Proceedings of ICFTBA 2026 Symposium: Strategic Management & Business Analytics: Data-Driven Performance Models
DOI: 10.54254/2754-1169/2026.LD38484

Statistical Methods for Measuring Financial Market Risk: A Literature Review

Feng Xie

College of International Education, South China Agricultural University, Guangzhou, China
18565166528@163.com

Abstract. Financial market risk arises from adverse changes in asset values and is commonly assessed through volatility, Value at Risk (VaR), and tail risk. This literature review compares the variance–covariance method, Generalized Autoregressive Conditional Heteroskedasticity (GARCH) models, historical simulation, and Extreme Value Theory (EVT). Quantile regression is considered as a supplementary approach to dynamic VaR estimation. The methods are evaluated according to their statistical assumptions, predictive accuracy, tail-risk identification, data requirements, computational complexity, and practical suitability. The literature indicates that the variance–covariance method is efficient but vulnerable to distributional and linearity assumptions. GARCH models respond to time-varying volatility, although their performance depends on model specification and the innovation distribution. Historical simulation is transparent and accommodates portfolio revaluation, but relies on the relevance of past observations. EVT provides a direct framework for extreme losses but is sensitive to threshold selection and limited tail data. Backtesting remains necessary because theoretical flexibility does not ensure accurate forecasts. No method consistently dominates across all market conditions. Reliable measurement requires model selection to reflect return characteristics, portfolio structure, forecast purpose, and data availability. In some applications, combined methods provide a more appropriate representation of volatility dynamics and tail severity.

Keywords: Value at Risk, volatility, GARCH, historical simulation, extreme value theory

1. Introduction

Financial market risk arises when adverse movements in asset prices, interest rates, exchange rates, or other market variables reduce the value of a financial position. Many statistical models of market risk analyse asset returns rather than price levels because returns provide a comparable representation of proportional price changes. However, financial returns often display heavy tails, volatility clustering, and asymmetric movements, making risk estimation more difficult than normal-distribution assumptions suggest [1].

Three interrelated metrics lie at the core of this analytical domain. Volatility describes the dispersion of returns and the general degree of market uncertainty. Value at Risk (VaR) estimates a loss threshold for a specified confidence level and holding period. Tail risk concerns the probability and severity of unusually large losses, including losses beyond the VaR threshold. VaR became

© 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: 10.54254/2754-1169/2026.LD38484

widely used because it presents portfolio risk as a single, interpretable quantity [2]. Nevertheless, it does not indicate the magnitude of losses after the threshold has been exceeded and does not satisfy all properties associated with coherent risk measurement [3].

Statistical methods address these features in different ways. The variance–covariance method provides a simple parametric estimate but commonly relies on distributional and linearity assumptions. Generalized Autoregressive Conditional Heteroskedasticity (GARCH) models allow volatility to change over time. Historical simulation uses the empirical distribution of past returns, while Extreme Value Theory (EVT) focuses on observations in the distributional tail. Quantile regression offers an additional way to estimate conditional risk thresholds directly.

This review compares these methods in terms of statistical assumptions, predictive accuracy, tail-risk identification, data requirements, computational complexity, and practical suitability. It contributes by clarifying the relative strengths and limitations of major statistical risk-measurement methods, thereby supporting more appropriate model selection in financial practice.

2. Foundations of market risk measurement

Market-risk models generally begin with a time series of asset prices and convert these prices into returns. For an asset with price (P_t) at time t, the simple return is

$$R_t = (P_t - P_{t-1})/P_{t-1} \tag{1}$$

while the logarithmic return is

$$r_t = \ln(P_t/P_{t-1}) \tag{2}$$

Returns standardise price changes relative to the initial price and therefore permit comparisons across assets and periods. Logarithmic returns are particularly convenient in time-series analysis because logarithmic returns over successive, non-overlapping periods are additive.

Volatility is commonly measured by the standard deviation of returns, while variance is often modelled directly. As a key indicator of uncertainty, volatility fails to tell between favourable and adverse price movements. Moreover, volatility is time-varying rather than constant over time. Large changes are often followed by further large changes, while relatively stable periods tend to persist. This pattern, known as volatility clustering, is a major reason for using models of time-varying conditional variance [1]. VaR translates the return distribution into a loss threshold. Specifically, VaR represents a selected quantile of the portfolio loss distribution at a given confidence level. A one-day 99% VaR of £1 million indicates that, under the model used, the probability that the next day's loss will not exceed £1 million is approximately 99% [2]. Its definition must therefore specify the portfolio, holding period, and confidence level. VaR estimates can differ when models impose different distributional assumptions or use different historical samples.

