International Journal of
Financial Studies

MDPI

Article

When Does Uncertainty Calibration Exhaust the Risk Budget? Mathematical Feasibility Analysis and a Convergence-Audited Neural–EWMA Comparison

Ho Thi Hien [1], Jinho Cha [2], Sahng Min Han [3], Long Pham [4,*] and Ho Thi Hoang Luong [1]

[1] Faculty of Economics and Business Administration, Nghe An University, Vinh 43134, Nghe An, Vietnam; hienht@nau.edu.vn (H.T.H.); luonghth@nau.edu.vn (H.T.H.L.)
[2] Department of Computer Science, Gwinnett Technical College, Lawrenceville, GA 30043, USA; jcha@gwinnetttech.edu
[3] Department of Computer Science, Middle Georgia State University, Macon, GA 31206, USA; sahngmin.han@mga.edu
[4] Department of Decision Sciences and Economics, College of Business, Texas A&M University–Corpus Christi, Corpus Christi, TX 78412, USA
* Correspondence: long.pham@tamucc.edu

Abstract

A binding portfolio risk mandate may leave no admissible allocation once an empirical uncertainty margin is charged to estimated risk. We study this authorization outcome across 437 previously explored five-session decisions involving two exchange-traded funds (ETFs), the SPDR S&P 500 ETF Trust (SPY) and the iShares 1–3 Year Treasury Bond ETF (SHY). At the 12% annual mandate, a common uniform margin yields 161 neural and 18 exponentially weighted moving average (EWMA) empty sets. Under the same 126-matured-score rule, Ledoit–Wolf and oracle-approximating shrinkage yield 117 and 99, while newly fitted dynamic conditional correlation–generalized autoregressive conditional heteroskedasticity (DCC–GARCH) and daily-return heterogeneous autoregressive (HAR)-type benchmarks yield 8 and 69. A risk-scaled simultaneous envelope eliminates estimated exclusions for all six model families, without establishing prospective risk-bound coverage; historical full-grid exceedance frequencies remain around 11–12%. The estimated exclusion gap varies with calibration-history length, including residuals from 2020. In 437 retrospective policy decisions, 12%/50%-cash minus 15%/no-cash changes annual certainty-equivalent return by +0.535 percentage points for neural and −0.034 for EWMA at $\gamma = 3$, with descriptive paired month-block intervals $[-0.648, 1.762]$ and $[-0.975, 0.995]$. A separate post-source-data chronological extension supplies seven matured decisions per original model, with no observed bound exceedance, but no outcomes yet qualify as advance-logged prospective validation. Our mathematical contribution is an exact two-asset authorization and repair audit, not general multi-asset theory or demonstrated investor-welfare dominance.

Keywords: uncertainty calibration; portfolio feasibility; calibrated risk budgets; simultaneous risk-scaled envelope; covariance shrinkage; cash-inclusive allocation; historical economic evaluation

Check for updates

Academic Editor: Giuseppe Orlando

Received: 23 October 2026
Revised: 9 November 2026
Accepted: 16 November 2026
Published: 20 November 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

1. Introduction

A portfolio risk limit is more than a volatility target: it determines whether an investment manager is authorized to hold a proposed allocation. Forecast uncertainty therefore creates a decision problem that is not resolved by improving prediction alone. When a

calibration margin is added to point-risk estimates, every candidate portfolio may breach the stated limit. An empty feasible set can be a justified warning, but it leaves the manager with unanswered questions: which part of the upper-risk calculation caused the rejection, what change would make an allocation admissible, and would that change preserve the investor's economic objective? The answers matter precisely because a manager cannot respond to every rejection by relaxing a binding mandate.

Robust portfolio research establishes how uncertainty enters investment constraints and why stronger protection can narrow the opportunity set (Bertsimas & Sim, 2004; Goldfarb & Iyengar, 2003). Financial studies of optimized portfolios shift attention from predicted performance to the risks and returns an investor can obtain after trading costs (DeMiguel et al., 2009; Jensen et al., 2026). Learning-assisted security selection and robust portfolio construction provide ways of generating investable candidates (Chaweewanchon & Chaysiri, 2022; Miri et al., 2025), while research on constraint-aware reasoning examines the rules governing portfolio recommendations (Mashayamombe et al., 2026). These developments bring into focus a logically prior question under a binding uncertainty-adjusted mandate: *is any candidate admissible in the first place?* The paper studies that feasibility boundary, the changes permitted when it closes, and the economic consequences of restoring investment opportunities.

For a positive-semidefinite five-session point-risk forecast $\widehat{Q}_t$, empirical margin q_t, and permissible allocation grid $\mathcal{W}$, the smallest risk budget that authorizes at least one allocation is $B_t^{\mathrm{crit}} = q_t + \min_{w \in \mathcal{W}} \rho_t(w; \widehat{Q}_t)$. This identity provides the organizing object of the paper. We first characterize the entire feasible interval and its discretization gap; we then quantify the minimum risk-budget repair and establish perturbation and dual certificates for an authorization decision. Finally, risk-budget-dependent component identities explain how the two fitted inputs—point risk and calibration—can generate nonlinear differences in binary feasible-set outcomes. The purpose of these results is to turn an unexplained rejection into an interpretable threshold and a measurable menu of remedies.

We investigate the empirical mechanism on 437 nonoverlapping five-session decisions involving two exchange-traded funds (ETFs), the SPDR S&P 500 ETF Trust (SPY) and the iShares 1–3 Year Treasury Bond ETF (SHY) from December 2017 through August 2026. Under the primary 12% annualized mandate, matched neural and exponentially weighted moving average (EWMA) controllers produce 161 and 18 empty estimated feasible sets. Exchanging the fitted point-risk and calibration components produces 9 and 165, respectively; the exact symmetric decomposition assigns +34.21 percentage points of the 32.72-point matched difference to the buffer and −1.49 points to the point-risk component. Because this diagnosis depends on the fitted upper-risk surface, we next ask whether the margin protects the actions at issue. Matured upper-bound scores, crisis localization, action-restricted margins, numerical convergence and alternative fitted specifications form a connected reliability assessment. The neural empty-set count, for example, ranges from 74 to 245 across ten random seeds.

The second managerial choice is whether the permissible investment domain can be expanded without changing the nominal risky-sleeve mandate. We introduce cash and prove an exposure-homogeneous upper-risk constraint and an exact minimum cash-allowance threshold. A 50% cash cap restores estimated feasibility for both reference models, although 40 of 100 alternative model–calibration specifications retain infeasible dates. We then follow the feasible decisions into transaction costs, realized risky-sleeve quadratic variation (QV) and certainty-equivalent (CE) returns. With five-basis-point costs and lagged historical three-month Treasury yields as a cash-return proxy, annual CE increases from 3.77% to 5.98% for neural control and from 4.37% to 6.03% for EWMA in the observed historical paths (without establishing welfare dominance). The same policies also generate

more ex-post five-session risk-budget exceedances. Thus the financial decision is not simply whether an allocation can be authorized, but whether the chosen authorization remedy offers an acceptable combination of exposure, protection and welfare.

