Proceedings of the 4th International Conference on Machine Learning and Automation
DOI: 10.54254/2755-2721/2026.38688

Adaptive Measurement of Portfolio Downside Risk Based on Quantile Machine Learning

Ying Hui

Faculty of Business and Economics, The University of Melbourne, Melbourne, Australia
huiying02589@outlook.com

Abstract. The risk of downside of the portfolio depends on volatility, drawdown, asset dependence, and regime in a nonlinear way which undermines the stability of risk measurement with fixed distribution. In this paper, we propose a machine-learning based adaptive downside risk measure that utilizes LightGBM quantile regression to predict the conditional tail of the portfolio made up of equally-weighted eight ETFs. The data is composed of daily SPY, IWM, EFA, EEM, TLT, GLD, DBC, and VNQ during the period of 2010 to 2025, and features are constructed by rolling window volatility, semivariance, drawdown, dependence, and market state. Conditional quantiles between 0.25% and 2.50% are forecast and used to construct one-day ahead 97.5% VaR and Expected Shortfall. The adaptive LightGBM in the out-of-sample test produces a pinball loss of 0.0187±0.0011, a VaR violation ratio of 1.04±0.08, and an ES backtest score of 0.021±0.006. Its joint tail loss score significantly outperforms Historical Simulation, Gaussian GARCH, and linear Quantile Regression. When the regime is volatile, the forecast improvement is still significant. Feature ablation suggests that downside-history and asset dependence variables contribute most.

Keywords: tail risk, extreme market conditions, extreme value theory, Expected Shortfall, GJR-GARCH

1. Introduction

Unconditional variance fails as a risk measure in case of asymmetric losses, fat-tailed distribution, and conditional dependence between assets. While VaR concentrates on a quantile of loss, ES represents the mean of losses above it. Recently, it was shown that ES can be recovered from a number of quantiles much better than from a parametric assumption about tail behavior [1]. Moreover, accounting for evolution of the relationship between VaR and ES helps to improve risk forecasting in case of tail regime switching [2]. Non-linear quantile models are especially relevant since the effect of volatility or dependence does not necessarily have to be constant along the return distribution; indeed, neural quantile regression has proven that non-linear tail mapping can enhance systemic-risk measurement [3].

The same is true at the portfolio level. Mixed-frequency quantile models demonstrate the improvement of VaR and ES predictions when state information is included [4]; Bayesian tail-risk models indicate the importance of dynamic quantile specification stability and initialization [5].

© 2026 The Authors. This is an open access article distributed under the terms of the Creative Commons Attribution License 4.0 (https://creativecommons.org/licenses/by/4.0/).

DOI: 10.54254/2755-2721/2026.38688

Option-implied information also suggests that tree ensembles and non-linear regressions are capable of extracting information in the lower tail of the distribution which cannot be captured by traditional linear methods [6]. Additionally, stateful recurrent models outperform a number of econometric benchmarks in joint VaR/ES forecasting due to the ability of capturing the serial dependence of the risk process [7]. Therefore, distributional machine learning becomes increasingly popular for portfolio optimization in the face of non-Gaussian and systemic risks [8]. Quantile-regression-forests recently confirm the capability of capturing the lower quantile calibration even in volatile energy-market environments [9]. Direct comparison of LightGBM and deep quantile architectures demonstrates that gradient boosting remains competitive lower-tail forecasting benchmark [10].

This paper develops an adaptive LightGBM quantile model for an eight-assets portfolio. Our contribution is in the combination of multi-horizon downside risk characteristics, dynamic cross-asset dependence, market state variables, rolling training, and multi-quantile ES recovery into a single leakage-free forecasting framework. It is tested using sequential out-of-sample forecasting, benchmark comparison, feature ablation, VaR/ES backtesting, and volatility regimes analysis.

2. Literature review

2.1. Portfolio downside risk measurement

Downside risk for portfolios is related to bad returns and not dispersion of returns. The variance class of methods treats bad and good deviations equally and thus does not capture capital loss. The value at risk captures a probability-dependent loss boundary, whereas the expected shortfall is the average of results outside the loss boundary [1-4]. The two methods rely on a conditional distribution of returns, which varies in terms of its volatility, skewness, dependence, and fat-tail characteristics. Portfolio estimation is particularly dependent on changes in dependence.

2.2. Quantile-based financial modeling

The quantile approach estimates chosen quantile ranges in the conditional distribution directly. As opposed to mean regression, it enables the impact of the predictors to vary between the center and tails of the conditional distribution. Extensions of machine learning techniques enable the estimation of the nonlinear interaction and threshold effects without imposing a particular distribution [2-6]. The use of gradient boosting is suitable for risk analysis because of the interaction between variance, drawdown, dependence, and the market characteristics.

