



Explainable Predictive Modeling for Value-at-Risk and Expected Shortfall: An Empirical Study of Investment Portfolios

Henry Sarpong

Morgan Stanley, Active and Affluent Trader Associate, USA

DOI : <https://doi.org/10.55248/gengpi.07.1026.3286>

ABSTRACT :

The increasing complexity of global financial markets, heightened asset price volatility, and evolving investment risk exposures have intensified the demand for accurate, transparent, and adaptive portfolio risk assessment frameworks. Traditional financial risk models, particularly Value-at-Risk (VaR) and Expected Shortfall (ES), remain fundamental to measuring potential investment losses but frequently encounter limitations when capturing nonlinear dependencies, extreme market events, and rapidly changing financial conditions. This study examines the application of explainable predictive modelling techniques to enhance the estimation and interpretation of VaR and ES across diversified investment portfolios. Particular emphasis is placed on econometric forecasting, machine learning algorithms, volatility modelling, and explainable artificial intelligence methods for identifying influential risk factors and predicting potential portfolio losses. The study further considers empirical evaluation through historical financial data, backtesting procedures, comparative model performance, and risk attribution analysis. Additionally, challenges involving model uncertainty, data quality, interpretability, and regulatory accountability are critically examined. Ultimately, the study highlights the potential of integrating predictive accuracy with transparent model explanations to strengthen portfolio risk management, support informed investment decisions, and improve financial resilience under uncertain market conditions.

Keywords: Value-at-Risk; Expected Shortfall; Explainable Artificial Intelligence; Predictive Modelling; Portfolio Risk Management; Financial Market Volatility.

1. RESEARCH CONTEXT, FINANCIAL RISK PROBLEM, AND STUDY CONTRIBUTIONS

1.1. The Changing Nature of Investment Portfolio Tail Risk

Contemporary investment portfolios operate within increasingly interconnected financial markets, where volatility clustering, contagion effects, and dynamic asset dependencies amplify exposure to extreme losses [1,2]. Periods of financial instability frequently exhibit abrupt changes in correlations, weakening diversification benefits and increasing systemic vulnerability. Conventional risk assessment frameworks often assume relatively stable return distributions, despite empirical evidence of skewness, excess kurtosis, and nonlinear market behaviour [3]. Consequently, normally distributed loss assumptions may underestimate the probability and magnitude of extreme adverse outcomes. Financial crises and sudden liquidity disruptions further demonstrate that historical relationships cannot reliably represent future market conditions. These challenges underscore the importance of adaptive forecasting frameworks capable of capturing time-varying volatility, asymmetric risk transmission, and evolving dependence structures. Investment managers therefore require forward-looking analytical approaches that quantify changing downside exposures while supporting robust portfolio allocation and risk mitigation decisions [4].

1.2. The VaR–ES Measurement Problem

Value-at-Risk (VaR) and Expected Shortfall (ES) represent complementary measures of portfolio downside exposure. VaR identifies the loss threshold associated with a specified confidence level, whereas ES quantifies the average severity of losses within the corresponding distributional tail [5]. Although widely applied, VaR provides limited information about losses exceeding its estimated threshold and may inadequately reflect extreme market deterioration. ES addresses this limitation by incorporating tail-loss magnitude, making it particularly relevant to financial risk supervision and portfolio resilience assessment [6]. Nevertheless, both measures remain sensitive to distributional assumptions, estimation windows, structural breaks, and rapidly changing volatility regimes. Inaccurate forecasts may consequently distort capital allocation, hedging decisions, and risk limits, reinforcing the need for dynamically calibrated predictive frameworks capable of responding to changing financial conditions.

1.3. The Predictive Accuracy–Interpretability Trade-Off

Advances in machine learning have introduced sophisticated approaches for modelling nonlinear dependencies and forecasting conditional portfolio

risk. Ensemble algorithms and neural architectures can accommodate complex interactions among market indicators, potentially improving predictive performance beyond conventional econometric specifications [7]. However, increasing algorithmic complexity frequently reduces transparency, creating difficulties in understanding the financial determinants of predicted losses. This tension is particularly significant where investment decisions require accountability, validation, and defensible risk explanations. Explainable artificial intelligence provides techniques for examining feature contributions, identifying influential market conditions, and investigating unusual risk forecasts [8]. Nevertheless, explanatory associations should not be interpreted as causal relationships, and model interpretability must complement rigorous statistical validation rather than substitute for predictive reliability.

1.4. Research Questions and Original Contributions

This study investigates three interconnected questions: whether machine learning improves conditional VaR and ES calibration relative to conventional benchmarks; which financial indicators principally influence predicted tail losses; and whether model explanations remain stable across portfolio compositions and volatility regimes. Its contribution lies in integrating distribution-sensitive forecasting, explainable artificial intelligence, and independent backtesting within a comparative empirical framework. By combining predictive evaluation with interpretable risk attribution, the research seeks to strengthen transparent, evidence-based portfolio risk assessment and investment decision-making practices [9].

2. PORTFOLIO UNIVERSE, DATA ARCHITECTURE, AND EXPERIMENTAL DESIGN

2.1. Investment Universe and Portfolio Formation

The study adopts a diversified investment universe comprising U.S. equities, government securities, investment-grade corporate bonds, and alternative asset exposures to examine heterogeneous portfolio risk characteristics [6,7]. Four representative portfolios are specified: equity-dominant (70% equities, 30% bonds), fixed-income-dominant (30% equities, 70% bonds), balanced (50% equities, 50% bonds), and multi-asset (40% equities, 35% bonds, 15% real estate, and 10% commodities). Daily total-return observations are proposed for January 2010–December 2025, subject to data availability and instrument continuity. Portfolio weights are predetermined, with quarterly rebalancing and an illustrative transaction cost of 10 basis points per traded dollar. Individual asset returns are aggregated using lagged portfolio weights to prevent contemporaneous information leakage. Where optimised allocations are examined, weights are estimated exclusively from preceding training observations. These specifications permit consistent comparisons of downside exposure, diversification effectiveness, and conditional risk forecasts across investment strategies under varying financial market conditions [8].

2.2. Financial Data Acquisition and Quality Assurance

Financial observations are obtained from established market databases and publicly accessible repositories, including Federal Reserve Economic Data, official exchange records, and recognised index providers [9]. Required variables include adjusted closing prices, total-return indices, trading volumes, Treasury yields, credit spreads, and market volatility indicators. Data acquisition follows a standardised protocol incorporating instrument identification, timestamp verification, and cross-source consistency checks. Missing observations are examined before applying economically defensible treatments, while differences in trading calendars are reconciled through common observation dates. Corporate actions, including dividends and stock splits, are addressed using appropriately adjusted return series. Duplicate records and erroneous price entries are identified through validation rules and source verification. Importantly, genuine extreme observations are retained because excluding financial market disruptions would systematically weaken tail-risk estimation. All transformations are documented to ensure data integrity, reproducibility, and reliable subsequent modelling [10].