The argument proceeds from the investment mandate to the information and decisions it governs. Section 2 connects uncertainty-aware portfolio construction with the separate question of admissibility. Section 3 defines the dated information and risk-measurement protocol; Section 4 develops the exact feasibility and repair results; Section 5 tests their empirical content and economic implications. Discussion then translates the findings into a risk-control reporting sequence and identifies conditions for extending the evidence to other trading settings.

2. Literature and the Portfolio Authorization Problem

2.1. The Economic Price of a Binding Risk Constraint

The mean–variance tradition establishes how investors trade expected return against estimated risk (Markowitz, 1952). When risk also serves as a binding investment mandate, however, the estimation problem becomes a decision problem: a forecast can determine whether *any* allocation is permitted, not merely which weight ranks first. Portfolio restrictions change the effects of noisy covariance inputs (Jagannathan & Ma, 2003), while robust formulations make estimation uncertainty part of the optimization problem (Goldfarb & Iyengar, 2003). The cost of that protection is central to the price-of-robustness analysis of Bertsimas and Sim (2004) and the treatment of overconservative robust counterparts by Roos and den Hertog (2020). Moment ambiguity and distributionally robust optimization provide richer ways of representing parameter uncertainty (Delage & Ye, 2010; Mohajerin Esfahani & Kuhn, 2018; Natarajan et al., 2009; Wiesemann et al., 2014).

For a portfolio manager, the distinction between an expensive feasible allocation and an *empty* allocation set is consequential. The former still permits optimization within the mandate; the latter requires a different decision before optimization can proceed. How much of the required risk budget is consumed by the forecast itself, and how much by the margin added to protect against its error? At what precise budget would an allocation reappear? Those questions motivate the critical-budget threshold and minimum-repair results in Section 4. They also make the empirical size and reliability of the uncertainty margin indispensable financial quantities rather than secondary implementation details.

2.2. From Risk Forecasts to Admissible Portfolio Choices

Financial risk forecasts use dynamic correlations, realized-volatility models, shrinkage estimators and machine-learning methods to construct estimates of conditional risk (Christensen et al., 2023; Corsi, 2009; Engle, 2002; Ledoit & Wolf, 2004). But forecast evaluation depends on the realized-risk measure (Patton, 2011), and economic outcomes also depend on how estimated risk is converted into exposure and trades. Volatility management changes the allocation itself (Moreira & Muir, 2017); optimized portfolios may struggle to outperform simple diversified rules after implementation (DeMiguel et al., 2009). Accounting for turnover and transaction costs is therefore central to assessing the investable efficient frontier (Jensen et al., 2026).

The connection from risk information to investment action is developed further in learning-assisted security selection (Chaweewanchon & Chaysiri, 2022), neural estimation combined with robust portfolio selection under ambiguity (Miri et al., 2025), and signal-to-policy learning for risk-aware allocations (Yu & Chang, 2026). Alternative entropy-based portfolio criteria further illustrate how the choice of risk functional changes admissible investment trade-offs (Giunta et al., 2024). Constraint-aware portfolio reasoning makes the role of investment restrictions more transparent (Mashayamombe et al., 2026). Together,

these approaches motivate examining the point at which a candidate allocation must actually satisfy a binding risk mandate. For a mandate augmented by an empirical uncertainty margin, this is a separate decision: even when the forecasting and ranking stages work as intended, *the admissible set may be empty*.

When every candidate fails the upper-risk constraint, the manager cannot solve the problem by ranking those same candidates again. The first question is the exact risk threshold at which an allocation becomes admissible. The second is how much of that threshold comes from the point-risk forecast and how much from the calibration margin. The third is which changes are permitted—for example, enlarging the allocation domain to include cash rather than automatically relaxing the risk limit—and whether the resulting portfolio provides credible risk protection and economic value. These questions require the feasibility and repair results of Section 4, and they make the reliability of the empirical calibration margin the next step in the analysis.

2.3. Decision-Relevant Calibration and the Cost of Conservatism

The authorization boundary cannot be assessed without examining the empirical buffer that enters it. In our reference construction, the largest positive five-session risk-forecast residual over the 201-weight grid produces one common score. A margin computed from past scores consequently applies to *every* permissible weight. This design has an attractive logical property: on the event that the next realized score does not exceed the margin, its upper-risk bound holds simultaneously throughout the grid. But the economic consequence is equally important. One unusually large error at a weight that the controller would never select can tighten the risk limit for every action the controller might use.

Sequential prediction and adaptive calibration research develops ways to reason about uncertainty under dependence and changing distributions (Gibbs & Candès, 2024; Xu & Xie, 2021; Zaffran et al., 2022). Conformal risk control and downstream-decision-oriented predictive sets connect the protected quantity to a specified loss or action (Angelopoulos et al., 2024; Wang & Dobriban, 2026). The resulting design question for our application is precise: *does an action-uniform margin protect the relevant investment decisions at a justifiable cost in foregone feasible opportunities?* We investigate its realized upper-bound exceedances, magnitudes and clustering, then examine alternative margins restricted to prespecified investment sleeves. A reduced margin is meaningful only together with the action set for which its protection is evaluated. Lemma 7 establishes the exact conditional implication connecting a score event to the realized risk of a feasible selected action; the empirical analysis determines how frequently the event occurs.

2.4. Contributions: Diagnosis, Stability, Repair and Valuation

The theoretical formulas below are exact but elementary specializations to a two-asset, positive-semidefinite quadratic model. Our empirical contribution is the audited sensitivity of investment authorization to the fitted covariance, simultaneous calibration score, calibration-history window, and permitted cash domain. The original uniform construction remains a clearly labeled reference case, not a recommended universal margin. The literature leads to four connected diagnostic components. *Diagnosis* locates the exact critical risk budget at which an allocation becomes admissible and accounts for the difference between continuous and executable grid weights. *Certification* identifies the smallest budget adjustment, tests sensitivity to perturbations and derives a dual certificate, while separating point-risk and calibration contributions along the budget frontier. *Repair* changes the allowable investment domain by permitting cash and determines the minimum cash capacity required for an admissible position under the exposure-based risk definition. *Valuation* connects the selected positions to subsequent prediction errors, realized quadratic variation,

trading costs and certainty-equivalent welfare. Each contribution produces information required for the next decision: identifying the source of exclusion precedes choosing a remedy, and evaluating a remedy requires following the resulting trade.

