Proceedings of the 5th International Conference on Financial Technology and Business Analysis
DOI: 10.54254/2754-1169/2026.39134

# *Measuring Extreme Risk in the CSI 300 Index: A GARCH-EVT Approach under the FRTB Framework*

**Manye Wang**

*Shenzhen MSU-BIT University, Shenzhen, China*
*1120240814@smbu.edu.cn*

***Abstract.*** The Fundamental Review of the Trading Book (FRTB) replaces Value-at-Risk (VaR) with 97.5% Expected Shortfall (ES) as the benchmark for market-risk internal models. Features of China's A-share market—including price limits, T+1 settlement, and substantial retail participation—produce return tails that deviate markedly from normal and Student-t distributions. Standard models may therefore overestimate extreme risk. Using the CSI 300 Index, this study applies a filtered extreme value theory (EVT) framework. An EGARCH model with Student-t innovations first filters conditional heteroscedasticity; a peaks-over-threshold/generalized Pareto distribution (POT/GPD) model is then fitted to the tails of the standardized residuals. Rolling out-of-sample ES forecasts are subsequently evaluated by backtesting. Four findings emerge. First, EGARCH-t provides the best fit among the GARCH-family specifications and captures both asymmetric volatility and heavy tails. Second, filtered EVT passes the Acerbi-Székely backtest for 97.5% ES, whereas historical simulation and GARCH-t do not. Third, filtered EVT is the only model retained in the Model Confidence Set (MCS). Fourth, under the FRTB framework, filtered EVT requires approximately 16% less capital than historical simulation. A decomposition attributes this reduction to improved tail modeling rather than volatility modeling.

***Keywords:*** GARCH-EVT, expected shortfall, extreme value theory, FRTB, CSI 300

## 1. Introduction

Implemented under a normality assumption, Value-at-Risk (VaR) responds slowly to tail losses, a limitation the 2008 global financial crisis made explicit. VaR identifies a loss threshold at a specified confidence level but provides no information about losses beyond that threshold. In its 2016 Minimum Capital Requirements for Market Risk (FRTB; BCBS d424), the Basel Committee on Banking Supervision (BCBS) replaced the 99% VaR benchmark with 97.5% Expected Shortfall (ES) under the internal models approach. As the conditional mean loss beyond the VaR threshold, ES addresses two limitations of VaR at once, namely the lack of subadditivity and insensitivity to tail severity [1].

However, a regulatory change does not ensure that the prescribed method is directly applicable to every market. The A-share market is characterized by daily price limits (±10% or ±20%), T+1 settlement, retail investors who account for approximately 60% of turnover, frequent policy intervention, and long-standing constraints on short selling. These institutional features distinguish

© 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: 10.54254/2754-1169/2026.39134

broad indices such as the CSI 300 from their developed-market counterparts, producing heavier tails, stronger volatility clustering, and extreme movements that often coincide with jumps [2]. This study addresses three questions: To what extent do normal- and Student-t-based estimates of FRTB ES misstate extreme risk in the A-share market? Does any discrepancy arise from volatility modeling or from the assumed tail distribution? Can a filtered EVT framework address the discrepancy?

This study draws on two strands of literature. The first is the volatility literature: the ARCH model [3], the GARCH model [4], and subsequently GJR-GARCH and EGARCH, which allow for the leverage effect [5], all seek to remove conditional heteroscedasticity from returns. The second strand is extreme value theory. Filtered EVT was introduced in [6], where GARCH first removes volatility and POT/GPD is then fitted to the standardized residuals; the theoretical underpinnings were supplied later in [7]. More recent work develops this literature in several directions at once, covering EVT models with covariates, hybrid and mixed-frequency estimators, Bayesian realized-volatility models, recurrent neural-network forecasts, and backtesting procedures that account for estimation error [8-14]. Two gaps remain. Evidence on GARCH-EVT models is drawn mostly from developed markets, so the bias associated with A-share institutional features has rarely been quantified systematically. Backtesting, moreover, continues to focus primarily on VaR, leaving ES calibration under the FRTB largely unexamined.