Tail risk concerns extreme losses in the adverse region of the distribution. VaR identifies the boundary of this region but does not measure the size of losses beyond it, which contributes to its limitations as a risk measure [3]. Expected Shortfall (ES) addresses this issue by measuring the average loss conditional on losses exceeding the VaR threshold [4].

DOI: 10.54254/2754-1169/2026.LD38484

3. Variance–covariance approaches

The variance–covariance approach, also called the parametric or delta-normal method, estimates portfolio risk from the means, variances, and correlations of asset returns. Portfolio volatility is obtained by combining individual volatilities with a covariance matrix, which records how pairs of returns move together. Under the common assumptions of normally distributed returns and a portfolio whose value changes approximately linearly with its risk factors, VaR can be calculated from a normal-distribution quantile, estimated portfolio volatility, and position value. The RiskMetrics framework helped establish this approach in practical market-risk management [2].

Its advantage is computational efficiency. Once the covariance matrix has been estimated, the risk of a large portfolio can be summarised without revaluing every position under numerous scenarios. Correlations also allow the method to represent diversification, since losses in one asset may be partly offset by movements in another. It is useful for routine reporting and portfolios dominated by linear instruments, including equities, bonds and foreign-exchange positions.

The simplicity of the method is also the source of its main limitations. Empirical returns frequently exhibit heavy tails and asymmetry. For this reason, a normal distribution may assign too little probability to extreme losses [1]. Estimates based on a fixed historical window can respond slowly when volatility and correlations change abruptly. The linear approximation is also less reliable for options and other instruments with nonlinear payoffs. VaR may consequently understate exposure during market stress. These limitations do not make the method uniformly inaccurate. Performance depends on the portfolio, estimation window, confidence level, and procedure used to update volatility. Hendricks compared twelve VaR approaches across 1,000 foreign-exchange portfolios, illustrating why estimates should be assessed against observed outcomes rather than judged from theoretical simplicity alone [5]. The variance–covariance method is best treated as an efficient benchmark whose assumptions must be checked.

4. GARCH-based models

The assumption of constant volatility is difficult to reconcile with the clustering observed in financial returns. Engle introduced the Autoregressive Conditional Heteroskedasticity (ARCH) model to allow conditional variance to depend on previous shocks [6]. Bollerslev later developed the GARCH model by allowing current variance to depend on both past squared shocks and past conditional variance [7]. This extension provides a more parsimonious representation of persistent volatility than an ARCH model containing many lagged terms.

The widely used GARCH(1,1) specification contains one lag of each component. A large unexpected return increases the next estimate of conditional variance, while the influence of earlier variance gradually declines rather than disappearing immediately. The model can therefore reproduce periods in which high volatility is followed by high volatility and tranquil conditions persist for some time. Forecast conditional volatility may then be used as an input to a VaR calculation. In contrast with a variance–covariance estimate based on an unchanging historical variance, GARCH updates the risk estimate as new information becomes available.

While this responsiveness is useful, it does not eliminate distributional assumptions. A standard GARCH model is often estimated with normally distributed innovations, even though standardised financial returns may remain heavy-tailed. Student's t or skewed distributions can be used instead, although the resulting forecasts depend on the selected specification. The basic GARCH model is also symmetric: positive and negative shocks of equal size have the same effect on estimated variance. This can be restrictive when adverse returns cause a stronger volatility response, a pattern

DOI: 10.54254/2754-1169/2026.LD38484

usually described as the leverage effect. More elaborate variants address asymmetry, but introduce additional parameters and model-selection decisions.

Evidence does not support treating greater complexity as automatically superior. Hansen and Lunde compared 330 ARCH-type models using exchange-rate and IBM return data [8]. They found no evidence that GARCH(1,1) was inferior to more sophisticated alternatives for the exchange-rate series. For IBM returns, however, models that accommodated a leverage effect performed better. The comparison shows that forecast performance can vary with both the asset and the volatility feature being modelled. GARCH models are therefore valuable when volatility is dynamic, but their accuracy depends on the innovation distribution, treatment of asymmetry, estimation sample, and forecast horizon. They provide a flexible framework rather than a single specification that is suitable in every market.

5. Historical simulation

Historical simulation estimates VaR from the empirical distribution of past market movements. Changes in relevant risk factors over a selected historical window are applied to the current portfolio, producing hypothetical profits and losses. These outcomes are ranked, and the loss at the required percentile becomes the VaR estimate. A 99% VaR, for example, is taken from the most adverse one per cent of simulated outcomes.

Unlike the variance–covariance approach, historical simulation does not require an explicit normal-distribution assumption. It preserves dependence patterns that occurred in the data and, under full revaluation of positions, can accommodate nonlinear instruments. Its steps are also comparatively easy to explain. Hendricks included historical approaches in his comparison of twelve VaR models, reflecting their practical importance [5]. The absence of a specified parametric distribution does not make the method assumption-free. Equal-weighted historical simulation treats observations within the selected window as equally relevant and assumes that their empirical distribution is informative about future risk. A window dominated by calm observations may underestimate risk when conditions deteriorate. Conversely, an extreme event can continue to influence the estimate until it leaves the window, even after current conditions have stabilised.