The data protocol must therefore preserve one chronological decision calendar and clearly distinguish information available at the decision date from outcomes observed afterwards. Section 3 implements that protocol; Section 4 derives the decision rules; the empirical results then evaluate whether the rules reveal informative, economically consequential differences rather than merely producing different forecasts.

An additional decision implication follows from the new cross-estimator comparisons: the empirical authorization gap is not a universal ranking of covariance forecasts. The risk limit applies to a jointly specified forecast, calibration rule, permissible asset set, and dated information record. Accordingly, our contribution is to audit the *marginal authorization consequences* of those choices, not to infer that any one fitted forecast is universally less reliable. Section 6, specifically the Section *Marginal Authorization and Managerial Implications*, translates this distinction into an operational sequence, with separate checks for estimation, authorization and realized economic outcomes.

3. Data and Evaluation Protocol

3.1. Market Inputs and the Common Decision Calendar

We study the SPDR S&P 500 ETF Trust (SPY) and the iShares 1–3 Year Treasury Bond ETF (SHY). Their historical adjusted-opening-price observations provide a consistent return series for the same two assets, allowing the feasible-set geometry to be examined without changes in the investment universe. The series share 2964 dates from 1 December 2014 through 15 September 2026. The dated predictors include lagged Cboe Volatility Index (VIX), realized market volatility and S&P 500 momentum. To establish a conservative decision-time information convention, these market-state covariates are shifted by one trading session; the additional trade-execution lag is specified below. Because vendor publication times and opening-auction fills were not independently authenticated, the reported returns and transaction costs represent a consistently timed historical execution proxy, whose implications for implementation are discussed in Section 6.

Evaluation starts on 12 December 2017, after the training and calibration warmup, and ends on 18 August 2026. It consists of *437 nonoverlapping five-session allocation decisions* and 2185 subsequent daily return sessions. This common calendar is essential: changing the sampled dates while comparing forecasting methods could turn a difference in authorization decisions into a difference in market exposure. All primary forecast swaps, cash policies and risk diagnostics therefore retain the same 437 decision dates. The calendar has been used in earlier design exploration; the additional analyses are retrospective sensitivity and mechanism investigations within that history, not an independent prospective or untouched out-of-sample confirmation.

3.2. A Five-Session Risk Measure and a Strictly Ordered Information Set

Let $r_s \in \mathbb{R}^2$ be the SPY/SHY adjusted-open-to-next-open simple-return vector, and let t denote a portfolio signal date. After the five subsequent holding intervals have ended, the realized second-moment target is

$$Q_t^{\text{real}} = \sum_{h=1}^{5} r_{t+h} r_{t+h}^{\top}. \tag{1}$$

The execution protocol uses a one-session lag and includes the completed five-session outcome in a calibration or training history only after its final return has matured. The matrix Q_t^{real} is the uncentered sum of five outer products, so its portfolio quadratic form

measures the realized second-moment contribution of the five holding intervals. With SPY weight $w \in [0,1]$, SHY weight $1-w$ and $v(w) = (w, 1-w)^\top$, define

$$\rho_t(w;Q) = \sqrt{v(w)^\top Q v(w)}. \tag{2}$$

The quantity is a five-session risk *proxy* expressed in return units. The later empirical budget-exceedance indicators refer to this same proxy, which keeps the theory and observed test target aligned.

Two families supply positive-semidefinite (PSD) five-session forecasts $\widehat{Q}_t^m$, where $m \in \{N, E\}$ denotes neural and EWMA. The multilayer perceptron (MLP) predicts the three distinct normalized elements of Q_t^{real} from lagged market and ETF features; its symmetric output is projected onto the PSD cone. The reference neural model is implemented using the `MLPRegressor` class from scikit-learn. Its architecture has one hidden layer of 12 units with the hyperbolic tangent (tanh) activation function. The parameter $\alpha = 2$ controls the ℓ_2 weight penalty, and a fixed random seed is used.

This deliberately small network is a historically fixed benchmark, not a state-of-the-art forecasting architecture. The convergence-audited fit uses a 1000-iteration ceiling. The preliminary 140-iteration configuration generated warnings in 37 fits; increasing only the ceiling to 1000 removes those reported iteration-limit warnings, while preserving the earlier fit as a numerical-sensitivity comparator. The EWMA reference uses decay $\lambda = 0.94$.

At each decision date, only *fully matured* prior outcomes enter estimation. The first eligible fit requires at least 504 matured training observations and 35 past calibration scores; the rolling training record has at most 1260 rows. Calibration uses preceding nonoverlapping prequential forecast epochs. The design therefore separates the information used to authorize an allocation from its future realized risk and return. The execution lag and maturity restrictions create a consistent, auditable historical information cut at every decision date.

3.3. From One Empirical Risk Margin to a Feasible Investment Set

The reference controller places a single safety margin on all SPY/SHY weights, so its calibration target is defined on the full admissible grid. For model m the nonnegative action-uniform score is

$$S_t^m = \max\left\{0, \max_{w \in \mathcal{W}} \left[\rho_t(w; Q_t^{\text{real}}) - \rho_t\left(w; \widehat{Q}_t^m\right)\right]\right\}, \tag{3}$$

where $\mathcal{W} = \{0, 0.005, \ldots, 1\}$ contains 201 weights. The next margin q_t^m is the empirical 90th percentile of no more than the 126 most recent *matured* scores, subject to at least 35 eligible scores. The margin is applied after, rather than inside, the covariance forecast; keeping these components separate makes the two-by-two exchange audit possible. The 90th percentile specifies a 10% nominal upper-tail design target, while the rolling window contains at most 126 matured score observations; it is the 90th percentile of those scores, not their maximum, that supplies q_t^m. This reference window is evaluated alongside 35-, 63- and 252-epoch alternatives, rather than treated as an empirically optimal length. Section 5.2 evaluates the resulting score exceedances on the full grid and in stressed subperiods.

The principal annualized risk mandate is 12%, corresponding to the five-session budget

$$B = 0.12\sqrt{5/252} \approx 0.0169. \tag{4}$$

An action is *estimated feasible* when its forecast plus empirical margin fits within this budget:

$$\mathcal{F}_t(\widehat{Q}, q) = \left\{w \in \mathcal{W} : \rho_t(w; \widehat{Q}) + q \le B\right\}. \tag{5}$$

The square-root-of-time factor is a stated annual-to-five-session mandate convention; it is not evidence that ETF returns are independent and identically distributed or that the ex-post

five-day risk measure obeys exact diffusion scaling. We take 12% as the fixed reference risk mandate and evaluate $B(b) = b\sqrt{5/252}$ at annual mandates $b \in \{0.10, 0.12, 0.15\}$ without changing the fitted forecasts and margins. These three values reveal how a stricter or more permissive authorization boundary alters the same historical decisions. The controller's 80% strategic SPY weight supplies a transparent tracking reference; turnover is measured relative to the drifted pre-trade portfolio. When $\mathcal{F}_t$ is empty, the reference simulation uses a minimum-*estimated*-risk fallback and records the date as a mandate exception. The economic analysis separately tests 0-, 5- and 10-basis-point trading charges and reports the impact of cash allowances under all three mandates.