This study addresses these gaps with CSI 300 data through a sequence of volatility filtering, extreme-value fitting, ES estimation, and regulatory benchmarking. Three findings emerge. First, among the GARCH-family specifications an EGARCH model with Student-t innovations provides the best fit, and CSI 300 volatility exhibits both asymmetry and heavy tails. Second, at the 97.5% ES level emphasized by the FRTB, filtered EVT is the only model retained in the Model Confidence Set (MCS), with an FRTB capital charge approximately 16% below that of historical simulation; a decomposition attributes this reduction to improved tail modeling (GPD rather than the Student-t distribution) rather than to volatility modeling. Third, filtered EVT significantly outperforms GARCH-t at the 97.5% ES level (Diebold-Mariano $p = 0.026$). The two models share the same EGARCH volatility equation, so their difference isolates the importance of the assumed tail distribution.

This study makes three contributions. First, it quantifies the structural bias of standard risk measures under A-share institutional conditions: conventional methods overestimate ES and increase the capital charge by approximately 16%. Second, an attribution analysis shows that this bias arises from inadequate tail modeling rather than an overly simple volatility specification, indicating that the principal improvement should be made in the tail model. Third, the study develops an integrated evaluation framework that combines GARCH selection, ES backtesting, and an assessment of economic value. By applying the Acerbi-Székely ES backtest rather than relying only on VaR tests, the framework is aligned with the requirements of the FRTB internal models approach.

## 2. Data and methodology

### 2.1. Data source and sample description

Daily closing prices of the CSI 300 Index (ticker: sh000300) from January 7, 2002, through August 14, 2026 make up the sample, which covers 5,970 trading days. The starting date is not arbitrary: by 2002 the institutional structure of the A-share market was largely in place. The window also contains several periods of severe market stress, among them the 2008 global financial crisis, the 2015

DOI: 10.54254/2754-1169/2026.39134

market crash and the 2020 COVID-19 shock, which is what supplies sufficient tail observations for the analysis. Daily log returns are calculated as r_t = ln(P_t/P_{t-1}) × 100 and are expressed as percentages.

Over the sample period the CSI 300 generated an average daily return of approximately 0.02%, a standard deviation of 1.56%, skewness of -0.38 and kurtosis of 7.68. Such kurtosis stands well above the normal-distribution benchmark of 3, so the return distribution carries excess kurtosis and heavy tails. Figure 1 illustrates these features in two ways. The left panel shows an empirical histogram with visibly heavier tails than the fitted normal distribution, while the Q-Q plot in the right panel deviates from the reference line at both ends: extreme returns occur more frequently than normality would allow.

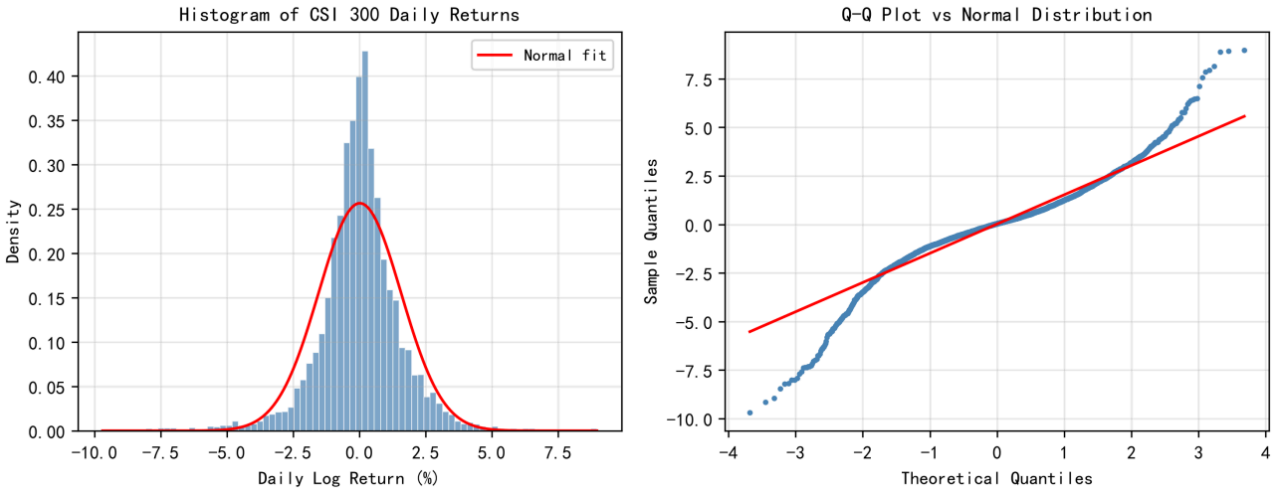



Figure 1. Distribution of daily CSI 300 returns and the corresponding normal Q-Q plot