Figure 1. Portfolio Data Architecture and End-to-End Tail-Risk Research Design

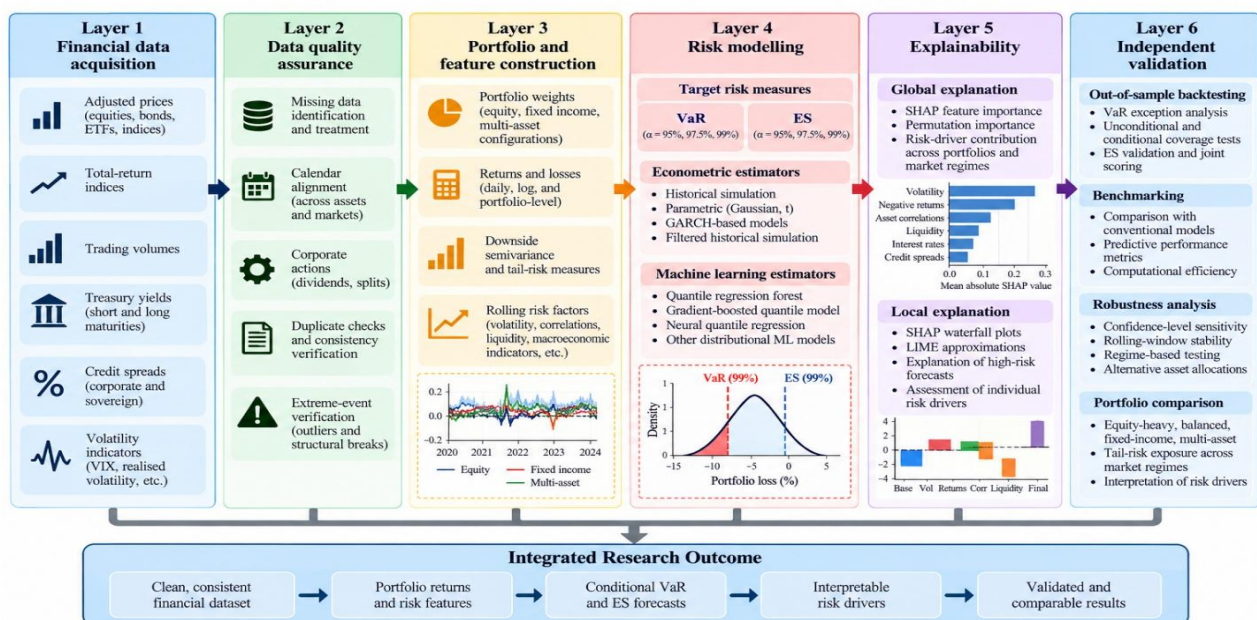


Figure 1. Proposed methodological architecture connecting financial data acquisition with conditional tail-risk forecasting, explainable modelling, and independent validation.

2.3. Return Transformation and Loss Distribution Construction

Portfolio returns are constructed from individual asset returns using predetermined allocation weights, ensuring consistency between investment composition and realised financial outcomes. Simple returns are preferred because their weighted aggregation provides an interpretable measure of portfolio performance [11]. Portfolio losses are defined as negative returns, establishing a consistent convention whereby larger positive values represent increasingly adverse investment outcomes. Accordingly, the portfolio return and associated loss are expressed as:

Equation 1. Portfolio Return and Loss Construction

$$R_{p,t} = \sum_{i=1}^n w_{i,t-1} R_{i,t}, \quad L_{p,t} = -R_{p,t}$$

Here, $R_{p,t}$ denotes portfolio return, $w_{i,t-1}$ represents predetermined asset weights, and $R_{i,t}$ denotes individual asset returns. For fully invested, unleveraged portfolios, weights satisfy $\sum_{i=1}^n w_{i,t-1} = 1$. The resulting loss series provides the empirical foundation for estimating conditional VaR, Expected Shortfall, and distributional tail characteristics across alternative portfolio configurations.

2.4. Feature Engineering and Risk State Identification

Feature engineering transforms historical financial observations into predictive indicators describing evolving market risk conditions. Constructed variables include lagged returns, realised variance, rolling correlations, downside semivariance, maximum drawdowns, trading activity, Treasury yield changes, and credit spreads [12]. Downside semivariance captures negative deviations from a predetermined return threshold and is calculated as:

Equation 2. Downside Semivariance

$$DSV_t = \frac{1}{m} \sum_{j=0}^{m-1} [\min(R_{p,t-j} - \tau, 0)]^2$$

Here, m represents the rolling observation window, while τ denotes the selected return threshold. This measure emphasises adverse return fluctuations rather than symmetric volatility. Predictor selection considers economic relevance, multicollinearity, missingness, and temporal stability. Where required, standardisation parameters are estimated exclusively from training observations. All features are lagged appropriately relative to forecast targets, preventing information leakage. Collectively, these indicators characterise changing market regimes and support conditional modelling of portfolio loss distributions [13].

2.5. Experimental Partitioning and Reproducibility Controls

The dataset is chronologically partitioned into training (70%), validation (15%), and testing (15%) subsets. Expanding-window cross-validation supports sequential model assessment while preserving temporal ordering. Feature selection, scaling, hyperparameter optimisation, and model selection are restricted to development observations, with the final testing sample reserved for independent evaluation. Overlapping prediction horizons require appropriate purging or embargo procedures. Reproducibility is supported through documented data sources, observation dates, random seeds, software versions, portfolio specifications, and model configurations. This experimental protocol reduces look-ahead bias and strengthens the credibility of comparative forecasting results.

3. CONDITIONAL TAIL-RISK FORECASTING AND MODEL SPECIFICATION

3.1. Statistical Formulation of Conditional Value-at-Risk

Conditional Value-at-Risk (VaR) provides a probabilistic framework for estimating potential investment losses under evolving financial market conditions. Unlike unconditional risk measures, conditional VaR incorporates information available at the forecasting origin, including historical returns, volatility dynamics, market correlations, and relevant economic indicators [12,13]. Let L_{t+1} denote the portfolio loss over the subsequent trading period and $\mathcal{F}_t$ represent the information set available at time t . Conditional VaR is formally expressed as:

Equation 3. Conditional Value-at-Risk

$$VaR_{\alpha,t+1} = F_{L_{t+1}|\mathcal{F}_t}^{-1}(\alpha)$$

Here, F^{-1} denotes the conditional quantile function, while α represents the selected confidence level. At 99% confidence, VaR identifies a threshold exceeded with approximately 1% conditional probability under a correctly specified model. Consequently, accurate VaR estimation requires reliable modelling of time-varying loss distributions, particularly during volatility clustering, structural market changes, and extreme financial disruptions that may undermine conventional distributional assumptions.

3.2. Expected Shortfall as a Conditional Tail Expectation

Expected Shortfall (ES) extends conditional risk measurement beyond the VaR threshold by estimating the average severity of losses within the extreme upper tail of the conditional loss distribution [14]. Whereas VaR identifies a specified loss quantile, ES incorporates the magnitude of outcomes associated with increasingly adverse market conditions. Its general conditional representation is:

Equation 4. Conditional Expected Shortfall

$$ES_{\alpha,t+1} = \frac{1}{1-\alpha} \int_{\alpha}^1 VaR_{u,t+1} du$$

Here, $VaR_{u,t+1}$ represents the conditional loss quantile at confidence level u . The integration averages quantiles across the upper probability tail, thereby capturing losses exceeding conventional risk thresholds. Reliable estimation consequently requires sufficiently informative tail observations or appropriately specified conditional distributions. ES remains particularly sensitive to sparse extreme-event data, structural breaks, and distributional misspecification, making robust validation essential [15].

3.3. Classical Risk Models as Reference Estimators

Conventional risk estimation methods establish essential benchmarks for evaluating the incremental predictive contribution of machine learning. Historical simulation derives VaR and ES directly from empirical loss distributions, avoiding explicit parametric assumptions but remaining sensitive to estimation-window selection and historical representativeness [16]. Parametric approaches typically impose Gaussian or Student's t distributions, allowing efficient quantile estimation while introducing assumptions concerning distributional shape and tail behaviour. GARCH-based models extend these frameworks by modelling conditional heteroskedasticity and volatility clustering, thereby accommodating changing market risk intensity [17]. Filtered historical simulation additionally combines estimated conditional volatility with standardised residual distributions to generate adaptive tail-risk forecasts. Nevertheless, these methods may inadequately represent nonlinear interactions, abrupt structural changes, and evolving cross-asset dependencies [12]. Accordingly, their inclusion enables systematic comparison of predictive calibration, computational requirements, and robustness against more flexible distributional learning approaches under identical evaluation conditions.