3.4. Allocation Mechanics and Within-Horizon Weight Accounting

The investment decision is made once per nonoverlapping five-session episode. A candidate target is $u_t = (u_{S,t}, u_{H,t}, u_{C,t})^\top$, where the nonnegative weights sum to one, $u_{C,t} \le c_{\max}$ and $s_t = u_{S,t} + u_{H,t}$ is the risky ETF-sleeve exposure. When $s_t > 0$, the within-sleeve SPY fraction is $w_t = u_{S,t}/s_t$; when $s_t = 0$ the sleeve contributes zero risk. For the common-buffer reference, the pre-trade risk test is

$$s_t\{\rho_t(w_t; \widehat{Q}_t) + q_t\} \le B. \tag{6}$$

The controller chooses among feasible grid candidates according to the explicit tracking–turnover objective

$$\min_{u_t:\,\text{estimated feasible}} \left\{0.25\|u_t - (0.8, 0.2, 0)^\top\|_2^2 + 0.005\|u_t - \tilde{u}_t\|_1\right\}, \tag{7}$$

where $\tilde{u}_t$ is the drifted portfolio immediately before the first execution trade. If no candidate satisfies the estimated limit, a minimum-estimated-upper-risk action is taken as an explicitly labeled *mandate exception*, not as a feasible trade.

The selected target u_t is *held as a target through daily constant-mix rebalancing during the five-session holding episode*; the optimizer does not choose new targets after observing intra-episode returns. At each of its five daily executions, asset-return drift creates pre-trade weights $\tilde{u}_{t,h}$; the simulation trades back to u_t, charges $c_{t,h} = 0.0005\|u_t - \tilde{u}_{t,h}\|_1$ at the five-basis-point reference cost, and records the portfolio return

$$r_{t,h}^{\text{net}} = (1 - c_{t,h})\left(1 + u_t^\top R_{t,h}\right) - 1, \qquad \tilde{u}_{t,h+1,j} = \frac{u_{t,j}(1 + R_{t,h,j})}{1 + u_t^\top R_{t,h}}, \tag{8}$$

where $R_{t,h} = (r_{S,t,h}, r_{H,t,h}, r_{C,t,h})^\top$. Thus the reported turnover captures the drift induced by each preceding return, not an assumption that weights are unchanged without trading. The five-session realized risk measure for the risky sleeve is

$$\rho_{t,\text{real}}^{\text{sleeve}}(u_t) = \left[\sum_{h=1}^{5}\left(u_{S,t}r_{S,t,h} + u_{H,t}r_{H,t,h}\right)^2\right]^{1/2} = s_t\rho_t(w_t; Q_t^{\text{real}}). \tag{9}$$

The equality follows from daily rebalancing to the *same* target and the definition of Q_t^{real}; it would not be exact for un-rebalanced buy-and-hold weights that drift freely throughout the episode. Cash return is valued separately through the lagged Treasury-yield proxy and is not included in this ETF-sleeve risk measure. Consequently, the estimated risk constraint is an authorization calculation for the chosen rebalanced exposure, not a bound on a continuously drifting buy-and-hold or full-portfolio tail-loss process. Buy-and-hold within-period drift, slippage and realized transaction prices remain external-validity limitations.

3.5. Why the Bound Can Fail in Sharp Market Moves

The risk margin is the rolling empirical 90th percentile of historical, fully matured five-session residual scores; it is not a distribution-free conditional guarantee. Large abrupt

changes in volatility, covariance or correlation can exceed a forecast and its lagged calibration margin before the new residual has matured and entered the next window. Conversely, an extreme past score can remain in a long rolling window and overconstrain later decisions even after contemporaneous point risk declines. The February–June 2020 retrospective score-removal exercise and the 35/63/126/252-score sensitivity quantify this historical dependence; they do not establish a causal crisis effect. In the earlier dated 2020 stress subsample, cash-inclusive neural and EWMA reference policies show 7/21 and 5/21 realized ETF-sleeve budget exceedances, respectively. Cash permission may also increase SPY exposure, so fewer estimated empty sets need not imply fewer realized breaches. These observations motivate reporting both authorization and subsequent risk protection, and motivate truly advance-logged future evaluation.

3.6. A Deliberately Simple Cash-Allowing Reference

The earlier archived simple-cash comparison also evaluated an unconditional daily constant-mix allocation of 50% SHY and 50% cash, without covariance forecasts or a calibrated allocation rule. To avoid conflating cash-rate assumptions, we compare that rule only with the *same* archived zero-cash-carry policy simulations, on the same 437 decisions with five-basis-point turnover costs and $\gamma = 3$; these values are separate from the later D/J results using lagged observed Treasury yields. In that zero-carry setting, the simple SHY/cash allocation has an annual certainty equivalent of 0.956%, versus 5.235% and 5.457% for the neural and EWMA 12%/50%-cash policies. The average SPY exposure is zero for the simple policy but approximately 43.4% and 49.6% for the two forecast-based policies, respectively. This contrast is *not* a risk-matched performance test, a claim of risk-budget compliance by the unconditional reference, or evidence of a causally superior allocation rule. Its main interpretation is that cash permission and equity exposure must be evaluated jointly. We additionally report the 12%/50%-cash versus 15%/zero-cash comparison with directly paired descriptive CE intervals, both of which cross zero.

3.7. Retrospective Risk-Scaled Simultaneous Envelope and Classical Estimators

The reference rule charges a single positive margin q_t to all 201 weights. We additionally assess a simultaneous, but weight-dependent, risk-scaled rule. Let $\hat{\rho}_t^m(w) = \rho_t(w; \widehat{Q}_t^m)$ and define the fixed scale

$$h_t^m(w) = \frac{\max\{\hat{\rho}_t^m(w), 0.002\}}{0.01} > 0, \qquad S_t^{m,\mathrm{sc}} = \max_{w \in \mathcal{W}} \frac{[\rho_t(w; Q_t^{\mathrm{real}}) - \hat{\rho}_t^m(w)]_+}{h_t^m(w)}. \tag{10}$$

The scaled margin $q_t^{m,\mathrm{sc}}$ is the rolling empirical 90th percentile of up to L fully matured historical $S_t^{m,\mathrm{sc}}$ scores, with $L \in \{35, 63, 126, 252\}$ and a minimum history of 35. The resulting upper bound at weight w is $\hat{\rho}_t^m(w) + q_t^{m,\mathrm{sc}} h_t^m(w)$. The scale constants 0.002 and 0.01 are fixed protocol choices, not independently selected or optimized on a prospective holdout. No separate weight-by-weight empirical quantile is fitted; the maximum normalized score still imposes simultaneous protection across all 201 weights.