## 2.2. GARCH-family model selection

Volatility modeling precedes EVT estimation because the tail fit assumes approximately independent and identically distributed (i.i.d.) residuals; conditional heteroscedasticity must therefore be removed before any tail distribution is fitted. Four GARCH-family specifications (GARCH, GJR-GARCH, EGARCH and APARCH) are compared under four residual distributions: normal, Student-t, skewed-t, and generalized error distribution (GED). Selection rests primarily on the Akaike information criterion (AIC), supplemented by the Ljung-Box Q test for residual autocorrelation, the ARCH-LM test for remaining heteroscedasticity, and the Engle-Ng sign-bias test for asymmetric effects.

## 2.3. Extreme value theory and POT/GPD fitting

Institutional features of the A-share market, including price limits, T+1 settlement, and substantial retail participation, may produce tails that are heavier than those accommodated by a Student-t distribution and are therefore not fully captured by GARCH alone. After returns are filtered to obtain approximately i.i.d. standardized residuals, the tails are modeled using the peaks-over-threshold (POT) method. The threshold u is selected jointly from the mean excess (ME) plot and the stability path of the shape parameter ξ (Figure 2), reducing reliance on any single selection rule. Exceedances are then fitted with the generalized Pareto distribution (GPD), from which VaR and ES are estimated

DOI: 10.54254/2754-1169/2026.39134

at the 97.5% and 99% confidence levels. The 97.5% level corresponds to the FRTB internal models standard, while the 99% level is included for comparison.

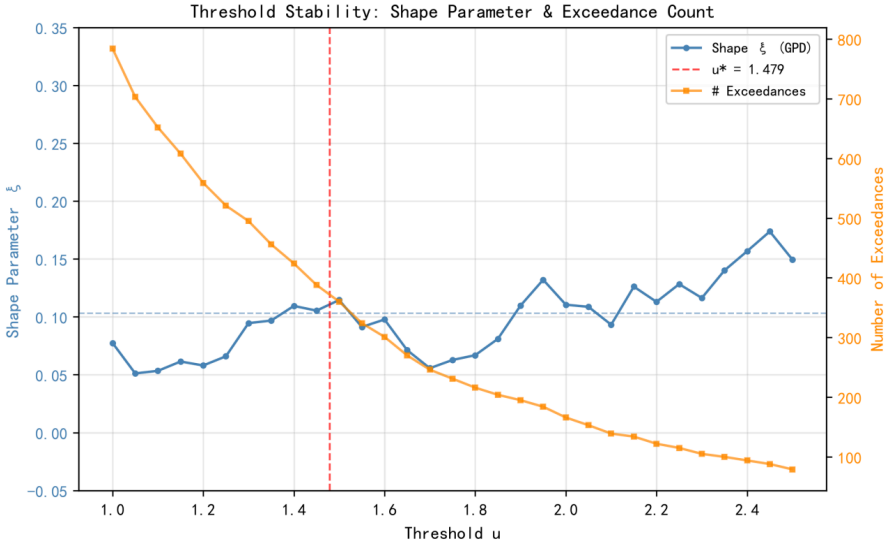



Figure 2. Stability of the GPD shape parameter ξ across thresholds; the red dashed line denotes the selected threshold (u* = 1.479)

### 2.4. Backtesting and model comparison

Backtesting proceeds at two levels. VaR forecasts are evaluated using the proportion-of-failures test [15] and the conditional-coverage test [16]. ES forecasts are evaluated using the z1 unbiasedness test and z2 conditional-coverage test of [1]. The z1 test examines whether the mean ES forecast error is zero, whereas the z2 test examines whether forecast errors cluster over time. The three candidate models are compared pairwise using the Diebold-Mariano test and jointly using the Model Confidence Set (MCS) procedure [17]. A rolling window of 750 trading days is advanced one day at a time to preserve the integrity of the out-of-sample evaluation.

## 3. Empirical results

### 3.1. GARCH-family selection results

Table 1 reports the eight best-performing models according to AIC, with EGARCH-t first at 20,124.12 and EGARCH-skew-t close behind ($\Delta AIC < 2$). The skewed-t specification adds a parameter without a meaningful improvement in fit, so EGARCH-t is retained on grounds of parsimony; EGARCH-GED follows in third place at 20,124.42. Two features of CSI 300 volatility lie behind this ranking. The first is asymmetry, which the leverage component of EGARCH captures, and the second is heavy tails, which the Student-t distribution represents more effectively than the normal distribution.

