Median-Anchored Adjustment of Joint VaR–ES Forecasts

Dingli Wang* Tae-Hwy Lee†

August 1, 2026

Abstract

Risk managers often work with a fitted forecasting model they cannot replace even when its tail forecasts need adjustment. We propose a median-anchored rule that multiplies the distances from the fitted median to Value-at-Risk (VaR) and Expected Shortfall (ES) by a common positive multiplier. The rule preserves VaR–ES ordering, and minimizing VaR check loss gives a closed-form weighted-quantile estimator. We establish consistency, give an asymptotic distribution under high-level conditions accounting for baseline estimation, and derive a VaR coverage-error bound at the forecast origin. When the same multiplier correctly adjusts both VaR and ES, the adjusted pair has lower conditional zero-homogeneous Fissler–Ziegel (FZ0) risk at that origin. Monte Carlo experiments examine this condition and several forms of misspecification. In rolling S&P 500 forecasts the adjustment lowers FZ0 loss in all four GARCH cells, each pairwise significant, with the largest and most robust gain, 10.4%, for Gaussian GARCH at the 1% tail; the 5% gains do not survive the most conservative family-wise adjustments. A quantile-regression baseline marks the boundary: when the fitted tail is already flexible, one multiplier adds little. Filtered historical simulation quantifies what standardized residuals and conditional scales add when available.

Keywords: Backtesting; Elicitability; Fissler–Ziegel Score; Forecast Evaluation; Model Risk; Quantile Regression.

*Corresponding author. Department of Economics, University of California, Riverside, CA 92521, USA. E-mail: dwang177@ucr.edu.

†Department of Economics, University of California, Riverside, CA 92521, USA. E-mail: taelee@ucr.edu.

1 Introduction

Risk forecasts are often adjusted after the underlying model has been approved. The Basel backtesting rules are a familiar example: exception counts have long affected supervisory responses and capital multipliers, while the current market-risk rules place Expected Shortfall at the center of the internal-models approach and retain backtesting requirements (Basel Committee on Banking Supervision, 1996, 2019). The common feature is that the model remains in place while its reported tail forecasts are adjusted in light of realized outcomes. The rule studied here is an adjustment of that kind, and its multiplier is not a regulatory capital multiplier.

The same constraint arises in banks, vendor platforms, and internal model reviews. A user may have an archive of past median, VaR, and ES forecasts together with their subsequent realizations, or may be able to evaluate the current fitted model on past states. The standardized residuals, the filter state, and the density needed to re-estimate the tail model may be unavailable in the setting considered here. Our main implementation evaluates the current fitted model on past states; a supplementary comparison uses the forecast vintage generated at each past origin. When the reported tail is too thin or too thick, a one-parameter adjustment is attractive; the statistical question is how to estimate its multiplier without discarding useful location and volatility dynamics or breaking the VaR–ES ordering.

VaR and ES must be evaluated jointly. VaR is a conditional quantile, so quantile methods, including Mincer–Zarnowitz (MZ) regressions, can be used to adjust it (Mincer and Zarnowitz, 1969; Gaglianone et al., 2011; Fosten et al., 2024). ES is not elicitable on its own; VaR and ES are jointly elicitable and can be ranked by Fissler–Ziegel (FZ) scores (Fissler and Ziegel, 2016; Nolde and Ziegel, 2017; Patton et al., 2019; Dimitriadis and Bayer, 2019). We use the zero-homogeneous FZ0 score defined in Section 3.2. An unrestricted affine adjustment of VaR can have a nonpositive slope. If the fitted affine map is also applied to ES, a nonpositive slope collapses or reverses the VaR–ES ordering and can move the pair outside the FZ0 domain. In the S&P 500 1% skew-t QR application, 46.9% of the rolling affine-MZ fits have negative slopes.

We propose a median-anchored rule. It fixes the fitted median and multiplies the median-to-VaR and median-to-ES distances by a common positive multiplier. VaR and ES therefore move together, and an ordered pair remains ordered. The multiplier minimizes VaR check loss and has

a closed-form representation as a spread-weighted empirical quantile. The method is designed for settings in which the fitted median provides a stable location reference and a common multiplier is a reasonable approximation for the two lower-tail distances. More general tail misspecification requires a richer model.

Several related methods use more information or allow more parameters. Quantile MZ directly adjusts VaR but does not by itself give an ordered joint VaR–ES update. Model-risk buffers use backtesting outcomes to adjust risk forecasts (Boucher et al., 2014). Forecast combinations and joint VaR–ES regressions estimate richer transformations or combine several forecast pairs (Taylor, 2020; Storti and Wang, 2023; Dimitriadis and Bayer, 2019). Probability-integral-transform diagnostics require the full predictive distribution (Diebold et al., 1998; Berkowitz, 2001), while filtered historical simulation (FHS) requires standardized residuals and conditional scales (Barone-Adesi et al., 1999; McNeil and Frey, 2000). Dynamic VaR–ES models estimate new tail equations rather than adjust a reported forecast pair (Taylor, 2019; D'Innocenzo et al., 2026; Liu and Luger, 2026; Gatta et al., 2024; Gerlach et al., 2025). Closest to our setting, Zhang et al. (2023) rescale the reported VaR and ES levels by two positive multipliers estimated through rolling FZ0 minimization under an ordering constraint, so the two coordinates can move independently. We study the parsimonious case in which a single multiplier rescales both distances measured from the fitted median, which holds the location fixed and leaves only the tail spread free. This structure preserves ES below VaR without an additional constraint, gives a closed-form estimator, and permits direct results for coverage and FZ0 risk at the forecast origin. Our focus is the one-parameter rule: how to estimate its multiplier, how that estimate carries over from past forecast–realization pairs to a new origin, and when the rule improves the reported pair. More flexible adjustments address a different problem.

The paper makes three contributions. First, it defines the median-anchored rule and shows how it changes quantiles and ES while preserving their ordering. Second, it derives the weighted-quantile estimator and studies its two-step plug-in implementation. Third, it separates sampling error in the estimated multiplier from the difference between the historical estimation window and the current state, and it shows that when the same multiplier correctly adjusts both VaR and ES, the adjusted pair has lower conditional FZ0 risk at the forecast origin with probability approaching one. The analysis also identifies three ways in which the common-multiplier approximation may

fail: median error may vary with the state, the required multiplier may change over time, or VaR and ES may require different adjustments.

The empirical analysis uses Gaussian GARCH, Student-t GARCH, and a skew-t quantile-regression density. The primary sample contains 6,800 daily S&P 500 forecasts; NASDAQ and FTSE 100 results are reported in the supplement. The strongest result is at the 1% tail: for Gaussian GARCH, FZ0 loss falls by 10.4%, and the gain remains significant under every reported block length. Conditional-coverage tests still reject or are borderline in that cell. At the 5% tail, both GARCH baselines improve under pairwise tests, but the gains do not survive the 100–500-day family-wise block adjustments. The skew-t results show that a common multiplier can add little when the fitted tail is already flexible or unstable. Comparisons with an equal-weight estimator, an FZ0-targeted estimator, proportional scaling, a nonnegative-slope MZ regression, and the historical forecast vintages show how the results change when the main restrictions are altered. Filtered historical simulation shows what additional information on residuals and conditional scales can provide.

Section 2 defines the rule and estimator. Section 3 gives the theory results. Sections 4 and 5 present the Monte Carlo and empirical results. Section 6 concludes. The supplement contains proofs, implementation details, and additional results.

2 Method

Proposition 1 describes the rule. Lemma 1 characterizes the state-specific population minimizer, Lemma 2 gives the weighted-quantile estimator, and Proposition 2 establishes consistency. Theorems 1 and 2 study coverage and FZ0 risk.

2.1 Median-Anchored Adjustment Rule

Fix a left-tail level $\tau_0 \in (0, 0.5)$ and use the median $\bar{\tau} = 0.5$ as a pre-specified anchor. Fixing the median separates a location shift from a change in lower-tail spread and keeps the anchor away from the target tail. The supplement reports median specification checks and a sensitivity analysis that uses the 0.25 quantile as the anchor. We focus on the left tail; the right-tail case reverses the direction.

Let $\widehat{F}_t(\cdot \mid X_t)$ denote a baseline one-step-ahead conditional cumulative distribution function (CDF) for $Y_{t+1} \mid X_t$, produced by any forecasting engine that reports the anchor quantile, the target lower-tail quantile, and ES.[1] Define

$$\widehat{q}_{\tau,t} := \inf\{y \in \mathbb{R} : \widehat{F}_t(y \mid X_t) \geq \tau\}, \qquad q^{\star}_{\tau,t} \text{ analogously from the true } F^{\star}(\cdot \mid X_t),$$

and assume $\widehat{F}_t$ and $F^{\star}$ are continuous and strictly increasing in neighborhoods of the relevant quantiles, so these quantiles are uniquely defined. The CDF is a device for describing the rule; the estimator below uses only the reported triple.

Adjustment rule. For each t and $\kappa > 0$, define

$$\mathcal{A}_{\kappa,t}(z) := \begin{cases} z, & z \geq \widehat{q}_{\bar{\tau},t}, \\ \widehat{q}_{\bar{\tau},t} - \kappa\,(\widehat{q}_{\bar{\tau},t} - z), & z < \widehat{q}_{\bar{\tau},t}. \end{cases} \tag{1}$$

We call κ the multiplier. The rule leaves the distribution unchanged above the anchor and rescales the distance to the anchor below it: with the median held fixed, $\kappa > 1$ widens the lower tail and $\kappa < 1$ narrows it. If $\widetilde{Y}_{t+1}$ has conditional law $\widehat{F}_t$ and $\widetilde{Y}^{\text{adj}}_{t+1} := \mathcal{A}_{\kappa,t}(\widetilde{Y}_{t+1})$, then the adjusted CDF is

$$F_{\text{adj}}(y \mid X_t) = \Pr(\widetilde{Y}^{\text{adj}}_{t+1} \leq y \mid X_t) = \widehat{F}_t(\mathcal{A}^{-1}_{\kappa,t}(y) \mid X_t).$$

Throughout, the superscript adj denotes quantities produced by the adjustment rule.

Proposition 1 (Properties of the adjustment rule)**.** *For any $\kappa > 0$, $\mathcal{A}_{\kappa,t}(\cdot)$ is strictly increasing and continuous. If $\widehat{F}_t(\cdot \mid X_t)$ is continuous, then $F_{\text{adj}}(\cdot \mid X_t)$ is continuous. For $y < \widehat{q}_{\bar{\tau},t}$, $\kappa \geq 1$ implies $F_{\text{adj}}(y \mid X_t) \geq \widehat{F}_t(y \mid X_t)$, and $\kappa \leq 1$ implies $F_{\text{adj}}(y \mid X_t) \leq \widehat{F}_t(y \mid X_t)$. Quantiles transform under the same rule: since $\tau_0 < \bar{\tau}$, the target quantile lies below the anchor, and*

$$q^{\text{adj}}_{\tau_0,t}(\kappa) = \widehat{q}_{\bar{\tau},t} - \kappa(\widehat{q}_{\bar{\tau},t} - \widehat{q}_{\tau_0,t}). \tag{2}$$

[1] Examples include GARCH models with parametric innovations, skew-t quantile regression, CAViaR-type quantile models paired with an ES or density component (Engle and Manganelli, 2004), and other engines that report an ordered median–VaR–ES triple. A VaR-only fit must first be supplemented with an ES forecast before the joint pair can be updated.

If the baseline left-tail expected shortfall $\widehat{e}_{\tau_0,t} := \mathbb{E}[\widetilde{Y}_{t+1} \mid \widetilde{Y}_{t+1} \le \widehat{q}_{\tau_0,t}, X_t]$ *exists, the same affine branch applies to it,*

$$e^{\mathrm{adj}}_{\tau_0,t}(\kappa) = \widehat{q}_{\bar{\tau},t} - \kappa\big(\widehat{q}_{\bar{\tau},t} - \widehat{e}_{\tau_0,t}\big),$$

so VaR and ES move affinely with the same multiplier once the anchor is fixed. Finally, if the baseline triple is ordered, $\widehat{e}_{\tau_0,t} \le \widehat{q}_{\tau_0,t} < \widehat{q}_{\bar{\tau},t}$*, then* $e^{\mathrm{adj}}_{\tau_0,t}(\kappa) \le q^{\mathrm{adj}}_{\tau_0,t}(\kappa) < \widehat{q}_{\bar{\tau},t}$ *for every* $\kappa > 0$*.*

Order preservation is therefore automatic: no restriction on κ beyond positivity is required. The FZ0 sign restriction $e^{\mathrm{adj}}_{\tau_0,t} < 0$ is separate and is checked in the empirical work.

2.2 Population Minimizers

To define the population minimizers and the plug-in estimator, suppose that the baseline forecasts come from a parameterized family. Let θ_0 be the pseudo-true baseline parameter and write

$$m^0_t := q_{\bar{\tau}}(X_t;\theta_0), \qquad q^0_t := q_{\tau_0}(X_t;\theta_0), \qquad e^0_t := e_{\tau_0}(X_t;\theta_0), \qquad S^0_t := m^0_t - q^0_t.$$

More generally, write $q^0_{u,t} := q_u(X_t;\theta_0)$. All population objects in this subsection are defined from this pseudo-true path. Sample fitted paths use hats and the origin-window subscript $s \mid T, n$, which denotes the value at state s computed from the model fitted on the window of length n ending at origin T. This convention keeps the population minimizer separate from the retrospective path generated by the estimated current-window model.

Pooled population minimizer. Define the pseudo-true anchored quantile path

$$q^{\mathrm{adj},0}_t(\kappa) := m^0_t - \kappa S^0_t.$$

For a fixed tail level τ_0, let $\mathcal{K} = [\kappa_L, \kappa_U] \subset (0,\infty)$ be the compact multiplier interval and define

$$\kappa_0 \in \arg\min_{\kappa\in\mathcal{K}} \mathbb{E}\Big[\rho_{\tau_0}\{Y_{t+1} - q^{\mathrm{adj},0}_t(\kappa)\}\Big], \qquad \rho_\tau(u) := u\{\tau - \mathbf{1}(u<0)\}. \tag{3}$$

The check loss is strictly consistent for the quantile functional (Koenker and Bassett, 1978; Giacomini and Komunjer, 2005; Gneiting, 2011). The expectation integrates over the historical distribution of states, so κ_0 is a pooled pseudo-true multiplier; it need not equal the state-specific

minimizer at every value of X_t.

State-specific population minimizer. At a fixed state X_t, define

$$\kappa_t^\star \in \arg\min_{\kappa \in \mathcal{K}} \mathbb{E}\Big[\rho_{\tau_0}\{Y_{t+1} - q_t^{\mathrm{adj},0}(\kappa)\} \mid X_t\Big]. \tag{4}$$

This is the conditional population minimizer for the pseudo-true baseline forecasts.

Lemma 1 (State-specific population minimizer)**.** *Suppose the conditional check-loss criterion is integrable and $F^\star(\cdot \mid X_t)$ is continuous and strictly increasing near $q^\star_{\tau_0,t}$. If $S_t^0 > 0$ and $q^\star_{\tau_0,t} < m_t^0$, define the unconstrained value*

$$\kappa_{t,\mathrm{int}} := \frac{m_t^0 - q^\star_{\tau_0,t}}{m_t^0 - q_t^0}.$$

If $\kappa_{t,\mathrm{int}} \in \mathcal{K}$, the argmin in (4) is the singleton $\{\kappa_{t,\mathrm{int}}\}$ and $q_t^{\mathrm{adj},0}(\kappa_t^\star) = q^\star_{\tau_0,t}$. If $\kappa_{t,\mathrm{int}}$ lies below or above the compact interval $\mathcal{K}$, the constrained state-specific population minimizer is the corresponding boundary point; in general, the constrained forecast then does not equal the true quantile.

When $\kappa_{t,\mathrm{int}} \in \mathcal{K}$, the state-specific minimizer sets the adjusted VaR equal to the true conditional quantile while leaving the baseline median fixed. Otherwise, the minimizer is the nearest boundary point and the adjusted VaR generally cannot reach the true quantile. The minimizer may vary with X_t. This variation determines whether a multiplier estimated from the historical window is suitable at the current forecast origin.

Define the normalized median error and the normalized true median-to-VaR distance

$$a_t := \frac{m_t^0 - q^\star_{\bar\tau,t}}{S_t^0}, \qquad b_t := \frac{q^\star_{\bar\tau,t} - q^\star_{\tau_0,t}}{S_t^0}.$$

Both are population quantities, not additional estimated parameters.

The two terms sum to the unconstrained value, $\kappa_{t,\mathrm{int}} = a_t + b_t$, and the state-specific minimizer is its projection onto $\mathcal{K}$: $\kappa_t^\star = \Pi_{\mathcal{K}}(a_t + b_t)$, where $\Pi_{\mathcal{K}}(x) := \min\{\max\{x, \kappa_L\}, \kappa_U\}$; the supplement proves that the constrained minimizer is this projection. When $a_t + b_t$ is interior, the adjusted VaR equals the true VaR. If the baseline median equals the true median, then $a_t = 0$.

For an estimation window $\mathcal{W}_{T,n} = \{T-n, \ldots, T-1\}$ of length n ending at origin T, define the

window-specific population minimizer

$$\kappa_{T,n}^{0} \in \arg\min_{\kappa \in \mathcal{K}} \mathbb{E}\left[\frac{1}{n}\sum_{s \in \mathcal{W}_{T,n}} \rho_{\tau_0}\{Y_{s+1} - q_s^{\mathrm{adj},0}(\kappa)\}\right]. \tag{5}$$

When the minimizer set is not a singleton, $\kappa_{T,n}^{0}$ denotes the smallest minimizer; the estimator uses the same convention. Under stationarity this minimizer reduces to κ_0 in (3). When the distribution changes over time, the notation keeps the estimation window explicit. The coverage result below depends on $|\kappa_{T,n}^{0} - \kappa_T^{\star}|$, the difference between the window minimizer and the current-state minimizer. The theory concerns one forecast origin and its estimation window; the empirical application repeats the procedure over rolling origins.

If every state-specific minimizer in the window lies within $\delta_{T,n}$ of a common value $\kappa^{\dagger}$, then $|\kappa_{T,n}^{0} - \kappa^{\dagger}| \le \delta_{T,n}$. The supplement states and proves this result. The VaR-side hypothesis of Theorem 2 is the special case $\delta_{T,n} = 0$; that theorem also requires the same multiplier to match the ES coordinate. The decomposition $\kappa_{t,\mathrm{int}} = a_t + b_t$ also shows why the state-specific minimizers may differ: a median error that is constant relative to the baseline spread is absorbed by the common multiplier, whereas a state-dependent median error cannot be removed by one multiplier.

Even when the adjusted VaR equals the true VaR, the adjusted ES need not equal the true ES. With the interior state-specific minimizer in place, the remaining ES error is

$$B_{e,t} = (m_t^0 - e_{\tau_0,t}^{\star}) - \frac{m_t^0 - q_{\tau_0,t}^{\star}}{m_t^0 - q_t^0}\,(m_t^0 - e_t^0),$$

which vanishes only under additional compatibility of the lower-tail path; the supplement states this formally. We accordingly rank pairs by joint scores, treat ES backtests as specification checks, and do not infer ES adequacy from VaR coverage.

A location–scale sufficient condition. Suppose the true and the baseline law share a location and scale path and differ only in the innovation distribution: $Y_{t+1} = \mu_t + \sigma_t \varepsilon_{t+1}$ with ε_{t+1} distributed as G, while the baseline reports the quantiles of $\mu_t + \sigma_t \widehat{\varepsilon}_{t+1}$ with $\widehat{\varepsilon}_{t+1}$ distributed as $\widehat{G}$. Then the required spread ratio is the same in every state,

$$b_t = \frac{G^{-1}(\bar{\tau}) - G^{-1}(\tau_0)}{\widehat{G}^{-1}(\bar{\tau}) - \widehat{G}^{-1}(\tau_0)},$$

because the state-specific scale σ_t cancels. The same cancellation makes a_t state-free, so the unconstrained VaR-side minimizer $\kappa_{t,\text{int}} = a_t + b_t$ does not vary with the state; if the true and the baseline innovation medians also coincide, then $a_t = 0$ and the displayed ratio is that minimizer. The argument concerns the VaR coordinate: correct adjustment of ES still requires the compatibility condition above. The supplement also gives a local version: if the location–scale representation errors are uniformly $o(1)$, the same ratio approximates b_t with $o(1)$ error.

For a symmetric location–scale baseline whose median equals its location, write z_{τ_0} and w_{τ_0} for the baseline innovation VaR and ES. Then $q^{\text{adj}}_{\tau_0,t} = \mu_t + \kappa\sigma_t z_{\tau_0}$ and $e^{\text{adj}}_{\tau_0,t} = \mu_t + \kappa\sigma_t w_{\tau_0}$. In the VaR coordinate, the method is therefore a one-sided scale adjustment.[2] Even here the rule does more than rescale VaR: the same multiplier moves ES, and the ordering is preserved for every $\kappa > 0$.

This condition covers a location–scale model with a misspecified but time-invariant innovation distribution. A state-dependent location error, a change in the innovation distribution, or a state-dependent tail shape violates the condition. The Monte Carlo section studies these cases.

The population quantities are unobserved. We therefore report median hit rates, rolling and subperiod estimates of the multiplier, a stability check, and score-domain checks. These results cannot establish the maintained conditions, but they can indicate when a richer model is needed.

2.3 The Estimator

Estimation proceeds in two steps, only the second of which is the method's own: the first is the baseline fit that produces the forecasts to be adjusted. A user who inherits a fitted model does not perform that fit, but the retrospective path is still generated by the fitted parameter, so the multiplier is a two-step plug-in estimator either way. At forecast origin T, let $\mathcal{W}_{T,n} = \{T-n, \ldots, T-1\}$ be the estimation window. First estimate the baseline parameter $\widehat{\theta}_{T,n}$ on that window. Then evaluate the fitted model at each historical state $s \in \mathcal{W}_{T,n}$:

$$\widehat{q}_{u,s|T,n} := q_u(X_s; \widehat{\theta}_{T,n}), \qquad u \in \{\tau_0, \bar{\tau}\},$$

[2]For daily returns the fitted median is usually small relative to the tail quantiles, so the median-anchored rule is numerically close to proportional scaling. The median anchor makes the rule translation equivariant and preserves the baseline distances among the median, VaR, and ES. These properties matter more when the median is not small relative to the tail quantiles, as at longer horizons or for variables in levels.

and define $\widehat{e}_{\tau_0,s|T,n} := e_{\tau_0}(X_s; \widehat{\theta}_{T,n})$. These are historical fitted values from the current model, not forecasts that were issued at date s.

For the in-window fitted values at origin T, define

$$\widehat{S}_{s|T,n} := \widehat{q}_{\bar{\tau},s|T,n} - \widehat{q}_{\tau_0,s|T,n}, \qquad R_{s|T,n} := \frac{\widehat{q}_{\bar{\tau},s|T,n} - Y_{s+1}}{\widehat{S}_{s|T,n}}.$$

The plug-in estimator is any minimizer of

$$\widehat{\kappa}_{T,n} \in \arg\min_{\kappa \in \mathcal{K}} \frac{1}{n} \sum_{s \in \mathcal{W}_{T,n}} \rho_{\tau_0}\left(Y_{s+1} - \left[\widehat{q}_{\bar{\tau},s|T,n} - \kappa \widehat{S}_{s|T,n}\right]\right). \tag{6}$$

Because $\tau_0 < \bar{\tau}$ and the fitted quantiles are ordered, $\widehat{S}_{s|T,n} > 0$. Positive homogeneity of check loss gives the equivalent weighted problem

$$\widehat{\kappa}_{T,n} \in \arg\min_{\kappa \in \mathcal{K}} \frac{1}{n} \sum_{s \in \mathcal{W}_{T,n}} \widehat{S}_{s|T,n}\, \rho_{\tau_0}(\kappa - R_{s|T,n}). \tag{7}$$

The same window is used in both steps. Because all historical fitted values depend on the common first-stage estimate $\widehat{\theta}_{T,n}$, consistency and inference must account for baseline estimation. Section 3.1 states the required conditions, and the supplement gives the criterion decomposition. The supplement also reports a two-block sample split as a finite-sample sensitivity check; it is much noisier at $n = 1{,}000$ and is not the implemented estimator.

Lemma 2 (Weighted-quantile representation)**.** *Fix $\tau_0 < \bar{\tau}$ and let $\mathcal{Q}_{T,n}$ denote the set of unconstrained weighted empirical $(1-\tau_0)$-quantiles of $\{R_{s|T,n} : s \in \mathcal{W}_{T,n}\}$ with weights $\{\widehat{S}_{s|T,n} : s \in \mathcal{W}_{T,n}\}$; that is, $k \in \mathcal{Q}_{T,n}$ if*

$$\sum_{s: R_{s|T,n} \le k} \widehat{S}_{s|T,n} \ge (1-\tau_0) \sum_{s \in \mathcal{W}_{T,n}} \widehat{S}_{s|T,n} \ge \sum_{s: R_{s|T,n} < k} \widehat{S}_{s|T,n}.$$

The unconstrained minimizers of (7) over $\mathbb{R}$ are the elements of $\mathcal{Q}_{T,n}$. Over $\mathcal{K} = [\kappa_L, \kappa_U]$, the minimizer set is the projection of $\mathcal{Q}_{T,n}$ onto $\mathcal{K}$. Thus, if an unconstrained weighted quantile is interior, the implemented selection is the smallest k satisfying the displayed inequalities.

The quantile level is $(1-\tau_0)$ because the criterion is $\rho_{\tau_0}(\kappa - R)$ rather than $\rho_{\tau_0}(R - \kappa)$. Computing

the estimator requires only sorting the ratios, cumulating the spread weights, and applying the fixed bounds. The weights are implied by the original check-loss criterion: a change in κ moves the fitted VaR at date s by $\widehat{S}_{s|T,n}$. At an interior population minimizer, and when the threshold has no probability atom, the corresponding first-order condition is

$$\mathbb{E}\big[S_t^0\,\{\tau_0 - \mathbf{1}\big(Y_{t+1} < m_t^0 - \kappa_0 S_t^0\big)\}\big] = 0.$$

Thus the criterion centers a spread-weighted hit moment rather than the ordinary hit rate. An equal-weight variant that drops these weights uses the ordinary empirical quantile of the same ratios and centers the unweighted hit rate instead; Section 5 reports it. In finite samples, ties and the weighted-quantile kink are handled by the subgradient inequalities in Lemma 2. The asymptotic-normality result is stated for an interior population minimizer. Empirically, the implemented bounds $\mathcal{K} = [0.25, 4]$ never bind in the 122,556 adjusted equity forecast rows.[3]

Information-set interpretation. At origin T, both $\widehat{\theta}_{T,n}$ and $\widehat{\kappa}_{T,n}$ are functions of information available by time T. The current fitted pair $(\widehat{q}_{\tau_0,T|T,n}, \widehat{e}_{\tau_0,T|T,n})$ and its anchored update are therefore known before Y_{T+1} is observed:

$$\mathcal{I}_T \longrightarrow (\widehat{\theta}_{T,n}, \widehat{\kappa}_{T,n}) \longrightarrow \big(\widehat{q}^{\,\mathrm{adj}}_{\tau_0,T|T,n}(\widehat{\kappa}_{T,n}), \widehat{e}^{\,\mathrm{adj}}_{\tau_0,T|T,n}(\widehat{\kappa}_{T,n})\big).$$

The rolling evaluation is therefore out of sample at every forecast origin, even though the criterion is built from current-fit retrospective values rather than from forecasts issued at the corresponding historical dates.

