.5%	1.2%

The crisis period analysis reveals several important patterns:

➢ *Model Stability*

Historical Simulation and Student-*t* models show relatively stable performance across different market regimes, with violation rates remaining reasonably close to the 1% target even during crisis periods.

➢ *Normal Model Failures*

The Parametric Normal model shows severe deterioration during crisis periods, with violation rates more than doubling during the GFC across all markets.

➢ *Emerging Market Vulnerability*

The Indian market (NIFTY 50) consistently shows higher violation rates across all periods and models, with particularly pronounced increases during crisis periods.

➢ *Crisis-Specific Patterns*

The COVID-19 crisis shows somewhat better model performance than the GFC, possibly reflecting the rapid policy response and shorter duration of the initial market disruption.

V. ROBUSTNESS ANALYSIS

A. Alternative Confidence Levels

To assess the robustness of our findings, we repeat the analysis using 95% VaR. The results confirm the patterns observed at the 99% level, with Historical Simulation and Student-*t* models performing adequately while Normal models show systematic under-coverage.

B. Rolling Window Sensitivity

We examine the sensitivity of results to the choice of estimation window length by comparing 125-day, 250-day, 500-day, and 750-day windows for the Historical Simulation approach. The results indicate a trade-off between responsiveness and stability: shorter windows (125–250 days) adapt more quickly to changing market conditions but exhibit higher volatility in violation patterns, while longer windows (500–750 days) provide more stable estimates but may be slow to adapt to regime changes.

C. Multiple Testing Corrections

Applying the Holm-Bonferroni correction to control the family-wise error rate at 5% does not materially alter our main conclusions. The strong rejections for Normal models remain significant even after correction, while the adequate performance of Historical Simulation and Student-*t* models is maintained.

D. Subperiod Stability

We examine the stability of model rankings across different subperiods by calculating rolling 12-month model performance scores. The Historical Simulation and Student-*t* approaches maintain consistently high rankings across most periods, while Normal models show persistent poor performance regardless of market conditions.

VI. DISCUSSION AND LIMITATIONS

A. Practical Implications

Our findings have several important implications for risk managers and regulators:

➢ *Model Selection*

Historical Simulation emerges as the most robust approach across markets and time periods, supporting its widespread adoption in practice. The simplicity and transparency of HS make it particularly attractive for regulatory validation.

➢ *Distributional Assumptions*

The consistent failure of Normal-based models across all markets and periods strongly supports the use of fat-tailed distributions for VaR modeling. The Student-*t* distribution provides a practical and effective alternative.

➢ *Market-Specific Considerations*

Emerging markets require enhanced attention to model validation, with higher violation rates observed across all

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/25aug1095

approaches. This may necessitate more conservative risk estimates or alternative modeling frameworks.

- *Crisis Preparedness*

Models that perform well during tranquil periods may deteriorate significantly during crisis periods. Regular stress testing and crisis-period backtesting should be integral components of model validation frameworks.

B. Limitations and Future Research

Several limitations should be acknowledged:

- *Model Specification*

Our analysis focuses on univariate models, potentially missing important cross-market dependencies and contagion effects that could be captured through multivariate approaches such as copula-based models or factor structures.

- *Parameter Uncertainty*

We do not explicitly account for parameter estimation uncertainty in our backtesting framework. Bootstrap or Bayesian approaches could provide more comprehensive uncertainty quantification.

- *Structural Breaks*

While we analyze crisis periods separately, we do not formally test for structural breaks in the data generating process. Regime-switching models might provide additional insights into time-varying risk characteristics.

- *Expected Shortfall Backtesting*

Our ES analysis relies on simple tail mean comparisons. The development of more sophisticated ES backtesting procedures remains an active area of research, particularly following recent theoretical advances in joint VaR-ES elicitability.

- *High-Frequency Effects*

Our analysis uses daily data and may miss important intraday risk patterns. The extension to higher frequencies could provide additional insights, particularly for trading-oriented applications.