For a nonconstant envelope, the estimated critical budget changes from the reference identity $q_t + \min_w \hat{\rho}_t^m(w)$ to

$$B_{t,\mathrm{sc}}^{\mathrm{crit}} = \min_{w \in \mathcal{W}} \{\hat{\rho}_t^m(w) + q_t^{m,\mathrm{sc}} h_t^m(w)\}. \tag{11}$$

Therefore the exact two-asset common-buffer formula is retained for the reference specification and is not incorrectly applied to the risk-scaled alternative. Both rules use the same matured prequential chronology and evaluate subsequent realized five-session ETF risk on the same grid; none of the historical comparisons is described as prospective conditional coverage.

To test estimator dependence, we retain the frozen convergence-audited neural and EWMA forecasts and add Ledoit–Wolf and oracle-approximating shrinkage (OAS) sample-covariance estimators, each formed using a trailing 504 daily ETF returns prior to the decision and converted to the specified five-session matrix scale. The pipeline produces 473 historical forecast entries and 437 matched evaluated decisions per estimator. A complementary retrospective analysis fits two additional dynamic covariance benchmarks using the same dated evaluation calendar (Section 3.8).

3.8. Additional Dynamic and Heterogeneous-Autoregressive Benchmark Protocol

Beyond the shrinkage estimates, the additional dynamic-benchmark analysis fits a Gaussian generalized autoregressive conditional heteroskedasticity (GARCH(1,1)) model with dynamic conditional correlation (DCC) using the preceding 504 daily SPY/SHY returns and refits its parameters every 63 trading sessions. It sums the forecast conditional covariance matrices over the five holding intervals specified by the original one-session execution lag. A separate daily-return heterogeneous autoregressive (HAR)-type alternative uses lag-1, lag-5, and lag-22 daily-return outer-product covariance predictors in a ridge regression with training responses matured strictly before each signal. The forecast matrix is projected onto the positive-semidefinite cone when needed. This HAR-type proxy uses daily returns, *not* intraday realized-volatility data, and must not be called a high-frequency HAR–RV test. Both models produce 473 dated forecast histories and are evaluated on the same 437-date calendar under the uniform and risk-scaled 90th-percentile calibration schemes. The computational consistency audit reports 241,813 checks passed and zero failed; this is a computational-reconciliation result, not a guarantee of validity or independent verification.

For policy differences, the direct paired analysis compares the daily returns of 12%/50%-cash against 15%/no-cash directly using *paired* calendar-month resampling with 2000 draws over 105 observed months. The comparison is repeated for risk-aversion coefficients $\gamma \in \{1, 3, 5, 10\}$. These exploratory intervals summarize the previously reused historical calendar and should not be interpreted as a new untouched experiment. A chronological post-archive assessment connects additional dated SPY/SHY and market-state observations to a frozen model protocol and separates historical post-source extension from signals first logged before execution. The primary 437-date history and input artifacts remain unchanged.

3.9. Audit Sequence and Economic Endpoints

We use one decision chronology to connect four measurement layers. First, the two-by-two exchange of fitted covariance forecasts and empirical buffers attributes differences in the *estimated empty-set indicator*; exchanged pairs are arithmetic diagnostics evaluated on the same dates. Second, fully matured residual scores and action-restricted alternatives examine the reliability and economic cost of the calibration margin. Third, numerical sensitivity includes ten neural seeds and a 100-specification matrix spanning training and calibration windows, model architecture, regularization, EWMA decay and covariance estimation. Fourth, we extend the admissible portfolio to SPY, SHY and cash while keeping the original tracking and turnover convention, then measure the realized consequences of the selected policy.

The cash extension is designed to evaluate how a broader admissible investment set changes portfolio authorization while preserving the original five-session SPY/SHY risk mandate. Accordingly, the constraint measures risk attributable to the ETF sleeve: when the fraction of wealth invested in that sleeve is s, both its fitted quadratic-variation risk and its empirical calibration margin are scaled by s. Cash contributes zero to this *ETF-sleeve exposure measure*. This consistent scaling produces an exact minimum cash-allowance

threshold without changing the original two-ETF risk forecasts. All subsequent five-session budget-exceedance statistics use the same risky-sleeve definition.

Cash also contributes to investment returns. We value dated carry using the historical three-month U.S. Treasury yield (Federal Reserve Economic Data (FRED) DGS3MO) after a two-ETF-session information lag, so the daily proxy does not presume knowledge of a same-session published yield. Quoted annual yields are converted to $r_{c,s} = (1+y_{s-2})^{1/252}-1$, where y is expressed as an annual decimal yield. This conversion approximates daily cash carry from a bond-equivalent quoted Treasury yield rather than applying the exact money-market instrument accrual convention; it is not the observed total return on a separately traded cash instrument. It enters net portfolio returns and certainty-equivalent welfare; the calibrated ETF-sleeve risk mandate and the portfolio weights selected by the original controller are held fixed when the assumed cash carry is replaced.

The economic calculation charges the stated five-basis-point cost to the *two-leg* L^1 turnover of portfolio weights relative to drifted holdings. With daily post-cost portfolio returns r_s^{net}, we report annual mean–variance certainty equivalent and volatility using

$$\text{CE}_\gamma = 252\left(\bar{r}^{\text{net}} - \frac{\gamma}{2}\widehat{\text{Var}}(r^{\text{net}})\right), \quad \sigma_{\text{ann}} = \sqrt{252\widehat{\text{Var}}(r^{\text{net}})}, \quad \gamma = 3,$$

where the sample variance uses the same one-degree-of-freedom adjustment in both statistics. We separately record *(i)* estimated empty-set events, *(ii)* ex-post realized risky-sleeve budget exceedances, *(iii)* upper-bound score exceedances, and *(iv)* net investor welfare. Estimated feasibility determines which trades can be authorized; matured residual scores evaluate calibration; realized quadratic variation measures the subsequent risky-sleeve exposure; and CE captures net investor welfare.

Sampling frequency and market windows can change observed return distributions (Orlando & Bufalo, 2021). We therefore keep the same 437 decisions in every paired comparison and report calendar-year concentration and month-block CE sensitivity. Because that calendar also informed earlier research development, these diagnostics measure the stability of conclusions within the observed history, where return dependence and crisis concentration affect their interpretation. An independently selected later calendar would be needed to estimate transportability to future trading decisions. These sampling considerations qualify empirical generalization, while the exact feasibility identities depend only on their stated mathematical assumptions.