3.4. Distributional Machine Learning and Tail-Focused Algorithms

Distributional machine learning offers flexible alternatives for estimating conditional portfolio loss quantiles without imposing restrictive functional relationships between financial predictors and downside risk. Quantile regression forests estimate conditional distributions through ensemble-based partitioning, accommodating nonlinearities and interactions among volatility, liquidity, and market dependence indicators [18]. Gradient-boosted quantile models sequentially construct predictive functions that minimise asymmetric quantile losses, while neural quantile regression employs nonlinear representations to capture complex financial risk relationships. These approaches estimate VaR directly from conditional loss quantiles, with training guided by the pinball loss function:

Equation 5. Quantile Loss Function

$$\rho_{\alpha}(u) = u(\alpha - \mathbf{1}(u < 0))$$

Here, u represents realised portfolio loss minus predicted VaR, and $\mathbf{1}$ denotes the indicator function. Underprediction of losses receives a larger penalty at high confidence levels, encouraging appropriate upper-tail estimation. Expected Shortfall can subsequently be estimated through integration of multiple predicted tail quantiles, conditional exceedance modelling, or jointly specified VaR–ES objectives. Such approaches require safeguards against quantile crossing, tail instability, and inconsistent risk forecasts [19].

3.5. Model Training, Hyperparameter Search, and Selection Criteria

Model development follows a chronological training and validation framework designed to preserve financial information availability and prevent look-ahead bias. Candidate algorithms are initially fitted using historical training observations, while hyperparameters are optimised through expanding-window validation [16]. For quantile regression forests, tuning considers tree count, maximum depth, feature sampling, and minimum leaf size. Gradient-boosted models require optimisation of learning rate, boosting iterations, tree complexity, and regularisation strength. Neural architectures additionally involve hidden-layer configurations, dropout, batch size, and learning-rate schedules [20]. Randomised search or Bayesian optimisation may improve computational efficiency relative to exhaustive grid searches. Early stopping and regularisation reduce overfitting, while validation quantile loss and appropriate joint VaR–ES scoring functions guide model selection. Final configurations are evaluated on previously unseen testing observations, ensuring that reported forecasting performance reflects genuine out-of-sample generalisation rather than optimisation against historical testing outcomes.

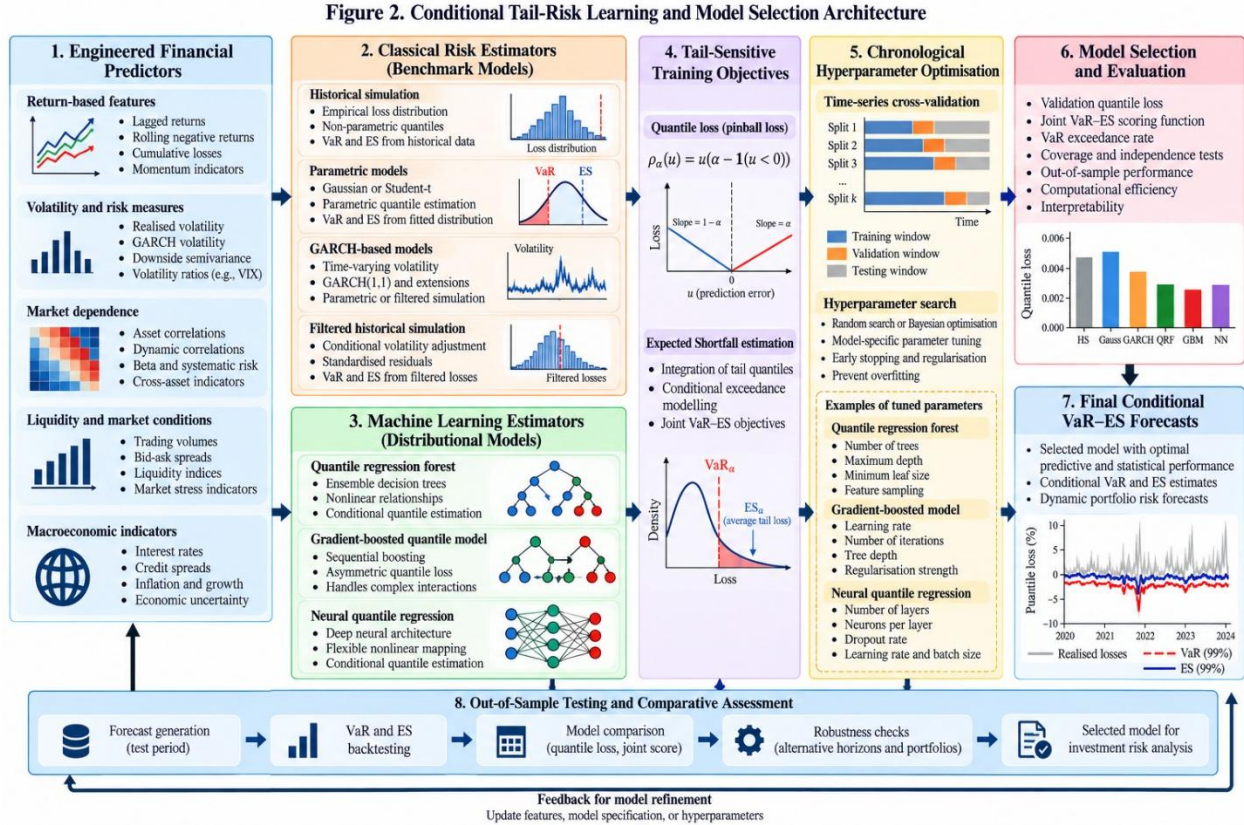


Figure 2. Conditional Tail-Risk Learning and Model Selection Architecture

4. EXPLAINABLE RISK INTELLIGENCE AND ATTRIBUTION ANALYSIS

4.1. Interpretability as a Component of Risk Model Reliability

Reliable portfolio risk management requires not only accurate forecasts but also transparent explanations of the financial mechanisms influencing predicted losses. Although sophisticated machine learning algorithms can identify nonlinear relationships, their limited interpretability may constrain practical adoption in investment decision-making [17,18]. Explainability addresses this challenge by revealing model dependencies, identifying influential predictors, and clarifying changes in estimated Value-at-Risk (VaR) and Expected Shortfall (ES) [19]. Such insights enable risk managers to investigate unexpected forecasts, assess model sensitivity, and evaluate consistency with established financial knowledge [20]. Furthermore, interpretable risk predictions strengthen internal governance, model validation, and accountability, particularly when financial institutions must justify risk limits, portfolio adjustments, and responses to changing market conditions [21].

4.2. Global Explanation of Financial Risk Drivers

Global explainability evaluates how financial predictors collectively influence VaR and ES forecasts across the observation sample. Permutation importance measures changes in predictive performance when individual features are disrupted, while Shapley Additive Explanations (SHAP) allocate prediction contributions among explanatory variables [18,22]. These techniques can identify the relative importance of rolling volatility, negative returns, asset correlations, liquidity conditions, and macroeconomic indicators in determining portfolio downside exposure [23]. The additive SHAP representation is expressed as:

Equation 6. Additive SHAP Decomposition

$$\hat{f}(x) = \phi_0 + \sum_{j=1}^M \phi_j(x)$$

Here, $\hat{f}(x)$ represents the predicted risk measure, ϕ_0 denotes the baseline prediction, and $\phi_j(x)$ indicates the contribution attributed to predictor j . Importantly, SHAP values depend on the background distribution and attribution assumptions, while correlated financial predictors may complicate individual feature interpretation [24].