2.3. Adaptive risk measurement

Risk models estimated using fixed-sample calibrations may become obsolete due to regime shifts. The use of rolling estimation allows for controlled forgetting, with new data entering the estimation period and old data being dropped. Market state indicators including implied volatility, momentum, and interest rate changes can aid in the adaptive measurement process in the sense that the model may adapt to the portfolio's weakening situation as well as to the general market stress [7, 10]. Evaluation must involve strictly chronological training and testing due to potential information leakage of volatility regimes.

DOI: 10.54254/2755-2721/2026.38688

3. Methodology

3.1. Portfolio data and construction

This portfolio is made up of SPY, IWM, EFA, EEM, TLT, GLD, DBC, and VNQ, with weights for large-cap stocks of the United States, small-cap stocks of the United States, developed foreign equities, emerging markets, long-duration bonds, gold, commodity index, and real estate. The adjusted closing price on a daily basis from January 2010 to December 2025 is sourced from Refinitiv Workspace and is analyzed using Python 3.11. Each of these securities is given a target weight of 12.5%, and the portfolio is re-balanced on the first trading day of the month. Within the period between two consecutive re-balances, the weights float with respect to asset returns so that the portfolio's path is a realistic one without daily re-balancing. Daily portfolio returns are based on log returns calculated using lagged weights, thus excluding contemporaneous information from the prediction target.

3.2. Adaptive downside-risk features

All features are based on information as of the end of day tt and are used to predict return on day t+1t+1. The set of portfolio state features includes realized volatility over 5, 20, and 60 trading days; downside semivariance for 20 and 60 trading days; cumulative returns; maximum drawdown; skewness; excess kurtosis; and the proportion of negative return days. Dependency is measured through the 20-day average pairwise correlation between all eight ETFs and the explained variance ratio of the first principal component, rising as returns are dominated by the common market factor. Market-state features external to the portfolio are the VIX index value and change, S&P 500 momentum over 20 days, and 5-day change in the U.S. 10-year Treasury yield. All features are winsorized to the training window extremes and scaled for the linear benchmark, while LightGBM gets the unscaled values.

3.3. Quantile machine learning model

For LightGBM 4.3 quantile estimates of conditional returns of the portfolio, separate models are built for 0.25%, 0.50%, …, 2.50%. Rolling window of 1,000 trading days and retraining every 20 trading days are used. Using Optuna, 80 time series validation experiments are conducted for the following parameters: learning rate, tree depth, number of leaves, feature fraction, minimum leaf size, regularization type, and number of boosting iterations. The tuning range is set to be 0.01-0.08 for learning rate, 3-8 for tree depth, 15-63 for the number of leaves, 0.

For quantile level qq, the boosted predictor and regularized training criterion are as shown in Equation (1) - (3):

$$\widehat{F}_q^{(M)}(x) = \sum_{m=1}^{M} f_{m,q}(x) \tag{1}$$

$$\theta_q^* = \arg\ min_{\{f_{m,q}\}} \left\{ \sum_{t=1}^{N} \rho_q \left(y_t - \sum_{m=1}^{M} f_{m,q}(\mathbf{x}_t) \right) + \sum_{m=1}^{M} \left[\gamma T_m + \frac{\lambda}{2} \left\| \mathbf{w}_m \right\|_2^2 \right] \right\} \tag{2}$$

$$\rho_q(u) = u(q - \mathbb{I}[u < 0]) \tag{3}$$

DOI: 10.54254/2755-2721/2026.38688

Here, T_m is the number of leaves in tree m, and $\boldsymbol{w}_m$ contains its terminal-node scores. This objective allows asymmetric tail errors to determine each tree sequence while regularization limits overly specific partitions during rare stress observations.

3.4. Dynamic VaR and expected shortfall

One-day-ahead 97.5% VaR is the amount of loss in the positive direction that corresponds to the forecasted 2.50% quantile of portfolio returns. On the other hand, ES can be computed from the whole lower tail forecast and not just the VaR factor. To be specific, the average of ten forecasted quantiles ranging from 0.25% to 2.50% is calculated using the monotonic projection technique with 0.25 percentage point equal probability spacing in between to approximate the conditional lower tail integral. Thus, ES will differ if the shape of the estimated tail changes even with similar 2.50% quantile.