The diagnostic tests support this selection. At lag 20 the Ljung-Box p-value of the standardized residuals from the preferred model is 0.397 and the ARCH-LM p-value is 0.091, so neither test rejects the null of i.i.d. residuals and the volatility filter is adequate. The sign-bias p-values are similarly reassuring, at 0.103 for negative shocks against 0.914 for positive shocks, which is what should be observed once the leverage effect has been captured.

DOI: 10.54254/2754-1169/2026.39134

Table 1. GARCH-family model selection results (top eight by AIC)

| Rank | Model | Distribution | Order | AIC | LB(20) p | ARCH-LM p |
|---|---|---|---|---|---|---|
| 1 | EGARCH | t | (1,0) | 20,124.12 | 0.397 | 0.091 |
| 2 | EGARCH | skew-t | (1,0) | 20,124.29 | 0.429 | 0.104 |
| 3 | EGARCH | GED | (1,0) | 20,124.42 | 0.597 | 0.206 |
| 4 | EGARCH | t | (0,0) | 20,126.01 | 0.391 | 0.089 |
| 5 | APARCH | GED | (1,0) | 20,130.50 | 0.873 | 0.555 |
| 6 | GJR | GED | (1,0) | 20,140.74 | 0.975 | 0.884 |
| 7 | GARCH | GED | (1,0) | 20,142.45 | 0.972 | 0.899 |
| 8 | GARCH | t | (1,0) | 20,149.38 | 0.958 | 0.839 |

Note: LB(20) denotes the p-value from the Ljung-Box test of standardized residuals at lag 20; ARCH-LM denotes the p-value from the test for remaining ARCH effects.

### 3.2. Extreme value fitting and threshold stability

Using the standardized residuals from the EGARCH-t model, the threshold is selected jointly from the ME plot and the stability path of $\xi$. As shown in Figure 2, $\xi$ remains between approximately 0.06 and 0.12 as the threshold varies from 1.3 to 1.7, suggesting that the GPD fit is not highly sensitive to the selected cutoff. The chosen threshold is $u^* = 1.4788$, corresponding to approximately the 93.8th percentile and yielding 369 exceedances (6.18% of the sample). The fitted GPD has a shape parameter of $\xi = 0.103$ and a scale parameter of $\beta = 0.602$. Because $\xi$ lies within the plausible interval (0, 0.35), the estimated tail is heavy.

Across the full sample the 97.5% EVT-VaR and EVT-ES estimates are 2.99% and 4.06%. Historical simulation (HS) produces corresponding estimates of 3.34% and 4.95%, both above the EVT figures. The gap has a straightforward source. HS relies entirely on the empirical distribution of historical observations, and with so few extreme observations available its representation of the tail is coarse, which yields conservative estimates. Fitting the GPD directly to exceedances instead allows EVT to pick up both the rate and the shape of tail decay, and the resulting tail-risk estimates are more precise.

### 3.3. Backtesting results

The rolling out-of-sample backtest in Table 2 is built on a 750-day window and 5,198 forecast observations. At the 97.5% level the EVT VaR exceedance rate is 2.64%, close to the nominal rate of 2.5%; the Kupiec test returns a p-value of 0.535, so the null hypothesis of correct unconditional coverage cannot be rejected. On the ES side the Acerbi-Székely test gives $z1 = -0.094$ and $z2 = -0.038$, both inside the 5% acceptance interval [-0.86, +0.70], which is no evidence of systematic bias. A p-value of 0.179 from the conditional-coverage test likewise points to no significant clustering of exceedances. Figure 3 displays the VaR and ES exceedances over the backtest period.