Tail estimation creates an additional data problem. With 250 daily observations, a 99% estimate is determined by only a few of the largest losses, making the result sensitive to individual observations and window length. Pritsker's analysis of historical simulation and its extensions identifies weaknesses arising from time-varying volatility and the statistical properties of the resulting estimates [9]. Weighted historical simulation gives more importance to recent observations, while Filtered Historical Simulation (FHS) adjusts past returns for changes in volatility. These extensions improve responsiveness but require additional modelling choices. Historical simulation is therefore attractive when transparency and portfolio revaluation are priorities, but its reliability depends on sample size and the relevance of past market conditions.

6. Extreme value methods

Extreme Value Theory (EVT) examines observations in the far tail of a distribution rather than attempting to fit the entire return series. This focus is relevant because the largest financial losses often occur more frequently than a normal model predicts. Embrechts, et al. provide the theoretical foundation for applying extreme-value methods to insurance and finance [10].

Two main approaches are available. The block-maxima method divides the sample into periods and retains the largest loss from each block. The Peaks Over Threshold (POT) method uses all losses

DOI: 10.54254/2754-1169/2026.LD38484

above a selected high threshold and commonly models the excesses with a Generalized Pareto Distribution (GPD). POT can make more efficient use of scarce tail data because it may include several extreme observations from the same period. The fitted tail can then be extrapolated to obtain extreme VaR and Expected Shortfall estimates. Danielsson and de Vries used a semi-parametric approach in which the centre of the distribution was estimated empirically and the most extreme returns were modelled parametrically [11].

Extrapolation is both the main advantage and the main source of uncertainty. Conventional historical simulation is bounded by the scenarios contained in its sample, whereas EVT can estimate quantiles outside the observed range. The result, however, is sensitive to threshold selection. A low threshold supplies more observations but may include data that do not follow the tail distribution. A very high threshold reduces this bias but leaves few observations, increasing estimation variance. Structural changes and dependence between extreme returns create further difficulties.

Conditional heteroskedasticity also complicates a direct application of EVT. McNeil and Frey addressed this problem through a two-stage method [12]. A GARCH model first estimates current volatility, after which EVT is applied to the tail of the standardised residuals. Their approach shows that volatility and tail models can perform complementary tasks. GARCH captures time-varying volatility conditions, while EVT describes the severity of extreme innovations. Although this combination is relevant to high-confidence VaR, it requires more data, modelling choices, and diagnostic testing than simpler approaches. EVT is therefore most valuable when the purpose is explicit tail measurement, rather than routine volatility reporting or computational simplicity.

7. Comparative assessment

Comparing market-risk methods requires more than identifying which model produces the largest or smallest VaR. A useful model should generate thresholds with an appropriate rate of exceptions, where an exception occurs when realised loss exceeds predicted VaR. Kupiec's test examines whether the observed proportion of exceptions is consistent with the stated confidence level [13]. Christoffersen extended this logic by testing conditional coverage, which also considers whether exceptions occur independently rather than in clusters [14]. A model may produce the correct total number of exceptions while still reacting inadequately to changing conditions.

The four approaches differ substantially in their assumptions and operational characteristics. The variance–covariance method has the lowest computational burden but relies strongly on distributional and linearity assumptions. GARCH improves responsiveness to volatility changes, although tail estimates depend on the innovation distribution. Historical simulation offers intuitive full revaluation and avoids a fixed parametric distribution, but its estimates depend on the relevance and size of the historical sample. EVT provides the most direct treatment of extreme quantiles, at the cost of threshold uncertainty and limited tail observations.

Predictive accuracy is therefore conditional on both data and purpose. During stable periods, a simple model may provide adequate estimates without the estimation risk introduced by additional parameters. When volatility changes rapidly, GARCH-based forecasts are more responsive. Historical simulation is useful when portfolio nonlinearity and transparent scenario construction are important, provided that the historical window remains representative. For very high confidence levels, EVT is better aligned with the statistical problem, as it models the tail directly.

These differences provide a practical basis for context-dependent method selection. The entries do not imply a fixed ranking, they indicate relative suitability under each stated condition. The choice should also reflect forecast horizon, available observations, portfolio composition, and the cost of model failure. Backtesting must be interpreted cautiously because high-confidence VaR

DOI: 10.54254/2754-1169/2026.LD38484

produces few exceptions, limiting the statistical power of tests. Model selection is an iterative process involving estimation, diagnostic checking, backtesting, and revision.