4. Mathematical Analysis of Calibrated Feasibility

The preceding section turns a practical authorization problem into four mathematical tasks. First, determine exactly when the permissible set contains an allocation. Second, compute the least adjustment capable of reopening it. Third, certify whether that feasibility judgment survives a change in the fitted risk inputs. Fourth, identify how the point-risk forecast and calibration margin move the decision boundary, and how permission to hold cash changes the answer. Each result below supplies a concrete output to one of these tasks. The geometry and perturbation claims are deterministic under $Q \succeq 0$ and $q \geq 0$; the link to subsequent realized risk is established conditional on the specified calibration-score event.

4.1. Exact Continuous Geometry and Discrete Feasibility

The first decision is whether the constraint itself permits any allocation, before the controller chooses among allocations. We solve the continuous two-asset problem and then account explicitly for the 0.005 trading grid. This distinction matters because a narrow mathematically feasible interval may contain no executable grid point.

For the moment let $\mathcal{W}_c = [0,1]$ denote the continuous fully invested two-asset simplex, with $v(w) = (w, 1-w)^\top$. Write a generic symmetric positive-semidefinite 2×2 matrix as

$$Q = \begin{pmatrix} A & C \\ C & D \end{pmatrix} \succeq 0, \qquad \eta = A + D - 2C = (1,-1)Q(1,-1)^\top \geq 0, \qquad R = B - q. \tag{12}$$

The annualized budget is $b = B\sqrt{252/5}$. Both risk $\rho(w;Q) = \sqrt{v(w)^\top Q v(w)}$ and q are in five-session risk units, not in return or annualized percentage units.

Lemma 1 (Exact feasible-interval geometry)**.** *For $Q \succeq 0$, $q \geq 0$ and $B \geq 0$, define*

$$f(w) = v(w)^\top Q v(w) = \eta w^2 + 2(C-D)w + D. \tag{13}$$

If $R < 0$, $\mathcal{F}_c(Q,q;B) = \emptyset$. If $\eta = 0$, $f(w) = D$ for all $w \in [0,1]$ and

$$\mathcal{F}_c(Q,q;B) = \begin{cases} [0,1], & D \leq R^2 \text{ and } R \geq 0, \\ \emptyset, & \text{otherwise.} \end{cases} \tag{14}$$

If $\eta > 0$, put

$$z = \frac{D-C}{\eta}, \qquad f_0 = D - \frac{(D-C)^2}{\eta} \geq 0, \qquad \nu = \sqrt{\frac{R^2 - f_0}{\eta}} \quad \text{when } R \geq 0,\ R^2 \geq f_0. \tag{15}$$

When $R \geq 0$ and $R^2 \geq f_0$, the exact continuous feasible set is the possibly empty interval

$$\mathcal{F}_c(Q,q;B) = [z - \nu, z + \nu] \cap [0,1]. \tag{16}$$

If $R^2 < f_0$, the set is empty. In all cases the point-risk-minimizing continuous weight and risk are

$$w_c^* = \Pi_{[0,1]}(z), \qquad a_c(Q) = \sqrt{f(w_c^*)} \quad (\eta > 0), \tag{17}$$

with any $w_c^* \in [0,1]$ optimal if $\eta = 0$. Consequently

$$\mathcal{F}_c(Q,q;B) \neq \emptyset \iff B \geq q + a_c(Q). \tag{18}$$

Proof. The expansion in Equation (13) follows by substituting $v(w)$. Positive semidefiniteness gives $f(w) \geq 0$ for every real w and $\eta \geq 0$. If $\eta = 0$, then $Q^{1/2}(1,-1)^\top = 0$, hence f is constant and equals $f(0) = D$. If $\eta > 0$, completing the square yields $f(w) = f_0 + \eta(w-z)^2$. For $R \geq 0$ the inequality $\sqrt{f(w)} \leq R$ is equivalent to $f(w) \leq R^2$, producing Equation (16); no feasible point exists when $R < 0$. The convex quadratic attains its minimum on $[0,1]$ at the projection of its vertex z, which proves the last equivalence. □

The implemented policy uses the *grid* $\mathcal{W}_h = \{kh : k = 0, \ldots, M\}$ with $h = 1/M = 0.005$ and $M = 200$; thus the continuous interval need not contain an admissible grid point.

Lemma 2 (Exact grid optimizer and controlled discretization loss)**.** *For $\eta > 0$ the grid-risk minimum is attained at*

$$w_h^* = h\,\Pi_{\{0,\ldots,M\}}(\mathrm{round}(z/h)), \qquad a_h(Q) = \sqrt{f(w_h^*)}, \tag{19}$$

where ties in rounding may be broken arbitrarily. For $\eta = 0$ all grid weights give the same risk. With $a_c(Q)$ as in Lemma 1,

$$0 \leq a_h(Q) - a_c(Q) \leq \frac{h}{2}\sqrt{\eta}. \tag{20}$$

Accordingly the discrete and continuous critical budgets satisfy

$$0 \le B_h^{\mathrm{crit}} - B_c^{\mathrm{crit}} \le \frac{h}{2}\sqrt{\eta}, \qquad B_j^{\mathrm{crit}} = q + a_j(Q), \quad j \in \{h, c\}. \tag{21}$$

If $\mathcal{F}_c \neq \emptyset$ and its interval endpoints in Equation (16) are $L \le U$ after intersection with $[0,1]$, then the grid set is nonempty exactly when

$$\lceil L/h \rceil \le \lfloor U/h \rfloor. \tag{22}$$

Proof. For $\eta > 0, f(w) = f_0 + \eta(w - z)^2$, so minimizing over the uniform grid is equivalent to choosing its closest point to z, clipped to the endpoints. If $z \notin (0,1)$ the continuous minimum occurs at the same endpoint, hence $a_h = a_c$. Otherwise a grid point lies within $h/2$ of z, so $a_h \le \sqrt{f_0 + \eta h^2/4} \le \sqrt{f_0} + h\sqrt{\eta}/2 = a_c + h\sqrt{\eta}/2$. Nonnegativity follows from $\mathcal{W}_h \subset [0,1]$. The final statement is the integer-point condition on a closed interval. □

Corollary 1 (Boundary buffer cases)**.** *On either domain $\mathcal{W}_j$, $j \in \{c, h\}$, the risk constraint is infeasible exactly when $q + a_j(Q) > B$. In particular, $q > B$ guarantees emptiness. However, $q = B$* alone is not sufficient*: at equality the set is nonempty if and only if an allowed portfolio has zero estimated risk. When $a_j(Q) > 0$, $q \ge B$ guarantees emptiness.*

Proof. This is the threshold $B_j^{\mathrm{crit}} = q + a_j(Q)$ together with $a_j(Q) \ge 0$. □

A constructed rank-one example shows why the grid restriction matters even with an exact PSD forecast. For $Q = \left(\begin{smallmatrix} 4 & -2 \\ -2 & 1 \end{smallmatrix}\right)$, $\rho(w; Q) = |3w - 1|$. The continuous choice $w = 1/3$ has zero risk, whereas the nearest 0.005-grid choice $w = 0.335$ has risk 0.005. Thus the hypothetical residual budget $B - q = 0.002$ permits a continuous portfolio but no grid portfolio. The example isolates a discretization-induced rejection that arises entirely from the prescribed trading grid.