DOI: 10.54254/2754-1169/2026.39134

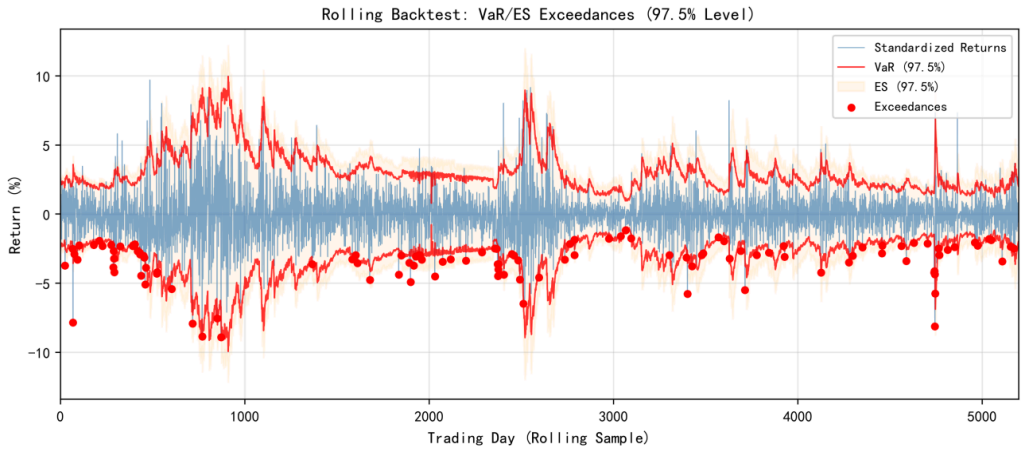



Figure 3. Rolling 97.5% VaR and ES exceedances (red dots indicate exceedance observations)

Table 2. Rolling out-of-sample backtest results for the filtered EVT model

| Confidence | Exceedance Rate | Expected | Kupiec p | AS $z_1$ | AS $z_2$ | Cond. cov. p |
|---|---|---|---|---|---|---|
| 97.5% | 2.64% | 2.50% | 0.535 | -0.094 | -0.038 | 0.179 |
| 99.0% | 1.15% | 1.00% | 0.275 | -0.204 | -0.043 | 0.034* |

Note: * indicates significance at the 5% level; AS denotes the Acerbi-Székely ES backtest, and Cond. cov. the conditional-coverage test. The 5% acceptance interval for the Acerbi-Székely statistics is [-0.86, +0.70].

Once the confidence level moves to 99%, the exceedance rate falls to 1.15% and the Kupiec test yields a p-value of 0.275, so unconditional coverage remains adequate. Conditional coverage behaves differently: its p-value drops to 0.034 and is rejected at the 5% level, meaning that exceedances cluster further into the tail. Institutional features of the A-share market are consistent with this pattern, since price limits can reduce liquidity during extreme episodes and thereby produce consecutive large movements such as limit-down days. GARCH captures part of the resulting volatility clustering but does not fully represent institution-induced jumps.

### 3.4. Model comparison and economic value

The incremental performance of filtered EVT is assessed with pairwise Diebold-Mariano tests among EVT, HS and GARCH-t at the 97.5% and 99% levels. Two results stand out. At the 97.5% VaR level the DM statistic for EVT relative to HS is 4.08 ($p < 0.001$), a significant difference; at the ES level the DM statistic for EVT relative to GARCH-t is 1.94 ($p = 0.026$), significant at the 5% level. The superior predictive ability (SPA) test does not reject the null hypothesis in any of the four settings (VaR and ES at 97.5% and 99%), consistent with EVT performing at least as well as the best competing model.

The MCS results are more discriminating. At the 97.5% ES level, EVT is the only model retained in the final set, whereas GARCH-t ($p = 0.088$) and HS ($p = 0.077$) are both excluded. At the 99% VaR level, EVT and GARCH-t are retained, while HS is excluded. Thus, at the 97.5% ES level emphasized by the FRTB, filtered EVT is uniquely retained among the three models.

In economic terms the average FRTB internal-models capital charge is 3.56% for EVT against 4.23% for HS, a reduction of approximately 16%. EVT also records the lowest ES exceedance rate, at 0.96%. Figure 4 compares capital charges and ES exceedance rates across the three models, and

DOI: 10.54254/2754-1169/2026.39134

the statistical and the economic evidence point in the same direction, toward the filtered EVT specification.