Why check loss rather than FZ0. One could estimate κ by minimizing FZ0 numerically after imposing the score domain. We use VaR check loss for three reasons. First, it gives a closed-form weighted quantile and an explicit first-order condition, and it places the estimator in a familiar quantile-estimation structure that supports the limit theory of Section 3. Second, when the same multiplier correctly adjusts both VaR and ES, check loss and FZ0 have the same population minimizer (Theorem 2). Third, when this condition fails, the two criteria may select different multipliers because FZ0 trades off VaR and ES errors. The empirical section reports this

[3] $6 \times 6{,}800$ S&P rows, $6 \times 6{,}799$ NASDAQ Composite rows, and $6 \times 6{,}827$ FTSE 100 rows across the three baselines and two tail levels.

comparison directly.

Relation to quantile-MZ adjustment. In the VaR coordinate alone, the anchored update can be written as

$$q^{\mathrm{adj}}_{\tau_0,t}(\kappa) = (1-\kappa)\widehat{q}_{\bar{\tau},t} + \kappa\widehat{q}_{\tau_0,t}.$$

It is a median-anchored affine restriction with a positive slope, which keeps the adjusted pair ordered. Quantile Mincer–Zarnowitz (MZ) regression, the closest classical VaR-side adjustment rule (Mincer and Zarnowitz, 1969; Giacomini and Komunjer, 2005; Gaglianone et al., 2011; Fosten et al., 2024), fits an unrestricted affine map whose slope can be nonpositive and whose ES extension can leave the domain of lower-tail joint scoring. The restriction $\kappa > 0$ rules out the ordering reversal, though not the separate sign restriction on ES. The empirical analysis therefore reports Raw, affine MZ, Anchored, and FHS as separate comparisons.

3 Large-Sample Results and Specification Checks

The large-sample analysis separates estimation-window asymptotics from the forecast-origin index. For a generic sequence of windows with $n \to \infty$, write $\widehat{\theta}_n$ and $\widehat{\kappa}_n$ for the two plug-in stages and suppress the origin T. The one-step results then restore the origin notation $\widehat{\kappa}_{T,n}$. This distinction avoids treating the number of rolling evaluation origins as the sample size of a single multiplier estimate. Table 1 collects the recurring notation.

Table 1: Main notation

Symbol	Meaning	Where
(m_t^0, q_t^0, e_t^0); S_t^0	Pseudo-true median–VaR–ES path and its spread $m_t^0 - q_t^0$.	Section 2.2
$\mathcal{K} = [\kappa_L, \kappa_U]$	Compact multiplier interval; $[0.25, 4]$ in the application.	Section 2.2
$\kappa_{t,\mathrm{int}}$; $\kappa_t^\star$	Unconstrained state-specific multiplier and its projection onto $\mathcal{K}$.	Lemma 1
a_t; b_t	Normalized median error and normalized true tail distance; $\kappa_{t,\mathrm{int}} = a_t + b_t$.	Section 2.2
$\kappa^0_{T,n}$	Window-specific population minimizer at origin T.	Eq. (5)
κ_0	Pooled population minimizer of the check-loss criterion.	Eq. (3)
$\widehat{S}_{s\|T,n}$; $R_{s\|T,n}$	In-window fitted spread and the ratio it weights.	Section 2.3
$\widehat{\kappa}_{T,n}$	Plug-in spread-weighted quantile estimator.	Section 2.3
$\kappa^\dagger$	Multiplier that applies to both VaR and ES in Theorem 2.	Theorem 2

3.1 Estimator Asymptotics

Assumption 1 (Two-step plug-in regularity)**.** *Let $\widehat{\theta}_n$ denote the baseline parameter estimated on an estimation window of length n, and let θ_0 be its pseudo-true limit.*

(i) The underlying data process is stationary and α-mixing with moments sufficient for a uniform law of large numbers for the fixed-θ_0 criterion.

(ii) $\widehat{\theta}_n \to_P \theta_0$. The fitted median and target quantile paths are locally Lipschitz in θ, with a stationary integrable envelope, so replacing θ_0 by $\widehat{\theta}_n$ changes the criterion uniformly by $o_P(1)$.

(iii) The population criterion in (3) has a unique interior minimizer $\kappa_0 \in \mathrm{int}(\mathcal{K})$, and the spread-weighted distribution of the ratios has a positive density at κ_0.

Assumption 1 is a high-level two-step condition. Item (ii) controls the effect of using one estimated baseline parameter throughout the retrospective fitted path. The positive density in item (iii) makes the unconstrained sample weighted quantile interior with probability approaching one. The supplement gives the consistency decomposition and discusses sufficient regularity for the GARCH and quantile-regression paths.

Proposition 2 (Consistency)**.** *Under Assumption 1, $\widehat{\kappa}_n \xrightarrow{P} \kappa_0$ as $n \to \infty$.*

Asymptotic distribution. Under high-level two-step conditions stated in the supplement—an asymptotically linear first stage, a uniform local linearization of the second-stage score with a derivative limit bounded away from zero, a central limit theorem for the complete two-step influence process, and a maximal-weight condition on the spreads—the plug-in estimator satisfies

$$\sqrt{n}(\widehat{\kappa}_n - \kappa_0) \Rightarrow \mathcal{N}(0, J^{-2}\Omega),$$

where J is the limiting derivative of the second-stage score and Ω is the long-run variance of the complete two-step influence function.

A heteroskedasticity-and-autocorrelation-consistent (HAC) estimate computed from one realized fitted path measures conditional-on-fit variation and need not recover the common first-stage component. The supplement therefore treats the reported HAC standard errors for the retrospective pooled multiplier as descriptive, conditional-on-fit uncertainty. A fully first-stage-aware bootstrap would have to resample the underlying dated series, refit the baseline, rebuild the fit-

ted path, and re-estimate the multiplier under a separately specified resampling design. None of the forecast-comparison conclusions depends on the limiting distribution of $\widehat{\kappa}_n$: the rolling procedures are compared as forecasting methods with a fixed estimation window (Giacomini and White, 2006), using Diebold and Mariano (1995) statistics, and the supplement reports the multiple-testing adjustments separately.

The large-n statements refer to the estimation-window length. The empirical design fixes that length at 1,000 days; the 6,800 evaluation origins increase the precision of the out-of-sample comparisons, not the consistency of any one fixed-window estimate.

3.2 Coverage and FZ0 Risk at the Forecast Origin

Proposition 2 and the asymptotic distribution concern the estimator. We next study the forecast made at origin T. The first result bounds the VaR coverage error. The remaining two compare the conditional FZ0 risk of the raw and adjusted pairs: a local comparison, and then a direct result when the same multiplier corrects both coordinates. The joint comparison requires more because matching VaR does not by itself make the ES forecast accurate.

Coverage at the forecast origin. Let T denote the current forecast origin and let $\mathcal{I}_T$ be the information set at T, including X_T, the baseline forecast, and the historical sample used to estimate $\widehat{\kappa}_{T,n}$. We assume X_T is a sufficient forecasting state, $F^\star(\cdot \mid \mathcal{I}_T) = F^\star(\cdot \mid X_T)$ a.s., so that the state-specific population minimizer is also the minimizer conditional on $\mathcal{I}_T$. Define

$$r_{T,n} := |\widehat{q}_{\bar{\tau},T|T,n} - m_T^0| + \kappa_U |\widehat{S}_{T|T,n} - S_T^0|.$$

Theorem 1 (Coverage error at the forecast origin)**.** *Assume $\kappa_{T,\mathrm{int}} \in \mathcal{K}$, with the continuity and ordering conditions of Lemma 1, so the state-specific population minimizer matches the true conditional quantile, $S_T^0 = O_P(1)$, and the conditional density is bounded by $\bar{f}$ on $[q^\star_{\tau_0,T} - \delta_0, q^\star_{\tau_0,T} + \delta_0]$ for some $\delta_0 > 0$. Define*

$$\Delta_{T,n} := r_{T,n} + S_T^0 \left(|\widehat{\kappa}_{T,n} - \kappa_{T,n}^0| + |\kappa_{T,n}^0 - \kappa_T^\star| \right).$$

Then

$$\left|\Pr\left\{Y_{T+1} \le \widehat{q}^{\,\mathrm{adj}}_{\tau_0,T|T,n}(\widehat{\kappa}_{T,n}) \mid \mathcal{I}_T\right\} - \tau_0\right| \le \bar{f}\,\Delta_{T,n}\mathbf{1}\{\Delta_{T,n} \le \delta_0\} + \mathbf{1}\{\Delta_{T,n} > \delta_0\}.$$

If $\Delta_{T,n} = o_P(1)$*, the right-hand side equals* $\bar{f}\,\Delta_{T,n} + o_P(1)$ *and one-step state-conditional coverage follows.*

Theorem 1 separates three terms: estimation of the current fitted values, estimation of the multiplier around the window minimizer, and the difference between the window minimizer and the current-state minimizer. Under stationarity, $\kappa^0_{T,n} = \kappa_0$, Proposition 2 controls the estimation term, and under the additional limit conditions in the supplement the estimation component is $O_P(n^{-1/2})$. Under local change, the theorem remains a deterministic error decomposition around the window minimizer in (5); root-n control of the estimation component would require a separately stated triangular-array or locally stationary analogue of the stationary assumptions and is not claimed here. The result concerns the baseline state X_T; omitted predictive information can still produce rejections of dynamic-quantile (DQ) checks (Engle and Manganelli, 2004), as the empirical checks and the third data-generating process of Section 4 illustrate.

Expected shortfall and FZ0 risk. Because the adjustment rule is affine below the anchor, ES transforms with the same multiplier as VaR (Proposition 1), and we evaluate the pair with a strictly consistent Fissler–Ziegel score (Fissler and Ziegel, 2016). For daily equity returns, whose left-tail ES is negative on the return scale ($e < 0$), we use the zero-homogeneous score

$$s_{\tau_0}(q, e; y) = -\frac{1}{\tau_0 e}\mathbf{1}\{y \le q\}(q - y) + \frac{q}{e} + \log(-e) - 1, \qquad e < 0. \tag{8}$$

Lower scores are better; we refer to (8) as FZ0. For brevity, write $(\widehat{q}_T, \widehat{e}_T) := (\widehat{q}_{\tau_0,T|T,n}, \widehat{e}_{\tau_0,T|T,n})$ for the raw pair at origin T, and $(\widehat{q}^{\,\mathrm{adj}}_T, \widehat{e}^{\,\mathrm{adj}}_T)$ for its adjusted counterpart at $\widehat{\kappa}_{T,n}$.

Error in the adjusted pair. The distance between the adjusted pair and the true pair has five sources: error in the current fitted values, estimation error in $\widehat{\kappa}_{T,n}$, the difference between the window minimizer and the multiplier that applies at the current state, median error, and error in the lower-tail quantile curve. The supplement gives a deterministic bound for these terms.

Let $\mathcal{R}_T(q, e)$ be conditional FZ0 risk and let $(q_T^\star, e_T^\star)$ be the true conditional pair. Suppose the raw and adjusted pairs lie in a common convex neighborhood within the score domain on which the Hessian eigenvalues are between fixed constants $0 < \lambda_{\min} \leq \lambda_{\max} < \infty$. Define

$$D_T := \frac{\lambda_{\min}}{4} \left\| \begin{pmatrix} \widehat{q}_T - q_T^\star \\ \widehat{e}_T - e_T^\star \end{pmatrix} \right\|^2, \qquad \mathcal{E}_T^{\text{adj}} := (\widehat{q}_T^{\text{adj}} - q_T^\star)^2 + (\widehat{e}_T^{\text{adj}} - e_T^\star)^2.$$

Proposition 3 (Local FZ0 risk comparison at the forecast origin)**.** *Maintain the score-domain and moment conditions stated in the supplement (Assumption A.1, future-origin score-comparison conditions). Suppose also that both forecast pairs remain in the stated common neighborhood with probability approaching one. If $\mathcal{E}_T^{\text{adj}} = o_P(D_T)$, then*

$$\mathcal{R}_T(\widehat{q}_T, \widehat{e}_T) - \mathcal{R}_T(\widehat{q}_T^{\text{adj}}, \widehat{e}_T^{\text{adj}}) \geq D_T + o_P(D_T).$$

By the pair-error bound in the supplement, a squared five-source error of smaller order than D_T is sufficient for the rate condition. The adjusted pair therefore has lower conditional expected FZ0 loss whenever the raw risk gap is nondegenerate and the explicit pair-error terms are of smaller order. The result is local and does not infer a joint-risk improvement from VaR coverage alone. The next theorem gives a direct result when the same multiplier correctly adjusts both VaR and ES.

Theorem 2 (FZ0 risk when the same multiplier adjusts VaR and ES)**.** *Let Assumption 1 hold, and maintain the score-domain and moment conditions of Proposition 3 (Assumption A.1 in the supplement), including its common-neighborhood requirement. Suppose there is a multiplier $\kappa^\dagger$ in the interior of $\mathcal{K}$ with $|\kappa^\dagger - 1| \geq c > 0$ such that, almost surely,*

$$(q_t^\star, e_t^\star) = \big(m_t^0 - \kappa^\dagger(m_t^0 - q_t^0),\ m_t^0 - \kappa^\dagger(m_t^0 - e_t^0)\big),$$

and that the baseline tail distances $m_t^0 - q_t^0$ and $m_t^0 - e_t^0$ are bounded and bounded away from zero. If the current fitted median, VaR, and ES converge in probability to their pseudo-true values, then

$\kappa_0 = \kappa^\dagger$, $\widehat{\kappa}_{T,n} \to_P \kappa^\dagger$, $\mathcal{E}_T^{\mathrm{adj}} = o_P(1)$, *and* D_T *is bounded away from zero in probability. Consequently*

$$P\big\{\mathcal{R}_T(\widehat{q}_T^{\mathrm{adj}}, \widehat{e}_T^{\mathrm{adj}}) < \mathcal{R}_T(\widehat{q}_T, \widehat{e}_T)\big\} \to 1,$$

and the risk improvement exceeds a fixed positive constant with probability approaching one.

Theorem 2 gives the main risk result. Under its maintained condition, the check-loss minimizer is $\kappa^\dagger$, the estimator is consistent for it, and the adjusted pair converges to the true VaR–ES pair. Because $|\kappa^\dagger - 1|$ and the baseline tail distances are bounded away from zero, the raw pair remains a fixed distance from the truth. The adjusted pair therefore has lower conditional FZ0 risk with probability approaching one. The condition concerns the VaR–ES pair at the target level and does not require the true median to equal the baseline median. The lower-tail quantile-curve condition in the supplement is sufficient for the ES part of the result. The condition is restrictive and is not meant to describe all forecast errors. Section 4 reports a simulation that imposes this relation and then studies several forms of misspecification.

Scope. The results are stated for one tail level at a time, and separate estimates at different tail levels need not be monotone across quantiles. A one-parameter adjustment that preserves the within-tail VaR–ES ordering cannot correct errors that require different adjustments for VaR and ES, repair a misspecified median, or protect against abrupt breaks that the stability check may miss; those cases call for a richer model.

3.3 Stability Check for the Multiplier

A constant multiplier should not produce persistent drift in the check-loss score residuals. We therefore report a Nyblom-type κ-stability check on the partial sums of the check-loss score residuals, in the parameter-instability tradition of Nyblom (1989); Hansen (1992); Andrews (1993). The statistic evaluates the score residuals at $\widehat{\kappa}^{\mathrm{in}}$, the weighted-quantile estimator of Section 2.3 computed on the retrospective path from a single full-span refit. This full-span fit is used only for the check and is not the rolling fit that issues the forecasts. Because $\widehat{\kappa}^{\mathrm{in}}$ solves the pooled first-order condition, the terminal score sum is centered up to the weighted-quantile subgradient and a finite-sample discretization term. The partial-sum process is then compared with the Brownian-bridge reference. Writing the score residual as $\psi_t(\widehat{\kappa}^{\mathrm{in}})$ (the weighted check-loss subgradient) and

estimating its long-run variance by a Bartlett HAC estimator $\widehat{\sigma}^2$, we report

$$\mathrm{Ny}_n = \frac{1}{n^2\widehat{\sigma}^2} \sum_{s=1}^{n} \left(\sum_{t=1}^{s} \psi_t(\widehat{\kappa}^{\mathrm{in}}) \right)^2 . \tag{9}$$

A rejection provides evidence against a constant adjustment multiplier; non-rejection does not provide evidence in favor of constancy. The supplement reports the size and power of this check: size is conservative, power against smooth drift is moderate, and power against moderate breaks is limited. We therefore treat the statistic as a specification check and interpret it together with hit rates, FZ0 losses, median checks, and MZ results.

4 Monte Carlo Results

The Monte Carlo study uses four data-generating processes (DGPs) that progressively relax the conditions of Section 3. DGP 1 imposes the condition of Theorem 2, under which one multiplier applies to both VaR and ES, and asks whether the estimated multiplier recovers it and improves an independent future sample. DGP 2 keeps the innovation misspecification fixed over time and asks whether a stable pooled multiplier exists and helps. DGP 3 adds a state-dependent error in the median, and DGP 4 also lets the innovation distribution vary with the state; both ask what survives when the maintained conditions fail, and they are reported together below. DGPs 2–4 share the specification

$$Y_{t+1} = \mu_t + \sigma_t \varepsilon_{t+1}, \qquad \mu_t = X_t'\beta, \qquad \sigma_t = \exp\{0.5(a_0 + X_t'a)\},$$

where X_t contains four lags of an observed autoregressive state Z_t. DGP 2 sets $\beta = 0$ and uses one skew-t innovation distribution. DGP 3 adds the state-dependent mean that the constant-mean GARCH baselines omit. DGP 4 retains that mean and lets the innovation distribution depend on the state. The supplement gives the parameter values and innovation distributions. The estimator is the same plug-in estimator used in the empirical application, and each reported cell uses $B = 500$ replications.

4.1 DGP 1: Same Multiplier for VaR and ES

DGP 1 treats the baseline forecasts as given and imposes the condition in Theorem 2. Returns are $Y_{t+1} = \sigma_t h_{\kappa^\dagger}(\varepsilon_{t+1})$ with $m_t \equiv 0$, where ε_{t+1} has distribution function G and $h_{\kappa^\dagger}$ multiplies negative innovations by a known multiplier $\kappa^\dagger$ and leaves positive innovations unchanged. The baseline reports the median, VaR, and ES of the untransformed innovation distribution. Thus the baseline median is correct and the same multiplier relates the baseline and true VaR, ES, and lower-tail quantiles. The baseline forecasts are not estimated in this design, so the experiment isolates estimation of κ and its use in a later sample. The multiplier is estimated once on an estimation block and held fixed on an independent continuation of length 1,000.

Table 2 reports the heavy-tailed dynamic-volatility cells at $\tau_0 = 0.05$. The estimator recovers $\kappa^\dagger$ (bias at most 0.008, with root mean squared error, RMSE, falling in n), future coverage of the adjusted VaR sits at the nominal level, the mean future FZ0 loss of the adjusted pair is below the raw loss in every cell, and the share of replications with a strict improvement lies between 0.94 and 1.00. Two reference calculations use the known data-generating process: the raw future hit rate has the closed form $G\{G^{-1}(\tau_0)/\kappa^\dagger\}$ cell by cell, and in constant-volatility cells the Monte Carlo FZ0 risk gap matches the population value obtained by numerical integration (both reported in the supplement). These numerical results agree with the conclusion of Theorem 2 without relying on the sufficient curvature condition in Proposition 3. The supplement reports the full 96-cell grid, both tail levels, Gaussian innovations, and the finite-sample behavior of the 1% order statistic to which the multiplier reduces in constant-weight cells.

Table 2: Monte Carlo results, DGP 1: out-of-sample performance when the same multiplier adjusts VaR and ES (t_5 innovations, dynamic volatility, $\tau_0 = 0.05$)

	$\widehat{\kappa}$		Hit rate	FZ0 loss		
n	Bias	RMSE	Anchored	Raw	Anchored	Improvement share
$\kappa^\dagger = 0.75$						
250	−0.003	0.080	0.052	−0.390	−0.472	0.962
1000	0.002	0.041	0.050	−0.388	−0.478	0.992
2000	0.000	0.028	0.050	−0.388	−0.480	0.998
$\kappa^\dagger = 1.25$						
250	0.007	0.135	0.052	0.116	0.047	0.942
1000	0.005	0.067	0.050	0.111	0.034	0.988
2000	0.002	0.048	0.050	0.115	0.034	0.986
$\kappa^\dagger = 1.75$						
250	0.008	0.197	0.051	0.974	0.377	1.000
1000	−0.002	0.090	0.050	0.986	0.368	1.000
2000	−0.001	0.066	0.050	0.992	0.370	1.000

Notes: Given-baseline design in which the same multiplier adjusts the true VaR and ES ($Y_{t+1} = \sigma_t h_{\kappa^\dagger}(\varepsilon_{t+1})$, $m_t \equiv 0$); $B = 500$ replications per cell; the multiplier is estimated once on the estimation block and evaluated on an independent future continuation of length 1,000 (the dynamic state chain continues). Improvement share is the fraction of replications in which the anchored future-mean FZ0 loss is strictly below raw. Full factorial results, closed-form raw hit rates, and population FZ0 risk gaps computed by numerical integration are in the supplement.

4.2 DGP 2: Misspecified Innovation Distribution

This design combines stochastic volatility with a fixed standardized skew-t innovation distribution. The skew-t QR baseline is close to the true innovation distribution, while Gaussian and Student-t GARCH retain useful volatility dynamics but impose the wrong tail shape. The experiment asks whether the estimator recovers a stable population multiplier and whether the adjustment improves VaR coverage and FZ0 loss. For the GARCH baselines, the design does not require the same multiplier to be correct for the entire lower-tail quantile curve, so matching VaR can still leave an ES error.

Table 3: Monte Carlo results, DGP 2: in-window estimates under a misspecified innovation distribution

		Estimated multiplier		VaR hit rate		FZ0 loss	
Baseline	n	$\bar{\widehat{\kappa}}_n$	$\sqrt{n}\cdot$Std	Raw	Anchored	Raw	Anchored
skew-t QR ($\kappa_0 = 1.000$; correct innovation distribution)							
	100	0.999	0.23	0.047	0.044	−0.183	−0.193
	400	0.999	0.15	0.049	0.049	−0.052	−0.052
	1600	1.000	0.10	0.050	0.050	−0.021	−0.021
Student-t GARCH ($\kappa_0 = 1.251$)							
	100	1.261	1.64	0.080	0.045	0.015	−0.064
	400	1.254	1.60	0.081	0.049	0.053	−0.029
	1600	1.251	1.58	0.081	0.050	0.065	−0.017
Gaussian GARCH ($\kappa_0 = 1.125$)							
	100	1.173	1.42	0.069	0.044	−0.029	−0.083
	400	1.140	1.40	0.066	0.049	0.028	−0.017
	1600	1.126	1.45	0.064	0.050	0.034	−0.007

Notes: Nominal $\tau_0 = 0.05$. This table reports quantities from the fitted estimation sample, not from an independent forecast sample. The table uses the plug-in estimator, and the displayed κ_0 values are rounded long-run pooled check-loss minimizers under DGP 2, which has no state-dependent mean error. The symbol $\bar{\widehat{\kappa}}_n$ is the mean estimate across replications, and Std is its standard deviation across replications; multiplying by $\sqrt{n}$ makes the column approximately constant in n under the root-n rate. Here n is the simulated sample length, matching the estimation-window notation of the main text. The GARCH rows report effects on VaR coverage and average FZ0 loss; they do not assume that the same multiplier is correct for the entire lower-tail quantile curve. FZ0 levels are in-window averages and should be compared across methods within a given n, not across sample lengths.

Table 3 shows a regular pattern. The estimated multiplier converges toward its pooled population minimizer, the GARCH hit rates move toward 5%, and average FZ0 loss falls when the baseline tail is too thin. For the innovation-shape-matched skew-t baseline, the estimated correction becomes negligible. In the shortest skew-t sample, the estimated multiplier is already close to one, and the Raw and Anchored losses become equal as the sample grows. The small scaled standard deviation occurs because the first-stage quantile condition and the second-stage weighted-hit condition are nearly the same in this design; it does not imply a faster convergence rate.

4.3 DGPs 3 and 4: Median and Innovation Misspecification

This subsection reports DGPs 3 and 4. Both use an independent future sample, so the table reports forecast performance rather than fit in the estimation sample.

The skew-t QR baseline includes the state directly, so there is little remaining error for a common multiplier to correct, and it shows no gain in Table 4; the combined misspecification reduces the GARCH gains. Under median misspecification the multiplier still brings unconditional coverage close to the nominal level and lowers average loss, while conditional coverage is monotone

in the state: at the 5% tail, coverage under the pooled multiplier falls from 0.084 in the lowest state quintile to 0.027 in the highest around an unconditional rate of 0.050 (Table 5). A common multiplier shifts the coverage profile but cannot flatten it. In the design where the same multiplier is correct for VaR and ES, the profile is flat because the volatility scale cancels from the hit event; DGP 2, in which the innovation distribution does not vary with the state, keeps the state-specific minimizer approximately constant, apart from baseline estimation and volatility-tracking error. The monotone pattern therefore reflects the state-dependent median error. Unconditional coverage and average loss do not separate these cases; the median, conditional-coverage, and stability checks reported alongside them do. A useful average correction need not be a correct state-by-state model.

Table 4: Monte Carlo results, DGPs 3 and 4: out-of-sample performance under median and innovation misspecification

DGP	Baseline	Raw hit	Anchored hit	FZ0 loss, Raw minus Anchored
DGP 3 (median)	Gaussian GARCH	0.0616	0.0502	0.0258 (0.0010)
	Student-t GARCH	0.0729	0.0499	0.0471 (0.0017)
	skew-t QR	0.0511	0.0512	−0.0001 (0.0000)
DGP 4 (median + innovation)	Gaussian GARCH	0.0543	0.0507	0.0067 (0.0006)
	Student-t GARCH	0.0648	0.0506	0.0195 (0.0010)
	skew-t QR	0.0503	0.0504	−0.0003 (0.0002)

Notes: $n_{\text{est}} = 1{,}600$, future evaluation length 1,000, and $B = 500$ replications. Each cell reports the across-replication average of the replication-level future-mean loss difference; Monte Carlo standard errors, in parentheses, are the across-replication standard deviation of that mean divided by $\sqrt{B}$. Positive differences favor Anchored. These simulations study misspecification rather than the assumptions of Proposition 3. The 1% tail and replication-level test frequencies are reported in the supplement.

Table 5: Monte Carlo results, DGP 3: conditional coverage by state quintile at the pooled multiplier

			State quintile					
Baseline	τ_0	κ	1 (low)	2	3	4	5 (high)	Uncond.
Gaussian GARCH	0.05	pooled $\widehat{\kappa}$	0.084	0.057	0.046	0.035	0.027	0.050
		1 (raw)	0.105	0.071	0.057	0.044	0.033	0.062
Gaussian GARCH	0.01	pooled $\widehat{\kappa}$	0.015	0.011	0.009	0.007	0.006	0.010
		1 (raw)	0.041	0.029	0.023	0.018	0.014	0.025
Student-t GARCH	0.05	pooled $\widehat{\kappa}$	0.083	0.057	0.046	0.035	0.027	0.050
		1 (raw)	0.124	0.084	0.067	0.052	0.039	0.073
Student-t GARCH	0.01	pooled $\widehat{\kappa}$	0.015	0.011	0.009	0.007	0.006	0.010
		1 (raw)	0.034	0.024	0.019	0.015	0.012	0.021

Notes: $B = 500$ replications at $n = 1{,}600$ (the largest-n protocol of Table 3); each evaluation origin is assigned to a quintile of the state Z_t, and conditional coverage is pooled across replications within each quintile. The pooled criterion centers a spread-weighted hit condition; in this design the unweighted hit rate is also near the nominal level. Under correct state-conditional coverage every quintile entry would equal τ_0. The monotone pattern across quintiles reflects the state-dependent median error.