8. Quantile regression extensions

Quantile regression provides an alternative to models that estimate the complete return distribution. It estimates a selected conditional quantile directly from available information. Engle and Manganelli developed the Conditional Autoregressive Value at Risk (CAViaR) framework, in which the VaR quantile evolves according to its past values and observed returns [15]. This approach can represent changing risk without imposing normality on the entire distribution.

CAViaR is relevant because it treats VaR as a dynamic quantity rather than deriving it solely from a volatility forecast. Its performance nevertheless depends on the chosen quantile specification, and estimation becomes less precise at very high confidence levels where exceedances are scarce. It also provides limited information about the severity of losses beyond the estimated quantile unless combined with an additional tail measure. Quantile regression therefore broadens the modelling framework but does not remove the need for backtesting or tail-risk analysis.

9. Conclusion

This paper compares the principal statistical methods used to measure financial market risk through volatility, VaR, and tail risk. The variance-covariance approach is efficient. However, it is sensitive to distributional and linearity assumptions. GARCH models capture time-varying volatility. Still, they depend on model specification. Historical simulation is transparent. Yet it relies on the relevance of past observations. EVT directly addresses extreme losses. Nevertheless, it is constrained by threshold selection and limited tail data. Quantile regression provides an additional framework for estimating dynamic risk thresholds.

The review suggests that no method excels universally. Model selection should reflect return characteristics, portfolio composition, forecast horizon, confidence level, data availability, and the intended use of the estimate, while performance should be evaluated through appropriate backtesting. Combined approaches such as GARCH-EVT may represent both volatility dynamics and tail severity, although they require additional estimation and diagnostic testing. Because existing studies differ in assets, sample periods, and evaluation criteria, this review cannot establish a universal ranking. Reliable market-risk measurement ultimately depends on explicit assumptions, context-sensitive model selection, and continuing validation.

References

[1] Cont, R. (2001). Empirical properties of asset returns: Stylized facts and statistical issues. *Quantitative Finance, 1*(2), 223–236. https://doi.org/10.1080/713665670

[2] J. P. Morgan/Reuters. (1996). RiskMetrics—Technical document (4th ed.). Available at: https://www.msci.com/documents/10199/5915b101-4206-4ba0-aee2-3449d5c7e95a (Accessed: 28 August 2026).

[3] Artzner, P., Delbaen, F., Eber, J.-M., & Heath, D. (1999). Coherent measures of risk. *Mathematical Finance, 9*(3), 203–228. https://doi.org/10.1111/1467-9965.00068

[4] Basel Committee on Banking Supervision. (2019). Minimum capital requirements for market risk. Bank for International Settlements. Available at: https://www.bis.org/bcbs/publ/d457.htm (Accessed: 28 August 2026).

[5] Hendricks, D. (1996). Evaluation of Value-at-Risk models using historical data. *Federal Reserve Bank of New York Economic Policy Review, 2*(1), 39–69.

[6] Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica, 50*(4), 987–1007. https://doi.org/10.2307/1912773

DOI: 10.54254/2754-1169/2026.LD38484

[7] Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics, 31*(3), 307–327. https://doi.org/10.1016/0304-4076(86)90063-1
[8] Hansen, P. R., & Lunde, A. (2005). A forecast comparison of volatility models: Does anything beat a GARCH(1,1)? *Journal of Applied Econometrics, 20*(7), 873–889. https://doi.org/10.1002/jae.800
[9] Pritsker, M. (2006). The hidden dangers of historical simulation. *Journal of Banking & Finance, 30*(2), 561–582.
[10] Embrechts, P., Klüppelberg, C., & Mikosch, T. (1997). *Modelling extremal events for insurance and finance*. Springer. https://doi.org/10.1007/978-3-642-33483-2
[11] Danielsson, J., & de Vries, C. G. (2000). Value-at-risk and extreme returns. *Annales d'Économie et de Statistique, 60*, 239–270. https://doi.org/10.2307/20076262
[12] McNeil, A. J., & Frey, R. (2000). Estimation of tail-related risk measures for heteroscedastic financial time series: An extreme value approach. *Journal of Empirical Finance, 7*(3–4), 271–300.
[13] Kupiec, P. H. (1995). Techniques for verifying the accuracy of risk measurement models. *The Journal of Derivatives, 3*(2), 73–84. https://doi.org/10.3905/jod.1995.407942
[14] Christoffersen, P. F. (1998). Evaluating interval forecasts. *International Economic Review, 39*(4), 841–862.
[15] Engle, R. F., & Manganelli, S. (2004). CAViaR: Conditional autoregressive value at risk by regression quantiles. *Journal of Business & Economic Statistics, 22*(4), 367–381. https://doi.org/10.1198/073500104000000370