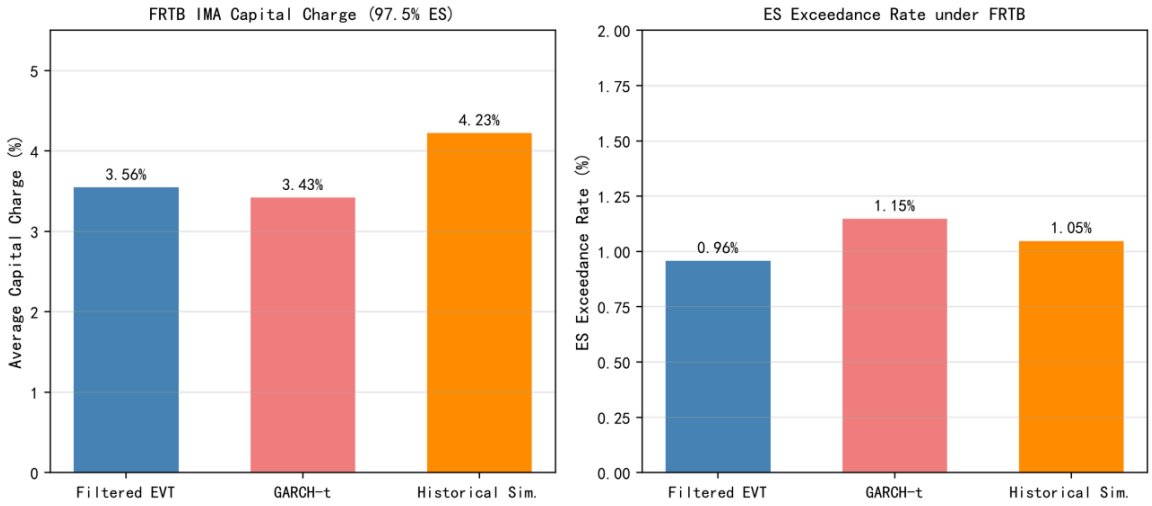



Figure 4. FRTB capital charges and ES exceedance rates across the three models

## 4. Discussion

### 4.1. Implications of volatility model selection

EGARCH-t provides the best fit among the GARCH-family specifications, with two implications. First, the logarithmic EGARCH specification permits asymmetric volatility responses to positive and negative shocks. Its selection indicates a leverage effect in CSI 300 returns: negative shocks increase volatility more than positive shocks of equal magnitude. This result is consistent with A-share market microstructure, in which frequent policy intervention and substantial retail participation may intensify volatility during sell-offs relative to rallies. Second, the Student-t distribution outperforms the normal distribution, indicating that the standardized residuals retain excess kurtosis after volatility filtering and that normality provides an inadequate representation of the tails. Although APARCH-skew-t is more flexible, it yields no meaningful improvement in AIC; the additional parameters are therefore not warranted.

### 4.2. Incremental value of tail modeling: attribution analysis

Section 3 shows that EVT reduces the capital charge by approximately 16% relative to HS without identifying where the reduction originates. Decomposing the difference requires the three models to be laid out along their volatility and tail components. HS parameterizes neither conditional volatility nor the tail distribution and relies entirely on the empirical history. GARCH-t models volatility with EGARCH but assumes Student-t residuals. Filtered EVT keeps that same EGARCH volatility filter and models the residual tail with the GPD instead.

Such an arrangement makes attribution reasonably clean. What separates GARCH-t from HS mixes two things at once, better volatility dynamics and better tail modeling: conditional variance is allowed to vary over time, so volatility clustering is captured, yet the tail remains constrained by the parametric form of the Student-t distribution. What separates filtered EVT from GARCH-t isolates a single element, the tail specification, because the two models use the same EGARCH volatility equation and only the distribution of the residual tail differs.

DOI: 10.54254/2754-1169/2026.39134

The Diebold-Mariano results support this attribution indirectly. At the 97.5% ES level, EVT against GARCH-t gives DM = 1.94 ($p = 0.026$), a statistically significant difference. Since the volatility equation is shared, the performance gap is attributable to the tail specification: the GPD estimates the rate of tail decay through $\xi$, whereas the Student-t decays at a rate governed by its degrees of freedom and need not track the heavy A-share tail closely. At the 97.5% VaR level, EVT against HS gives DM = 4.08 ($p < 0.001$), and this larger value reflects volatility and tail improvements contributing together. Improved tail modeling accounts for most of the capital reduction.

The policy implication is clear. Under the FRTB, improving risk measurement does not necessarily require a more complex volatility model; moving from GARCH to EGARCH yields limited gains. Greater improvement instead comes from modeling the tail more flexibly. Although the Student-t distribution is heavier-tailed than the normal distribution, its decay is constrained by its parametric form. Through the estimated shape parameter $\xi$, the GPD adapts to the empirical tail and represents extreme losses more accurately.