5 Empirical Application

We study one-step-ahead left-tail forecasts for daily S&P 500 returns at $\tau_0 \in \{0.01, 0.05\}$. The underlying return series runs from January 4, 1995 to December 31, 2025; the sample contains 6,800 rolling forecast origins beginning December 18, 1998, with realized evaluation dates through December 31, 2025; returns are multiplied by 100. The application does not use an external production archive: we generate one-step forecasts recursively and then treat the resulting path as the inherited input. At each origin, the preceding 1,000 observations are used both to estimate the baseline model and to estimate the multiplier. The three baselines are Gaussian GARCH, Student-t GARCH, and a skew-t quantile-regression density. They represent a thin-tailed parametric model, a heavy-tailed parametric model, and a flexible semiparametric model. We use these deliberately different baselines to illustrate when the adjustment is useful and when it is not.

The fitted medians are well centered in the primary sample. The fractions of returns below the fitted median are 0.504, 0.509, and 0.502 for skew-t QR, Student-t GARCH, and Gaussian GARCH, with two-sided binomial reference p-values 0.528, 0.152, and 0.771. We use these values as a descriptive check of the median, not as a test of correct conditional median specification.

We label the original forecasts Raw and the median-anchored forecasts Anchored. The primary comparison is Raw versus Anchored. Forecast pairs are ranked by the zero-homogeneous FZ score in (8), on the common dates for which both pairs satisfy $e \leq q$ and $e < 0$. Following Giacomini and White (2006), we compare the complete rolling procedures, including estimation of the baseline model and the multiplier, as forecasting methods. Raw is the special case $\kappa = 1$.

We report pairwise Diebold–Mariano (DM) tests with Bartlett Newey–West standard errors (Diebold and Mariano, 1995; Newey and West, 1987). Let N_{eval} be the number of scored out-of-sample loss differences in a pairwise comparison. The bandwidth is $\lfloor 4(N_{\text{eval}}/100)^{2/9} \rfloor$ and the reference distribution is standard normal. The supplement reports longer HAC bandwidths, fixed-b critical values, and Romano–Wolf adjustments for the six Raw-versus-Anchored comparisons (Romano and Wolf, 2005). We also report an affine MZ adjustment for VaR and apply the same fitted map to ES; the fraction of forecasts that remain in the FZ0 domain is reported separately.

For the skew-t QR baseline, linear quantile regressions at $\tau \in \{0.05, 0.25, 0.50, 0.75, 0.95\}$ are matched to a skew-t density by nonlinear least squares (Koenker and Bassett, 1978; Azzalini and

Capitanio, 2003). Its 1% VaR and ES are therefore fitted-tail extrapolations rather than direct 1% quantile regressions.

Table 6: Daily S&P 500 left-tail forecasting

		FZ0 loss		Anchored vs. Raw	Hit rate		Kupiec p		CC p	
Tail	Baseline	Raw	Anchored	DM p	Raw	Anchored	Raw	Anchored	Raw	Anchored
$\tau_0 = 0.01$	skew-t QR	2.015	2.036	0.615	0.013	0.015	0.011	< 0.001	< 0.001	< 0.001
	Student-t GARCH	1.259	1.237	0.026	0.015	0.013	< 0.001	0.020	< 0.001	0.022
	Gaussian GARCH	1.389	1.245	< 0.001	0.022	0.012	< 0.001	0.060	< 0.001	0.048
$\tau_0 = 0.05$	skew-t QR	1.198	1.194	0.038	0.055	0.054	0.055	0.124	< 0.001	< 0.001
	Student-t GARCH	0.849	0.828	0.003	0.065	0.054	< 0.001	0.187	< 0.001	0.392
	Gaussian GARCH	0.858	0.840	0.002	0.060	0.053	< 0.001	0.295	0.001	0.559

Notes: Raw and Anchored are compared on identical score-valid origins within each baseline. For skew-t QR, FZ0 means and the DM test use $N_{\text{score}} = 6{,}797$ after three positive-ES exclusions, while hit rates and the Kupiec and Christoffersen tests use all $N_{\text{hit}} = 6{,}800$ origins. For both GARCH baselines, $N_{\text{score}} = N_{\text{hit}} = 6{,}800$. CC is Christoffersen's conditional-coverage test. DQ checks for all cells appear in the supplement. DM p is the two-sided pairwise Diebold–Mariano test for equal FZ0 loss between Anchored and Raw. Lower loss is better.

Table 6 gives the main evidence. At the 5% tail, Anchored lowers FZ0 loss from 0.849 to 0.828 for Student-t GARCH and from 0.858 to 0.840 for Gaussian GARCH. The pairwise DM p-values are 0.003 and 0.002. Hit rates move from 6.5% to 5.4% and from 6.0% to 5.3%; the Kupiec tests (Kupiec, 1995) no longer reject, and the Christoffersen (Christoffersen, 1998) conditional-coverage p-values rise to 0.392 and 0.559. The DQ check still rejects in these two cells ($p = 0.004$ and 0.008; the full conditional-coverage and DQ panel is reported in the supplement), so the multiplier improves the unconditional-coverage and Christoffersen checks without removing the predictability detected by the DQ regression.

At 1%, the Gaussian GARCH FZ0 loss falls by 10.4% and the hit rate moves from 2.2% to 1.2%, but the conditional-coverage test still rejects ($p = 0.048$). The Student-t gain is smaller, and its coverage tests still reject. At 5% the skew-t update is economically negligible (1.198 to 1.194) despite a nominally significant DM p-value that is not robust to longer HAC bandwidths. The 1% skew-t row does not improve, consistent with the instability and fitted-tail extrapolation documented in the supplement. A rejection of the stability check provides evidence against a constant multiplier, but it does not determine the sign of the average Raw-minus-Anchored loss difference.

The 5% GARCH improvements also appear on NASDAQ, where both pairwise DM p-values are below 0.001. They are weaker on the FTSE 100: the p-values are 0.126 for Student-t GARCH and 0.050 for Gaussian GARCH.

The supplement reports the affine MZ comparison. When the affine slope is positive and the transformed pair remains in the FZ0 domain, MZ can be competitive. For the skew-t baseline, however, the unrestricted affine map often has a negative slope, and the transformed pair then leaves the FZ0 domain. Anchored adjustment instead fixes the median and preserves the VaR–ES ordering for every positive multiplier.

Figure 1 shows the rolling multipliers and cumulative Raw-minus-Anchored FZ0 loss differences for the two GARCH models at 5%. The gain is not confined to a single market subperiod; the supplement reports the corresponding subperiod averages and DM tests. It also reports shared-index Romano–Wolf adjustments: on the common $N_{\text{fam}} = 6{,}797$-date family calendar, the two 5% GARCH comparisons reject under the 19-day convention $\lceil N_{\text{fam}}^{1/3} \rceil$ but do not reject under the reported 100-, 250-, or 500-day blocks; by contrast, the Gaussian GARCH 1% loss gain remains significant at every reported block length. Longer HAC bandwidths (100 and 250 days) and fixed-b critical values leave the pairwise GARCH conclusions unchanged at both tails, while the skew-t 5% p-value moves to 0.21–0.26.

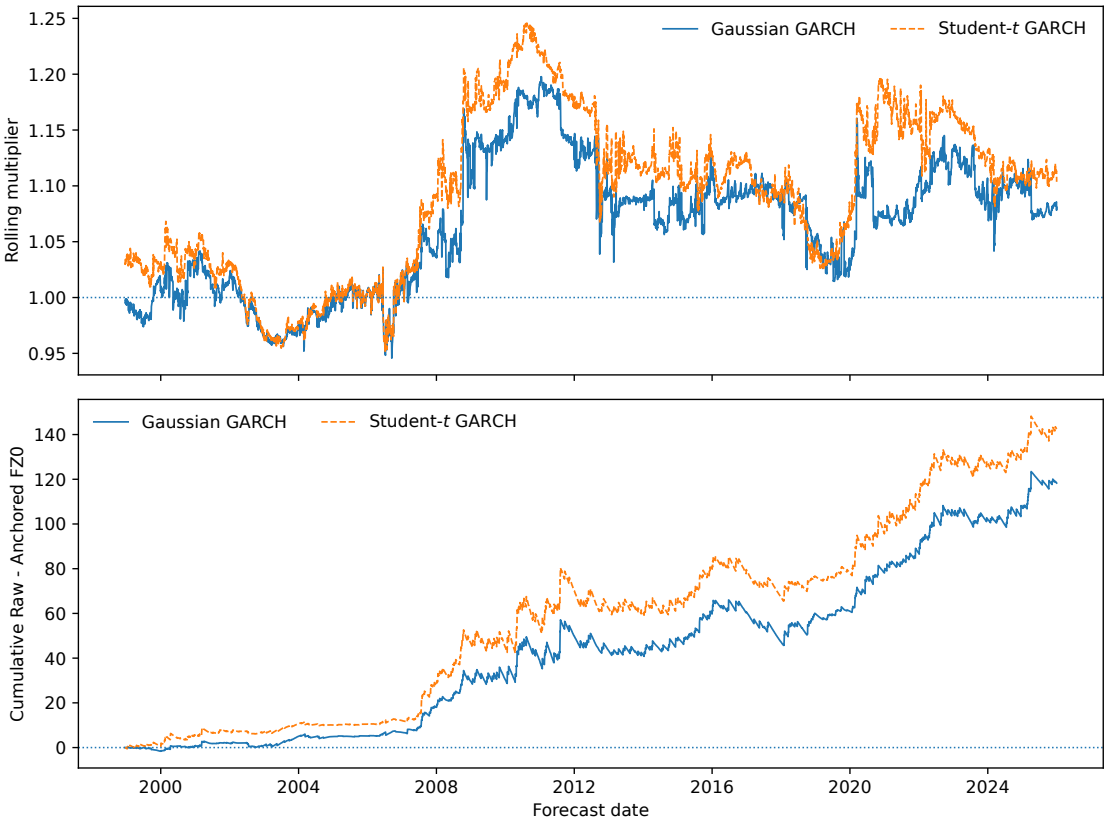

Figure 1: Rolling S&P 500 results, $\tau_0 = 0.05$

Notes: The upper panel plots the rolling multiplier for Student-t and Gaussian GARCH. The lower panel plots cumulative Raw-minus-Anchored FZ0 loss; positive values favor Anchored. Units are daily FZ0 loss accumulated over forecast origins.

FHS is the relevant comparison when the model's standardized residuals and conditional-scale path are available. It can change the residual-tail shape rather than only the distances from the median. At the 5% tail, its FZ0 loss is essentially the same as Anchored under Student-t GARCH (0.8286 versus 0.8283) and lower under Gaussian GARCH (0.8326 versus 0.8404). The supplement reports the full common-date comparison. These results show the role of information availability: Anchored is useful when the fitted median–VaR–ES path is available but the standardized residuals or the full predictive density are not.

Table 7 compares Anchored with four alternatives based on the same historical fitted values: an equal-weight estimator, an FZ0-targeted estimator, a nonnegative-slope affine MZ regression, and a zero-anchored proportional estimator. Panel B instead uses the forecast vintage generated at each historical origin.

The equal-weight estimator removes the spread weights from (7):

$$\widehat{\kappa}_{T,n}^{\mathrm{EW}} \in \arg\min_{\kappa\in\mathcal{K}} \frac{1}{n} \sum_{s\in\mathcal{W}_{T,n}} \rho_{\tau_0}(\kappa - R_{s|T,n}).$$

Because $Y_{s+1} \le \widehat{q}_{\tau,s|T,n} - \kappa\widehat{S}_{s|T,n}$ if and only if $R_{s|T,n} \ge \kappa$, this ordinary empirical $(1-\tau_0)$-quantile targets the unweighted in-window hit rate, up to ties and projection onto $\mathcal{K}$. We report it only to show what changes when the spread weights are removed. It is statistically indistinguishable from Anchored for the GARCH baselines and significantly worse for skew-t at the 5% tail. No difference is significant at the 1% tail.

The FZ0-targeted estimator κ^{FZ} gives similar results for the GARCH baselines, with one marginal difference for Gaussian GARCH at 5% ($p = 0.056$). For skew-t, it reaches the parameter bounds in 80 windows at 5% and has higher loss. The nonnegative-slope MZ regression does not significantly improve on Anchored and is significantly worse for Student-t GARCH at 5%. Its lower point estimate of FZ0 loss for skew-t at 1% ($p = 0.107$) occurs because the slope constraint binds in 47% of the windows, reducing the fit to an unconditional historical quantile.

The zero-anchored proportional estimator is estimated by check loss. It uses a weighted-quantile formula when all fitted VaRs are negative and direct convex minimization otherwise. It is statistically indistinguishable from Anchored in every cell. Thus the median anchor provides translation equivariance and a closed-form estimator that is available in every window; it does not produce a statistically detectable average-loss gain over proportional scaling in this application.

Panel B estimates the same weighted quantile from the forecast vintage generated at each past origin, with no access to the current fit. On the common origins the retrospective implementation is better at the 5% tail ($p < 0.001$ for Student-t GARCH and $p = 0.010$ for Gaussian GARCH against the historical-vintage variant), and the two are statistically indistinguishable at 1%, where the historical-vintage variant retains the Gaussian GARCH gain over Raw. Re-evaluating the historical states under the current parameters improves performance at the 5% tail, while a user with only archived forecasts still obtains a similar correction at the 1% tail.

Table 7: Adjustment comparisons

		Anchored	Equal-weight	FZ0-targeted	MZ ($\beta_{\mathrm{MZ}} \geq 0$)	Proportional
Panel A: full sample; FZ0 loss, DM p-values against Anchored in parentheses						
$\tau_0 = 0.05$	skew-t QR	1.194	1.196 (0.011)	1.252 (0.315)	1.202 (0.510)	1.194 (0.643)
	Student-t GARCH	0.828	0.830 (0.159)	0.828 (0.602)	0.839 (0.012)	0.829 (0.234)
	Gaussian GARCH	0.840	0.841 (0.428)	0.834 (0.056)	0.847 (0.147)	0.841 (0.095)
$\tau_0 = 0.01$	skew-t QR	2.036	2.000 (0.109)	2.112 (0.280)	1.803 (0.107)	2.015 (0.063)
	Student-t GARCH	1.237	1.234 (0.559)	1.237 (0.946)	1.230 (0.657)	1.239 (0.332)
	Gaussian GARCH	1.245	1.243 (0.705)	1.241 (0.649)	1.242 (0.848)	1.245 (0.358)
Panel B: historical-vintage variant, common origins 1,001–6,800						
		Raw	Anchored	Historical vintage		
$\tau_0 = 0.05$	skew-t QR	1.215	1.211	1.279 (0.033)		
	Student-t GARCH	0.810	0.787	0.800 (<0.001)		
	Gaussian GARCH	0.818	0.798	0.810 (0.010)		
$\tau_0 = 0.01$	skew-t QR	2.121	2.136	2.173 (0.734)		
	Student-t GARCH	1.230	1.201	1.220 (0.150)		
	Gaussian GARCH	1.372	1.208	1.217 (0.436)		

Notes: Equal-weight (κ^{EW}) is the ordinary empirical $(1-\tau_0)$-quantile of the in-window ratios; it gives each date equal weight instead of the spread weight used by Anchored. FZ0-targeted (κ^{FZ}) minimizes in-window average FZ0 loss over the score-valid parameter range. MZ ($\beta_{\mathrm{MZ}} \geq 0$) fits a nonnegative-slope affine quantile regression in each window and applies the same transformation to ES. Proportional is the zero-anchored estimator obtained from check loss; when its positive-weight quantile formula is unavailable, the convex problem is solved directly. The historical-vintage version applies the spread-weighted quantile to the forecast vintage generated at each past origin, with Raw and Anchored recomputed on the same origins; parentheses in Panel B are DM p-values for the historical-vintage version against Anchored. Every DM comparison uses the common score-valid origins of the two forecasts compared. The supplement gives the definitions and full results.

Joint loss ranks forecast pairs; it does not establish absolute ES adequacy or remove all dynamics from the VaR hit sequence. An exceedance-residual bootstrap check speaks to the ES side directly: for Student-t GARCH at the 5% tail the p-value rises from 0.025 for Raw to 0.700 after anchoring, so the adjustment removes the ES rejection in that cell, while the Gaussian rejection remains. The check uses an iid bootstrap and is reported as supporting evidence rather than as the forecast-ranking criterion; formal ES backtests are developed by Du and Escanciano (2017), Kratz et al. (2018), and Bayer and Dimitriadis (2022). Backtest p-values here are computed for forecasts that embed estimation error, which by itself distorts test size (Escanciano and Olmo, 2010), one more reason to read coverage tests as specification checks rather than as the ranking criterion.

The supplement reports the complete S&P specification checks, NASDAQ and FTSE results, uncertainty and stability of the estimated multiplier, block-length and multiplicity checks, adjustment comparisons, direct 1% QR, affine MZ, FHS, and window sensitivities.

6 Conclusion

Median-anchored adjustment addresses a common practical problem. It keeps the fitted median fixed, multiplies the distances from the median to VaR and ES by a common positive multiplier, and preserves the VaR–ES ordering. The multiplier has a closed-form weighted-quantile estimator and can be applied without replacing the baseline model's location or volatility dynamics.

The evidence is strongest when the baseline has useful scale dynamics but a reported tail that is systematically too thin. For the GARCH baselines, the adjustment lowers FZ0 loss and improves unconditional coverage, although the DQ tests still reject in some cells. The largest and most robust gain occurs for Gaussian GARCH at the 1% tail. The skew-t results show that a common multiplier adds little when the fitted tail is already flexible, while the misspecification designs show that it still lowers average loss for the GARCH baselines when the maintained conditions fail, though it shifts rather than flattens the state-conditional coverage profile.

Two conditions govern its use: the fitted median should be a reasonably well-centered location reference, and one multiplier should serve for both VaR and ES. The median check assesses the location reference, whereas the multiplier-stability check assesses time variation in the VaR-targeted multiplier. Whether the same multiplier also works for ES remains an additional condition. Ordering is preserved for every positive multiplier, and the multiplier can be estimated from reported median–VaR–ES forecasts and their subsequent realizations. Under the conditions in Section 3, the method also has coverage and FZ0 risk results at the forecast origin. When median error varies with the state, or when VaR and ES require different adjustments, a richer model is preferable.

Generative AI disclosure. During the preparation of this work, the authors used the generative AI tool ChatGPT to assist with language polishing. After using this tool, the authors reviewed and edited the content as needed and take full responsibility for the content of the published article.

Data availability statement. The data were derived from the following resource available in the public domain: daily closing price series for the S&P 500 (`^GSPC`), NASDAQ Composite (`^IXIC`), and FTSE 100 (`^FTSE`) from Yahoo Finance, https://finance.yahoo.com.

Supplemental Appendix

This supplement has four parts. Section A contains the proofs and supporting results. Section B describes the skew-t quantile-regression density, the data, rolling estimation, score and backtest conventions, numerical settings, reproducibility, and uncertainty in the estimated multiplier. Section C gives the Monte Carlo designs and full tables. Section D reports the additional empirical results, including filtered historical simulation and the alternative adjustment rules. Notation follows the main text; equations, tables, and figures are numbered within sections.

A Proofs and Technical Results

This section proves the properties of the adjustment rule, the state-specific and window minimizer results, the weighted-quantile representation, the two-step plug-in results, the coverage bound, and the FZ0 risk comparison at a new forecast origin. Population objects are defined from the pseudo-true path (m_t^0, q_t^0, e_t^0) generated by θ_0. For the implemented estimator, use the shorthand

$$\widehat{m}_{s|T,n} := \widehat{q}_{\bar{\tau},s|T,n}, \qquad \widehat{q}_{s|T,n} := \widehat{q}_{\tau_0,s|T,n}, \qquad \widehat{e}_{s|T,n} := \widehat{e}_{\tau_0,s|T,n},$$

and define

$$\widehat{S}_{s|T,n} := \widehat{m}_{s|T,n} - \widehat{q}_{s|T,n}, \qquad R_{s|T,n} := \frac{\widehat{m}_{s|T,n} - Y_{s+1}}{\widehat{S}_{s|T,n}}.$$

The compact multiplier interval is $\mathcal{K} = [\kappa_L, \kappa_U] \subset (0, \infty)$. All population statements below are written on the pseudo-true baseline path; hats are reserved for the implemented plug-in estimator.

A.1 Right-Tail Version ($\tau_0 > \bar{\tau}$)

When $\tau_0 > \bar{\tau}$, the adjustment applies to the right tail. Define

$$\mathcal{A}^R_{\kappa,t}(y) := \begin{cases} y, & y \le \widehat{q}_{\bar{\tau},t}, \\ \widehat{q}_{\bar{\tau},t} + \kappa\,(y - \widehat{q}_{\bar{\tau},t}), & y > \widehat{q}_{\bar{\tau},t}, \end{cases} \qquad \kappa > 0.$$

We record the changes rather than appeal only to symmetry.

Step 1: Monotonicity and inverse. Both pieces have positive slope, they agree at the

anchor, and any point below (or at) the anchor maps no higher than any point above it. Hence $\mathcal{A}^R_{\kappa,t}$ is continuous and strictly increasing. On the right tail,

$$(\mathcal{A}^R_{\kappa,t})^{-1}(y) = \widehat{q}_{\bar{\tau},t} + \kappa^{-1}(y - \widehat{q}_{\bar{\tau},t}), \qquad y > \widehat{q}_{\bar{\tau},t}.$$

Consequently, if $\widetilde{Y}_{t+1}$ has baseline conditional CDF $\widehat{F}_t$, the transformed CDF satisfies

$$F^R_{\text{adj}}(y \mid X_t) = \widehat{F}_t\big((\mathcal{A}^R_{\kappa,t})^{-1}(y) \mid X_t\big)\,.$$

For $y > \widehat{q}_{\bar{\tau},t}$, $\kappa > 1$ makes the inverse point smaller than y and therefore places more probability in the right tail; $\kappa < 1$ has the opposite effect.

Step 2: Quantile and upper-tail mean. Monotonicity gives

$$q^{R,\text{adj}}_{\tau_0,t}(\kappa) = \widehat{q}_{\bar{\tau},t} + \kappa\big(\widehat{q}_{\tau_0,t} - \widehat{q}_{\bar{\tau},t}\big).$$

If the right-tail ES convention is $e^R_{\tau_0,t} = \mathbb{E}[Y_{t+1} \mid Y_{t+1} \geq q_{\tau_0,t}]$, continuity implies that the conditioning event is preserved by the increasing map. Affinity above the anchor then yields

$$e^{R,\text{adj}}_{\tau_0,t}(\kappa) = \widehat{q}_{\bar{\tau},t} + \kappa\big(\widehat{e}^R_{\tau_0,t} - \widehat{q}_{\bar{\tau},t}\big).$$

Thus an ordered right-tail pair remains ordered for every $\kappa > 0$.

Step 3: Weighted-quantile estimator. For an in-window observation define the positive right-tail spread and ratio

$$S^R_{s|T,n} := \widehat{q}_{\tau_0,s|T,n} - \widehat{q}_{\bar{\tau},s|T,n}, \qquad R^R_{s|T,n} := \frac{Y_{s+1} - \widehat{q}_{\bar{\tau},s|T,n}}{S^R_{s|T,n}}.$$

The right-tail adjusted quantile is $\widehat{q}_{\bar{\tau},s|T,n} + \kappa S^R_{s|T,n}$, so

$$Y_{s+1} - q^{R,\text{adj}}_{\tau_0,s}(\kappa) = S^R_{s|T,n}\big(R^R_{s|T,n} - \kappa\big).$$

Positive homogeneity of check loss gives

$$\sum_s \rho_{\tau_0}\Big(Y_{s+1} - q^{R,\mathrm{adj}}_{\tau_0,s}(\kappa)\Big) = \sum_s S^R_{s|T,n}\,\rho_{\tau_0}\big(R^R_{s|T,n} - \kappa\big).$$

The subgradient condition is therefore

$$\sum_{s:R^R_{s|T,n}\le k} S^R_{s|T,n} \ge \tau_0 \sum_s S^R_{s|T,n} \ge \sum_{s:R^R_{s|T,n}<k} S^R_{s|T,n}.$$

Hence the unconstrained right-tail estimator is the spread-weighted empirical τ_0-quantile of R^R, rather than the $(1-\tau_0)$-quantile that arises in the left-tail parameterization. Projection onto $\mathcal{K}$ and the consistency, local limit, and future-origin coverage arguments then proceed with the same steps as in the left-tail case after replacing the spread and ratio by S^R and R^R.

A.2 Proof of Proposition 1

Proof. Fix X_t and write $a := \widehat{q}_{\bar{\tau},t}$ for brevity.

Step 1 (Strict monotonicity). On $[a,\infty)$, $\mathcal{A}_{\kappa,t}(y) = y$ hence slope 1. On $(-\infty,a)$, $\mathcal{A}_{\kappa,t}(y) = a - \kappa(a-y) = (1-\kappa)a + \kappa y$ hence slope $\kappa > 0$. Therefore $\mathcal{A}_{\kappa,t}(\cdot)$ is strictly increasing on each region. At the junction $y = a$ both branches coincide, so the full map is strictly increasing.

Step 2 (Continuity). At $y = a$,

$$\lim_{y\uparrow a} \mathcal{A}_{\kappa,t}(y) = (1-\kappa)a + \kappa a = a = \lim_{y\downarrow a} \mathcal{A}_{\kappa,t}(y),$$

hence $\mathcal{A}_{\kappa,t}$ is continuous.

Step 3 (Continuity of F_{adj}). If $\widehat{F}_t(\cdot \mid X_t)$ is continuous and $\mathcal{A}^{-1}_{\kappa,t}(\cdot)$ is continuous (it is, as the inverse of a strictly increasing continuous function), then the composition $F_{\mathrm{adj}}(y \mid X_t) = \widehat{F}_t(\mathcal{A}^{-1}_{\kappa,t}(y) \mid X_t)$ is continuous.

Step 4 (Tail dominance). For $y < a$, the relevant inverse branch is

$$\mathcal{A}^{-1}_{\kappa,t}(y) = a - \kappa^{-1}(a-y).$$

If $\kappa \geq 1$, then $\kappa^{-1} \leq 1$ so $\mathcal{A}_{\kappa,t}^{-1}(y) \geq a - (a - y) = y$. Since $\widehat{F}_t(\cdot \mid X_t)$ is nondecreasing,

$$F_{\text{adj}}(y \mid X_t) = \widehat{F}_t(\mathcal{A}_{\kappa,t}^{-1}(y) \mid X_t) \geq \widehat{F}_t(y \mid X_t).$$

If $0 < \kappa \leq 1$, then $\kappa^{-1} \geq 1$ so $\mathcal{A}_{\kappa,t}^{-1}(y) \leq y$ and the inequality reverses.