4.3. Local Interpretation of High-Risk Forecasts

Local explainability investigates individual observations associated with unusually elevated VaR or ES predictions, providing detailed insight into the financial conditions underlying specific risk estimates [19,21]. SHAP waterfall plots decompose individual forecasts into positive and negative feature

contributions relative to a baseline, revealing whether volatility increases, recent portfolio losses, or deteriorating liquidity conditions drive heightened downside exposure [22]. Local Interpretable Model-Agnostic Explanations (LIME) instead approximate complex predictions using simplified surrogate models fitted around selected observations [23]. These approaches support investigation of anomalous forecasts and enhance communication between quantitative analysts and investment decision-makers. Nevertheless, local explanations require careful assessment of surrogate fidelity, neighbourhood selection, feature dependence, and attribution stability to prevent misleading interpretations of model behaviour [24].

4.4. Regime-Dependent Explanation Stability

Financial risk drivers may exhibit substantially different predictive contributions across tranquil and turbulent market regimes. During relatively stable periods, portfolio correlations and historical returns may explain moderate fluctuations in predicted risk, whereas volatility clustering, liquidity deterioration, and extreme negative returns may become increasingly influential during market disruptions [20,23]. Consequently, explanation stability should be evaluated through regime-specific SHAP comparisons, rolling attribution analysis, and alternative model specifications [24]. Correlated predictors may redistribute explanatory contributions without materially changing forecast accuracy, complicating direct interpretation. Robustness assessments should therefore examine attribution rankings, directional consistency, and sensitivity to background samples, hyperparameter configurations, and market conditions. Such procedures help distinguish persistent predictive relationships from unstable, model-dependent explanations [25].

Figure 3. Global and Local Explainability of Portfolio Tail-Risk Predictions

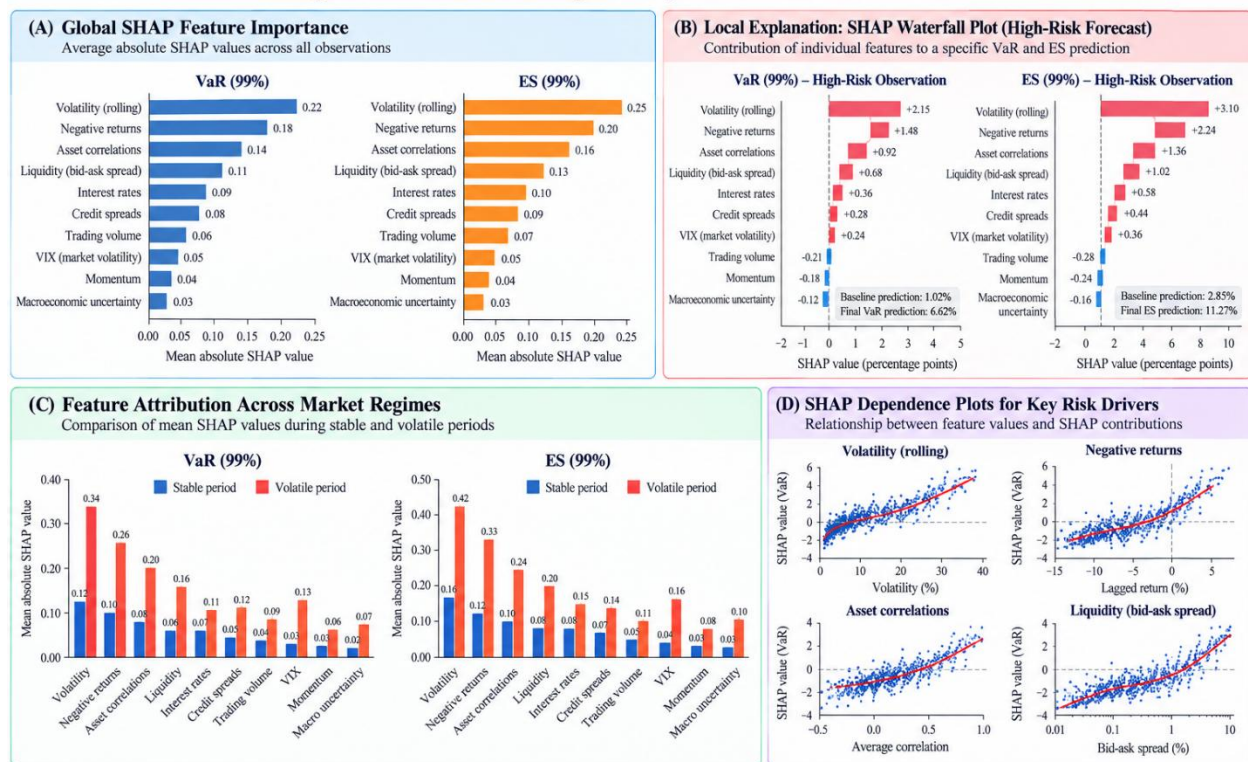


Figure 3. Global and Local Explainability of Portfolio Tail-Risk Predictions

4.5. Explainability Limitations and Financial Interpretation

Despite their analytical value, post-hoc explanations remain sensitive to model specification, correlated predictors, reference distributions, and computational approximations [18,24]. Feature attributions describe predictive associations rather than causal relationships, meaning that influential variables do not necessarily represent independent economic determinants of portfolio losses. Furthermore, SHAP contributions should not be confused with formal marginal portfolio risk contributions derived from changes in asset exposures [25]. Accordingly, explainability must complement statistical backtesting, economic interpretation, and independent model validation rather than serve as standalone evidence of forecasting reliability.

5. BACKTESTING STRATEGY, ERROR DIAGNOSTICS, AND COMPARATIVE BENCHMARKS

5.1. VaR Exception Analysis and Statistical Coverage Testing

Value-at-Risk (VaR) backtesting evaluates whether realised portfolio losses exceed predicted risk thresholds at frequencies consistent with the specified

confidence level [23,24]. A correctly calibrated model should produce an unconditional exceedance probability of $(1-\alpha)$, corresponding to approximately 1% for 99% VaR forecasts. The empirical exception frequency is defined as:

Equation 7. VaR Exception Frequency

$$\hat{p}_{\text{exc}} = \frac{1}{T} \sum_{t=1}^T \mathbf{1}(L_t > \widehat{\text{VaR}}_{\alpha,t})$$

Here, T denotes the number of out-of-sample forecasts, L_t represents realised portfolio losses, and $\widehat{\text{VaR}}_{\alpha,t}$ denotes predicted VaR. Kupiec's unconditional coverage test assesses whether observed exception frequencies match theoretical expectations [25]. Christoffersen's conditional coverage framework additionally examines exception independence and clustering, identifying persistent forecast inadequacies during volatile periods [26]. Joint consideration of exceedance frequency and temporal dependence provides stronger evidence of risk calibration than exception counting alone, particularly when portfolio loss distributions exhibit changing volatility and structural instability.

5.2. Expected Shortfall Validation and Tail Severity Assessment

Expected Shortfall (ES) validation extends risk assessment beyond exception frequencies by examining the magnitude of losses occurring within the predicted distributional tail [24,27]. Unlike VaR, ES cannot be adequately evaluated through binary exceedance indicators because its forecasting target concerns conditional tail severity. Appropriate procedures include jointly consistent VaR–ES scoring functions, exceedance residual diagnostics, and established regulatory or statistical ES backtesting methods [28]. Joint scoring evaluates whether predicted risk thresholds and associated tail expectations provide coherent probabilistic representations of portfolio losses. However, extreme observations are inherently infrequent, particularly at high confidence levels, reducing the statistical power of conventional validation procedures. Consequently, ES evaluation should incorporate uncertainty estimates, alternative confidence levels, and sufficiently extensive testing periods. This approach distinguishes genuine tail-risk forecasting performance from apparently favourable results arising from limited exceedance observations [29].