The logged quantities "buffer alone $\ge B$" count boundary-or-exceedance events; without the underlying values they must not be conflated with a verified count of strict $q > B$ events. In the empirical analysis we use the *actual grid feasible-set indicators*, not that shortcut.

4.2. Minimum Budget Repair, Stability, and Dual Certificates

When the feasible set is empty, the decision system needs a quantitative remedy rather than a binary rejection. The next result measures the exact budget shortfall and provides a sufficient stability margin: it distinguishes decisions that remain unchanged under an input perturbation from decisions near the boundary that merit closer review.

The feasibility condition has an operational interpretation as a *minimum risk-budget repair*. Define $a_h(Q) = \min_{w \in \mathcal{W}_h} \rho(w; Q)$ and

$$\begin{aligned} G(B; Q, q) &= B - q - a_h(Q), \\ \delta^*(B; Q, q) &= \min\{\delta \ge 0 : \exists w \in \mathcal{W}_h,\ \rho(w; Q) + q \le B + \delta\}. \end{aligned} \tag{23}$$

Lemma 3 (Exact repair and perturbation robustness)**.** *For any two positive-semidefinite matrices Q, Q' and nonnegative buffers q, q', the minimum additional five-session budget needed to restore grid feasibility is*

$$\delta^*(B; Q, q) = [q + a_h(Q) - B]_+ = [-G(B; Q, q)]_+, \qquad [x]_+ = \max\{x, 0\}. \tag{24}$$

Moreover, with the spectral norm $\|\cdot\|_{\mathrm{op}}$,

$$\left|[a_h(Q) + q] - [a_h(Q') + q']\right| \le |q - q'| + \sqrt{\|Q - Q'\|_{\mathrm{op}}} =: \varepsilon. \tag{25}$$

Therefore, if $|G(B;Q,q)| > \varepsilon$, *changing* (Q,q) *to* (Q',q') *cannot reverse the original grid feasibility classification. For fixed* Q, *the repair is piecewise linear and nondecreasing in* q, *nonincreasing in* B, *with slopes* $\partial\delta^*/\partial q = 1$ *and* $\partial\delta^*/\partial B = -1$ *whenever* $\delta^* > 0$ *and away from the kink.*

Proof. A budget increment δ permits some grid allocation exactly if $B + \delta \geq q + a_h(Q)$; minimizing the nonnegative increment proves Equation (24). For every $w \in \mathcal{W}_h$, $\|v(w)\|_2^2 = w^2 + (1-w)^2 \leq 1$. The elementary inequality $|\sqrt{x} - \sqrt{y}| \leq \sqrt{|x-y|}$ for nonnegative x, y, combined with the quadratic-form norm bound, gives

$$|\rho(w;Q) - \rho(w;Q')| \leq \sqrt{|v(w)^\top (Q - Q')v(w)|} \leq \sqrt{\|Q - Q'\|_{\mathrm{op}}}.$$

Taking minima preserves this uniform bound and adding $|q - q'|$ gives Equation (25). The risk-budget margin changes by at most ε, so its strict sign cannot change if its initial absolute value exceeds ε. The slope statements follow directly from the positive-part formula. □

The perturbation inequality is a deterministic certificate: for any supplied tolerance on the covariance matrix and empirical margin, it identifies decisions whose feasibility classification is stable. Turning it into an empirical stability frequency requires credible bounds on the size of those input errors. The empirical section therefore complements this structural result with observed variation across neural initializations and alternative forecast and calibration specifications.

Lemma 4 (Geometric dual certificate for continuous infeasibility)**.** *Let* $Q = L^\top L \succeq 0$, $e_1 = (1,0)^\top$ *and* $e_2 = (0,1)^\top$. *The minimum continuous point risk admits the exact dual representation*

$$a_c(Q) = \min_{0 \leq w \leq 1} \|L(we_1 + (1-w)e_2)\|_2 = \max_{\|u\|_2 \leq 1} \min\{u^\top L e_1,\ u^\top L e_2\}. \tag{26}$$

If a vector u *with* $\|u\|_2 \leq 1$ *satisfies*

$$\min\{u^\top L e_1, u^\top L e_2\} > B - q, \tag{27}$$

then the continuous feasible set, and therefore every subset grid, is empty. Conversely, continuous infeasibility admits such a dual witness.

Proof. Use $\|x\|_2 = \max_{\|u\|_2 \leq 1} u^\top x$ and the compact convex sets $[0,1]$ and the Euclidean unit ball. The bilinear objective permits the minimax interchange by the standard convex–concave minimax theorem. The minimum of an affine function of $w \in [0,1]$ occurs at an endpoint, proving Equation (26). If Equation (27) holds, $a_c(Q) > B - q$ and feasibility fails. Conversely $a_c(Q) > B - q$ and compactness ensure a maximizer satisfying the strict inequality. Grid infeasibility by itself does *not* imply that the continuous dual condition holds because a nonempty continuous interval may miss the grid. □

4.3. Risk-Budget Frontiers and Exact Covariance–Buffer Decomposition

The minimum-repair threshold is additive in the point-risk minimum and the buffer. The corresponding yes/no investment authorization is not: a threshold indicator can switch abruptly even when its inputs change smoothly. This distinction explains why a comparison at a fixed institutional mandate can look quite different from one conducted after the mandate is relaxed. We therefore decompose the threshold and its binary authorization consequence separately.

For each decision epoch t define $a_{N,t} = a_h(\widehat{Q}_t^N)$, $a_{E,t} = a_h(\widehat{Q}_t^E)$ and $q_{N,t}$, $q_{E,t}$ from their *own* prequential calibration histories. The four swapped *critical budget levels* and their empty-set indicators are

$$T_{ij,t} = a_{i,t} + q_{j,t}, \qquad I_{ij,t}(B) = \mathbf{1}\{T_{ij,t} > B\}, \qquad (i,j) \in \{N,E\}^2. \tag{28}$$

For the time-averaged indicator define $f_{ij}(B) = n^{-1}\sum_{t=1}^{n} I_{ij,t}(B)$, with $n = 437$. This *risk-budget feasibility profile* counts how often the observed critical thresholds exclude all actions as the mandate varies; the conventional efficient frontier instead compares achievable returns and risks of selected portfolios.