Step 5 (Quantile and ES transformation, and order preservation). These remaining claims of Proposition 1 are proved in Section A.7. □

A.3 Proof of Lemma 1

Proof. Fix X_t and consider the conditional objective

$$\mathcal{M}_t(\kappa) := \mathbb{E}\Big[\rho_{\tau_0}\Big(Y_{t+1} - q_{\tau_0,t}^{\text{adj},0}(\kappa)\Big) \;\Big|\; X_t\Big].$$

Step 1 (Check loss identifies conditional quantiles). For any real q, strict consistency of the check loss implies that $q \mapsto \mathbb{E}[\rho_{\tau_0}(Y_{t+1} - q) \mid X_t]$ is uniquely minimized at $q_{\tau_0,t}^\star$ provided $F^\star(\cdot \mid X_t)$ is continuous and strictly increasing near $q_{\tau_0,t}^\star$.

Step 2 (The adjusted τ_0-quantile is a strictly monotone reparameterization). Assume $\tau_0 < \bar{\tau}$, $q_t^0 < m_t^0$, and $q_{\tau_0,t}^\star < m_t^0$. Then the left-tail branch applies and

$$q_{\tau_0,t}^{\text{adj},0}(\kappa) = m_t^0 - \kappa\big(m_t^0 - q_t^0\big) = m_t^0 - \kappa S_t^0, \qquad S_t^0 := m_t^0 - q_t^0 > 0.$$

Hence $\kappa \mapsto q_{\tau_0,t}^{\text{adj},0}(\kappa)$ is continuous and strictly decreasing. The unconstrained multiplier that reaches the true quantile is

$$\kappa_{t,\text{int}} = \frac{m_t^0 - q_{\tau_0,t}^\star}{m_t^0 - q_t^0} > 0.$$

Step 3 (Constrained characterization). For $\mathcal{K} = [\kappa_L, \kappa_U]$, define the projection

$$\Pi_{\mathcal{K}}(x) := \min\{\max\{x, \kappa_L\}, \kappa_U\}.$$

Because $\mathcal{M}_t(\kappa)$ equals the strictly consistent check-loss objective evaluated at a strictly monotone

threshold, its minimizer over the compact interval $\mathcal{K}$ is

$$\kappa_t^\star = \Pi_{\mathcal{K}}(\kappa_{t,\text{int}}).$$

If $\kappa_{t,\text{int}} \in \mathcal{K}$, the adjusted threshold equals $q^\star_{\tau_0,t}$. If the ratio lies outside $\mathcal{K}$, the nearest boundary is the unique constrained minimizer and the true quantile is generally not reached. □

A.4 State-Specific and Window Minimizers

This subsection proves the decomposition of the state-specific minimizer and the result for the window minimizer summarized in Section 2.2 of the main text.

Proposition (State-specific minimizer decomposition)**.** *Let the constrained state-specific minimizer be defined on the same compact set $\mathcal{K}$ as the estimator. Since $\kappa_{t,\text{int}} = a_t + b_t$, the constrained minimizer satisfies*

$$\kappa_t^\star = \Pi_{\mathcal{K}}(a_t + b_t),$$

where $\Pi_{\mathcal{K}}(x) := \min\{\max\{x, \kappa_L\}, \kappa_U\}$ denotes projection onto $\mathcal{K} = [\kappa_L, \kappa_U]$. If $a_t + b_t \in \mathcal{K}$, this reduces to $\kappa_t^\star = a_t + b_t$ and the adjusted VaR equals the true VaR. If the baseline and true medians coincide, then $a_t = 0$.

Lemma (Window minimizer near a common multiplier)**.** *Suppose that, almost surely, for every $s \in \mathcal{W}_{T,n}$, the conditional check-loss criterion is integrable and has a unique interior minimizer $\kappa_s^\star$ with $|\kappa_s^\star - \kappa^\dagger| \le \delta_{T,n}$. Then the window minimizer defined in the main text satisfies*

$$|\kappa^0_{T,n} - \kappa^\dagger| \le \delta_{T,n}.$$

Thus the window minimizer is close to a common multiplier whenever the state-specific minimizers are uniformly close to it on the estimation window.

Proof. The unconstrained reachability ratio satisfies

$$\kappa_{t,\text{int}} = \frac{m_t^0 - q^\star_{\tau_0,t}}{S_t^0} = \frac{m_t^0 - q^\star_{\tilde{\tau},t}}{S_t^0} + \frac{q^\star_{\tilde{\tau},t} - q^\star_{\tau_0,t}}{S_t^0} = a_t + b_t.$$

The constrained state-specific minimizer is therefore $\Pi_{\mathcal{K}}(a_t + b_t)$.

For the window-minimizer claim, let $M_s(\kappa)$ denote the state-s conditional check-loss criterion. Each M_s is convex and has its unique minimizer $\kappa_s^\star$ in $[\kappa^\dagger - \delta_{T,n}, \kappa^\dagger + \delta_{T,n}]$. By convexity and uniqueness, every conditional subgradient is strictly negative below this interval and strictly positive above it. Under the stated integrability conditions, the expectation of a measurable conditional subgradient is a subgradient of the averaged criterion and has the same sign outside the interval. Hence the averaged criterion cannot be minimized below or above the interval. Any minimizer therefore lies in the same interval, proving

$$|\kappa_{T,n}^0 - \kappa^\dagger| \leq \delta_{T,n}.$$

□

A.5 ES Error after VaR Adjustment

This subsection gives the ES-error expression displayed in Section 2.2 of the main text.

Proposition (Correct VaR does not imply correct ES)**.** *Suppose the interior state-specific minimizer $\kappa_t^\star$ is used. The remaining ES error is*

$$B_{e,t} := e_t^{\mathrm{adj},0}(\kappa_t^\star) - e_{\tau_0,t}^\star = (m_t^0 - e_{\tau_0,t}^\star) - \frac{m_t^0 - q_{\tau_0,t}^\star}{m_t^0 - q_t^0}(m_t^0 - e_t^0),$$

where $e_t^{\mathrm{adj},0}(\kappa) := m_t^0 - \kappa(m_t^0 - e_t^0)$. A constant multiplier can therefore give correct VaR coverage while ES remains misspecified. Correct ES requires an additional restriction on the lower-tail quantile curve.

Proof. The state-specific minimizer identity gives

$$\kappa_t^\star = \frac{m_t^0 - q_{\tau_0,t}^\star}{m_t^0 - q_t^0}.$$

Substituting this value into the ES transformation yields

$$e_{\tau_0,t}^{\mathrm{adj},0}(\kappa_t^\star) = m_t^0 - \frac{m_t^0 - q_{\tau_0,t}^\star}{m_t^0 - q_t^0}(m_t^0 - e_t^0).$$

Subtracting $e_{\tau_0,t}^\star$ gives the expression in the main text. The expression is zero if and only if the

same multiplier correctly adjusts the baseline VaR and ES. A median error that is constant relative to the baseline spread may still yield a stable VaR minimizer, but it does not impose the additional restriction needed for ES. Thus correct VaR does not imply correct ES. □

A.6 A Location–Scale Sufficient Condition

This subsection states the sufficient condition behind the location–scale discussion in Section 2.2 of the main text.

Proposition (Location–scale sufficient condition)**.** *Suppose* $Y_{t+1} = \mu_t + \sigma_t \varepsilon_{t+1}$ *with* $\sigma_t > 0$ *and true innovation law* G*, while the pseudo-true baseline uses the same location and scale but innovation law* $\widehat{G}$*:*

$$q^{\star}_{u,t} = \mu_t + \sigma_t G^{-1}(u), \qquad q^{0}_{u,t} = \mu_t + \sigma_t \widehat{G}^{-1}(u).$$

If $\widehat{G}^{-1}(\bar{\tau}) \neq \widehat{G}^{-1}(\tau_0)$*, then substituting the two representations into the definitions of* a_t *and* b_t *gives*

$$a_t = \frac{\widehat{G}^{-1}(\bar{\tau}) - G^{-1}(\bar{\tau})}{\widehat{G}^{-1}(\bar{\tau}) - \widehat{G}^{-1}(\tau_0)}, \qquad b_t = \frac{G^{-1}(\bar{\tau}) - G^{-1}(\tau_0)}{\widehat{G}^{-1}(\bar{\tau}) - \widehat{G}^{-1}(\tau_0)}$$

for every state t*, because the state-specific scale* σ_t *cancels in both ratios. Hence* $\kappa_{t,\mathrm{int}} = a_t + b_t$ *does not vary with the state. If in addition* $G^{-1}(\bar{\tau}) = \widehat{G}^{-1}(\bar{\tau})$*, then* $a_t = 0$ *and* $\kappa_{t,\mathrm{int}} = b_t$*. If, along a local-misspecification or triangular-array sequence, the representation holds up to errors that are uniformly* $o(1)$ *over the relevant window at* $u \in \{\tau_0, \bar{\tau}\}$*, and the baseline spread is bounded away from zero, the same ratio approximates* b_t *uniformly with* $o(1)$ *error.*

Proof. Consider first the maintained location–scale case. Under the true law and the baseline law,

$$q^{\star}_{u,t} = \mu_t + \sigma_t G^{-1}(u), \qquad q^{0}_{u,t} = \mu_t + \sigma_t \widehat{G}^{-1}(u).$$

Hence

$$m^0_t - q^{\star}_{\bar{\tau},t} = \sigma_t\{\widehat{G}^{-1}(\bar{\tau}) - G^{-1}(\bar{\tau})\}, \qquad S^{\star}_t(\tau_0, \bar{\tau}) = q^{\star}_{\bar{\tau},t} - q^{\star}_{\tau_0,t} = \sigma_t\{G^{-1}(\bar{\tau}) - G^{-1}(\tau_0)\},$$

and

$$S^0_t(\tau_0, \bar{\tau}) = m^0_t - q^0_t = \sigma_t\{\widehat{G}^{-1}(\bar{\tau}) - \widehat{G}^{-1}(\tau_0)\}.$$

When the baseline innovation spread is nonzero, the state-dependent scale cancels in both normalized quantities:

$$a_t = \frac{m_t^0 - q_{\bar{\tau},t}^\star}{S_t^0(\tau_0,\bar{\tau})} = \frac{\widehat{G}^{-1}(\bar{\tau}) - G^{-1}(\bar{\tau})}{\widehat{G}^{-1}(\bar{\tau}) - \widehat{G}^{-1}(\tau_0)}, \qquad b_t(\tau_0,\bar{\tau}) = \frac{S_t^\star(\tau_0,\bar{\tau})}{S_t^0(\tau_0,\bar{\tau})} = \frac{G^{-1}(\bar{\tau}) - G^{-1}(\tau_0)}{\widehat{G}^{-1}(\bar{\tau}) - \widehat{G}^{-1}(\tau_0)}.$$

Both ratios are constant in the state, and they telescope to

$$\kappa_{t,\mathrm{int}} = a_t + b_t = \frac{\widehat{G}^{-1}(\bar{\tau}) - G^{-1}(\tau_0)}{\widehat{G}^{-1}(\bar{\tau}) - \widehat{G}^{-1}(\tau_0)},$$

which is also constant in the state. If the innovation medians coincide, so that $G^{-1}(\bar{\tau}) = \widehat{G}^{-1}(\bar{\tau})$, then $a_t = 0$ and $\kappa_{t,\mathrm{int}} = b_t$. If the common value lies in the interior of $\mathcal{K}$, then at this value the adjusted target quantile equals $q_{\tau_0,t}^\star$ for every state t; by conditional quantile optimality, it minimizes each state-specific check-loss criterion and therefore also the pooled criterion that defines κ_0. The volatility scale cancels, so the result allows time-varying conditional heteroskedasticity.

Now consider a local-misspecification or triangular-array sequence in which the location–scale quantile representation holds up to errors that are uniformly $o(1)$ over $u \in \{\tau_0, \bar{\tau}\}$ and states in the relevant window. These errors perturb both $S_t^\star$ and S_t^0 by uniformly $o(1)$ terms. If the baseline tail spread is uniformly bounded away from zero, the denominator of $b_t = S_t^\star / S_t^0$ remains bounded away from zero, and a first-order ratio expansion gives

$$\sup_{t \in \mathcal{W}_{T,n}} \left| b_t(\tau_0,\bar{\tau}) - \frac{G^{-1}(\bar{\tau}) - G^{-1}(\tau_0)}{\widehat{G}^{-1}(\bar{\tau}) - \widehat{G}^{-1}(\tau_0)} \right| = o(1),$$

which is the asserted approximate constant-ratio result. □

A.7 Transformation of Quantiles and Expected Shortfall

Lemma A.1 (Quantile and ES transformation under $\mathcal{A}_{\kappa,t}$)**.** *Assume $\widehat{F}_t(\cdot \mid X_t)$ is continuous and strictly increasing at the relevant quantiles, including the lower-tail range $u \in (0, \tau_0]$, so that $\mathcal{A}_{\kappa,t}(\cdot)$ is strictly increasing and continuous and the quantile representation for ES is well defined. Then:*

(i) For any $\tau \in (0,1)$,

$$q_{\tau,t}^{\mathrm{adj}}(\kappa) = \mathcal{A}_{\kappa,t}(\widehat{q}_{\tau,t}).$$

In particular, if $\tau \le \bar{\tau}$, then

$$q^{\mathrm{adj}}_{\tau,t}(\kappa) = \widehat{q}_{\bar{\tau},t} - \kappa\big(\widehat{q}_{\bar{\tau},t} - \widehat{q}_{\tau,t}\big).$$

(ii) If $\tau_0 < \bar{\tau}$ and $\widehat{e}_{\tau_0,t}$ exists, then

$$e^{\mathrm{adj}}_{\tau_0,t}(\kappa) = \widehat{q}_{\bar{\tau},t} - \kappa\big(\widehat{q}_{\bar{\tau},t} - \widehat{e}_{\tau_0,t}\big).$$

Proof. Fix X_t and use $\mathcal{A}_{\kappa,t}(\cdot)$ as defined in the main text.

Step 1 (Distribution identity). Because $\widetilde{Y}^{\mathrm{adj}}_{t+1} = \mathcal{A}_{\kappa,t}(\widetilde{Y}_{t+1})$ and $\mathcal{A}_{\kappa,t}$ is strictly increasing,

$$F_{\mathrm{adj}}(y \mid X_t) = \Pr\big(\mathcal{A}_{\kappa,t}(\widetilde{Y}_{t+1}) \le y \mid X_t\big) = \Pr\big(\widetilde{Y}_{t+1} \le \mathcal{A}^{-1}_{\kappa,t}(y) \mid X_t\big) = \widehat{F}_t(\mathcal{A}^{-1}_{\kappa,t}(y) \mid X_t).$$

Step 2 (Quantiles commute with strictly increasing maps). Assume $\widehat{F}_t(\cdot \mid X_t)$ is continuous and strictly increasing in a neighborhood of $\widehat{q}_{\tau,t}$, so the τ-quantile is unique. Let $\widehat{q}_{\tau,t}$ be the τ-quantile of $\widehat{F}_t(\cdot \mid X_t)$ and let $q^{\mathrm{adj}}_{\tau,t}(\kappa)$ be the τ-quantile of $F_{\mathrm{adj}}(\cdot \mid X_t)$. Using Step 1 and strict monotonicity,

$$F_{\mathrm{adj}}(\mathcal{A}_{\kappa,t}(\widehat{q}_{\tau,t}) \mid X_t) = \widehat{F}_t(\widehat{q}_{\tau,t} \mid X_t) = \tau,$$

and similarly, for any $y < \mathcal{A}_{\kappa,t}(\widehat{q}_{\tau,t})$ we have $\mathcal{A}^{-1}_{\kappa,t}(y) < \widehat{q}_{\tau,t}$, hence $F_{\mathrm{adj}}(y \mid X_t) = \widehat{F}_t(\mathcal{A}^{-1}_{\kappa,t}(y) \mid X_t) < \tau$ by continuity/strict increase near the quantile. Therefore $q^{\mathrm{adj}}_{\tau,t}(\kappa) = \mathcal{A}_{\kappa,t}(\widehat{q}_{\tau,t})$.

For $\tau \le \bar{\tau}$, we have $\widehat{q}_{\tau,t} \le \widehat{q}_{\bar{\tau},t}$ so the left-tail branch applies:

$$\mathcal{A}_{\kappa,t}(\widehat{q}_{\tau,t}) = \widehat{q}_{\bar{\tau},t} - \kappa\big(\widehat{q}_{\bar{\tau},t} - \widehat{q}_{\tau,t}\big),$$

which proves part (i).

Step 3 (ES via quantile representation). Assume $\tau_0 < \bar{\tau}$ so that the affine quantile formula above holds for all $u \in (0, \tau_0]$:

$$q^{\mathrm{adj}}_{u,t}(\kappa) = \widehat{q}_{\bar{\tau},t} - \kappa\big(\widehat{q}_{\bar{\tau},t} - \widehat{q}_{u,t}\big).$$

When $\widehat{F}_t(\cdot \mid X_t)$ is continuous and $\widehat{e}_{\tau_0}(X_t)$ exists, we can use the standard quantile representation

$$e_{\tau_0}(X_t) = \frac{1}{\tau_0} \int_0^{\tau_0} q_{u,t}\, du.$$

Thus

$$\begin{aligned} e^{\mathrm{adj}}_{\tau_0}(X_t;\kappa) &= \frac{1}{\tau_0} \int_0^{\tau_0} q^{\mathrm{adj}}_{u,t}(\kappa)\, du \\ &= \frac{1}{\tau_0} \int_0^{\tau_0} \Big\{ \widehat{q}_{\bar{\tau},t} - \kappa(\widehat{q}_{\bar{\tau},t} - \widehat{q}_{u,t}) \Big\}\, du \\ &= \widehat{q}_{\bar{\tau},t} - \kappa\Big(\widehat{q}_{\bar{\tau},t} - \frac{1}{\tau_0} \int_0^{\tau_0} \widehat{q}_{u,t}\, du \Big) \\ &= \widehat{q}_{\bar{\tau},t} - \kappa\big(\widehat{q}_{\bar{\tau},t} - \widehat{e}_{\tau_0}(X_t) \big), \end{aligned}$$

which proves part (ii).

Finally, subtracting the two displays gives

$$e^{\mathrm{adj}}_{\tau_0,t}(\kappa) - q^{\mathrm{adj}}_{\tau_0,t}(\kappa) = \kappa\big(\widehat{e}_{\tau_0,t} - \widehat{q}_{\tau_0,t}\big) \le 0, \qquad q^{\mathrm{adj}}_{\tau_0,t}(\kappa) - \widehat{q}_{\bar{\tau},t} = -\kappa\big(\widehat{q}_{\bar{\tau},t} - \widehat{q}_{\tau_0,t}\big) < 0,$$

for every $\kappa > 0$ whenever the baseline triple is ordered, $\widehat{e}_{\tau_0,t} \le \widehat{q}_{\tau_0,t} < \widehat{q}_{\bar{\tau},t}$. This is the order-preservation claim of Proposition 1. □

A.8 Proof of Lemma 2

Proof for the plug-in estimator. Fix a forecast origin T, a window length n, and $\mathcal{W}_{T,n} = \{T - n, \ldots, T-1\}$. Put $w_s := \widehat{S}_{s|T,n} > 0$ and $R_s := R_{s|T,n}$. The sample criterion is

$$\widehat{M}_{T,n}(\kappa) = \frac{1}{n} \sum_{s \in \mathcal{W}_{T,n}} \rho_{\tau_0}\{w_s(\kappa - R_s)\}.$$

Step 1: Positive homogeneity. For $a > 0$, $\rho_{\tau_0}(au) = a\rho_{\tau_0}(u)$. Therefore

$$\widehat{M}_{T,n}(\kappa) = \frac{1}{n} \sum_{s \in \mathcal{W}_{T,n}} w_s \rho_{\tau_0}(\kappa - R_s).$$

Step 2: Subgradient. The subgradient of check loss is

$$\partial\rho_{\tau_0}(u) = \begin{cases} \{\tau_0 - 1\}, & u < 0, \\ [\tau_0 - 1, \tau_0], & u = 0, \\ \{\tau_0\}, & u > 0. \end{cases}$$

Hence

$$\partial_\kappa \widehat{M}_{T,n}(\kappa) = \frac{1}{n} \sum_{s \in \mathcal{W}_{T,n}} w_s \, \partial\rho_{\tau_0}(\kappa - R_s).$$

Step 3: Quantile inequalities. An unconstrained minimizer $\widetilde{\kappa}$ satisfies $0 \in \partial_\kappa \widehat{M}_{T,n}(\widetilde{\kappa})$. Separating observations below, above, and tied at $\widetilde{\kappa}$ gives

$$\sum_{s:R_s \le \widetilde{\kappa}} w_s \ge (1 - \tau_0) \sum_s w_s \ge \sum_{s:R_s < \widetilde{\kappa}} w_s.$$

These are the defining inequalities for a weighted empirical $(1 - \tau_0)$-quantile. Conversely, the inequalities imply that zero belongs to the subgradient, so every such weighted quantile is a minimizer.

Step 4: Ties, bounds, and the implemented selection. If the quantile set contains an interval because of ties, the objective is flat on that interval. Minimization over $\mathcal{K} = [\kappa_L, \kappa_U]$ projects the unconstrained quantile set onto $\mathcal{K}$. The implementation selects the smallest admissible weighted quantile, matching the convention of the main text. □

Corollary 1 (Cross-fitted sensitivity)**.** *If w_s and R_s are replaced by positive cross-fitted spreads and their associated ratios, the same finite-sample argument applies verbatim. This algebraic fact does not make the cross-fitted and full-window estimators first-order equivalent.*

A.9 Proof of Proposition 2

This proof treats the estimator used in the empirical analysis: the baseline is fitted once on the current window, that fitted model generates the in-window median and target-quantile paths, and the multiplier is estimated from those fitted paths.

Let n be the estimation-window length, $\widehat{\theta}_n$ the baseline estimator, and θ_0 its pseudo-true prob-

ability limit. Write

$$m_s(\theta) := q_{\bar{\tau},s}(\theta), \qquad q_s(\theta) := q_{\tau_0,s}(\theta), \qquad S_s(\theta) := m_s(\theta) - q_s(\theta),$$

and

$$q_s^{\mathrm{adj}}(\kappa, \theta) = m_s(\theta) - \kappa S_s(\theta).$$

The implemented criterion is

$$\widehat{M}_n(\kappa, \widehat{\theta}_n) = \frac{1}{n} \sum_{s=1}^{n} \rho_{\tau_0}\{Y_{s+1} - q_s^{\mathrm{adj}}(\kappa, \widehat{\theta}_n)\}.$$

Define the fixed-first-stage population criterion

$$M(\kappa, \theta_0) = \mathbb{E}\left[\rho_{\tau_0}\{Y_{t+1} - q_t^{\mathrm{adj}}(\kappa, \theta_0)\}\right]$$

and let κ_0 be its unique minimizer on $\mathcal{K}$.

Proof of consistency. **Step 1: Lipschitz control of the generated path.** The check loss is one-Lipschitz. For $\kappa \in \mathcal{K}$,

$$\left|\rho_{\tau_0}\{Y_{s+1} - q_s^{\mathrm{adj}}(\kappa, \theta)\} - \rho_{\tau_0}\{Y_{s+1} - q_s^{\mathrm{adj}}(\kappa, \theta_0)\}\right| \leq \left|q_s^{\mathrm{adj}}(\kappa, \theta) - q_s^{\mathrm{adj}}(\kappa, \theta_0)\right|.$$

Because $\mathcal{K}$ is compact,

$$\sup_{\kappa \in \mathcal{K}} \left|q_s^{\mathrm{adj}}(\kappa, \theta) - q_s^{\mathrm{adj}}(\kappa, \theta_0)\right| \leq C\left\{|m_s(\theta) - m_s(\theta_0)| + |q_s(\theta) - q_s(\theta_0)|\right\}$$

for a finite constant C depending only on the bounds of $\mathcal{K}$.

Step 2: Uniformly negligible first-stage replacement. Under the local Lipschitz condition in the main-text two-step regularity assumption, there is a stationary integrable envelope L_s such that the right-hand side in Step 1 is at most $L_s\|\theta - \theta_0\|$ in a neighborhood of θ_0. Hence

$$\sup_{\kappa \in \mathcal{K}} \left|\widehat{M}_n(\kappa, \widehat{\theta}_n) - \widehat{M}_n(\kappa, \theta_0)\right| \leq \|\widehat{\theta}_n - \theta_0\| \frac{1}{n} \sum_{s=1}^{n} L_s = o_P(1),$$

because $\widehat{\theta}_n \to_P \theta_0$ and $n^{-1}\sum_s L_s = O_P(1)$.

Step 3: Uniform law for the fixed-first-stage criterion. Write $\ell_s(\kappa,\theta) := \rho_{\tau_0}\{Y_{s+1} - q_s^{\mathrm{adj}}(\kappa,\theta)\}$ for the per-observation criterion. For fixed θ_0, the loss is Lipschitz in κ with envelope $S_s(\theta_0)$:

$$|\ell_s(\kappa,\theta_0) - \ell_s(\kappa',\theta_0)| \le |\kappa - \kappa'| S_s(\theta_0).$$

The maintained mixing and moment conditions therefore give

$$\sup_{\kappa\in\mathcal{K}} \left| \widehat{M}_n(\kappa,\theta_0) - M(\kappa,\theta_0) \right| = o_P(1).$$

Step 4: Uniform convergence of the implemented criterion. By the triangle inequality and Steps 2–3,

$$\sup_{\kappa\in\mathcal{K}} \left| \widehat{M}_n(\kappa,\widehat{\theta}_n) - M(\kappa,\theta_0) \right| = o_P(1).$$

Step 5: Argmin. Continuity of $M(\cdot,\theta_0)$ on compact $\mathcal{K}$ and uniqueness of its minimizer imply strict separation outside every neighborhood of κ_0. The uniform convergence in Step 4 and the standard argmin theorem (van der Vaart and Wellner, 1996) then yield

$$\widehat{\kappa}_n \to_P \kappa_0.$$

□

Model-specific conditions. For a GARCH baseline, Step 2 requires consistency of the quasi-maximum likelihood estimator (QMLE) or pseudo-QMLE, stable filtering, and a uniform bound on the effect of initialization on the fitted conditional scale. For the quantile-regression / skew-t baseline, it requires convergence of the fitted quantile coefficients and continuity of the density-matching step. These are sufficient readings of the high-level path-replacement condition in Step 2; the proposition does not treat the generated fitted path as an ordinary stationary sequence without this argument.

A.10 Two-Step Asymptotic Normality

This subsection states the high-level limit theory summarized in Section 3.1 of the main text.

Assumption (High-level local limit conditions)**.** *In addition to the two-step plug-in regularity conditions (Assumption 1 of the main text), suppose: (i) the first-stage estimator has an asymptotically linear representation; (ii) the second-stage score admits a uniform local linearization around κ_0; (iii) the derivative limit is finite and bounded away from zero; (iv) the complete two-step influence process satisfies a central limit theorem with finite long-run variance; and (v)*

$$\max_{1\le s\le n} S_s(\theta_0)/\sqrt{n} = o_P(1).$$

A stationary $S_s(\theta_0)$ with a finite $(2+\delta)$ moment for some $\delta > 0$ is a sufficient maximal-weight condition.

Proposition (High-level asymptotic normality)**.** *Under the main text's Assumption 1 and the conditions above,*

$$\sqrt{n}(\widehat{\kappa}_n - \kappa_0) \Rightarrow \mathcal{N}(0, J^{-2}\Omega),$$

where J is the limiting derivative of the second-stage score and Ω is the long-run variance of the complete two-step influence function.

The main-text proposition is intentionally high-level. This section records the two-step expansion that the assumptions imply and explains why a single fitted-path HAC estimate need not be a complete variance estimator.