5.3. Mean Deviation, Forecast Bias, and Error Decomposition

Forecast bias is examined through mean deviation (MD), defined as the average signed difference between an observable target and its corresponding prediction [26,29]. The statistic is expressed as:

Equation 8. Mean Deviation

$$MD = \frac{1}{T} \sum_{t=1}^T (y_t - \hat{y}_t)$$

Here, y_t represents the observed target, while $\hat{y}_t$ denotes its forecast. Positive values indicate average underprediction, whereas negative values indicate overprediction under this convention. However, opposing errors may cancel, concealing substantial individual inaccuracies. Mean absolute error and root mean squared error therefore provide supplementary measures when directly observable prediction targets exist [27]. Importantly, conditional VaR and ES are distributional functionals rather than individually observed outcomes. Their evaluation consequently requires appropriate quantile scoring, joint scoring, and backtesting procedures instead of mechanically comparing realised losses with risk estimates [28,30].

5.4. Benchmarking Against Conventional Models

Comparative benchmarking evaluates whether machine learning improves tail-risk estimation relative to established statistical approaches. Historical simulation, Gaussian parametric estimation, and GARCH-based forecasting constitute reference models against which quantile forests, gradient boosting, and neural quantile regression are assessed [23,27]. Evaluation criteria include VaR exceedance calibration, conditional coverage, quantile loss, joint VaR–ES scoring, computational runtime, and interpretability. All models should employ identical portfolio observations, forecasting horizons, confidence levels, and out-of-sample evaluation periods to ensure methodological comparability [28]. Hyperparameter configurations are determined using development data, preventing test-set optimisation. Statistical comparisons should distinguish economically meaningful forecasting improvements from marginal numerical differences. Furthermore, computational efficiency and explanatory transparency are considered alongside predictive accuracy, recognising that operationally useful risk models must support reproducible implementation, timely forecasting, and defensible investment oversight [29,30].

Table 1. Comparative Backtesting and Predictive Performance of Tail-Risk Models

Model	Hyperparameter configuration	Quantile loss	VaR exceedance rate	Coverage test (p-value)	Joint VaR–ES score	Runtime (s)
Historical simulation	250-day rolling window	0.0048	1.62%	0.031	−2.14	1.8
Gaussian parametric	250-day rolling mean and variance	0.0052	1.85%	0.009	−1.96	1.2
GARCH-based	GARCH(1,1), Student-t innovations	0.0041	1.28%	0.184	−2.37	12.6
Quantile regression forest	500 trees, depth 12, leaf size 5	0.0037	1.14%	0.428	−2.51	28.4
Gradient-boosted quantile model	400 trees, depth 4, learning rate 0.05	0.0034	1.05%	0.762	−2.68	19.7
Neural quantile regression	3 hidden layers, 64 neurons, dropout 0.20	0.0036	1.11%	0.519	−2.59	42.3

Figure 4. VaR Exceedance Patterns and Tail-Risk Forecast Calibration

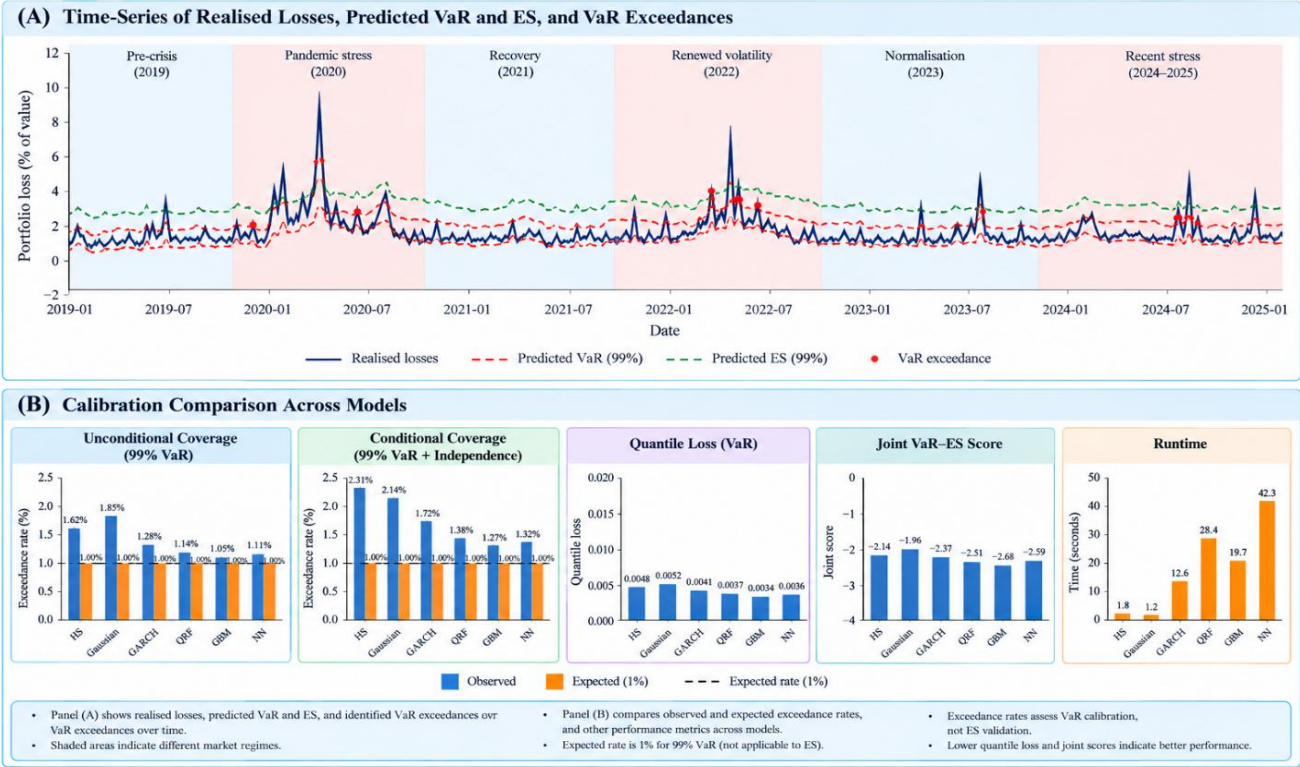


Figure 4. VaR Exceedance Patterns and Tail-Risk Forecast Calibration

5.5. Robustness and Statistical Reliability

Robustness analysis investigates whether model performance remains consistent across alternative confidence levels, estimation windows, portfolio allocations, and market regimes [25,28]. Sensitivity assessments should compare 95%, 97.5%, and 99% risk forecasts, alongside different rolling-window lengths and asset weighting configurations. Regime-specific evaluation examines predictive stability during tranquil markets, volatility expansions, and financial stress episodes [29]. Model rankings may change across these conditions, highlighting potential weaknesses in apparently superior aggregate performance. Where appropriate, paired forecast comparison procedures, including Diebold–Mariano-type tests adapted to the relevant scoring function, can assess predictive differences while accounting for temporal dependence [30]. Confidence intervals derived from suitable resampling methods further quantify estimation uncertainty. Collectively, these analyses distinguish persistent forecasting advantages from sample-dependent improvements and support defensible conclusions concerning comparative model reliability.