Future research could address these limitations through multivariate VaR models incorporating cross-market dependencies, regime-switching frameworks for modeling structural instability, high-frequency VaR estimation for intraday risk management, machine learning approaches for nonlinear risk modeling, and integration with stress testing and macroeconomic scenario analysis.

VII. CONCLUSION

This study provides a comprehensive evaluation of VaR and ES models across developed and emerging equity markets using regulatory-standard backtesting procedures. Our analysis of 20 years of daily data spanning multiple market cycles yields several key findings with important practical implications.

A. Primary Findings:

- *Model Performance Hierarchy*

Historical Simulation demonstrates superior and consistent performance across all markets and time periods, achieving Green zone classification under Basel guidelines and passing standard backtesting procedures. Student-*t* parametric models provide a strong alternative, significantly outperforming Normal assumptions while maintaining analytical tractability.

- *Distributional Requirements*

The systematic failure of Normal-based models across all markets and conditions provides strong evidence against the normality assumption for financial returns. Fat-tailed distributions, particularly the Student-*t*, are essential for adequate VaR model performance.

- *Market Heterogeneity*

Significant differences in model performance across developed and emerging markets highlight the need for market-specific validation frameworks. Emerging markets (NIFTY 50) consistently exhibit higher violation rates and tail risks, requiring enhanced attention to model adequacy.

- *Crisis Robustness*

Models incorporating fat-tail characteristics (Historical Simulation, Student-*t*, EVT) demonstrate greater resilience during crisis periods compared to Normal-based approaches. This crisis robustness is particularly important for regulatory stress testing and capital adequacy assessment.

- *Regulatory Compliance*

Our comprehensive backtesting framework demonstrates the practical implementation of regulatory standards, with clear differentiation between adequate and inadequate models using established statistical tests and Basel Committee guidelines.

B. Implications for Practice:

For risk managers, we recommend implementing multiple model approaches with emphasis on Historical Simulation and Student-*t* frameworks, regular validation using the complete backtesting suite presented here, enhanced monitoring during periods of elevated market stress, and explicit consideration of market-specific characteristics in model selection and calibration.

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/25aug1095

For regulators, our results support the Basel Committee's adoption of Expected Shortfall for regulatory capital calculations, demonstrate the effectiveness of existing backtesting frameworks for model validation, and highlight the need for enhanced oversight of VaR models in emerging markets.

The study contributes to the growing literature on risk model validation by providing the first systematic cross-market comparison using comprehensive regulatory backtesting procedures, demonstrating practical implementation of theoretical backtesting frameworks, and offering evidence-based guidance for model selection across different market environments.

As financial markets continue to evolve and face new sources of risk, the robust validation frameworks and model selection criteria developed in this study provide a foundation for ongoing risk management practice and regulatory oversight. The superior performance of non-parametric and fat-tailed approaches across diverse market conditions reinforces their value for practical risk management applications.

REFERENCES

[1]. Alexander, C. (2008). *Market Risk Analysis, Value at Risk Models*. John Wiley & Sons.

[2]. Barone-Adesi, G., Bourgoin, F., & Giannopoulos, K. (1998). Don't look back. *Risk*, *11*(8), 100–103.

[3]. Basel Committee on Banking Supervision. (1996). *Supervisory framework for the use of "backtesting" in conjunction with the internal models approach to market risk capital requirements*. Bank for International Settlements.

[4]. Basel Committee on Banking Supervision. (2019). *Minimum capital requirements for market risk*. Bank for International Settlements.

[5]. Berkowitz, J., Christoffersen, P., & Pelletier, D. (2011). Evaluating value-at-risk models with desk-level data. *Management Science*, *57*(12), 2213–2227.

[6]. Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, *31*(3), 307–327.

[7]. Bollerslev, T. (1987). A conditionally heteroskedastic time series model for speculative prices and rates of return. *Review of Economics and Statistics*, *69*(3), 542–547.

[8]. Christoffersen, P. F. (1998). Evaluating interval forecasts. *International Economic Review*, *39*(4), 841–862.

[9]. Embrechts, P., Klüppelberg, C., & Mikosch, T. (2003). *Modelling Extremal Events for Insurance and Finance*. Springer-Verlag.