Let

$$\psi_s(\kappa, \theta) := S_s(\theta)\left[\mathbf{1}\{R_s(\theta) \le \kappa\} - (1-\tau_0)\right]$$

be the second-stage score, with $R_s(\theta) = \{m_s(\theta) - Y_{s+1}\}/S_s(\theta)$. Suppose the first-stage estimator satisfies

$$\sqrt{n}(\widehat{\theta}_n - \theta_0) = \frac{1}{\sqrt{n}}\sum_{s=1}^{n} IF_{\theta,s} + o_P(1), \tag{A.1}$$

where $IF_{\theta,s}$ is the first-stage influence function, and the local score expansion is

$$0 = \frac{1}{n}\sum_{s=1}^{n}\psi_s(\kappa_0, \theta_0) + J(\widehat{\kappa}_n - \kappa_0) + \Gamma'(\widehat{\theta}_n - \theta_0) + o_P(n^{-1/2}), \tag{A.2}$$

where $J > 0$ and Γ is the derivative of the population score with respect to the first-stage parameter.

Derivation of the two-step representation. **Step 1: Subgradient at the estimated quantile.** The weighted-quantile first-order condition makes the sample score zero up to the largest normalized observation weight. The main-text maximal-weight condition states the required remainder directly. A convenient sufficient condition is $\mathbb{E}[S_t(\theta_0)^{2+\delta}] < \infty$ for some $\delta > 0$: by the union bound and Markov's inequality,

$$\Pr\left(\max_{1\le s\le n} S_s(\theta_0) > \varepsilon\sqrt{n}\right) \le \frac{\mathbb{E}[S_t(\theta_0)^{2+\delta}]}{\varepsilon^{2+\delta} n^{\delta/2}} \longrightarrow 0.$$

Thus the normalized maximal weight is $o_P(n^{-1/2})$.

Step 2: Local expansion. The high-level stochastic differentiability condition expands the sample score jointly in κ and θ around (κ_0, θ_0), giving (A.2).

Step 3: Substitute the first-stage influence function. Substituting (A.1) into (A.2) and multiplying by $\sqrt{n}$ gives

$$\sqrt{n}(\widehat{\kappa}_n - \kappa_0) = -J^{-1}\frac{1}{\sqrt{n}}\sum_{s=1}^{n}\left\{\psi_s(\kappa_0, \theta_0) + \Gamma' IF_{\theta,s}\right\} + o_P(1).$$

Step 4: Central limit theorem. If the complete influence process in braces satisfies a time-series CLT with long-run variance Ω, Slutsky's theorem yields

$$\sqrt{n}(\widehat{\kappa}_n - \kappa_0) \Rightarrow N(0, J^{-2}\Omega).$$

□

Variance estimation. A Newey–West calculation on the realized fitted-path score $\psi_s(\widehat{\kappa}_n, \widehat{\theta}_n)$ conditions on the common fitted parameter and generally omits the term $\Gamma' IF_{\theta,s}$. It is therefore a conditional-on-fit uncertainty measure. A fully first-stage-aware bootstrap would resample the underlying dated series, refit the baseline in each draw, reconstruct the fitted path, and re-estimate κ under a clearly specified resampling design. The table below reports only the reproducible conditional-on-fit HAC calculation, and no empirical conclusion relies on these standard errors.

Cross-fitting. A two-block sample split changes the first-stage estimation problem and is much noisier in a 1,000-day window. We therefore use it only as a deliberately demanding sensitivity, not as the main estimator or as a substitute for the plug-in theory. In known-target simulations

at the 5% tail, the full-window plug-in estimator has lower RMSE in all twelve design cells at $n \in \{500, 1000\}$. At the 1% tail its sampling standard deviation is lower in eleven of twelve cells; in the Gaussian design, the two-block mean is about 1.82 against a known target near 1.36. These results show the cost of halving the first-stage sample and explain why the split exercise is reported as a demanding sensitivity check rather than as the preferred construction.

A.11 Proof of Theorem 1

Let $\mathcal{I}_T$ be the information set at the current forecast origin. The state-sufficiency condition is $F^\star(\cdot \mid \mathcal{I}_T) = F^\star(\cdot \mid X_T)$ almost surely. Assume the conditional density is bounded by $\bar{f}$ on a neighborhood of $q^\star_{\tau_0,T}$. Define

$$\widehat{m}_T := \widehat{q}_{\bar{\tau},T|T,n}, \qquad \widehat{q}_T := \widehat{q}_{\tau_0,T|T,n}, \qquad \widehat{S}_T := \widehat{m}_T - \widehat{q}_T,$$

and

$$\widehat{q}^{\,\mathrm{adj}}_T := \widehat{m}_T - \widehat{\kappa}_{T,n}\widehat{S}_T.$$

On the pseudo-true path, $q^\star_{\tau_0,T} = m^0_T - \kappa^\star_T S^0_T$ by the state-specific minimizer identity.

Proof. **Step 1: Decompose the forecast-threshold error.** Add and subtract the pseudo-true path:

$$\widehat{q}^{\,\mathrm{adj}}_T - q^\star_{\tau_0,T} = (\widehat{m}_T - m^0_T) - \widehat{\kappa}_{T,n}(\widehat{S}_T - S^0_T) - (\widehat{\kappa}_{T,n} - \kappa^\star_T)S^0_T.$$

Since $\widehat{\kappa}_{T,n} \in [\kappa_L, \kappa_U]$,

$$|\widehat{q}^{\,\mathrm{adj}}_T - q^\star_{\tau_0,T}| \le r_{T,n} + S^0_T\left(|\widehat{\kappa}_{T,n} - \kappa^0_{T,n}| + |\kappa^0_{T,n} - \kappa^\star_T|\right),$$

where $r_{T,n} := |\widehat{m}_T - m^0_T| + \kappa_U|\widehat{S}_T - S^0_T|$. The displayed terms separate current-path estimation, estimation of the multiplier around the window minimizer, and the difference between the window minimizer and the current-state minimizer. Write their sum as

$$\Delta_{T,n} := r_{T,n} + S^0_T\left(|\widehat{\kappa}_{T,n} - \kappa^0_{T,n}| + |\kappa^0_{T,n} - \kappa^\star_T|\right).$$

Step 2: Convert threshold error into coverage error. By state sufficiency, $F^\star(q^\star_{\tau_0,T} \mid \mathcal{I}_T) = \tau_0$. On $\{\Delta_{T,n} \le \delta_0\}$, the adjusted threshold remains in the density-bounded neighborhood, so the mean-value theorem gives

$$\left| F^\star(\widehat{q}^{\,\mathrm{adj}}_T \mid \mathcal{I}_T) - \tau_0 \right| \le \bar{f}\, \Delta_{T,n}.$$

On the complementary event the probability difference is bounded by one. Therefore

$$\left| \Pr(Y_{T+1} \le \widehat{q}^{\,\mathrm{adj}}_T \mid \mathcal{I}_T) - \tau_0 \right| \le \bar{f}\, \Delta_{T,n} \mathbf{1}\{\Delta_{T,n} \le \delta_0\} + \mathbf{1}\{\Delta_{T,n} > \delta_0\}.$$

If $\Delta_{T,n} = o_P(1)$, the exit-event indicator is $o_P(1)$ and the asymptotic bound stated in the main text follows. □

Under the stationary high-level conditions of the main-text consistency and CLT propositions, the estimation term is $O_P(n^{-1/2})$; under local change, the theorem gives the displayed decomposition without assigning a root-n rate, consistent with the main-text discussion.

A.12 A Bound for the Adjusted Pair

This subsection states the deterministic error bound used after the FZ0 risk proposition. The maintained condition requires a common approximation to the lower-tail quantile curve.

A convenient sufficient condition is a common approximation to the lower-tail segment. Suppose there is a positive constant $\kappa^\dagger$ and a sequence $\zeta_T \ge 0$ such that, for $u \in (0, \tau_0]$,

$$\kappa^\dagger(m^0_T - q^0_{u,T}) = q^\star_{\tilde{\tau},T} - q^\star_{u,T} + \xi_{u,T}, \qquad \sup_{u \in (0,\tau_0]} |\xi_{u,T}| \le \zeta_T. \tag{A.3}$$

This condition is stronger than VaR alignment because ES averages the lower quantile segment. It is a sufficient condition for the score result below, not a consequence of correct VaR coverage.

Let $\alpha_T := |m^0_T - q^\star_{\tilde{\tau},T}|$ and define the current fitted-path error

$$r^z_{T,n} := |\widehat{m}_T - m^0_T| + |\widehat{q}_T - q^0_T| + |\widehat{e}_T - e^0_T|.$$

If the baseline tail distances are locally bounded, the affine map and the quantile representation of

ES give the following deterministic error decomposition.

Proposition (Error bound for the adjusted pair)**.** *Under* (A.3)*, there is a finite constant* C *such that*

$$\left\| \begin{pmatrix} \widehat{q}_T^{\,\mathrm{adj}} - q_T^\star \\ \widehat{e}_T^{\,\mathrm{adj}} - e_T^\star \end{pmatrix} \right\| \le C\, \varepsilon_T^{\mathrm{pair}}, \qquad \varepsilon_T^{\mathrm{pair}} := r_{T,n}^z + |\widehat{\kappa}_{T,n} - \kappa_{T,n}^0| + |\kappa_{T,n}^0 - \kappa^\dagger| + \alpha_T + \zeta_T. \tag{A.4}$$

The bound separates error in the fitted values, estimation of the multiplier, the difference between the window and current state, median error, and approximation error in the lower-tail quantile curve.

Proof. For this standalone proof, write

$$(\widehat{m}_T, \widehat{q}_T, \widehat{e}_T) := (\widehat{q}_{\bar{\tau},T|T,n}, \widehat{q}_{\tau_0,T|T,n}, \widehat{e}_{\tau_0,T|T,n})$$

and

$$r_{T,n}^z := |\widehat{m}_T - m_T^0| + |\widehat{q}_T - q_T^0| + |\widehat{e}_T - e_T^0|.$$

Write the current pseudo-true lower-tail distances as $S_{q,T}^0 = m_T^0 - q_T^0$ and $S_{e,T}^0 = m_T^0 - e_T^0$. At $u = \tau_0$, the main-text lower-tail segment approximation gives

$$|m_T^0 - \kappa^\dagger S_{q,T}^0 - q_T^\star| \le \alpha_T + \zeta_T.$$

The quantile representation of ES and the uniform segment bound imply the analogous inequality

$$|m_T^0 - \kappa^\dagger S_{e,T}^0 - e_T^\star| \le \alpha_T + \zeta_T.$$

Add and subtract the pseudo-true pair evaluated at $\kappa^\dagger$, then add and subtract the window minimizer $\kappa_{T,n}^0$. Bounded lower-tail distances and $\widehat{\kappa}_{T,n} \in \mathcal{K}$ give

$$|\widehat{q}_T^{\,\mathrm{adj}} - q_T^\star| \le C\{r_{T,n}^z + |\widehat{\kappa}_{T,n} - \kappa_{T,n}^0| + |\kappa_{T,n}^0 - \kappa^\dagger| + \alpha_T + \zeta_T\},$$
$$|\widehat{e}_T^{\,\mathrm{adj}} - e_T^\star| \le C\{r_{T,n}^z + |\widehat{\kappa}_{T,n} - \kappa_{T,n}^0| + |\kappa_{T,n}^0 - \kappa^\dagger| + \alpha_T + \zeta_T\}.$$

Combining the two coordinate bounds gives the stated pair-error bound. □

A.13 Proofs of Proposition 3 and Theorem 2

The score comparison is stated at a new forecast origin. The estimated multiplier and the raw and adjusted forecasts are $\mathcal{I}_T$-measurable, while Y_{T+1} is not used in their construction. This measurability condition is required before strict consistency can be applied to a random forecast.

Let

$$\mathcal{R}_T(q,e) := \mathbb{E}\left[s_{\tau_0}(q,e;Y_{T+1}) \mid \mathcal{I}_T\right]$$

be conditional FZ0 risk, and let $(q_T^\star, e_T^\star)$ be the true conditional VaR–ES pair. We first derive the local curvature used below. Define

$$A_T(q) := \mathbb{E}[(q - Y_{T+1})\mathbf{1}\{Y_{T+1} \le q\} \mid \mathcal{I}_T]\,.$$

If $F_T^\star$ is differentiable with a continuous conditional density $f_T^\star$ in a neighborhood of $q_T^\star$, then $A_T'(q) = F_T^\star(q)$ and $A_T''(q) = f_T^\star(q)$ on that neighborhood. Taking the conditional expectation of FZ0 yields

$$\mathcal{R}_T(q,e) = -\frac{A_T(q)}{\tau_0 e} + \frac{q}{e} + \log(-e) - 1.$$

The first derivatives are

$$\begin{aligned}
\partial_q \mathcal{R}_T(q,e) &= \frac{1 - F_T^\star(q)/\tau_0}{e}, \\
\partial_e \mathcal{R}_T(q,e) &= \frac{A_T(q)/\tau_0 - q}{e^2} + \frac{1}{e}.
\end{aligned}$$

At the true pair, $F_T^\star(q_T^\star) = \tau_0$ and $A_T(q_T^\star) = \tau_0(q_T^\star - e_T^\star)$, so both derivatives vanish. Differentiating once more gives

$$\begin{aligned}
\partial^2_{qq} \mathcal{R}_T(q,e) &= -\frac{f_T^\star(q)}{\tau_0 e}, \\
\partial^2_{qe} \mathcal{R}_T(q,e) &= -\frac{1 - F_T^\star(q)/\tau_0}{e^2}, \\
\partial^2_{ee} \mathcal{R}_T(q,e) &= -\frac{2\{A_T(q)/\tau_0 - q\}}{e^3} - \frac{1}{e^2}.
\end{aligned}$$

Substitution of the true-pair identities therefore gives

$$H_T(q_T^\star, e_T^\star) = \begin{pmatrix} \dfrac{f_T^\star(q_T^\star)}{\tau_0(-e_T^\star)} & 0 \\ 0 & \dfrac{1}{(e_T^\star)^2} \end{pmatrix}.$$

Both diagonal elements are positive when $e_T^\star < 0$ and the conditional density is positive at the quantile. Thus conditional FZ0 risk is locally quadratic with positive curvature.

Define the current raw fitted pair $(\widehat{q}_T, \widehat{e}_T) = (\widehat{q}_{\tau_0,T|T,n}, \widehat{e}_{\tau_0,T|T,n})$ and the corresponding adjusted pair $(\widehat{q}_T^{\mathrm{adj}}, \widehat{e}_T^{\mathrm{adj}})$. Let

$$D_T := \frac{\lambda_{\min}}{4}\{(\widehat{q}_T - q_T^\star)^2 + (\widehat{e}_T - e_T^\star)^2\}, \qquad \mathcal{E}_T^{\mathrm{adj}} := (\widehat{q}_T^{\mathrm{adj}} - q_T^\star)^2 + (\widehat{e}_T^{\mathrm{adj}} - e_T^\star)^2.$$

The second term includes the ES bias derived in Section A.5; it is not removed merely because the VaR minimizer is stable.

Assumption A.1 (Future-origin score-comparison conditions)**.** *With probability approaching one: (i) the raw and adjusted pairs are in a common convex neighborhood of the true pair on which the conditional FZ risk has a Hessian whose eigenvalues are uniformly bounded above and bounded below away from zero; (ii) all forecasts are $\mathcal{I}_T$-measurable and score-domain valid; and (iii) neighborhood-exit contributions are $o_P(D_T)$. The rate condition $\mathcal{E}_T^{\mathrm{adj}} = o_P(D_T)$ is stated as an explicit hypothesis of the risk-gap proposition; by the error bound for the adjusted pair, $(\varepsilon_T^{\mathrm{pair}})^2 = o_P(D_T)$ is sufficient for it.*

Proof of the future-origin comparison. **Step 1: Strict consistency.** Conditional strict consistency gives

$$\mathcal{R}_T(q, e) - \mathcal{R}_T(q_T^\star, e_T^\star) \geq 0$$

for every admissible $\mathcal{I}_T$-measurable pair, with equality only at the true pair under uniqueness.

Step 2: Local upper bound for the adjusted pair. A second-order Taylor expansion and the bounded Hessian give

$$0 \leq \mathcal{R}_T(\widehat{q}_T^{\mathrm{adj}}, \widehat{e}_T^{\mathrm{adj}}) - \mathcal{R}_T(q_T^\star, e_T^\star) \leq C\mathcal{E}_T^{\mathrm{adj}} + o_P(D_T).$$

The error bound for the adjusted pair supplies the convenient sufficient inequality $\mathcal{E}_T^{\mathrm{adj}} \le C'(\varepsilon_T^{\mathrm{pair}})^2$.

Step 3: Local lower bound for the raw pair. Positive curvature gives

$$\mathcal{R}_T(\widehat{q}_T, \widehat{e}_T) - \mathcal{R}_T(q_T^\star, e_T^\star) \ge 2D_T + o_P(D_T).$$

Step 4: Compare risks. Subtracting the Step 2 bound from the Step 3 bound and using $\mathcal{E}_T^{\mathrm{adj}} = o_P(D_T)$ yields

$$\mathcal{R}_T(\widehat{q}_T, \widehat{e}_T) - \mathcal{R}_T(\widehat{q}_T^{\mathrm{adj}}, \widehat{e}_T^{\mathrm{adj}}) \ge D_T + o_P(D_T) > 0$$

with probability approaching one whenever the raw risk gap is nondegenerate. □

Theorem 2 (restated). *Maintain Assumption A.1 together with the main-text stationary sampling assumption. Suppose there is a multiplier $\kappa^\dagger$ in the interior of $\mathcal{K}$ with $|\kappa^\dagger - 1| \ge c > 0$ such that, almost surely,*

$$(q_t^\star, e_t^\star) = \big(m_t^0 - \kappa^\dagger(m_t^0 - q_t^0), m_t^0 - \kappa^\dagger(m_t^0 - e_t^0)\big),$$

and that the baseline tail distances $m_t^0 - q_t^0$ and $m_t^0 - e_t^0$ are bounded and bounded away from zero. If the current fitted path converges to its pseudo-true path, then $\kappa_0 = \kappa^\dagger$, $\widehat{\kappa}_{T,n} \to_P \kappa^\dagger$, $\mathcal{E}_T^{\mathrm{adj}} = o_P(1)$, and D_T is bounded away from zero in probability, so the future-origin risk comparison holds with an improvement that exceeds a fixed positive constant with probability approaching one.

Proof. Step 1 (Target identification). Fix a state with positive tail distances. The common-multiplier relation gives $q_t^\star = m_t^0 - \kappa^\dagger(m_t^0 - q_t^0)$ with $\kappa^\dagger$ interior, so the state-specific unconstrained ratio equals $\kappa^\dagger$ and, by the state-specific minimizer lemma, $\kappa^\dagger$ minimizes the conditional check-loss criterion at every state. Taking expectations, $\kappa^\dagger$ minimizes the pooled population criterion, and the uniqueness required by the main-text sampling assumption gives $\kappa_0 = \kappa^\dagger$.

Step 2 (Estimation). The consistency proposition then gives $\widehat{\kappa}_{T,n} \to_P \kappa_0 = \kappa^\dagger$.

Step 3 (Adjusted pair error). Write

$$\widehat{q}_T^{\mathrm{adj}} - q_T^\star = (\widehat{m}_T - m_T^0) - \widehat{\kappa}_{T,n}\big\{(\widehat{m}_T - \widehat{q}_T) - (m_T^0 - q_T^0)\big\} - (\widehat{\kappa}_{T,n} - \kappa^\dagger)(m_T^0 - q_T^0).$$

The first two terms are $O_P(r_{T,n}^z)$ because $\widehat{\kappa}_{T,n} \in \mathcal{K}$ is bounded; the third is $o_P(1)$ by Step 2 and

bounded tail distances. The ES coordinate is identical with $m_T^0 - e_T^0$ in place of $m_T^0 - q_T^0$. Hence $\mathcal{E}_T^{\mathrm{adj}} = o_P(1)$.

Step 4 (Raw separation). Under the common-multiplier relation, $q_T^0 - q_T^\star = (\kappa^\dagger - 1)(m_T^0 - q_T^0)$ and $e_T^0 - e_T^\star = (\kappa^\dagger - 1)(m_T^0 - e_T^0)$, so the pseudo-true raw pair error satisfies $\|(q_T^0 - q_T^\star,\, e_T^0 - e_T^\star)\|^2 \geq 2c^2\underline{s}^2$ for the lower spread bound $\underline{s}$. With $r_{T,n}^z = o_P(1)$, the fitted raw pair inherits this separation, and $D_T \geq (\lambda_{\min}/4)c^2\underline{s}^2$ with probability approaching one.

Step 5 (Conclusion). Steps 3–4 give $\mathcal{E}_T^{\mathrm{adj}} = o_P(D_T)$; items (i) and (iii) control the common neighborhood and the exit terms; the risk-gap proposition then yields $\mathcal{R}_T(\widehat{q}_T, \widehat{e}_T) - \mathcal{R}_T(\widehat{q}_T^{\mathrm{adj}}, \widehat{e}_T^{\mathrm{adj}}) \geq D_T + o_P(D_T)$, which exceeds $(\lambda_{\min}/8)c^2\underline{s}^2$ with probability approaching one. □

Remark A.1 (Relation to the segment-scale condition). The lower-tail quantile-curve condition in the main text is sufficient for making $\mathcal{E}_T^{\mathrm{adj}}$ small, because integrating the quantile-path error controls ES. It is not implied by successful VaR coverage. DGP 2 has no state-dependent mean error. It studies the estimator under misspecification rather than the maintained condition of Theorem 2; the simulation that imposes that condition is in Section C.1.

Remark A.2 (Local, not global, dominance). The result is a rate-separation statement at a future origin. It does not say that the multiplier lowers FZ0 risk under arbitrary tail-shape misspecification or when the raw pair is already nearly correct.

Remark (A global decomposition of the FZ0 risk gap)**.** *For $e < 0$ let $A_T(q) := \mathbb{E}[(q - Y_{T+1})\mathbf{1}\{Y_{T+1} \leq q\} \mid \mathcal{I}_T]$ and $\varphi_T(q) := q - A_T(q)/\tau_0$, so that $\mathcal{R}_T(q,e) = \varphi_T(q)/e + \log(-e) - 1$. Then φ_T is concave with $\varphi_T'(q) = 1 - F_T^\star(q)/\tau_0$, is maximized at $q_T^\star$, and satisfies $\varphi_T(q_T^\star) = e_T^\star$, so $\varphi_T \leq e_T^\star < 0$ everywhere. For every score-domain-valid pair,*

$$\mathcal{R}_T(q,e) - \mathcal{R}_T(q_T^\star, e_T^\star) = d\left(\frac{\varphi_T(q)}{e}\right) + \log\frac{\varphi_T(q)}{e_T^\star}, \qquad d(x) := x - 1 - \log x \geq 0,$$

with both terms nonnegative: the first vanishes if and only if $e = \varphi_T(q)$, the second if and only if $\varphi_T(q) = e_T^\star$. The identity bounds the conditional risk gap globally, with no neighborhood or curvature condition; the local expansion in the proofs above is retained only for the quantitative constant in D_T. The structure is the mixture representation of consistent scoring functions (Fissler and Ziegel, 2016; Ehm et al., 2016). To check the identity, $A_T(q_T^\star) = \tau_0 q_T^\star - \tau_0 e_T^\star$ gives $\varphi_T(q_T^\star) = e_T^\star$; substituting

$\mathcal{R}_T$ and using $\log\{\varphi_T(q)/e\} + \log(-e) = \log\{-\varphi_T(q)\}$ gives the display; concavity follows from $A'_T = F^{\star}_T$. In the cells where the same multiplier is imposed for VaR and ES the identity reproduces every analytic risk-gap value by direct evaluation.

B Implementation and Numerical Procedures

This section describes the numerical procedures used in the Monte Carlo exercises and empirical applications. Gaussian and Student-t GARCH have closed-form conditional quantiles and tail integrals. The skew-t quantile-regression density is analytic, but its CDF, quantiles, and tail integral are evaluated numerically as described below.

B.1 Skew-t Quantile-Regression Density

Let Y_{t+1} denote the target variable and X_t the predictor vector. Following the quantile-grid construction of Adrian et al. (2019), the equity application first fits linear quantile regressions at $0.05, 0.25, 0.50, 0.75$, and 0.95 using an intercept and five lagged returns, then matches a skew-t distribution to those fitted quantiles. The conditional distribution has parameters $(\mu_t, \sigma_t, \alpha_t, \nu_t)$. We use the skew-$t$ density representation of Azzalini and Capitanio (2003)

$$g(y;\mu,\sigma,\alpha,\nu) = \frac{2}{\sigma}\, t\!\left(\frac{y-\mu}{\sigma};\nu\right) T\!\left(\alpha\frac{y-\mu}{\sigma}\sqrt{\frac{\nu+1}{\nu+((y-\mu)/\sigma)^2}};\nu+1\right), \tag{B.1}$$

where $t(\cdot;\nu)$ and $T(\cdot;\nu)$ denote the probability density function (PDF) and cumulative distribution function (CDF) of the standard Student-t distribution with ν degrees of freedom.

The five quantile regressions are fitted by the statsmodels quantile-regression routine with at most 1,000 iterations and tolerance 10^{-6}. Forecast quantiles are made weakly increasing by a cumulative maximum; ties are separated by a deterministic increment of order 10^{-10}. The skew-t match uses an equally weighted squared-quantile objective. For each shape pair

$$\alpha \in \{-8,-6,-5,-4,-3.5,-3,-2.5,-2,-1.5,-1,-0.5,0,0.5,1,1.5,2,3,4\},$$

$$\nu \in \{3.5,4,5,6,8,10,14,20,35,60\},$$

location and scale are obtained by least squares against the five standardized skew-t quantiles; candidates with nonfinite or nonpositive scale are discarded, and the shape pair with the smallest mean squared quantile error is selected. If no admissible match exists, the origin is retained with missing forecasts and a failure flag.

The standardized skew-t CDF is constructed on a deterministic 12,001-point grid over $[-60, 60]$ by cumulative trapezoidal integration and normalization. Quantiles are obtained by linear inversion of that grid. Expected shortfall is computed by adaptive one-dimensional quadrature of the standardized density from -50 to the target quantile, with a maximum of 200 subdivisions. These rules, rather than an analytic skew-t inverse CDF, define the reported rolling forecasts.

B.2 Data, Regressors, and Rolling Estimation

Daily equity returns are 100 times the log difference of adjusted closing prices from archived Yahoo Finance data for S&P 500 (`^GSPC`), NASDAQ Composite (`^IXIC`), and FTSE 100 (`^FTSE`). Dates are sorted within each exchange calendar, nonnumeric or missing returns are removed, and the three series are not aligned across markets. The tickers are price indices, so returns exclude dividends; adjusted and unadjusted closing prices coincide for these series. The replication file list records the archived files, row counts, and SHA-256 checksums. Replication should use these files and their recorded download metadata rather than download the series again.

Simulated samples and conditioning set. Each simulation uses a scalar state process $\{Z_t\}$ and the four-lag conditioning vector

$$X_t = (Z_t, Z_{t-1}, Z_{t-2}, Z_{t-3}) \in \mathbb{R}^4.$$

The experiments use two distinct protocols. In DGP 2, each replication simulates a sample of length n, fits each baseline once to that sample, reconstructs its fitted path on the same sample, and estimates the full-window multiplier there. The reported hit rates and FZ0 values are therefore in-sample fitted-path summaries, not independent future-origin scores. In the two misspecification experiments, each replication instead uses an estimation block of length $n_{\text{est}} = 1{,}600$ and a separate future continuation of length 1,000. The baseline parameters and multiplier are estimated only on the estimation block and are held fixed in the continuation; for GARCH, the conditional variance recursion is updated with realized continuation returns without re-estimating the parameters. All

random seeds are fixed at the replication level.