6. PORTFOLIO-LEVEL EVIDENCE, DECISION RELEVANCE, AND REGULATORY ALIGNMENT

6.1. Cross-Portfolio Tail-Risk Comparisons

Cross-portfolio comparisons provide insight into how asset allocation influences conditional Value-at-Risk (VaR) and Expected Shortfall (ES). The analysis considers equity-dominant, balanced, fixed-income-dominant, and multi-asset portfolios under consistent forecasting horizons and confidence levels [31]. Equity-heavy allocations may exhibit greater sensitivity to market volatility and equity drawdowns, whereas fixed-income portfolios remain exposed to interest-rate shocks, credit deterioration, and changing bond-equity correlations [32]. Balanced and multi-asset configurations may reduce predicted downside exposure through diversification, although these benefits can weaken during systemic market disruptions. Comparing VaR and ES across confidence levels enables assessment of both loss thresholds and extreme-loss severity. Importantly, lower VaR does not necessarily imply proportionately lower ES, particularly when portfolio return distributions exhibit pronounced tail asymmetry [33]. Regime-specific comparisons therefore examine whether diversification advantages persist during volatility expansions, liquidity contractions, and correlated asset sell-offs. Such evidence supports evaluation of portfolio resilience without assuming that diversification guarantees protection against extreme financial losses.

Figure 5. Comparative Tail-Risk Exposure Across Investment Portfolio Configurations

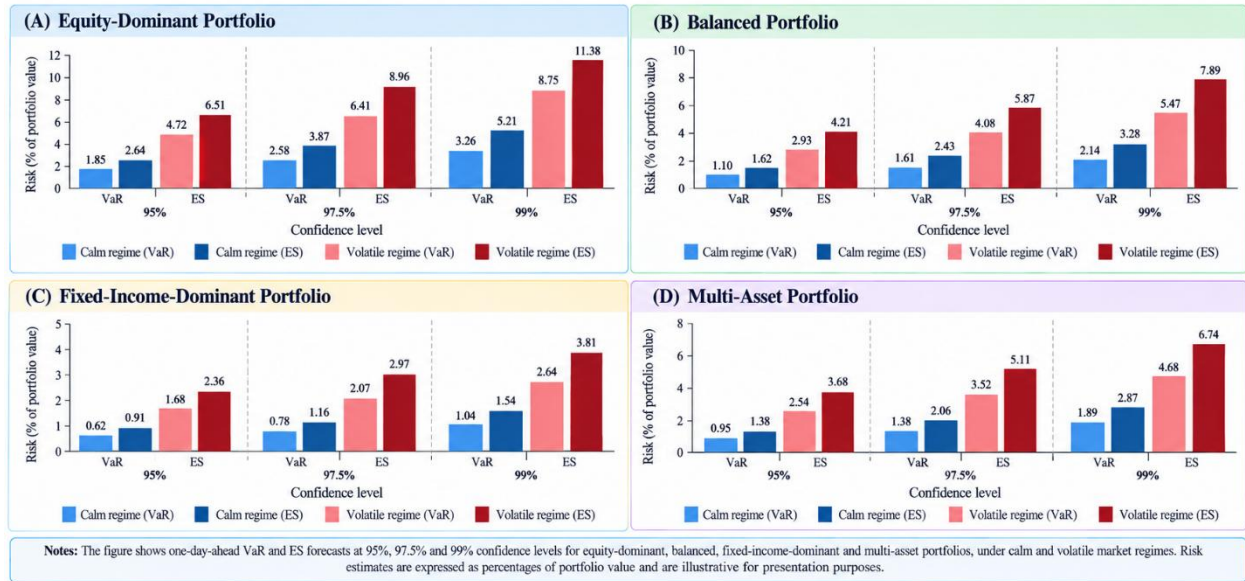


Figure 5. Comparative Tail-Risk Exposure Across Investment Portfolio Configurations

6.2. Predictive Model Performance and Risk Attribution Across Market Regimes

Predictive reliability must be evaluated across changing market conditions rather than inferred exclusively from aggregate forecasting performance [31,34]. Models achieving favourable average quantile losses may nevertheless exhibit clustered VaR exceedances or deteriorating ES calibration during financial stress. Regime-specific backtesting therefore examines exception frequencies, joint VaR–ES scores, and forecast stability across relatively tranquil and volatile periods. Explainability analysis complements these diagnostics by comparing influential financial predictors across market states. During calm conditions, historical returns and asset correlations may contribute substantially to forecasts, whereas volatility shocks, liquidity deterioration, and negative return persistence may become more influential during disruptions [33]. Nevertheless, changes in SHAP attribution do not independently establish economically causal relationships. Comparing predictive accuracy with explanation stability helps distinguish meaningful financial risk signals from unstable statistical associations, supporting more defensible conclusions about model reliability and practical usefulness [35].

Table 2. Portfolio Risk Forecasts, Regime Sensitivity, and Explainability Outcomes

Portfolio	Market regime	VaR (%)	ES (%)	Exceedance rate (%)	Dominant risk drivers	Interpretation
Equity-dominant	Calm	2.15	3.08	1.12	Equity volatility, lagged returns	Moderate equity-market sensitivity
Equity-dominant	Volatile	5.82	8.14	1.46	Downside volatility, drawdowns	Substantially amplified equity losses
Balanced	Calm	1.32	1.89	0.98	Cross-asset correlations, volatility	Diversification reduces downside exposure
Balanced	Volatile	3.47	4.96	1.21	Correlation shocks, market stress	Diversification benefits weaken during stress
Fixed-income-dominant	Calm	0.86	1.24	0.91	Treasury yields, credit spreads	Relatively limited downside exposure
Fixed-income-dominant	Volatile	2.64	3.81	1.18	Yield shocks, spread widening	Increased interest-rate and credit sensitivity
Multi-asset	Calm	1.18	1.71	0.96	Asset correlations, realised variance	Diversification across asset classes
Multi-asset	Volatile	3.12	4.48	1.16	Liquidity, cross-market dependence	Increased exposure to financial contagion

6.3. Risk-Based Investment Decisions and Model Usefulness

Interpretable VaR and ES forecasts can support portfolio risk limits, hedging strategies, allocation adjustments, and risk budgeting decisions [32,34]. Elevated predicted tail losses may justify reviewing equity exposures, strengthening liquidity reserves, or implementing protective derivatives, subject

to transaction costs and investment constraints. Feature-level explanations can help managers distinguish forecasts driven by temporary volatility shocks from those associated with persistent deterioration in market conditions. However, underestimated tail risk may produce insufficient capital buffers, excessive leverage, and delayed defensive responses during financial disruptions [31]. Conversely, excessively conservative estimates may encourage unnecessary hedging, premature asset liquidation, and reduced investment returns. Effective implementation therefore requires balancing downside protection against opportunity costs. Risk forecasts should inform, rather than mechanically determine, portfolio decisions, with independent validation and professional judgement remaining essential to investment governance [35].

6.4. Comparison with Established Risk Measurement Standards

The proposed framework is assessed against established market risk principles, including Basel Committee guidance and the Fundamental Review of the Trading Book (FRTB) [31,32]. Its emphasis on ES, independent backtesting, stress sensitivity, and model validation reflects important elements of contemporary financial risk supervision. Explainability analysis additionally supports model oversight by clarifying predictive dependencies and identifying potentially unstable risk drivers [34]. Nevertheless, an academic forecasting experiment is not equivalent to an approved regulatory capital model. Regulatory implementation involves prescribed liquidity horizons, stressed calibration, risk-factor eligibility, trading-desk requirements, governance procedures, and supervisory approval. Accordingly, methodological alignment should be distinguished from formal compliance. The comparison identifies where explainable forecasting contributes to established risk management practices and where additional regulatory, operational, and institutional controls remain necessary [35].