4. Experimental design and validation

4.1. Rolling out-of-sample experiment

The first 1,000 valid data points are set aside for preliminary modeling. After that, predictions will be made on a day-by-day basis until December 2025. Each 20 business days, the oldest data points are discarded, the newest data points are added, and all the quantile regression models are retrained. Hyperparameters are determined only in the current window of history, utilizing blocked validation. Therefore, the calculation of features, weights of the portfolio, scaling constraints, and model training is done based solely on data prior to the prediction date.

4.2. Benchmark and ablation tests

The new approach is benchmarked against the Historical Simulation for 1,000 days, the Gaussian GARCH(1,1), and Linear Quantile Regression, all using the same forecast dates. Four further benchmarks identify the underlying driver of performance: LightGBM with no cross-asset dependencies, no downside history, no exogenous market state variables, and expanding window LightGBM that never drops data. The former three test which information categories are responsible for improving lower tail predictions, while the latter disentangles nonlinear learning benefits from rolling sample dynamics.

4.3. Risk backtesting and market-state validation

The VaR specification is analyzed with violation frequency, violation ratio, Kupiec unconditional coverage, and Christoffersen conditional coverage tests. The ES estimation is studied with the Acerbi-Szekely test, mean excess loss, and mean multi-quantile pinball loss. The Diebold-Mariano tests are performed between the proposed approach and the benchmark approaches based on tail loss differential measures and variance-covariance matrix estimators with heteroskedasticity and autocorrelation correction for each day.

DOI: 10.54254/2755-2721/2026.38688

5. Results

5.1. Adaptive VaR and expected shortfall dynamics

The adaptive model generated pronounced but differentiated responses across market states. Mean 97.5% VaR increased from 1.21±0.18% in the low-volatility regime to 2.08±0.31% in the medium regime and 3.74±0.62% in the high regime. Corresponding ES estimates were 1.68±0.24%, 2.91±0.43%, and 5.36±0.88%. The ES–VaR spread expanded from 0.47±0.09 to 1.62±0.34 percentage points, indicating steeper lower-tail curvature. One-day ES revisions averaged 0.41±0.12 points during major drawdowns versus 0.08±0.03 in low-VIX periods (see Figure 1).

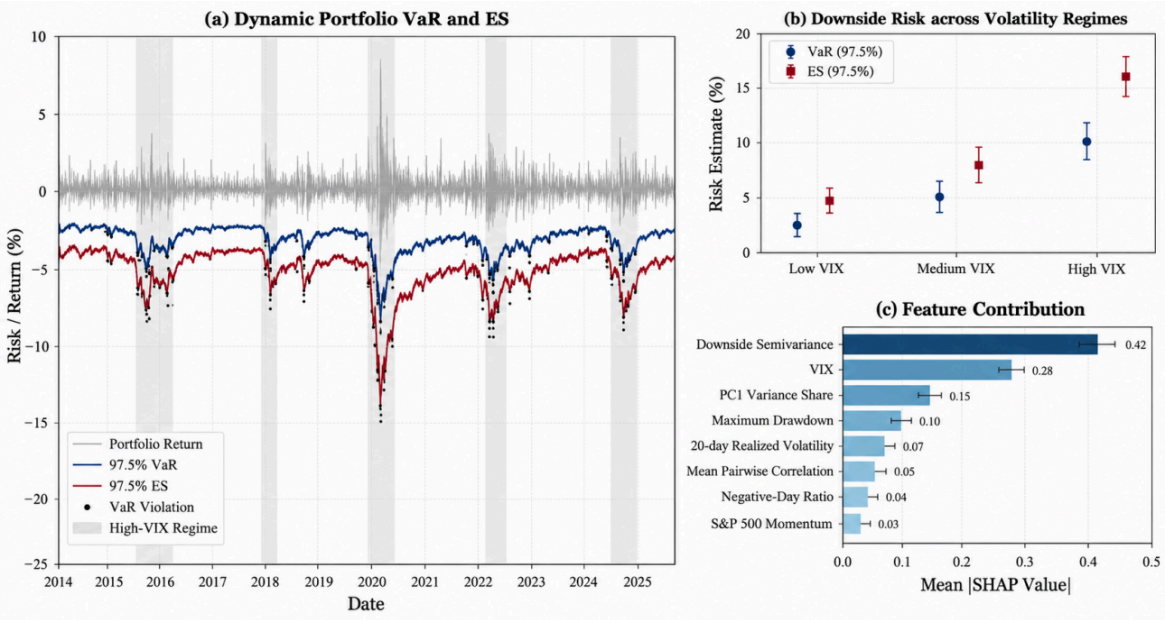

Figure 1. Adaptive portfolio downside-risk dynamics and feature contributions across market regimes

5.2. Out-of-sample performance and benchmark comparison

LightGBM produced the lowest mean pinball loss at 0.0187±0.0011*, compared with 0.0249±0.0016 for Historical Simulation, 0.0228±0.0014 for Gaussian GARCH, and 0.0216±0.0013 for linear Quantile Regression. Its VaR violation ratio was 1.04±0.08, and the conditional-coverage rejection frequency across rolling evaluation blocks was 0.07±0.03. The mean Acerbi-Szekely ES statistic fell to 0.021±0.006*, versus 0.058±0.012 for Historical Simulation and 0.043±0.010 for GARCH. Diebold-Mariano statistics against the three benchmarks were 4.62±0.41*, 3.78±0.36*, and 2.91±0.29*, respectively (see Table 1).