Empirical rolling estimation. The empirical application is different from both simulation protocols. At each forecast origin T, the baseline is refitted on the preceding $n = 1{,}000$ observations, the retrospective fitted path and multiplier are reconstructed on that same window, and the resulting pair is issued before Y_{T+1} is observed. Thus empirical baselines are retrained at every origin, whereas DGP 2 uses one fit per replication and the misspecification experiments use one estimation fit followed by a fixed-parameter future continuation.

Empirical applications. The equity samples and rolling-origin counts are documented below.

B.3 Sensitivity to the Estimation Window

The empirical analysis uses the same 1,000-day window to estimate the baseline and the multiplier. This section changes that common window to {500, 1,000, 2,000} days for the S&P GARCH-family cells at both tails while holding the model specification, retraining frequency, and evaluation rule fixed. Raw results therefore change across rows because the baseline is refitted under each window length.

Table B.1: Sensitivity to the estimation window: S&P 500 GARCH cells

			Rolling $\widehat{\kappa}$			Hit rate		Mean FZ0		
Baseline	τ_0	Window	mean	min	max	Raw	Anchored	Raw	Anchored	DM p
Gaussian GARCH	0.01	500	1.171	0.913	1.687	0.023	0.014	1.3961	1.2116	<0.001
		1000	1.157	0.946	1.459	0.023	0.013	1.3723	1.2083	<0.001
		2000	1.156	0.999	1.332	0.022	0.013	1.3600	1.2250	<0.001
Gaussian GARCH	0.05	500	1.076	0.893	1.264	0.062	0.053	0.8235	0.7966	<0.001
		1000	1.075	0.946	1.198	0.060	0.051	0.8183	0.7980	0.002
		2000	1.065	0.972	1.154	0.059	0.050	0.8136	0.7950	0.002
Student-t GARCH	0.01	500	1.063	0.837	1.382	0.017	0.014	1.2246	1.1974	0.083
		1000	1.048	0.910	1.254	0.016	0.014	1.2298	1.2008	0.009
		2000	1.040	0.947	1.141	0.015	0.013	1.2351	1.2182	0.116
Student-t GARCH	0.05	500	1.102	0.904	1.352	0.067	0.054	0.8138	0.7813	<0.001
		1000	1.104	0.951	1.246	0.064	0.052	0.8102	0.7868	0.004
		2000	1.091	0.993	1.192	0.063	0.051	0.8081	0.7882	0.006

Notes: The baseline estimation window and the multiplier window change together. All rows are tabulated on the common evaluation span on which each design is defined (origins from the 2,000-day burn-in onward, $N_{\text{eval}} = 5{,}800$). The 1,000-day row is recomputed on that common span from the stored per-origin output and reconciles with the primary rolling path to solver tolerance (2.63×10^{-8}). DM p compares Raw and Anchored FZ0 losses within each cell; family-wise inference applies to the primary 1,000-day design only.

The 5% results change little with the window: the mean multipliers differ by no more than 0.013, the adjusted hit rates remain between 0.050 and 0.054, and every DM comparison remains significant.

The main 5% findings are therefore not driven by the window choice. At 1%, the mean multiplier also changes little with the window, so the difference between the rolling and full-sample multipliers is not caused by a short estimation window. A longer window does not improve the results: the adjusted hit rates remain between 0.013 and 0.014, Gaussian GARCH has higher adjusted FZ0 loss at 2,000 days (1.2250 rather than 1.2083), and the Student-t GARCH comparison is significant at 1,000 days ($p = 0.009$) but not at 2,000 days ($p = 0.116$). We therefore retain the 1,000-day window.

B.4 VaR, ES, Backtests, and FZ0 Loss

All three main-text baselines (skew-t, Student-t GARCH, and Gaussian GARCH) supply the conditional quantiles and tail integrals needed for VaR, ES, and FZ0 evaluation. GARCH quantiles and tail integrals are available in closed form; skew-t quantiles are obtained by numerical CDF inversion and skew-t ES by one-dimensional numerical integration. Let $\widehat{f}_t$ denote the baseline conditional density at origin t and $\widehat{F}_t^{-1}$ its quantile function. Write τ_0 for the target tail level, $\bar{\tau}$ for the anchor, and y_{t+1} for the realization.

Tail functionals (VaR and ES). The baseline VaR is $\widehat{\mathrm{VaR}}_{\tau_0,t} = \widehat{F}_t^{-1}(\tau_0)$, evaluated in closed form for the GARCH baselines and by numerical inversion for skew-t. The baseline Expected Shortfall is

$$\widehat{\mathrm{ES}}_{\tau_0,t} = \frac{1}{\tau_0} \int_{-\infty}^{\widehat{\mathrm{VaR}}_{\tau_0,t}} y\, \widehat{f}_t(y)\, dy,$$

computed in closed form for Student-t and Gaussian location–scale forecasts and by adaptive Gauss–Kronrod quadrature for the skew-t.

Adjusted tail functionals. The adjusted VaR and ES follow the affine identities $q^{\mathrm{adj}}_{\tau_0,t}(\kappa) = \widehat{q}_{\bar{\tau},t} - \kappa(\widehat{q}_{\bar{\tau},t} - \widehat{q}_{\tau_0,t})$ and $e^{\mathrm{adj}}_{\tau_0,t}(\kappa) = \widehat{q}_{\bar{\tau},t} - \kappa(\widehat{q}_{\bar{\tau},t} - \widehat{e}_{\tau_0,t})$ (Lemma A.1), requiring only the baseline anchor quantile, target quantile, and ES as inputs. No sampling of the adjusted distribution is needed to compute $(q^{\mathrm{adj}}, e^{\mathrm{adj}})$.

Coverage tests and joint VaR–ES scores. The hit indicator $H_t = \mathbf{1}\{y_{t+1} \le \widehat{\mathrm{VaR}}_{\tau_0,t}\}$ feeds Kupiec's unconditional coverage and Christoffersen's independence likelihood-ratio tests (Kupiec, 1995; Christoffersen, 1998). For the daily S&P 500 return application and the theoretical results, the tail

score is the FZ0 loss of Patton et al. (2019):

$$s_{\tau_0}^{\mathrm{FZ0}}(q, e; y) = -\frac{1}{\tau_0\, e}\mathbf{1}\{y \le q\}(q - y) + \frac{q}{e} + \log(-e) - 1, \qquad e < 0.$$

The S&P 500 replication code implements the strict financial-return scoring restrictions. A forecast is scored by FZ0 only if the inputs are finite, the lower-tail pair is ordered ($e \le q$), and the ES forecast is negative ($e < 0$). Origins violating these restrictions are recorded as scoring failures rather than forced into the score. Averages and DM tests involving the natural affine MZ extension therefore use pairwise finite score differentials, and the corresponding scored fractions are reported below. Pairwise DM tests use a Bartlett Newey–West long-run variance. If N_{eval} denotes the number of scored out-of-sample loss differences, the bandwidth is $\lfloor 4(N_{\mathrm{eval}}/100)^{2/9} \rfloor$; there is no finite-sample rescaling, and the reference distribution is standard normal. The complete rolling procedures are compared as forecasting methods, following the perspective of Giacomini and White (2006).

B.5 Stability Check for the Multiplier

Reference distribution. The simulated p-values for the empirical stability statistics are referred to a simulated null: a GARCH$(1, 1)$ volatility process with skew-t innovations and a constant unit spread multiplier equal to one, matched to the empirical statistic on the sample size n alone. The null does not replicate the asset's own baseline parameters, ratio-series dependence, or spread paths, so the simulated p-values are reference values from a representative null simulation. They should not be interpreted as finite-sample p-values for the empirical application. Under the null the baseline is re-estimated in each replication, so the fitted ratio series also contains baseline estimation error.

The main article uses this check to assess whether the multiplier is constant over the historical estimation period. If κ changes over time, the partial sums of the check-loss score residuals should show systematic movement. On a historical estimation window of length n, let $\widehat{S}_t = \widehat{q}_{\bar{\tau},t} - \widehat{q}_{\tau_0,t} > 0$ and $\widehat{R}_t = (\widehat{q}_{\bar{\tau},t} - Y_{t+1})/\widehat{S}_t$ be computed from the same in-sample baseline quantile paths. The multiplier used in the check is

$$\widehat{\kappa}^{\mathrm{in}} \in \arg\min_{\kappa \in \mathcal{K}} \sum_{t=1}^{n} \widehat{S}_t\, \rho_{\tau_0}(\kappa - \widehat{R}_t), \tag{B.2}$$

with the same compact bounds $\mathcal{K}$ as the forecasting estimator. Define

$$\psi_t(\widehat{\kappa}^{\text{in}}) := \Big(\tau_0 - \mathbf{1}\{Y_{t+1} < q^{\text{adj}}_{\tau_0,t}(\widehat{\kappa}^{\text{in}})\}\Big)\widehat{S}_t, \qquad q^{\text{adj}}_{\tau_0,t}(\widehat{\kappa}^{\text{in}}) = \widehat{q}_{\bar{\tau},t} - \widehat{\kappa}^{\text{in}}\widehat{S}_t. \tag{B.3}$$

Because (B.2) is the same pooled check-loss objective whose score is evaluated in (B.3), the empirical subgradient condition centers the terminal score sum up to the usual finite-sample quantile-discretization convention; it does not force every intermediate partial sum to be zero. The check uses the retrospective path from a single full-span refit and estimates one multiplier over that path. The full-span fit is used only for this check; it is not the sequence of rolling fits or rolling multipliers used for one-step forecasts.

The single-parameter Nyblom statistic is

$$\text{Ny}_n = \frac{1}{n^2\widehat{\sigma}^2}\sum_{s=1}^{n}\left(\sum_{t=1}^{s}\psi_t(\widehat{\kappa}^{\text{in}})\right)^2. \tag{B.4}$$

The long-run variance is estimated with the Bartlett kernel,

$$\widehat{\sigma}^2 = \widehat{\gamma}_0 + 2\sum_{\ell=1}^{L}\left(1 - \frac{\ell}{L+1}\right)\widehat{\gamma}_\ell, \qquad L = \lfloor n^{1/4} \rfloor, \tag{B.5}$$

where $\widehat{\gamma}_\ell$ is the sample autocovariance of $\psi_t(\widehat{\kappa}^{\text{in}})$. The strict inequality in this hit indicator and the weak inequality in the estimation criterion differ only on the tie event $\{Y_{t+1} = \widehat{q}^{\text{adj}}_t\}$, which has probability zero under a continuous conditional distribution. In the classical parameter-stability framework of Nyblom (1989); Hansen (1992); Andrews (1993, 2003), a centered score partial sum has a Brownian-bridge limit and the corresponding 5% reference value is 0.461. Because the fitted path depends on estimated parameters, we treat the classical value only as a reference and place more weight on the simulated critical values described above. We use the statistic as a specification check: rejection provides evidence against a constant multiplier, while non-rejection does not provide evidence in favor of constancy.

Table B.2 reports the finite-sample size and power of the check.

B.6 Numerical Settings and Reproducibility

In the Monte Carlo designs the GARCH baselines are estimated with a constant conditional mean and GARCH$(1,1)$ conditional variance (`arch_model` with a constant-mean specification), matching the empirical implementation.

Table B.3 reports the baseline numerical settings used in the Monte Carlo experiments and empirical applications. These choices are held fixed across methods and across rolling windows unless explicitly stated otherwise, so that comparisons isolate modeling differences rather than differences in evaluation budgets.

Table B.3: Baseline numerical settings used in the experiments

Component	Baseline setting
Forecasting design	One-step ahead ($h = 1$); rolling window length n as specified per experiment
Baseline conditioning	Monte Carlo state variables are defined in Section C; empirical lag sets are documented in Section B.2
Anchor and tail levels	Anchor $\bar{\tau} = 0.5$; left-tail target $\tau_0 \in \{0.01, 0.05\}$ as specified by each experiment
Multiplier bounds	Compact set $\mathcal{K} = [0.25, 4.0]$; weighted-quantile solver is closed-form (sort + cumulative weights), with no iterative optimizer
VaR and ES	Closed-form VaR and ES for Student-t and Gaussian GARCH; numerical CDF inversion for skew-t VaR and adaptive Gauss–Kronrod quadrature of the skew-t density for ES
Adjusted VaR and ES	Affine identities $q^{\mathrm{adj}} = \widehat{q}_{\bar{\tau}} - \widehat{\kappa}_{T,n}(\widehat{q}_{\bar{\tau}} - \widehat{q}_{\tau_0})$ and the analogue for ES (Lemma A.1); no sampling of the adjusted distribution is needed
Joint VaR–ES score	FZ0, scored on common origins satisfying $e \le q$ and $e < 0$
VaR backtests	Kupiec unconditional-coverage and Christoffersen independence and conditional-coverage likelihood-ratio statistics, evaluated from the finite-sample hit sequence and compared with their standard asymptotic chi-square references
Stability check	κ-stability Nyblom statistic based on the in-sample plug-in score residual $\psi_t(\widehat{\kappa}^{\mathrm{in}})$ defined in Section B.5; Bartlett-kernel HAC with bandwidth $\lfloor n^{1/4} \rfloor$ and 5% critical value 0.461; used as a specification check rather than a decision rule

Replication files. The replication files archive the forecast-level rows used in the reported tables, including realizations, Raw and Anchored VaR–ES forecasts, score-domain flags, estimation inputs, configurations, and random seeds. These rows are sufficient to regenerate the reported scores, hit tests, checks, and tables. Some calculations were preserved at the forecast level rather than with a complete per-origin parameter and terminal-state record, so full raw-to-path reconstruction is not claimed for every baseline across software platforms. The replication materials identify the source,

checksum, and computing environment for each table-level evidence file, including the replication-level and summary outputs of the same-multiplier simulation.

B.7 Multiplier Uncertainty

The high-level expansion in Section A.10 follows standard two-step estimation logic (Newey, 1994). Its complete influence process contains both the second-stage weighted-quantile score and the common first-stage contribution. A HAC estimate from one realized fitted path conditions on the fitted first stage and need not recover the latter component. The object reported below is a pooled multiplier computed once from the retrospective path of the single full-span refit described in the table notes; it is different from the sequence of 1,000-day rolling multipliers. We therefore report its conditional-on-fit HAC standard error only as a descriptive uncertainty measure. No forecast-comparison conclusion in the article relies on this table.

Known-target simulations provide a limited finite-sample check for the conditional-on-fit approximation. In the GARCH-family cells, nominal 95% HAC intervals cover between 0.89 and 0.955 at the 1% tail and between 0.94 and 0.995 at the 5% tail. The skew-t QR HAC intervals undercover at the deeper tail (0.77–0.87). These results reinforce the descriptive interpretation of the table rather than establishing complete two-step coverage.

Comparing Tables B.1 and B.4 also explains the large 1% difference between the pooled and rolling multipliers. The full-span check uses the retrospective path of a single refit over the initial estimation period and the full evaluation span, whereas each rolling multiplier is re-estimated on a moving 1,000-day fitted path. Because the two objects use different fitted paths and objective spans, the pooled 1% value can lie outside the range of rolling estimates without contradiction. The full-span fit is not used for any forecast comparison. At 5%, the two summaries are much closer.

C Monte Carlo Designs and Full Results

This section gives the parameter values for the four Monte Carlo data-generating processes of the main text: DGP 1 (same multiplier for VaR and ES), DGP 2 (misspecified innovation distribution), DGP 3 (median misspecification), and DGP 4, which adds a state-dependent innovation distribution to the median error.

C.1 DGP 1: Same Multiplier for VaR and ES

Returns are $Y_{t+1} = \sigma_t\, h_{\kappa^\dagger}(\varepsilon_{t+1})$ with $m_t \equiv 0$, where $h_{\kappa^\dagger}(z) = \kappa^\dagger z$ for $z < 0$ and $h_{\kappa^\dagger}(z) = z$ otherwise. The innovation ε_{t+1} is iid standard normal or unit-variance Student-t_5. The baseline is given rather than estimated. It reports the median, VaR, and ES of the untransformed innovation distribution after multiplication by σ_t. Hence the baseline median equals the true median, and the known multiplier $\kappa^\dagger$ relates the baseline and true VaR, ES, and lower-tail quantiles. Volatility is either constant or follows the autoregressive state and volatility equation used in DGP 2 below, with the state process continuing into the future sample. The grid uses $\kappa^\dagger \in \{0.75, 1.25, 1.75\}$, $n \in \{250, 500, 1000, 2000\}$, $\tau_0 \in \{0.05, 0.01\}$, both innovation distributions, and both volatility cases. This gives 96 cells with $B = 500$ replications each. The multiplier is estimated once on the initial sample and held fixed over an independent continuation of length 1,000. Replication seeds are fixed by a common rule and stored with the replication-level results.

Four numerical checks compare the simulation with known reference quantities. The common-multiplier relation holds to machine precision on a dense quantile grid. Independent implementations reproduce the weighted-quantile and FZ0 calculations. The raw future hit rate matches $G\{G^{-1}(\tau_0)/\kappa^\dagger\}$ in all 96 cells. In the constant-volatility cells, the Monte Carlo FZ0 risk difference matches numerical integration. When the weights are constant, the multiplier is an order statistic. Its small finite-sample bias equals the corresponding Beta-integral calculation and declines at rate $O(1/n)$ along a fixed order-statistic sequence. At integer cutoffs, floating-point rounding can select an adjacent order statistic when the weights are equal; this issue does not arise for the continuously varying fitted spreads used in the empirical implementation. The mean out-of-sample loss difference is statistically significant in every group of cells. Table C.1 reports the full grid.

Table C.1: Monte Carlo results, DGP 1: out-of-sample performance when the same multiplier adjusts VaR and ES, full grid

$\kappa^\dagger$	n	τ_0	G	σ	Bias	RMSE	Hit raw	Hit raw (ref.)	Hit Anch.	Gap	Impr.	Risk gap (ref.)
0.75	250	0.05	N	St.	−0.000	0.059	0.014	0.014	0.051	0.1097	0.99	0.1213
0.75	250	0.05	N	Dyn.	0.000	0.061	0.014	0.014	0.052	0.1086	0.99	–
0.75	250	0.05	t_5	St.	−0.000	0.081	0.022	0.022	0.052	0.0783	0.95	0.0942
0.75	250	0.05	t_5	Dyn.	−0.003	0.080	0.021	0.022	0.052	0.0819	0.96	–
0.75	250	0.01	N	St.	0.003	0.074	0.001	0.001	0.011	0.1350	0.95	0.1680

$\kappa^\dagger$	n	τ_0	G	σ	Bias	RMSE	Hit raw	Hit raw (ref.)	Hit Anch.	Gap	Impr.	Risk gap (ref.)
0.75	250	0.01	N	Dyn.	−0.000	0.073	0.001	0.001	0.012	0.1346	0.95	–
0.75	250	0.01	t_5	St.	−0.007	0.123	0.003	0.003	0.013	0.0577	0.81	0.1150
0.75	250	0.01	t_5	Dyn.	0.001	0.132	0.003	0.003	0.012	0.0659	0.79	–
0.75	500	0.05	N	St.	−0.005	0.043	0.014	0.014	0.052	0.1133	1.00	0.1213
0.75	500	0.05	N	Dyn.	−0.001	0.044	0.014	0.014	0.051	0.1157	1.00	–
0.75	500	0.05	t_5	St.	0.008	0.058	0.022	0.022	0.050	0.0848	0.98	0.0942
0.75	500	0.05	t_5	Dyn.	0.001	0.056	0.022	0.022	0.051	0.0865	0.98	–
0.75	500	0.01	N	St.	0.012	0.058	0.001	0.001	0.010	0.1514	0.99	0.1680
0.75	500	0.01	N	Dyn.	−0.001	0.056	0.001	0.001	0.011	0.1499	0.98	–
0.75	500	0.01	t_5	St.	−0.013	0.089	0.003	0.003	0.012	0.0721	0.83	0.1150
0.75	500	0.01	t_5	Dyn.	0.000	0.094	0.003	0.003	0.011	0.0895	0.86	–
0.75	1000	0.05	N	St.	0.004	0.029	0.014	0.014	0.049	0.1217	1.00	0.1213
0.75	1000	0.05	N	Dyn.	0.000	0.030	0.014	0.014	0.050	0.1198	1.00	–
0.75	1000	0.05	t_5	St.	−0.004	0.039	0.021	0.022	0.051	0.0921	1.00	0.0942
0.75	1000	0.05	t_5	Dyn.	0.002	0.041	0.022	0.022	0.050	0.0905	0.99	–
0.75	1000	0.01	N	St.	0.004	0.038	0.001	0.001	0.010	0.1571	1.00	0.1680
0.75	1000	0.01	N	Dyn.	−0.003	0.036	0.001	0.001	0.011	0.1607	0.99	–
0.75	1000	0.01	t_5	St.	−0.007	0.063	0.003	0.003	0.011	0.0957	0.91	0.1150
0.75	1000	0.01	t_5	Dyn.	0.004	0.067	0.003	0.003	0.011	0.0931	0.89	–
0.75	2000	0.05	N	St.	−0.002	0.021	0.014	0.014	0.050	0.1210	1.00	0.1213
0.75	2000	0.05	N	Dyn.	−0.000	0.021	0.014	0.014	0.051	0.1184	1.00	–
0.75	2000	0.05	t_5	St.	0.002	0.029	0.022	0.022	0.050	0.0926	0.99	0.0942
0.75	2000	0.05	t_5	Dyn.	0.000	0.028	0.022	0.022	0.050	0.0923	1.00	–
0.75	2000	0.01	N	St.	0.001	0.027	0.001	0.001	0.010	0.1639	0.99	0.1680
0.75	2000	0.01	N	Dyn.	−0.000	0.027	0.001	0.001	0.010	0.1623	1.00	–
0.75	2000	0.01	t_5	St.	0.006	0.048	0.003	0.003	0.010	0.1086	0.93	0.1150
0.75	2000	0.01	t_5	Dyn.	0.001	0.047	0.003	0.003	0.010	0.1109	0.94	–
1.25	250	0.05	N	St.	−0.004	0.097	0.094	0.094	0.052	0.0956	0.99	0.1077
1.25	250	0.05	N	Dyn.	−0.004	0.097	0.094	0.094	0.052	0.0973	0.99	–
1.25	250	0.05	t_5	St.	−0.005	0.134	0.084	0.084	0.052	0.0644	0.94	0.0808
1.25	250	0.05	t_5	Dyn.	0.007	0.135	0.085	0.084	0.052	0.0691	0.94	–
1.25	250	0.01	N	St.	0.001	0.124	0.031	0.031	0.012	0.1902	0.96	0.2230
1.25	250	0.01	N	Dyn.	0.005	0.120	0.031	0.031	0.011	0.1880	0.97	–
1.25	250	0.01	t_5	St.	0.009	0.200	0.021	0.022	0.011	0.0656	0.80	0.1208
1.25	250	0.01	t_5	Dyn.	0.007	0.226	0.021	0.022	0.012	0.0684	0.79	–
1.25	500	0.05	N	St.	−0.007	0.072	0.094	0.094	0.052	0.1000	1.00	0.1077
1.25	500	0.05	N	Dyn.	−0.004	0.073	0.095	0.094	0.052	0.1047	1.00	–
1.25	500	0.05	t_5	St.	0.020	0.097	0.084	0.084	0.049	0.0733	0.95	0.0808
1.25	500	0.05	t_5	Dyn.	0.000	0.100	0.084	0.084	0.051	0.0715	0.97	–
1.25	500	0.01	N	St.	0.021	0.096	0.031	0.031	0.010	0.2068	0.99	0.2230
1.25	500	0.01	N	Dyn.	−0.003	0.091	0.032	0.031	0.011	0.2074	0.99	–

$\kappa^\dagger$	n	τ_0	G	σ	Bias	RMSE	Hit raw	Hit raw (ref.)	Hit Anch.	Gap	Impr.	Risk gap (ref.)
1.25	500	0.01	t_5	St.	−0.036	0.145	0.022	0.022	0.012	0.0883	0.86	0.1208
1.25	500	0.01	t_5	Dyn.	−0.010	0.155	0.022	0.022	0.011	0.0920	0.89	–
1.25	1000	0.05	N	St.	0.003	0.050	0.094	0.094	0.050	0.1070	1.00	0.1077
1.25	1000	0.05	N	Dyn.	−0.001	0.051	0.095	0.094	0.051	0.1051	1.00	–
1.25	1000	0.05	t_5	St.	−0.005	0.065	0.084	0.084	0.051	0.0770	0.98	0.0808
1.25	1000	0.05	t_5	Dyn.	0.005	0.067	0.084	0.084	0.050	0.0767	0.99	–
1.25	1000	0.01	N	St.	0.009	0.064	0.031	0.031	0.010	0.2063	0.99	0.2230
1.25	1000	0.01	N	Dyn.	0.003	0.065	0.031	0.031	0.010	0.2118	1.00	–
1.25	1000	0.01	t_5	St.	−0.016	0.108	0.022	0.022	0.011	0.1084	0.91	0.1208
1.25	1000	0.01	t_5	Dyn.	0.005	0.103	0.022	0.022	0.011	0.1108	0.90	–
1.25	2000	0.05	N	St.	−0.001	0.037	0.094	0.094	0.051	0.1059	1.00	0.1077
1.25	2000	0.05	N	Dyn.	0.002	0.036	0.094	0.094	0.050	0.1073	1.00	–
1.25	2000	0.05	t_5	St.	0.005	0.048	0.084	0.084	0.050	0.0786	0.97	0.0808
1.25	2000	0.05	t_5	Dyn.	0.002	0.048	0.084	0.084	0.050	0.0806	0.99	–
1.25	2000	0.01	N	St.	0.007	0.043	0.031	0.031	0.010	0.2216	0.99	0.2230
1.25	2000	0.01	N	Dyn.	−0.001	0.047	0.031	0.031	0.010	0.2182	0.99	–
1.25	2000	0.01	t_5	St.	0.011	0.080	0.022	0.022	0.010	0.1196	0.92	0.1208
1.25	2000	0.01	t_5	Dyn.	0.004	0.074	0.022	0.022	0.010	0.1187	0.92	–
1.75	250	0.05	N	St.	−0.000	0.139	0.174	0.174	0.052	0.8089	1.00	0.8208
1.75	250	0.05	N	Dyn.	−0.002	0.136	0.174	0.174	0.052	0.8106	1.00	–
1.75	250	0.05	t_5	St.	−0.007	0.195	0.151	0.151	0.052	0.5993	1.00	0.6247
1.75	250	0.05	t_5	Dyn.	0.008	0.197	0.150	0.151	0.051	0.5967	1.00	–
1.75	250	0.01	N	St.	−0.016	0.176	0.092	0.092	0.012	2.1086	1.00	2.1209
1.75	250	0.01	N	Dyn.	0.003	0.165	0.092	0.092	0.011	2.1008	1.00	–
1.75	250	0.01	t_5	St.	0.032	0.340	0.057	0.056	0.012	1.0663	1.00	1.1082
1.75	250	0.01	t_5	Dyn.	0.001	0.296	0.056	0.056	0.012	1.0621	1.00	–
1.75	500	0.05	N	St.	0.000	0.101	0.174	0.174	0.051	0.8118	1.00	0.8208
1.75	500	0.05	N	Dyn.	0.000	0.096	0.173	0.174	0.051	0.8141	1.00	–
1.75	500	0.05	t_5	St.	0.016	0.131	0.151	0.151	0.050	0.6219	1.00	0.6247
1.75	500	0.05	t_5	Dyn.	0.003	0.133	0.151	0.151	0.051	0.6180	1.00	–
1.75	500	0.01	N	St.	0.038	0.143	0.092	0.092	0.010	2.1054	1.00	2.1209
1.75	500	0.01	N	Dyn.	−0.002	0.128	0.092	0.092	0.011	2.1177	1.00	–
1.75	500	0.01	t_5	St.	−0.039	0.193	0.056	0.056	0.012	1.0760	1.00	1.1082
1.75	500	0.01	t_5	Dyn.	0.006	0.207	0.056	0.056	0.011	1.0926	1.00	–
1.75	1000	0.05	N	St.	0.004	0.070	0.173	0.174	0.050	0.8151	1.00	0.8208
1.75	1000	0.05	N	Dyn.	−0.006	0.070	0.175	0.174	0.051	0.8267	1.00	–
1.75	1000	0.05	t_5	St.	−0.012	0.095	0.152	0.151	0.052	0.6230	1.00	0.6247
1.75	1000	0.05	t_5	Dyn.	−0.002	0.090	0.151	0.151	0.050	0.6182	1.00	–
1.75	1000	0.01	N	St.	0.015	0.091	0.092	0.092	0.010	2.1119	1.00	2.1209
1.75	1000	0.01	N	Dyn.	0.007	0.099	0.091	0.092	0.010	2.0919	1.00	–
1.75	1000	0.01	t_5	St.	−0.017	0.152	0.056	0.056	0.011	1.1143	1.00	1.1082