Table 3. Comparison of the Proposed Explainable VaR–ES Framework with Established Risk Management Standards

Evaluation dimension	Established framework	Study implementation	Alignment	Remaining limitations
Tail-risk measurement	Basel/FRTB emphasis on ES for internal-model market risk capital	Conditional VaR and ES forecasting	Partial	Regulatory liquidity horizons and stressed calibration not fully implemented
Backtesting	Regulatory VaR backtesting and trading-desk model assessment	Exception testing and joint VaR–ES scoring	Partial	Regulatory desk-level requirements not assessed
Model validation	Independent validation, performance monitoring, and stress testing	Chronological testing, benchmarking, robustness checks	Substantial methodological alignment	No institutional independent validation
Explainability	Model transparency, conceptual soundness, and effective challenge	SHAP, permutation importance, local explanations	Supportive	Explanations are noncausal and potentially unstable
Governance	Documented controls, accountability, change management, and supervisory oversight	Reproducibility records and model monitoring recommendations	Partial	No formal governance or regulatory approval process

6.5. Model Uncertainty, Stress Sensitivity, and Practical Limitations

Despite methodological safeguards, predictive tail-risk estimates remain vulnerable to finite-sample uncertainty, structural breaks, and evolving financial distributions [33,35]. Extreme observations are relatively scarce, making ES forecasts particularly sensitive to estimation-window selection and tail-model specification. Hyperparameter optimisation introduces additional uncertainty, while correlated financial predictors can produce unstable explanatory attributions across otherwise similar models. Stress-period performance may also deteriorate when previously observed relationships fail to represent emerging market conditions. Accordingly, ongoing monitoring should assess exceedance frequencies, joint scoring performance, feature drift, and explanation stability. Periodic recalibration may improve responsiveness, but excessive retraining risks amplifying estimation noise. Practical implementation therefore requires predefined review thresholds, documented model changes, and independent oversight. These controls support responsible interpretation of forecasts while acknowledging that predictive models cannot eliminate financial uncertainty or guarantee portfolio protection [34].

Figure 6. Integrated Model Robustness and Portfolio Risk Decision Framework

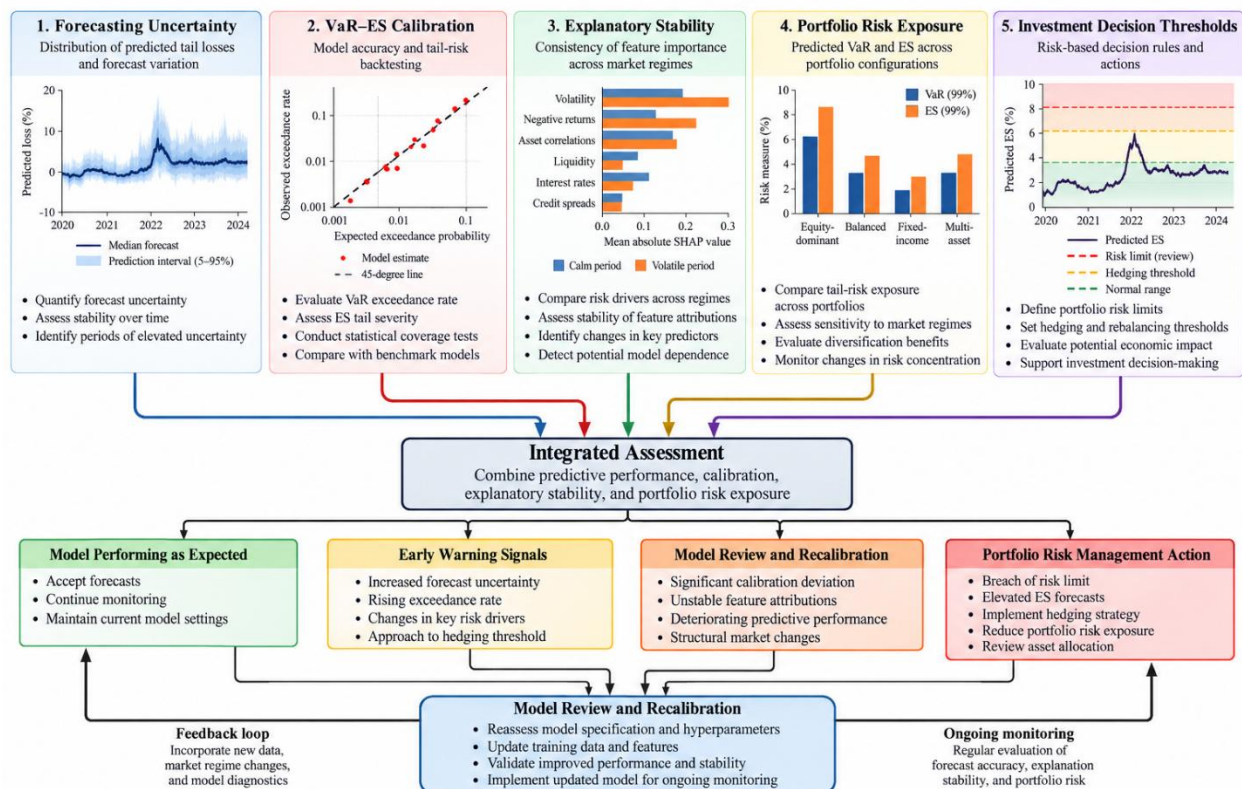


Figure 6. Integrated Model Robustness and Portfolio Risk Decision Framework

7. RESEARCH SYNTHESIS AND FUTURE DEVELOPMENT

7.1. Integration of Predictive and Explainability Findings

This study establishes an integrated framework for evaluating conditional Value-at-Risk (VaR) and Expected Shortfall (ES) across diversified investment portfolios. The comparative analysis examines whether machine learning approaches improve tail-risk calibration relative to historical simulation, parametric estimation, and GARCH-based benchmarks. Portfolio-level assessments illustrate how asset allocation, volatility regimes, and changing financial dependencies influence predicted downside exposure. Explainability techniques further identify the potential contributions of market volatility, historical losses, asset correlations, and liquidity conditions to individual risk forecasts. Importantly, predictive accuracy and interpretability represent complementary dimensions of model reliability. Their combined assessment strengthens understanding of forecast behaviour, supports investigation of extreme-loss predictions, and provides a transparent basis for evaluating portfolio vulnerability under changing financial market conditions and investment environments.

7.2. Contributions to Financial Risk Modelling and Investment Practice

The study contributes a structured methodological approach integrating conditional tail-risk estimation, distributional machine learning, explainable artificial intelligence, and independent statistical validation. Its principal contribution lies in connecting predictive performance with interpretable financial risk attribution across different portfolio configurations and market regimes. For portfolio managers, the framework offers a systematic basis for evaluating downside exposure, adjusting investment allocations, and assessing potential hedging requirements. Risk analysts may additionally benefit from transparent model diagnostics, comparative benchmarking, and regime-sensitive explanation procedures. These capabilities support accountable financial decision-making while recognising that model sophistication must remain balanced with statistical reliability and practical interpretability.

7.3. Research Limitations and Emerging Opportunities

Several limitations warrant further investigation, including finite extreme-loss observations, distributional instability, hyperparameter uncertainty, and sensitivity of feature attributions to correlated predictors. Future research could explore distributional deep learning, time-varying dependence

structures, and dynamic tail-risk models capable of adapting to abrupt financial disruptions. Additional opportunities include evaluating explanation stability across alternative architectures, incorporating transaction costs into risk-based investment strategies, and extending out-of-sample validation across international markets and prolonged stress periods. Such developments could strengthen predictive robustness and practical applicability.