Table 1. Out-of-sample downside-risk forecasting performance

Model	Pinball Loss	VaR Violation Ratio	ES Statistic	Mean Exceedance Loss	Joint Tail Score
Historical Simulation	0.0249±0.0016	1.34±0.12	0.058±0.012	5.91±0.73	0.087±0.006

DOI: 10.54254/2755-2721/2026.38688

Table 1. (continued)

Gaussian GARCH(1,1)	0.0228±0.0014	1.21±0.10	0.043±0.010	5.43±0.68	0.078±0.005
Linear Quantile Regression	0.0216±0.0013	1.14±0.09	0.036±0.008	5.17±0.61	0.071±0.004
Adaptive LightGBM	0.0187±0.0011*	1.04±0.08*	0.021±0.006*	4.61±0.55*	0.059±0.004*

** $p<0.01$ versus linear Quantile Regression after Holm adjustment.*

5.3. Market-regime robustness and feature contribution

The largest performance improvements were achieved in the high-volatility setting, in which case LightGBM pinball loss remained equal to 0.0281±0.0024 vs 0.0359±0.0029 for linear Quantile Regression, with the Diebold-Mariano statistic being 3.36±0.44*. Excluding the downside-history features led to an increase in loss to 0.0214±0.0013 and an increase in the ES statistic to 0.034±0.008. The expanding-window model reached 0.0206±0.0013, confirming value from discarding stale observations. Feature importance ranked downside semivariance at 0.142±0.018, VIX at 0.126±0.016, principal-component variance share at 0.109±0.015, and maximum drawdown at 0.103±0.014 (see Figure 1).

6. Conclusion

In addition, the paper shows that the conditional portfolio downside risk can be estimated in a more flexible way by constructing a set of conditional quantiles via rolling non-linear machine learning methods. The lightgbm method identified interactions between volatility, downside history, dependence, and market-state characteristics, whereas the construction of quantiles helped capture the changes in tail behavior of ES. The sequential tests confirmed the outperformance of the model compared to historical simulation, Gaussian garch and linear quantile regression in terms of calibration and tail losses, especially at times of elevated market volatility.

References

[1] Storti, G., & Wang, C. (2022). Nonparametric expected shortfall forecasting incorporating weighted quantiles. *International Journal of Forecasting, 38*(1), 224–239.
[2] Taylor, J. W. (2022). Forecasting value at risk and expected shortfall using a model with a dynamic omega ratio. *Journal of Banking & Finance, 140*, 106519.
[3] Keilbar, G., & Wang, W. (2022). Modelling systemic risk using neural network quantile regression. *Empirical Economics, 62*(1), 93–118.
[4] Candila, V., Gallo, G. M., & Petrella, L. (2023). Mixed-frequency quantile regressions to forecast value-at-risk and expected shortfall. *Annals of Operations Research*.
[5] Li, D., Clements, A., & Drovandi, C. (2023). A Bayesian approach for more reliable tail risk forecasts. *Journal of Financial Stability, 64*, 101098.
[6] Blom, H. M., de Lange, P. E., & Risstad, M. (2023). 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, 16*(7), 312.
[7] Qiu, Z., Lazar, E., & Nakata, K. (2024). VaR and ES forecasting via recurrent neural network-based stateful models. *International Review of Financial Analysis, 92*, 103102.
[8] Lin, W., & Taamouti, A. (2024). Portfolio selection under non-gaussianity and systemic risk: A machine learning based forecasting approach. *International Journal of Forecasting, 40*(3), 1179–1188.

DOI: 10.54254/2755-2721/2026.38688

[9] Candila, V., Petrella, L., & Andreani, M. (2025). Mixed-frequency quantile regression forests for value-at-risk forecasting. *Energy Economics, 149*, 108706.
[10] Vo, M. (2026). Deep quantile forecasting: Evaluating advanced neural networks for multi-horizon value-at-risk. *International Journal of Financial Studies, 14*(8), 218.