## 5. Conclusion

Using the CSI 300 Index, this study evaluates extreme-risk measurement in the A-share market through four stages: volatility filtering, extreme-value fitting, ES estimation, and regulatory benchmarking. The main conclusions are as follows.

First, the preferred volatility specification is an EGARCH model with Student-t innovations, and CSI 300 returns exhibit both asymmetric volatility and heavy tails. The leverage effect estimated for these returns is consistent with an asymmetric market response to policy news and to substantial retail participation, while excess kurtosis persists in the standardized residuals after volatility filtering.

Second, at the 97.5% ES level emphasized by the FRTB, filtered EVT is the only model the MCS retains, and its FRTB capital charge runs approximately 16% below that of historical simulation. According to the attribution analysis, this reduction arises from GPD-based tail modeling rather than from the volatility specification: filtered EVT and GARCH-t use the same EGARCH equation, so the significant difference in ES performance between them (DM = 1.94, $p = 0.026$) reflects only their different representations of the tail. A subsample analysis remains for future research.

Third, conventional risk measures exhibit structural bias under A-share institutional conditions: standard methods overestimate ES and increase the capital charge by approximately 16%. These findings favor a more flexible tail model over additional complexity in the volatility specification.

This study has several limitations. It examines only one broad market index, does not incorporate a jump-diffusion process, and does not evaluate performance across subsamples. Future research could extend the analysis to other indices and markets, introduce jump components to represent the extreme tail more accurately, conduct subsample backtests to assess time variation in EVT performance, and apply the framework to credit and operational risk.

## References

[1] Acerbi, C., & Székely, B. (2023). Backtestability and the ridge backtest. *Frontiers of Mathematical Finance, 2*(4), 497-521.

[2] Tong, B., Diao, X., & Li, X. (2024). Forecasting VaRs via hybrid EVT with normal and non-normal filters: A comparative analysis from the Chinese stock market. *Pacific-Basin Finance Journal, 83*, 102271.

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

DOI: 10.54254/2754-1169/2026.39134

[4] Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics, 31*(3), 307-327.
[5] Glosten, L. R., Jagannathan, R., & Runkle, D. E. (1993). On the relation between the expected value and the volatility of the nominal excess return on stocks. *Journal of Finance, 48*(5), 1779-1801.
[6] 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.
[7] Patton, A. J., Ziegel, J. F., & Chen, R. (2019). Dynamic semiparametric models for expected shortfall (and value-at-risk). *Journal of Econometrics, 211*(2), 388-413.
[8] Barendse, S., Kole, E., & van Dijk, D. (2023). Backtesting value-at-risk and expected shortfall in the presence of estimation error. *Journal of Financial Econometrics, 21*(2), 528-568.
[9] Luo, Y., Xue, X., & Izzeldin, M. (2024). When MIDAS meets LASSO: The power of low-frequency variables in forecasting value-at-risk and expected shortfall. *Journal of Financial Econometrics, 23*(1), Article nbae016.
[10] James, R., Leung, H., Leung, J. W. Y., & Prokhorov, A. (2023). Forecasting tail risk measures for financial time series: An extreme value approach with covariates. *Journal of Empirical Finance, 71*, 29-50.
[11] Chen, C. W. S., Watanabe, T., & Lin, E. M. H. (2023). Bayesian estimation of realized GARCH-type models with application to financial tail risk management. *Econometrics and Statistics, 28*, 30-46.
[12] Lyu, Y., Qin, F., Ke, R., Yang, M., & Chang, J. (2024). Forecasting the VaR of the crude oil market: A combination of mixed data sampling and extreme value theory. *Energy Economics, 133*, 107500.
[13] Krężołek, D., & Piontek, K. (2024). Volatility and models based on the extreme value theory for gold returns. *Statistics in Transition New Series, 25*(2), 1-22.
[14] 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.
[15] Kupiec, P. H. (1995). Techniques for verifying the accuracy of risk management models. *Journal of Derivatives, 3*(2), 73-84.
[16] Christoffersen, P. F. (1998). Evaluating interval forecasts. *International Economic Review, 39*(4), 841-862.
[17] Hansen, P. R., Lunde, A., & Nason, J. M. (2011). The model confidence set. *Econometrica, 79*(2), 453-497.