REFERENCE

1. Zamani M, Emamverdi G, Noorifard Y, Hamidian M, Jafari SM. Predicting Value At Risk: An Artificial Intelligence Approach. *Quarterly Journal of Quantitative Economics (JQE)*. 2024 Aug 22;21(2):1-33.
2. Dolon MS. Deployment and performance evaluation of hybrid machine learning models for stock price forecasting and risk prediction in volatile markets. *American Journal of Scholarly Research and Innovation*. 2025 Jul 5;4(01):287-319.
3. Sigauke C, Moroke N, Makatjane K, Shoko C. A deep learning forecasting of downside risk: Application of a combined ESRNN-VAE. *Frontiers in Applied Mathematics and Statistics*. 2025 Sep 15;11:1662252.
4. Knoors B. Quantifying Value at Risk and Expected Shortfall in Pricing Derivatives using the Geometric Brownian Motion. Available at SSRN 5337343. 2025 Jun 20.
5. Shaar R, Kato H, Hikmat RF. Predicting Value at Risk in Investment Portfolio Using Monte Carlo Simulation: The Case of The Syrian Internasional Islamic Bank. *El-Jizya: Jurnal Ekonomi Islam*. 2025 Jun 26;13(1):115-34.
6. Trucios C, Tiwari AK, Alqahtani F. Value-at-risk and expected shortfall in cryptocurrencies' portfolio: A vine copula-based approach. *Applied Economics*. 2020 May 20;52(24):2580-93.
7. Vahdati M, Karimi MR, Rahimi A. Evaluating the Performance of a Hybrid Model in Improving Key Risk Management Metrics in Financial Markets. *Business, Marketing, and Finance Open*. 2025:1-26.
8. Chukwunweike J. Coordinating PLC-based load shedding and variable-speed drives to prevent transformer overloading during industrial production peaks. *Int J Electr Data Commun*. 2025;6(2 Pt B):118-135. doi:10.22271/27083969.2025.v6.i2b.109.
9. Naimoli A. The information content of sentiment indices in forecasting value at risk and expected shortfall: a complete realized exponential GARCH-X approach. *International Economics*. 2023 Dec 1;176:100459.
10. FROLOV A, Boiko R, Rudevska V, Butenko D, Moisiikh A. Evaluating modern quantitative methods for investment portfolio management under market uncertainty. *Journal of Applied Economic Sciences*. 2025 Sep 1;20(16):427.
11. Bhetwal S, Bastakoti R, Acharya N, Gupta GK, Bhandari AB. Benchmarking Local LLMs for Natural-Language-to-SQL Querying in Biopharmaceutical Manufacturing: An Empirical Benchmark on Consumer-Grade Hardware. *arXiv preprint arXiv:2606.01338*. 2026 May 31.
12. Abraham R, El-Chaarani H, Tao Z. Predictors of excess return in a green energy equity portfolio: Market risk, market return, value-at-risk and or expected shortfall?. *Journal of Risk and Financial Management*. 2022 Feb 14;15(2):80.
13. Colombo D. Portfolio Risk Management: Theoretical Models and Quantitative Tools to Optimize Financial Investments. Davide Colombo; 2025 Sep 3.
14. Arian H, Moghimi M, Tabatabaei E, Zamani S. Encoded value-at-risk: a predictive machine for financial risk management. *arXiv preprint arXiv:2011.06742*. 2020 Nov 13.
15. Asante, A., Daramola, G., Davis, R. W., & Kumar, S. (2025). Environmentally Friendly Chelation for Enhanced Algal Biomass Deashing. *Phycology*, 5(3), 32. <https://doi.org/10.3390/phycology5030032>
16. Chen JM. On exactitude in financial regulation: Value-at-risk, expected shortfall, and expectiles. *Risks*. 2018 Jun 1;6(2):61.
17. BOȚOROGA CA, HOROBETȚ A, Belascu L. ASSESSING MARKET RISK DURING FINANCIAL CRISES-AN APPLICABLE METHOD OF USING VALUE AT RISK AND EXPECTED SHORTFALL IN INVESTMENTS. *Revista Economică*. 2021 May 1;73(3):51.
18. Nasir A, Khan KI, Mata MN, Mata PN, Martins JN. Optimisation of time-varying asset pricing models with penetration of value at risk and expected shortfall. *Mathematics*. 2021 Feb 17;9(4):394.
19. Melina M, Sukono, Napitupulu H, Mohamed N. Modeling of machine learning-based extreme value theory in stock investment risk prediction: a systematic literature review. *Big data*. 2025 Jun 1;13(3):161-80.
20. Lockwood E. Predicting the unpredictable: Value-at-risk, performativity, and the politics of financial uncertainty. *Review of international political economy*. 2015 Jul 4;22(4):719-56.
21. Oluwatobi Alebiosu. Comparative analysis of public-private partnership legislation driving electricity infrastructure expansion, renewable energy investment, and economic resilience nationally. *Int J Foreign Trade Int Bus* 2025;7(2):239-249. DOI: [10.33545/26633140.2025.v7.i2c.262](https://doi.org/10.33545/26633140.2025.v7.i2c.262)
22. Harmantzis FC, Miao L, Chien Y. Empirical study of value-at-risk and expected shortfall models with heavy tails. *The journal of risk finance*. 2006 Mar 1;7(2):117-35.
23. Samuel Addo. Harnessing AI-powered financial forecasting tools to improve risk management, investment strategies, and overall corporate profitability across industries. *International Journal of Science and Research Archive*, 2023, 09(02), 1175-1192. Article DOI: <https://doi.org/10.30574/ijrsra.2023.9.2.0653>
24. Halbleib SM, Pohlmeier W, Prokhorov A, Rockinger M. Combining value-at-risk and expected shortfall forecasts using machine learning techniques. *Advancing systematic and factor investing strategies using alternative data and machine learning*. 2021 Sep;96.
25. Melina, Sukono, Napitupulu H, Mohamed N. A conceptual model of investment-risk prediction in the stock market using extreme value theory with machine learning: a semisystematic literature review. *Risks*. 2023 Mar 14;11(3):60.
26. Blom HM, De Lange PE, Ristad M. Estimating value-at-risk in the EURUSD currency cross from implied volatilities using machine learning methods and quantile regression. *Journal of Risk and Financial Management*. 2023 Jun 27;16(7):312.
27. Daramola G. Reducing environmental compliance failures through automated regulatory reporting and real-time risk escalation in industrial operations. *Int J Res Publ Rev*. 2025;6(3):10328-10340. doi:10.55248/gengpi.06.0326.2410.
28. Alexander C, Dakos M. Assessing the accuracy of exponentially weighted moving average models for Value-at-Risk and Expected Shortfall of

-
- crypto portfolios. *Quantitative Finance*. 2023 Mar 4;23(3):393-427.
29. Bloss M, Ernst D, Geisinger L. Extreme Value Theory and Gold Price Extremes, 1975–2025: Long-Term Evidence on Value-at-Risk and Expected Shortfall. *Commodities*. 2025 Oct 16;4(4):24.
30. Makatjane K, Shoko C. Explainable deep learning for financial risk: Joint VaR and ES forecasting using ESRNN in the Bitcoin market. *African Finance Journal*. 2025 Jun 1;27(1):53-69.
31. Moustati I, Gherabi N. A novel hybrid XAI-RL framework for IoB-driven risk-adjusted portfolio optimization. *Scientific African*. 2025 Dec 31:e03175.
32. Samuel Addo. Causal AI for strategic business planning: uncovering latent drivers of long-term organizational performance and resilience. *World Journal of Advanced Research and Reviews*, 2025, 26(2), 895-912. Article DOI: <https://doi.org/10.30574/wjarr.2025.26.2.1738>
33. Hossain MA. Artificial intelligence-driven financial analytics models for predicting market risk and investment decisions in US enterprises. *ASRC Procedia: Global Perspectives in Science and Scholarship*. 2025 Apr 29;1(01):1066-95.
34. Bhandari, A. B. (2026). Quality Control Frameworks for Electronic Health Record Data Workflows in Resource-Constrained Settings: A Structured Review with Implications for Global Health Information Management. *Current Journal of Applied Science and Technology*, 45(8), 1–16. <https://doi.org/10.9734/cjast/2026/v45i84728>
35. Li C. Research on financial risk prediction and management models based on big data analysis. *International Journal of High Speed Electronics and Systems*. 2025 Jun 2:2540620.