$\kappa^\dagger$	n	τ_0	G	σ	Bias	RMSE	Hit raw	Hit raw (ref.)	Hit Anch.	Gap	Impr.	Risk gap (ref.)
1.75	1000	0.01	t_5	Dyn.	−0.006	0.149	0.057	0.056	0.011	1.1039	1.00	–
1.75	2000	0.05	N	St.	−0.003	0.049	0.173	0.174	0.050	0.8114	1.00	0.8208
1.75	2000	0.05	N	Dyn.	0.001	0.050	0.173	0.174	0.050	0.8141	1.00	–
1.75	2000	0.05	t_5	St.	0.005	0.066	0.153	0.151	0.051	0.6417	1.00	0.6247
1.75	2000	0.05	t_5	Dyn.	−0.001	0.066	0.151	0.151	0.050	0.6227	1.00	–
1.75	2000	0.01	N	St.	0.010	0.061	0.092	0.092	0.010	2.1278	1.00	2.1209
1.75	2000	0.01	N	Dyn.	−0.001	0.063	0.092	0.092	0.010	2.1187	1.00	–
1.75	2000	0.01	t_5	St.	0.013	0.107	0.057	0.056	0.010	1.1064	1.00	1.1082
1.75	2000	0.01	t_5	Dyn.	0.001	0.106	0.056	0.056	0.010	1.0949	1.00	–

Notes: Anch. = Anchored; n = estimation-block length; G = innovation law, with N the standard normal; St. = constant volatility; Dyn. = dynamic volatility; Gap = mean Raw-minus-Anchored future FZ0 loss; Impr. = improvement share; (ref.) = reference value computed from the known data-generating process. Full 96-cell factorial grid of the common-multiplier out-of-sample evaluation ($B = 500$ replications per cell, future length 1,000). Hit raw (ref.) is the closed-form raw hit rate $G\{G^{-1}(\tau_0)/\kappa^\dagger\}$; Risk gap (ref.) is the population FZ0 risk gap for static-volatility cells, obtained by numerical integration. With heavy-tailed innovations at the 1% tail, the average out-of-sample improvement is uniformly positive, while individual 1,000-day continuations can rank the two methods either way; the improvement-share column is descriptive. In the constant-weight cells the multiplier reduces to a single deep-tail order statistic. Its small finite-sample bias equals the analytical Beta-integral value for the order statistic selected at integer cutpoints, decays at the $O(1/n)$ rate along each fixed-selection subsequence, and is $o(n^{-1/2})$, so the consistency and root-n statements are unaffected.

C.2 DGP 2: Misspecified Innovation Distribution

Designs 2 to 4 share the same autoregressive state and volatility equation. We simulate a scalar state

$$Z_t = 0.85 Z_{t-1} + u_t, \qquad u_t \sim \mathcal{N}(0, 0.25^2),$$

and form the four-lag predictor vector

$$X_t = (Z_t, Z_{t-1}, Z_{t-2}, Z_{t-3})'.$$

The one-step-ahead outcome is

$$Y_{t+1} = \mu_t + \sigma_t \varepsilon_{t+1}, \qquad \mu_t = X_t'\beta, \qquad \sigma_t = \exp\{0.5(a_0 + X_t'a)\},$$

with

$$a_0 = -2.0, \qquad a = (0.10, 0, 0, 0)'.$$

DGP 2 sets $\beta = (0, 0, 0, 0)'$ (zero conditional mean). Both misspecification designs use $\beta = (0.30, 0, 0, 0)'$ to introduce state-dependent location variation. DGP 2 uses a standardized skew-t innovation,

$$\varepsilon_{t+1} \sim \text{SkewT}(\alpha = -3.0, \nu = 8.0),$$

where standardized means zero conditional mean and unit conditional variance. In this design, the skew-t quantile-regression baseline is correctly specified for the conditional innovation shape, while the GARCH-family baselines face a time-invariant error in the innovation distribution.

The standardized skew-$t(\alpha = -3, \nu = 8)$ innovation has median 0.1733. Because the GARCH baselines use symmetric innovation distributions, their fitted medians differ from the true median by an amount proportional to σ_t. For Gaussian GARCH, the difference is the full mean-to-median distance, $0.1733\,\sigma_t$. For Student-t GARCH, the pseudo-true location lies between the mean and the true median, so the difference is smaller. Relative to the baseline spread, this median error is approximately constant and is absorbed by the normalized median-error term a_t. With $\beta = 0$, both a_t and the spread ratio are constant up to volatility-tracking error. The state-specific minimizers are therefore approximately constant for the GARCH cells. The standardized median-to-tail spreads are 2.022 at the 5% tail and 3.344 at the 1% tail. Section C.3 reports the companion design with $\beta = (0.30, 0, 0, 0)'$, in which the conditional mean omitted by the constant-mean baselines makes the state-specific minimizer state-dependent.

C.3 DGP 3: Median Misspecification

DGP 3 introduces the state-dependent conditional mean, $\beta = (0.30, 0, 0, 0)'$, keeping every other element of DGP 2. The constant-mean GARCH baselines then omit a mean component that moves with the state but not in proportion to the baseline spread, so the normalized median error a_t is state-dependent and the state-specific minimizer varies across states in those cells even though the innovation shape error is unchanged. The design shows that the pooled criterion can center the spread-weighted hit condition even when the multiplier varies with the state. In this design the unweighted hit rate is also close to nominal, while state-conditional coverage remains tilted

across states; the quintile table is reported in the main-text misspecification section. The skew-t QR baseline includes the state directly and is much less affected by this omitted-location channel.

C.4 DGP 4: Combined Median and Innovation Misspecification

DGP 4 was introduced in the main-text Monte Carlo section. It preserves the shared state $Z_t = 0.85\, Z_{t-1} + u_t$ with $u_t \sim \mathcal{N}(0, 0.25^2)$, the lag structure $X_t = (Z_t, Z_{t-1}, Z_{t-2}, Z_{t-3})$, and the location–scale response

$$Y_{t+1} = \mu_t + \sigma_t\, \varepsilon_{t+1}, \qquad \mu_t = X_t'\beta, \quad \sigma_t = \exp(0.5(a_0 + X_t'a)),$$

with (β, a_0, a) as in DGP 3. The innovation distribution is regime-switching in the state:

$$\varepsilon_{t+1} \mid X_t \sim \begin{cases} \mathrm{SkewT}\big(\alpha_A(\gamma), \nu_A(\gamma)\big), \text{ standardized}, & Z_t \geq 0, \\ \mathrm{SkewT}\big(\alpha_B(\gamma), \nu_B(\gamma)\big), \text{ standardized}, & Z_t < 0, \end{cases}$$

with the regime-specific shape parameters

$$\alpha_A(\gamma) = -3.0 + 2.0\,\gamma, \quad \nu_A(\gamma) = 8.0 + 12.0\,\gamma, \qquad \alpha_B(\gamma) = -3.0 - 2.0\,\gamma, \quad \nu_B(\gamma) = 8.0 - 3.0\,\gamma,$$

and the gap parameter $\gamma \in [0, 1]$ interpolating between DGP 2 ($\gamma = 0$, both regimes collapse to the same standardized skew-t) and full regime switching ($\gamma = 1$, two clearly distinct skew-t shapes). Under this construction, the conditional normalized median-to-VaR distance

$$b_t(\tau_0, \bar{\tau}) = \frac{q^\star_{\bar{\tau},t} - q^\star_{\tau_0,t}}{\widehat{q}_{\bar{\tau},t} - \widehat{q}_{\tau_0,t}}$$

switches discretely across regimes for any baseline that fits a single within-window shape (such as skew-t QR or Student-t GARCH), so the common adjustment multiplier cannot concentrate b_t around a single constant κ_0. The common-multiplier location–scale condition fails by construction at $\gamma = 1$. The reported Monte Carlo tables in Section C.5 use $\gamma = 1$, the largest value considered in this family.

DGP 4 differs from DGP 3 only in the innovation distribution assigned to each observation; the state and volatility equations, including $\beta = (0.30, 0, 0, 0)'$, are unchanged. The simulation files

retain the regime labels and the regime-specific median-to-VaR ratios. Because the two regimes recur over time, this stationary state dependence need not produce a trend in the ordered score sequence. The time-stability statistic is therefore not designed to detect this alternative, even though the common-multiplier condition fails.

C.5 Out-of-Sample Results under Misspecification

Table C.2 reports the future-sample results used in the main-text misspecification table. The estimator is the same plug-in estimator used in DGP 2. DGP 3 isolates median misspecification. DGP 4 adds state-varying innovation shape to the same location-and-scale core, so it combines median and innovation misspecification rather than isolating either source of misspecification. Average coverage and loss can improve even when the common-multiplier condition fails at individual states.

D Detailed Empirical Tables and Specification Checks

This section reports the full equity results and the additional comparisons and sensitivity checks that support the empirical claims of the main text. Data-generating-process definitions and numerical settings for the simulations are in Section C; large grids based on alternative sample-splitting estimators are not repeated here.

D.1 Equity Index Results

This subsection collects the complete S&P 500 baseline comparison and the coverage, clustering, and dynamic-quantile panels referenced in Section 5 of the main text.

Table D.1 reports the complete S&P comparison on the samples and validity conventions stated in its notes.

Table D.2 reports the complete NASDAQ Composite comparison.

Table D.3 reports the complete FTSE 100 comparison.

Table D.4 reports the NASDAQ and FTSE 100 specification checks. In eight of the nine dual-level equity cells the adjusted 1% quantile stays below the adjusted 5% quantile at every origin; the lone

exception is a single FTSE 100 skew-t QR origin, where the two paths cross by less than 10^{-3} in return units.

The NASDAQ skew-t QR 1% cell shows the difference between the spread-weighted moment condition and ordinary coverage. Raw coverage is already close to nominal (hit rate 0.011, Kupiec $p = 0.233$), while Anchored raises the hit rate to 0.016 ($p < 0.001$); the full-span multiplier used in the stability check is 0.897, the stability check rejects (simulated $p = 0.010$), and mean FZ0 is not detectably different from Raw (DM $p = 0.500$). This is a case in which the multiplier is unhelpful for ordinary coverage. More generally, rejection of the stability check provides evidence against a constant multiplier; it does not predict the sign of the average Raw-minus-Anchored loss difference.

Table D.5 separates unconditional coverage, first-order hit dependence, Christoffersen conditional coverage, and the stronger DQ check.

Table D.6 reports the supporting ES specification check and its iid-inference limitation.

Table D.7 documents when the natural affine MZ extension leaves the joint score domain.

Table D.8 reports sensitivity to multiplicity adjustment and block length; it is not the primary pairwise comparison.

Table D.9 isolates direct 1% quantile estimation from the fitted skew-t tail extrapolation.

Table D.10 shows how the Raw-minus-Anchored loss differential varies across market episodes.

Table B.2: Finite-sample size and power of the multiplier stability check

Scenario	$n = 300$	$n = 600$	$n = 1200$
Constant multiplier = 1 (size)	0.028	0.038	0.026
Linear drift $0.8 \to 1.2$ (power)	0.076	0.104	0.194
Break $1.0 \to 1.5$ (power)	0.042	0.020	0.026
Break $1.0 \to 2.0$ (power)	0.008	0.062	0.538
Simulated 95% critical value	0.388	0.416	0.384
Classical 95% critical value		0.461	

Notes: Rejection rates use 500 replications and the classical 5% Nyblom critical value 0.461 (Nyblom, 1989). The simulated critical value is the null 95th percentile. The table is included in the supplement because it is a finite-sample check, not a central result. The null rejection rates are conservative in these simulations. The weak and nonmonotone rejection rates against moderate breaks are consistent with break-inflated long-run-variance estimates. The table shows why the main article treats rejection as evidence against constancy but does not treat non-rejection as evidence in favor of constancy, especially for moderate step breaks.

Table B.4: Retrospective full-span plug-in multiplier and conditional-on-fit HAC uncertainty

Index	Baseline	τ_0	$\widehat{\kappa}$	SE (HAC)
S&P 500	Gaussian GARCH	0.01	1.678	0.042
	Gaussian GARCH	0.05	1.047	0.020
	Student-t GARCH	0.01	1.727	0.044
	Student-t GARCH	0.05	1.067	0.021
NASDAQ	Gaussian GARCH	0.01	1.603	0.035
	Gaussian GARCH	0.05	1.084	0.019
	Student-t GARCH	0.01	1.636	0.034
	Student-t GARCH	0.05	1.097	0.019
FTSE 100	Gaussian GARCH	0.01	1.640	0.037
	Gaussian GARCH	0.05	1.047	0.019
	Student-t GARCH	0.01	1.676	0.037
	Student-t GARCH	0.05	1.069	0.020
S&P 500	skew-t QR	0.01	0.983	0.055
	skew-t QR	0.05	1.008	0.030
NASDAQ	skew-t QR	0.01	0.897	0.038
	skew-t QR	0.05	1.013	0.030
FTSE 100	skew-t QR	0.01	1.003	0.039
	skew-t QR	0.05	0.992	0.031

Notes: $\widehat{\kappa}$ is computed from the retrospective path of a single full-span refit and is used only for this check; it is not the rolling estimator used to issue forecasts. The HAC standard error conditions on that realized fitted path and may omit the common first-stage contribution. It is reported descriptively. The rolling multiplier distribution is summarized separately in the range and stability tables.

Table C.2: Monte Carlo results, DGPs 3 and 4: out-of-sample performance under misspecification at $n_{\text{est}} = 1{,}600$

DGP	Baseline	Raw hit	Anchored hit	FZ0 loss, Raw minus Anchored	DM reject
Panel A: 5% tail					
DGP 3 (median)	Gaussian GARCH	0.0616	0.0502	0.0258 (0.0010)	0.376
	Student-t GARCH	0.0729	0.0499	0.0471 (0.0017)	0.400
	skew-t QR	0.0511	0.0512	−0.0001 (0.0000)	0.164
DGP 4 (median + innovation)	Gaussian GARCH	0.0543	0.0507	0.0067 (0.0006)	0.284
	Student-t GARCH	0.0648	0.0506	0.0195 (0.0010)	0.206
	skew-t QR	0.0503	0.0504	−0.0003 (0.0002)	0.148
Panel B: 1% tail					
DGP 3 (median)	Gaussian GARCH	0.0253	0.0104	0.2402 (0.0062)	0.530
	Student-t GARCH	0.0209	0.0105	0.1055 (0.0041)	0.270
	skew-t QR	0.0114	0.0120	−0.0012 (0.0019)	0.150
DGP 4 (median + innovation)	Gaussian GARCH	0.0212	0.0104	0.1800 (0.0057)	0.380
	Student-t GARCH	0.0158	0.0102	0.0343 (0.0028)	0.144
	skew-t QR	0.0115	0.0115	−0.0015 (0.0018)	0.130

Notes: Future evaluation length 1,000 and $B = 500$ replications. Each cell reports the across-replication average of the replication-level future-mean loss difference; Monte Carlo standard errors, in parentheses, are the across-replication standard deviation of that mean divided by $\sqrt{B}$. Positive differences favor Anchored. "DM reject" is the fraction of replication-level paired DM tests that reject at 5%. These rows study misspecification; they do not impose the condition of Theorem 2.

Table D.1: Forecasting daily S&P 500 left-tail risk: complete baseline comparison

	FZ0 loss			**MZ valid**	**Hit**		**Kupiec** p		**DM** p	
Baseline	Raw	+MZ	Anchored	+MZ	Raw	Anchored	Raw	Anchored	Anch/Raw	Anch/MZ
Nominal tail $\tau_0 = 0.01$										
skew-t QR	2.015	1.842	2.036	0.531	0.013	0.015	0.011	< 0.001	0.615	0.142
Student-t GARCH	1.259	1.230	1.237	1.000	0.015	0.013	< 0.001	0.020	0.026	0.655
Gaussian GARCH	1.389	1.242	1.245	1.000	0.022	0.012	< 0.001	0.060	< 0.001	0.829
Nominal tail $\tau_0 = 0.05$										
skew-t QR	1.198	1.209	1.194	0.872	0.055	0.054	0.055	0.124	0.038	0.684
Student-t GARCH	0.849	0.840	0.828	1.000	0.065	0.054	< 0.001	0.187	0.003	0.012
Gaussian GARCH	0.858	0.847	0.840	1.000	0.060	0.053	< 0.001	0.295	0.002	0.144

Notes: Daily S&P 500 log returns are evaluated at $\tau_0 \in \{0.01, 0.05\}$. Lower FZ0 loss is better. FZ0 loss uses the strict financial-return scoring restrictions $e \leq q$ and $e < 0$. Origins that violate these restrictions are scored as missing. The MZ valid column reports the fraction of origins for which the natural affine MZ extension satisfies the scoring restrictions. The displayed +MZ mean is computed on its own score-valid subset; the Anchored-versus-MZ DM test uses their common valid dates. Raw and Anchored are the primary comparison; +MZ is a comparison based on an affine extension. Kupiec p-values evaluate unconditional coverage. DM p-values are unadjusted cell-level comparisons. Entries shown as < 0.001 are smaller than 0.001 after rounding.

Table D.2: Forecasting daily NASDAQ Composite left-tail risk: complete baseline comparison

	FZ0 loss			MZ valid	Hit		Kupiec p		DM p
Baseline	Raw	+MZ	Anchored	+MZ	Raw	Anchored	Raw	Anchored	Anch/Raw
Nominal tail $\tau_0 = 0.01$									
skew-t QR	3.475	1.963	3.495	0.645	0.011	0.016	0.233	< 0.001	0.500
Student-t GARCH	1.476	1.423	1.436	1.000	0.016	0.013	< 0.001	0.035	0.003
Gaussian GARCH	1.585	1.425	1.438	1.000	0.021	0.013	< 0.001	0.014	< 0.001
Nominal tail $\tau_0 = 0.05$									
skew-t QR	1.673	1.719	1.672	0.818	0.057	0.057	0.009	0.014	0.607
Student-t GARCH	1.099	1.075	1.073	1.000	0.069	0.054	< 0.001	0.099	< 0.001
Gaussian GARCH	1.102	1.080	1.078	1.000	0.066	0.055	< 0.001	0.079	< 0.001

Notes: Daily NASDAQ Composite log returns, 1995–2025, with the same rolling design as the S&P 500 application: 6,799 one-step evaluation origins, a 1,000-day rolling estimation window, refit at every origin, and the same three baselines. FZ0 loss uses the strict financial-return scoring restrictions $e \leq q$ and $e < 0$. The MZ valid column reports the fraction of origins for which the natural affine MZ extension satisfies the scoring restrictions. DM p-values are unadjusted pairwise comparisons. Entries shown as < 0.001 are smaller than 0.001 after rounding.

Table D.3: Forecasting daily FTSE 100 left-tail risk: complete baseline comparison

	FZ0 loss			MZ valid	Hit		Kupiec p		DM p
Baseline	Raw	+MZ	Anchored	+MZ	Raw	Anchored	Raw	Anchored	Anch/Raw
Nominal tail $\tau_0 = 0.01$									
skew-t QR	2.094	1.807	1.988	0.538	0.016	0.015	< 0.001	< 0.001	0.200
Student-t GARCH	1.196	1.179	1.168	1.000	0.015	0.011	< 0.001	0.569	0.074
Gaussian GARCH	1.281	1.186	1.176	1.000	0.020	0.011	< 0.001	0.492	0.001
Nominal tail $\tau_0 = 0.05$									
skew-t QR	1.173	1.169	1.173	0.914	0.052	0.051	0.485	0.632	0.816
Student-t GARCH	0.796	0.793	0.788	1.000	0.060	0.051	< 0.001	0.840	0.126
Gaussian GARCH	0.804	0.801	0.797	1.000	0.056	0.050	0.017	0.927	0.050

Notes: Daily FTSE 100 log returns, 1995–2025, with the same rolling design as the S&P 500 application: 6,827 one-step evaluation origins, a 1,000-day rolling estimation window, refit at every origin, and the same three baselines. Column definitions match the NASDAQ table. DM p-values are unadjusted pairwise comparisons. Entries shown as < 0.001 are smaller than 0.001 after rounding.

Table D.4: Median and multiplier-stability checks for the equity applications

		Median anchor		Nyblom, $\tau_0 = 0.01$		Nyblom, $\tau_0 = 0.05$	
Index	**Baseline**	Hit rate	p	Stat	Sim. p	Stat	Sim. p
S&P 500	skew-t QR	0.504	0.528	0.327	0.078	0.537	0.020
S&P 500	Student-t GARCH	0.509	0.152	0.212	0.194	0.149	0.312
S&P 500	Gaussian GARCH	0.502	0.771	0.157	0.296	0.145	0.316
NASDAQ	skew-t QR	0.496	0.552	0.645	0.010	2.396	0.000
NASDAQ	Student-t GARCH	0.501	0.894	0.681	0.010	0.045	0.838
NASDAQ	Gaussian GARCH	0.489	0.083	0.814	0.002	0.046	0.836
FTSE 100	skew-t QR	0.500	0.952	0.217	0.188	1.496	0.000
FTSE 100	Student-t GARCH	0.498	0.726	0.293	0.098	0.159	0.294
FTSE 100	Gaussian GARCH	0.493	0.240	0.341	0.072	0.158	0.294

Notes: The median-anchor hit rate is the fraction of evaluation origins with the realization at or below the recovered median anchor; the p-value is a two-sided normal-approximation binomial test of 0.5. The Nyblom statistics use the same in-sample plug-in multiplier-score construction at each tail level, with p-values simulated from the matched null design defined in Section B.5; a simulated p of 0.000 means no null replication exceeded the observed statistic.

Table D.5: S&P 500 conditional-coverage and dynamic-quantile checks

Baseline	Tail	Method	Hit	Indep. p	CC p	DQ p
Gaussian GARCH	5%	Raw	0.060	0.766	0.001	< 0.001
Gaussian GARCH	5%	Anchored	0.053	0.802	0.559	0.008
Gaussian GARCH	1%	Raw	0.022	0.390	< 0.001	< 0.001
Gaussian GARCH	1%	Anchored	0.012	0.111	0.048	< 0.001
Student-t GARCH	5%	Raw	0.065	0.782	< 0.001	< 0.001
Student-t GARCH	5%	Anchored	0.054	0.720	0.392	0.004
Student-t GARCH	1%	Raw	0.015	0.324	< 0.001	< 0.001
Student-t GARCH	1%	Anchored	0.013	0.141	0.022	< 0.001
skew-t QR	5%	Raw	0.055	< 0.001	< 0.001	< 0.001
skew-t QR	5%	Anchored	0.054	< 0.001	< 0.001	< 0.001
skew-t QR	1%	Raw	0.013	< 0.001	< 0.001	< 0.001
skew-t QR	1%	Anchored	0.015	< 0.001	< 0.001	< 0.001

Notes: The independence and CC columns are Christoffersen's independence and conditional-coverage tests. The DQ regression follows Engle and Manganelli (2004) and uses a constant, four lagged centered hit indicators, and the current VaR forecast; the reference distribution is chi-square with six degrees of freedom. DQ is a stronger dynamic check and is reported as a supplement check, not as the forecast-ranking criterion. The 1% rows are reported as a demanding tail case.

Table D.6: Exceedance-residual ES check for the daily S&P 500 application

Baseline	**Method**	**Sample**	n_{viol}	Mean z	t	p_t	p_{boot}
Nominal tail $\tau_0 = 0.01$							
skew-t QR	Raw	full	88	−0.306	−2.224	0.029	0.024
skew-t QR	Anchored	full	97	−0.297	−2.415	0.018	0.014
Student-t GARCH	Raw	full	105	−0.023	−0.836	0.405	0.398
Student-t GARCH	Anchored	full	88	−0.026	−0.929	0.356	0.353
Gaussian GARCH	Raw	full	151	−0.143	−5.615	< 0.001	< 0.001
Gaussian GARCH	Anchored	full	84	−0.130	−4.376	< 0.001	< 0.001
Nominal tail $\tau_0 = 0.05$							
skew-t QR	Raw	full	373	−0.120	−2.474	0.014	0.014
skew-t QR	Anchored	full	366	−0.119	−2.450	0.015	0.015
Student-t GARCH	Raw	full	441	−0.039	−2.286	0.023	0.025
Student-t GARCH	Anchored	full	364	−0.007	−0.380	0.704	0.700
Gaussian GARCH	Raw	full	409	−0.140	−7.524	< 0.001	< 0.001
Gaussian GARCH	Anchored	full	359	−0.112	−6.050	< 0.001	< 0.001

Notes: The test follows the exceedance-residual idea of McNeil and Frey (2000). On violation days, defined by $y_{t+1} \le q_t$, the standardized exceedance residual is $z_t = (y_{t+1} - e_t)/|e_t|$, where q_t and e_t are the method's VaR and ES forecasts; violations with $e_t \ge 0$ are dropped and counted (at most two per cell in this application). The Kupiec columns in the main-text results table use all violation days, so hit counts there can exceed n_{viol} by the number of dropped origins. Under adequate ES forecasts the residuals are centered at zero; regression-based ES backtests are developed by Bayer and Dimitriadis (2022). The t-statistic uses the iid standard error with $n_{\text{viol}} - 1$ degrees of freedom; p_{boot} is a two-sided centered iid bootstrap p-value with 9,999 resamples. Because violation residuals can be serially dependent, these columns are supporting specification checks rather than the primary forecast-ranking inference. Entries shown as < 0.001 are smaller than 0.001 after rounding.