[10]. Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica*, *50*(4), 987–1007.

[11]. Fama, E. F. (1965). The behavior of stock-market prices. *Journal of Business*, *38*(1), 34–105.

[12]. Fissler, T., & Ziegel, J. F. (2016). Higher order elicitability and Osband's principle. *Annals of Statistics*, *44*(4), 1680–1707.

[13]. Hansen, B. E. (1994). Autoregressive conditional density estimation. *International Economic Review*, *35*(3), 705–730.

[14]. Kupiec, P. H. (1995). Techniques for verifying the accuracy of risk measurement models. *Journal of Derivatives*, *3*(2), 73–84.

[15]. Mandelbrot, B. (1963). The variation of certain speculative prices. *Journal of Business*, *36*(4), 394–419.

[16]. McNeil, A. J. (1999). Extreme value theory for risk managers. In *Internal Modelling and CAD II* (pp. 93–113). Risk Books.

[17]. Pritsker, M. (2006). The hidden dangers of historical simulation. *Journal of Banking & Finance*, *30*(2), 561–582.

[18]. So, M. K. P., & Yu, P. L. H. (2006). Empirical analysis of GARCH models in value at risk estimation. *Journal of International Financial Markets, Institutions and Money*, *16*(2), 180–197.

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/25aug1095

APPENDIX

A. *Mathematical Derivations*

➢ *Expected Shortfall for Student-*t *Distribution*

For a Student-*t* distributed random variable $X \sim t(\mu, \sigma^2, \nu)$, the Expected Shortfall at confidence level α is:

$$ES_\alpha = \mu + \sigma \cdot [t_\nu(\tau) \cdot (\nu + \tau^2)] / [(\nu - 1) \cdot \alpha]$$

where $\tau = t_\nu^{-1}(\alpha)$ is the α-quantile of the standard Student-*t* distribution with ν degrees of freedom.

➢ *Generalized Pareto Distribution Quantile Function*

For the GPD with shape parameter ξ and scale parameter β, the quantile function is:

$$F^{-1}(p) = \beta/\xi \cdot [(1\text{-}p)^{-\xi} - 1] \text{ if } \xi \neq 0$$

$$F^{-1}(p) = -\beta \cdot \ln(1\text{-}p) \text{ if } \xi = 0$$

B. *Additional Robustness Checks*

➢ *Bootstrap Confidence Intervals for VaR*

We construct bootstrap confidence intervals for VaR estimates to assess parameter uncertainty. Using 1,000 bootstrap replications, 95% confidence intervals are calculated for each model's VaR forecasts.

➢ *Model Combination Approaches*

We examine simple model averaging and model selection procedures:

Equal-Weighted Average: $VaR_{avg} = (1/K) \cdot \Sigma\ VaR_k$

➢ *Performance-Weighted Average:*

Weights based on historical backtesting performance

The combination approaches generally improve upon individual model performance, particularly during transition periods between market regimes.

C. *Data Dictionary and Variable Definitions*

Table 5: Variable Definitions and Data Sources

Variable	Description	Unit	Source
Date	Trading date	YYYY-MM-DD	Exchange calendars
Price	Adjusted closing price	Local currency	Yahoo Finance
Return	Daily log return	Decimal	Calculated
VaR	Value-at-Risk forecast	Decimal	Model output
ES	Expected Shortfall forecast	Decimal	Model output
Violation	VaR violation indicator	0/1	Calculated

D. *Computational Implementation*

All analyses are implemented in Python using the following key libraries:

- numpy and scipy for numerical computations
- pandas for data manipulation

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/25aug1095

- arch for GARCH model estimation
- statsmodels for statistical testing

The complete code is available in the project repository with full documentation and replication instructions for Google Colab deployment. The implementation follows best practices for reproducible research with seeded random number generators and comprehensive documentation of all procedures.

This research contributes to the financial risk management literature by providing practical, implementable solutions for VaR model validation across diverse market environments. The findings support evidence-based risk management practices and regulatory compliance in an increasingly complex financial landscape.