Table D.7: FZ0 score-domain check for the affine MZ extension in the S&P 500 application

	$\tau_0 = 0.01$				$\tau_0 = 0.05$			
Baseline	$\Pr(e > q)$	$\Pr(e \geq 0)$	Joint invalid	Joint valid	$\Pr(e > q)$	$\Pr(e \geq 0)$	Joint invalid	Joint valid
skew-t QR	0.469	0.190	0.469	0.531	0.128	0.038	0.128	0.872
Student-t GARCH	0.000	0.000	0.000	1.000	0.000	0.000	0.000	1.000
Gaussian GARCH	0.000	0.000	0.000	1.000	0.000	0.000	0.000	1.000

Notes: At each forecast origin, the MZ intercept and slope are estimated by a τ_0-quantile regression of the 1,000 estimation-window realizations on the current-fit retrospective VaR path. The fitted affine map is then applied to the current VaR and ES coordinates. The strict S&P 500 FZ0 score is computed only when the lower-tail pair is ordered and the ES forecast is negative. The joint-invalid column is therefore the fraction of origins with either $e^{\mathrm{MZ}} > q^{\mathrm{MZ}}$ or $e^{\mathrm{MZ}} \geq 0$; because the baseline pair has $e < q$, the event $e^{\mathrm{MZ}} > q^{\mathrm{MZ}}$ is equivalent to a negative fitted MZ slope (a zero slope gives equality). In the skew-t QR cells the ordering failure is the dominant component. For the skew-t QR raw and anchored forecasts, the ES sign-restriction exclusion rate is only $3/6800 = 0.044\%$ at both tail levels. The anchored adjustment rule imposes $\kappa > 0$ and therefore preserves the VaR–ES ordering by construction.

Table D.8: Romano–Wolf multiplicity and block-length sensitivity: S&P 500 Raw versus Anchored

Tail	Baseline	Mean, Raw minus Anchored	RW 19	RW 100	RW 250	RW 500
5%	Gaussian GARCH	0.0172	0.0281	0.0767	0.0939	0.0837
5%	Student-t GARCH	0.0207	0.0344	0.0862	0.1053	0.0989
5%	skew-t QR	0.0040	0.1362	0.2517	0.2817	0.2823
1%	Gaussian GARCH	0.1447	0.0010	0.0080	0.0113	0.0088
1%	Student-t GARCH	0.0220	0.1202	0.2195	0.2434	0.2350
1%	skew-t QR	-0.0208	0.6672	0.6425	0.6377	0.6751

Notes: RW columns report shared-index Romano–Wolf stepdown p-values for the six S&P comparisons on their common 6,797-date calendar, under circular-block lengths 19, 100, 250, and 500 (Künsch, 1989; Politis and Romano, 1992). The pairwise DM tests in the main text use each baseline's complete Raw–Anchored score-valid sample; the RW columns instead use the common 6,797-date family calendar and shared bootstrap indices. The 19-day block is $\lceil 6{,}797^{1/3} \rceil$. The table is a sensitivity check for multiplicity and dependence, not the primary pairwise comparison.

Table D.9: Direct 1% quantile-regression comparison: S&P 500

Baseline	Method	N_{hit}	N_{score}	Mean FZ0	Hit rate	Mean $\widehat{\kappa}$
Direct QR01	Raw	6800	6800	1.7017	0.0138	–
Direct QR01	Anchored	6800	6800	1.7018	0.0138	1.000
Existing skew-t QR	Raw	6800	6797	2.0149	0.0132	–
Existing skew-t QR	Anchored	6800	6797	2.0356	0.0146	0.969

Descriptive loss differential	N_{score}	$\bar{d}$
Direct QR01 Raw minus Direct QR01 Anchored	6800	-0.0001
Direct QR01 Raw minus Existing skew-t QR Raw	6797	-0.3582
Direct QR01 Anchored minus Existing skew-t QR Anchored	6797	-0.3789

Notes: This sensitivity check fits rolling linear quantile regressions directly at $\tau = 0.01$ and at the median, using a 1000-day window. Because direct QR01 is VaR-only, ES is supplied by an empirical tail-gap plug-in, $\widehat{e}_t = \widehat{q}_{0.01,t} + \overline{(y_s - \widehat{q}_{0.01,s}) \mid y_s \le \widehat{q}_{0.01,s}}$, computed on the training window. Anchored uses the direct median QR as the anchor and the standard bounded weighted-quantile multiplier. The check therefore does not constitute a fully specified density forecast. Loss differentials are left minus right; negative values favor the left-hand method. Direct QR01 has substantially lower FZ0 loss than the skew-t QR 1% path, showing that the latter is weakened in part by tail extrapolation. Adjustment of direct QR01 is essentially neutral, with the mean anchored multiplier equal to one. This is partly mechanical when the median and target quantile regressions use the same sample and regressors: at $\kappa = 1$,

$$\sum_t S_t\{\tau_0 - \mathbf{1}(Y_{t+1} < q_t)\} = (\widehat{\beta}_{0.5} - \widehat{\beta}_{\tau_0})' \sum_t X_t\{\tau_0 - \mathbf{1}(Y_{t+1} < X_t'\widehat{\beta}_{\tau_0})\} = 0$$

up to ties and subgradient selection. The near-fixed point should therefore not be interpreted as independent evidence that no residual spread bias remains. Direct QR01 still has a hit rate above the nominal 1% level; the exercise is used to isolate tail extrapolation from adjustment and is not presented as a fully specified 1% density forecast. N_{hit} uses every available VaR forecast, whereas N_{score} uses only origins satisfying the FZ0 domain. The existing skew-t QR rows therefore use 6,800 origins for hit-rate calculations and 6,797 origins for FZ0 after three positive-ES exclusions. The method-level means in the first panel use each method's own score-valid dates, whereas the second panel recomputes differences on common dates. To displayed precision, the common-date Direct QR01 Raw mean is $2.0149 - 0.3582 = 1.6567$, not the all-date mean 1.7017; the three additional Direct QR01 dates carry unusually large FZ0 losses. The common-date differential is therefore the appropriate descriptive comparison. The primary HAC-based Raw-versus-Anchored DM result for the existing skew-t path is reported in the main table and is not duplicated here.

Table D.10: S&P 500 subperiod stability of Raw-versus-Anchored loss differences

Baseline	τ_0	Pre-GFC	GFC	Post-GFC	COVID
Gaussian GARCH	0.01	0.0348 (0.190)	0.3181 (0.002)	0.1453 (0.005)	0.2474 (0.008)
Gaussian GARCH	0.05	0.0057 (0.016)	0.0383 (0.046)	0.0118 (0.254)	0.0378 (0.010)
Student-t GARCH	0.01	0.0091 (0.441)	0.0399 (0.127)	0.0110 (0.463)	0.0549 (0.094)
Student-t GARCH	0.05	0.0098 (0.004)	0.0523 (0.062)	0.0127 (0.306)	0.0420 (0.029)
Skew-t QR	0.01	-0.0314 (0.223)	-0.5824 (0.048)	0.0078 (0.470)	0.1268 (0.004)
Skew-t QR	0.05	0.0030 (¡0.001)	0.0307 (¡0.001)	-0.0042 (¡0.001)	0.0117 (0.010)

Notes: Each entry reports the mean Raw-minus-Anchored FZ0 loss differential, computed under the strict financial-return scoring restrictions, with the unadjusted Diebold–Mariano p-value in parentheses. Positive values mean Anchored has lower mean loss. The S&P sample runs from 1998-12-18 through 2025-12-31. The subperiods are pre-GFC (1998-12-18–2007-07-31), GFC (2007-08-01–2009-06-30), post-GFC/pre-COVID (2009-07-01–2020-02-19), and COVID/post-COVID (2020-02-20–2025-12-31). The table is computed from the existing row-level observations that are valid for joint scoring without re-estimating the baselines. GARCH-family improvements are largest in crisis and post-COVID periods but are not generated by a single isolated subperiod. The skew-t QR rows are more mixed: the 1% row reverses during the GFC subperiod, while the 5% row has a small post-GFC/pre-COVID reversal. These reversals are consistent with the stability-check rejections and call for caution in those cells.

D.2 Filtered Historical Simulation

Filtered historical simulation (FHS) is the relevant richer comparison when the current GARCH fit, conditional-scale path, and standardized residuals are available. At each origin, the same GARCH class is refit on the same $n = 1{,}000$ observations used by the Raw forecast. Let $\widehat{\varepsilon}_{s|T} = (Y_{s+1} - \widehat{\mu}_T)/\widehat{\sigma}_{s|T}$ denote the resulting in-window standardized residuals, ordered as $\widehat{\varepsilon}_{(1)|T} \leq \cdots \leq \widehat{\varepsilon}_{(n)|T}$, and set $k = \lceil n\tau_0 \rceil$. FHS uses the inverted-CDF empirical quantile $\widehat{z}^{\mathrm{FHS}}_{\tau_0,T} = \widehat{\varepsilon}_{(k)|T}$ and averages all residuals not exceeding that quantile. If the boundary order statistic is tied, the implementation includes every residual at the boundary, so the finite-sample tail set can contain more than k observations; this is the stated tie rule used in the reported forecasts. With $\mathcal{T}_T := \{s : \widehat{\varepsilon}_{s|T} \leq \widehat{z}^{\mathrm{FHS}}_{\tau_0,T}\}$, let $\widehat{w}^{\mathrm{FHS}}_{\tau_0,T} = |\mathcal{T}_T|^{-1} \sum_{s \in \mathcal{T}_T} \widehat{\varepsilon}_{s|T}$. The one-step forecasts are

$$\widehat{q}^{\mathrm{FHS}}_{\tau_0,T} = \widehat{\mu}_T + \widehat{\sigma}_{T+1|T}\widehat{z}^{\mathrm{FHS}}_{\tau_0,T}, \qquad \widehat{e}^{\mathrm{FHS}}_{\tau_0,T} = \widehat{\mu}_T + \widehat{\sigma}_{T+1|T}\widehat{w}^{\mathrm{FHS}}_{\tau_0,T}.$$

It therefore re-estimates the residual-tail shape rather than applying one multiplier to the reported VaR–ES pair. The comparison shows the information trade-off: FHS is more flexible, while Anchored adjustment can be applied when a fitted historical tail path is available but standardized residuals and the conditional-scale path are not. The skew-t QR baseline is not included because the stored forecast archive does not contain the standardized-residual and scale paths required to construct FHS on the same rolling origins.

Table D.11: Filtered historical simulation comparison

Index	Baseline	Tail	Raw FZ0	Anchored FZ0	FHS FZ0	FHS hit	Kupiec p	ES p
S&P 500	Student-t GARCH	1%	1.2593	1.2374	1.2386	0.011	0.471	0.212
	Student-t GARCH	5%	0.8493	0.8283	0.8286	0.051	0.657	0.777
	Gaussian GARCH	1%	1.3892	1.2447	1.2364	0.011	0.283	0.131
	Gaussian GARCH	5%	0.8578	0.8404	0.8326	0.051	0.657	0.450
NASDAQ	Student-t GARCH	1%	1.4764	1.4363	1.4395	0.011	0.401	0.050
	Student-t GARCH	5%	1.0988	1.0729	1.0723	0.052	0.375	0.550
	Gaussian GARCH	1%	1.5846	1.4377	1.4339	0.011	0.282	0.025
	Gaussian GARCH	5%	1.1025	1.0778	1.0724	0.053	0.225	0.485
FTSE 100	Student-t GARCH	1%	1.1957	1.1677	1.1657	0.010	0.834	0.039
	Student-t GARCH	5%	0.7959	0.7883	0.7893	0.048	0.422	0.194
	Gaussian GARCH	1%	1.2807	1.1760	1.1739	0.011	0.492	0.060
	Gaussian GARCH	5%	0.8041	0.7966	0.7936	0.048	0.564	0.199

Notes: All three methods use common dates within each cell. Hit and Kupiec refer to the FHS VaR path. ES p is the exceedance-residual bootstrap check applied to the FHS pair. FHS can alter the residual-tail shape because it uses the standardized residual and conditional-scale paths; Anchored adjustment preserves the baseline's reported distances among the median, VaR, and ES. The table does not claim equivalence or dominance.

Table D.11 compares Anchored with the residual-based FHS forecasts from the same GARCH classes.

D.3 Bandwidth and Fixed-b Sensitivity of the Primary Comparisons

The Anchored-versus-Raw Diebold–Mariano comparisons in the main text use the baseline Bartlett bandwidth $\lfloor 4(n/100)^{2/9} \rfloor$. Table D.12 recomputes the six primary S&P comparisons with Bartlett bandwidths of 100 and 250 days and evaluates the 250-day statistic against fixed-b critical values (Kiefer and Vogelsang, 2005). Both GARCH conclusions are unchanged at both tails; the skew-t 5% comparison loses significance at long bandwidths, consistent with the block multiplicity adjustments.

Table D.12: Long-bandwidth and fixed-b DM checks, Anchored versus Raw

	Baseline	$M = 10$	$M = 100$	$M = 250$	Fixed-b reject (5%)
$\tau_0 = 0.05$	skew-t QR	0.038	0.212	0.244	no
	Student-t GARCH	0.003	0.001	<0.001	yes
	Gaussian GARCH	0.002	0.001	0.001	yes
$\tau_0 = 0.01$	skew-t QR	0.615	0.656	0.651	no
	Student-t GARCH	0.026	0.016	0.012	yes
	Gaussian GARCH	<0.001	<0.001	<0.001	yes

Notes: Entries are two-sided DM p-values with Bartlett bandwidth M; the last column applies fixed-b critical values at $M = 250$.

D.4 Adjustment Comparisons

This section defines the alternatives reported in the main text. The equal-weight estimator is included as a sensitivity comparison, not as a second estimator proposed by the paper. It removes the spread weights from the main check-loss criterion:

$$\widehat{\kappa}^{\mathrm{EW}}_{T,n} \in \arg\min_{\kappa \in \mathcal{K}} \frac{1}{n} \sum_{s \in \mathcal{W}_{T,n}} \rho_{\tau_0}(\kappa - R_{s|T,n}).$$

It is the projected ordinary empirical $(1 - \tau_0)$-quantile of the same ratios used by Anchored. Since

$$Y_{s+1} \le \widehat{m}_{s|T,n} - \kappa \widehat{S}_{s|T,n} \quad \Longleftrightarrow \quad R_{s|T,n} \ge \kappa,$$

it targets the ordinary in-window hit rate, up to empirical-quantile discreteness, ties, and projection onto $\mathcal{K}$. Anchored instead minimizes check loss in return units, which produces the spread weights $\widehat{S}_{s|T,n}$.

The FZ0-targeted estimator minimizes the in-window average FZ0 loss over

$$\mathcal{K}_{T,n}^{\mathrm{FZ}} = \left[\max\left\{0.25,\ \sup_{s\in\mathcal{W}_{T,n}} \frac{\widehat{m}_{s|T,n}}{\widehat{m}_{s|T,n} - \widehat{e}_{s|T,n}} + \varepsilon_{\mathrm{dom}}\right\}, 4\right], \qquad \varepsilon_{\mathrm{dom}} = 10^{-8}.$$

The small margin keeps every fitted ES strictly negative. The minimizer is computed on a deterministic grid with spacing 0.001 and then refined by golden-section search.

The nonnegative-slope MZ comparison solves $\min_{\alpha_{\mathrm{MZ}}\in\mathbb{R},\,\beta_{\mathrm{MZ}}\geq 0} \sum_s \rho_{\tau_0}(Y_{s+1} - \alpha_{\mathrm{MZ}} - \beta_{\mathrm{MZ}}\widehat{q}_s)$ in each window and applies the same affine transformation to ES. The intercept is unrestricted; when the unrestricted slope is negative, the constrained solution sets $\beta_{\mathrm{MZ}} = 0$ and the intercept is an in-window empirical τ_0-quantile. The slope constraint binds in 4,078 of 40,800 windows overall and in 47% of the skew-t 1% windows.

The historical-vintage version applies the spread-weighted quantile to the forecast vintage generated at each past origin, using one historical vintage per row, for origins 1,001–6,800. Raw and Anchored are recomputed on the same origins. All pairwise DM tests use the common dates on which both forecasts satisfy the FZ0 domain. The zero-anchored proportional estimator replaces $(\widehat{m}, \widehat{S})$ by $(0, -\widehat{q})$ in the check-loss problem. It uses a weighted-quantile formula when all weights are positive and direct one-dimensional convex minimization otherwise. Finally, the 0.25-anchor sensitivity check gives Gaussian GARCH 5% loss of 0.839 rather than 0.840 under the median anchor (DM $p = 0.0005$), a difference of about 0.001; Student-t GARCH is indistinguishable ($p = 0.931$), and the multiplier paths remain inside the parameter bounds.

Table D.13: Adjustment comparisons: full results

Baseline	Estimator	N_{score}	FZ0	Hit	DM p vs Raw	DM p vs Anch.	$\bar{\kappa}$
$\tau_0 = 0.05$							
skew-t QR	Raw	6797	1.198	0.055	–	0.038	–
	Anchored	6797	1.194	0.054	0.038	–	1.007
	κ^{EW}	6797	1.196	0.055	0.125	0.011	1.006
	κ^{FZ}	6798	1.252	0.055	0.390	0.315	0.990
	MZ ($\beta_{\text{MZ}} \geq 0$)	6797	1.202	0.055	0.784	0.510	–
	Proportional	6797	1.194	0.054	0.059	0.643	1.008
	Historical vintage	5797	1.279	0.056	0.050	0.033	1.003
Student-t GARCH	Raw	6800	0.849	0.065	–	0.003	–
	Anchored	6800	0.828	0.054	0.003	–	1.092
	κ^{EW}	6800	0.830	0.052	0.008	0.159	1.100
	κ^{FZ}	6800	0.828	0.051	0.004	0.602	1.105
	MZ ($\beta_{\text{MZ}} \geq 0$)	6800	0.839	0.051	0.249	0.012	–
	Proportional	6800	0.829	0.054	0.003	0.234	1.097
	Historical vintage	5800	0.800	0.052	0.209	<0.001	1.095
Gaussian GARCH	Raw	6800	0.858	0.060	–	0.002	–
	Anchored	6800	0.840	0.053	0.002	–	1.065
	κ^{EW}	6800	0.841	0.052	0.003	0.428	1.068
	κ^{FZ}	6800	0.834	0.047	0.002	0.056	1.104
	MZ ($\beta_{\text{MZ}} \geq 0$)	6800	0.847	0.052	0.144	0.147	–
	Proportional	6800	0.841	0.053	0.002	0.095	1.067
	Historical vintage	5800	0.810	0.052	0.226	0.010	1.066
$\tau_0 = 0.01$							
skew-t QR	Raw	6797	2.015	0.013	–	0.615	–
	Anchored	6797	2.036	0.015	0.615	–	0.969
	κ^{EW}	6797	2.000	0.015	0.735	0.109	0.973
	κ^{FZ}	6798	2.112	0.015	0.534	0.280	0.995
	MZ ($\beta_{\text{MZ}} \geq 0$)	6800	1.803	0.014	0.151	0.107	–
	Proportional	6797	2.015	0.015	0.990	0.063	0.967
	Historical vintage	5797	2.173	0.018	0.607	0.734	1.023
Student-t GARCH	Raw	6800	1.259	0.015	–	0.026	–
	Anchored	6800	1.237	0.013	0.026	–	1.045
	κ^{EW}	6800	1.234	0.012	0.057	0.559	1.070
	κ^{FZ}	6800	1.237	0.011	0.126	0.946	1.090
	MZ ($\beta_{\text{MZ}} \geq 0$)	6800	1.230	0.012	0.178	0.657	–
	Proportional	6800	1.239	0.013	0.030	0.332	1.046
	Historical vintage	5800	1.220	0.013	0.521	0.150	1.058
Gaussian GARCH	Raw	6800	1.389	0.022	–	<0.001	–
	Anchored	6800	1.245	0.012	<0.001	–	1.152
	κ^{EW}	6800	1.243	0.012	<0.001	0.705	1.170
	κ^{FZ}	6800	1.241	0.011	<0.001	0.649	1.211
	MZ ($\beta_{\text{MZ}} \geq 0$)	6800	1.242	0.012	<0.001	0.848	–
	Proportional	6800	1.245	0.013	<0.001	0.358	1.157
	Historical vintage	5800	1.217	0.013	<0.001	0.436	1.178

Notes: All rows use the archived forecast-level data described above. Historical vintage rows use origins 1,001–6,800 with Raw and Anchored re-summarized on the same origins. Proportional is estimated by check loss, with the convex problem solved directly where the positive-weight closed form is unavailable. $\bar{\kappa}$ is the mean multiplier path where a single multiplier exists.

References

Adrian, T., Boyarchenko, N., and Giannone, D. (2019). Vulnerable growth. *American Economic Review*, 109(4):1263–1289.

Andrews, D. W. K. (1993). Tests for parameter instability and structural change with unknown change point. *Econometrica*, 61(4):821–856.

Andrews, D. W. K. (2003). Tests for parameter instability and structural change with unknown change point: A corrigendum. *Econometrica*, 71(1):395–397.

Azzalini, A. and Capitanio, A. (2003). Distributions generated by perturbation of symmetry with emphasis on a multivariate skew t-distribution. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 65(2):367–389.

Barone-Adesi, G., Giannopoulos, K., and Vosper, L. (1999). VaR without correlations for portfolios of derivative securities. *Journal of Futures Markets*, 19(5):583–602.

Basel Committee on Banking Supervision (1996). Supervisory framework for the use of "back-testing" in conjunction with the internal models approach to market risk capital requirements. Technical report, Bank for International Settlements, Basel.

Basel Committee on Banking Supervision (2019). Minimum capital requirements for market risk. Technical report, Bank for International Settlements, Basel.

Bayer, S. and Dimitriadis, T. (2022). Regression-based expected shortfall backtesting. *Journal of Financial Econometrics*, 20(3):437–471.

Berkowitz, J. (2001). Testing density forecasts, with applications to risk management. *Journal of Business & Economic Statistics*, 19(4):465–474.

Boucher, C. M., Daníelsson, J., Kouontchou, P. S., and Maillet, B. B. (2014). Risk models-at-risk. *Journal of Banking & Finance*, 44:72–92.

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

Diebold, F. X., Gunther, T. A., and Tay, A. S. (1998). Evaluating density forecasts with applications to financial risk management. *International Economic Review*, 39(4):863–883.

Diebold, F. X. and Mariano, R. S. (1995). Comparing predictive accuracy. *Journal of Business & Economic Statistics*, 13(3):253–263.

Dimitriadis, T. and Bayer, S. (2019). A joint quantile and expected shortfall regression framework. *Electronic Journal of Statistics*, 13(1):1823–1871.

D'Innocenzo, E., Lucas, A., Schwaab, B., and Zhang, X. (2026). Joint extreme value-at-risk and expected shortfall dynamics with a single integrated tail shape parameter. *Journal of Business & Economic Statistics*. Advance online publication.

Du, Z. and Escanciano, J. C. (2017). Backtesting expected shortfall: Accounting for tail risk. *Management Science*, 63(4):940–958.

Ehm, W., Gneiting, T., Jordan, A., and Krüger, F. (2016). Of quantiles and expectiles: Consistent scoring functions, Choquet representations and forecast rankings. *Journal of the Royal Statistical Society: Series B*, 78(3):505–562.

Engle, R. F. and Manganelli, S. (2004). CAViaR: Conditional autoregressive value at risk by regression quantiles. *Journal of Business & Economic Statistics*, 22(4):367–381.

Escanciano, J. C. and Olmo, J. (2010). Backtesting parametric value-at-risk with estimation risk. *Journal of Business & Economic Statistics*, 28(1):36–51.

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

Fosten, J., Gutknecht, D., and Pohle, M.-O. (2024). Testing quantile forecast optimality. *Journal of Business & Economic Statistics*, 42(4):1367–1378.

Gaglianone, W. P., Lima, L. R., Linton, O., and Smith, D. R. (2011). Evaluating value-at-risk models via quantile regression. *Journal of Business & Economic Statistics*, 29(1):150–160.

Gatta, F., Lillo, F., and Mazzarisi, P. (2024). CAESar: Conditional autoregressive expected shortfall. Technical Report arXiv:2407.06619, arXiv.

Gerlach, R., Naimoli, A., and Storti, G. (2025). Using quantile time series and historical simulation to forecast financial risk multiple steps ahead. Technical Report arXiv:2502.20978, arXiv.

Giacomini, R. and Komunjer, I. (2005). Evaluation and combination of conditional quantile forecasts. *Journal of Business & Economic Statistics*, 23(4):416–431.

Giacomini, R. and White, H. (2006). Tests of conditional predictive ability. *Econometrica*, 74(6):1545–1578.

Gneiting, T. (2011). Quantiles as optimal point forecasts. *International Journal of Forecasting*, 27(2):197–207.

Hansen, B. E. (1992). Testing for parameter instability in linear models. *Journal of Policy Modeling*, 14(4):517–533.

Kiefer, N. M. and Vogelsang, T. J. (2005). A new asymptotic theory for heteroskedasticity-autocorrelation robust tests. *Econometric Theory*, 21(6):1130–1164.

Koenker, R. and Bassett, G. (1978). Regression quantiles. *Econometrica*, 46(1):33–50.

Kratz, M., Lok, Y. H., and McNeil, A. J. (2018). Multinomial VaR backtests: A simple implicit approach to backtesting expected shortfall. *Journal of Banking & Finance*, 88:393–407.

Künsch, H. R. (1989). The jackknife and the bootstrap for general stationary observations. *The Annals of Statistics*, 17(3):1217–1241.

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

Liu, X. and Luger, R. (2026). Quantile-based modeling of scale dynamics in financial returns for value-at-risk and expected shortfall forecasting. *International Journal of Forecasting*, 42(3):872–888.

McNeil, A. J. and 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.

Mincer, J. A. and Zarnowitz, V. (1969). The evaluation of economic forecasts. In *Economic forecasts and expectations: Analysis of forecasting behavior and performance*, pages 3–46. NBER.

Newey, W. K. (1994). The asymptotic variance of semiparametric estimators. *Econometrica*, 62(6):1349–1382.

Newey, W. K. and West, K. D. (1987). A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica*, 55(3):703–708.

Nolde, N. and Ziegel, J. F. (2017). Elicitability and backtesting: Perspectives for banking regulation. *The Annals of Applied Statistics*, 11(4):1833–1874.

Nyblom, J. (1989). Testing for the constancy of parameters over time. *Journal of the American Statistical Association*, 84(405):223–230.

Patton, A. J., Ziegel, J. F., and Chen, R. (2019). Dynamic semiparametric models for expected shortfall (and value-at-risk). *Journal of Econometrics*, 211(2):388–413.

Politis, D. N. and Romano, J. P. (1992). A circular block-resampling procedure for stationary data. In LePage, R. and Billard, L., editors, *Exploring the Limits of Bootstrap*, pages 263–270. Wiley, New York.

Romano, J. P. and Wolf, M. (2005). Stepwise multiple testing as formalized data snooping. *Econometrica*, 73(4):1237–1282.

Storti, G. and Wang, C. (2023). Modeling uncertainty in financial tail risk: A forecast combination and weighted quantile approach. *Journal of Forecasting*, 42(7):1648–1663.

Taylor, J. W. (2019). Forecasting value at risk and expected shortfall using a semiparametric approach based on the asymmetric Laplace distribution. *Journal of Business & Economic Statistics*, 37(1):121–133.

Taylor, J. W. (2020). Forecast combinations for value at risk and expected shortfall. *International Journal of Forecasting*, 36(2):428–441.

van der Vaart, A. W. and Wellner, J. A. (1996). *Weak Convergence and Empirical Processes*. Springer.

Zhang, N., Gong, Y., and Xue, X. (2023). Less disagreement, better forecasts: Adjusted risk measures in the energy futures market. *Journal of Futures Markets*, 43(10):1332–1372.