Lemma 5 (Additivity of critical budgets and threshold-induced interaction)**.** *For every epoch t,*

$$T_{NN,t} - T_{EE,t} = (a_{N,t} - a_{E,t}) + (q_{N,t} - q_{E,t}), \tag{29}$$

$$T_{NN,t} - T_{NE,t} - T_{EN,t} + T_{EE,t} = 0. \tag{30}$$

By contrast, the binary interaction

$$\mathcal{J}_t(B) = I_{NN,t}(B) - I_{NE,t}(B) - I_{EN,t}(B) + I_{EE,t}(B) \tag{31}$$

need not vanish, because the threshold indicator is nonlinear. The symmetric two-factor attribution

$$\phi_a(B) = \frac{1}{2}\{[f_{NN}(B) - f_{EN}(B)] + [f_{NE}(B) - f_{EE}(B)]\}, \tag{32}$$

$$\phi_q(B) = \frac{1}{2}\{[f_{NN}(B) - f_{NE}(B)] + [f_{EN}(B) - f_{EE}(B)]\} \tag{33}$$

is exact and satisfies

$$f_{NN}(B) - f_{EE}(B) = \phi_a(B) + \phi_q(B)\,. \tag{34}$$

Each $f_{ij}(B)$ is a nonincreasing, right-continuous step function of the budget. Neither the difference $f_{NN} - f_{EE}$ nor its two attributions is necessarily monotone in B.

Proof. Substitution of $T_{ij,t} = a_{i,t} + q_{j,t}$ proves both critical-budget identities. Nonlinearity of $x \mapsto \mathbf{1}\{x > B\}$ can preserve neither additivity nor a zero interaction. Adding the definitions of ϕ_a and ϕ_q cancels both cross terms, yielding Equation (34). Finally, as B increases a fixed strict-threshold indicator can only change from one to zero, whereas differences of such indicators may change in either direction. □

The names "covariance contribution" and "buffer contribution" denote the symmetric two-order arithmetic allocation defined in Equations (32) and (33). Because each empirical margin is estimated from its own forecast errors, swapping it onto the other forecast identifies the *fixed-input threshold effect* of that margin, while evaluating the swapped pair as a stand-alone trading policy would require recalibration to its new error process. Similarly, the sign of ϕ_a concerns the admission boundary and is distinct from a forecast-accuracy ranking. This separation tells the empirical analysis exactly which question the component exchange answers and which requires matured residual evaluation.

Lemma 6 (Budget-integrated accounting identity)**.** *If all fitted point-risk minima and buffers are nonnegative, then the area under the full five-session budget-feasibility curves is*

$$\int_0^\infty [f_{NN}(B) - f_{EE}(B)]\,\mathrm{d}B = \frac{1}{n}\sum_{t=1}^{n}[(a_{N,t} - a_{E,t}) + (q_{N,t} - q_{E,t})]. \tag{35}$$

Moreover, integrating the two symmetric contributions separately gives

$$\int_0^\infty \phi_a(B)\,\mathrm{d}B = \overline{a_N - a_E}, \qquad \int_0^\infty \phi_q(B)\,\mathrm{d}B = \overline{q_N - q_E}\,. \tag{36}$$

Proof. For every $x \geq 0$, $\int_0^\infty \mathbf{1}\{x > B\}\,\mathrm{d}B = x$. Apply this layer-cake identity separately to the four nonnegative critical budgets $T_{ij,t}$, average over the finite epochs, and use Equations (29), (32) and (33). □

Equations (35) and (36) explain why an attribution measured at one risk mandate can be much larger than at another: the binary effect depends on the placement of critical budgets around that threshold, whereas integrating across all nonnegative mandates recovers the average differences in point-risk and buffer levels. The 10%, 12% and 15% comparisons in Section 5 use three economically interpretable cuts through that exact budget-dependent structure. A full empirical area calculation would use the entire recorded set of critical budgets and would address the contribution averaged over all possible risk mandates rather than the specific institutional choices studied here.

4.4. A Conditional Link Between Calibration Scores and Realized Risk

An exact estimated-feasibility condition is not by itself a statement about the next five trading sessions. We now identify the event on which the calibrated bound actually controls the subsequently observed risky-sleeve measure. This is the logical step needed before interpreting the empirical score and crisis diagnostics.

The scoring rule in Equation (3) is uniform in w by construction. It yields an exact event-wise implication for the risk realized by any authorized grid action; the frequency of the event itself is subsequently measured using matured outcomes.

Lemma 7 (Conditional upper-risk implication from a uniform calibration score)**.** *Let $\rho_t^{\mathrm{real}}(w) = \rho_t(w; Q_t^{\mathrm{real}})$, $\hat{\rho}_t(w) = \rho_t(w; \widehat{Q}_t)$, and S_t be defined by Equation (3). On the event $S_t \le q_t$, the simultaneous bound*

$$\rho_t^{\mathrm{real}}(w) \le \hat{\rho}_t(w) + q_t \quad \text{for every } w \in \mathcal{W}_h \tag{37}$$

holds. Any action $\widehat{w}_t$ selected from a nonempty estimated feasible set therefore obeys $\rho_t^{\mathrm{real}}(\widehat{w}_t) \le B$. In any probability model where such objects are defined,

$$\Pr\big(\mathcal{F}_t \neq \emptyset,\ \rho_t^{\mathrm{real}}(\widehat{w}_t) > B\big) \le \Pr(S_t > q_t), \tag{38}$$

where $\widehat{w}_t \in \mathcal{F}_t$ whenever $\mathcal{F}_t \neq \emptyset$. The implication provides no numerical upper bound without a separately justified bound on $\Pr(S_t > q_t)$.

Proof. By the definition of the maximum positive residual, $\rho_t^{\mathrm{real}}(w) - \hat{\rho}_t(w) \le S_t \le q_t$ simultaneously for all grid weights on the stated event. If $\widehat{w}_t$ is feasible, $\hat{\rho}_t(\widehat{w}_t) + q_t \le B$, so the realized quadratic-variation risk is also at most B. Thus the joint event on the left of Equation (38) is a subset of $\{S_t > q_t\}$. □

The conditional protection result applies when the controller selects a weight satisfying the fitted risk constraint. On an empty-set date, its minimum-estimated-risk fallback instead has a positive fitted constraint violation and must be reported as an exception. Moreover, the historical rolling 90th-percentile score is a calibration design target under possible serial dependence; Section 5.2 evaluates its observed exceedance frequency and stress-period concentration before interpreting the protection delivered to authorized actions.

4.5. Cash Exposure as an Exact Authorization Repair

The preceding results show how to diagnose an empty fully invested set and how far the risk budget lies from admitting an allocation. Yet a manager may be prohibited from raising the mandate while still being authorized to hold cash. We therefore ask a more operational question: *what minimum permitted cash capacity reopens the set at the unchanged risk budget?*

The fully invested two-ETF constraint can be restrictive for an institutional investor whose mandate also permits cash. To determine how much of the investment domain must

change before a trade becomes available, let $s \in [0,1]$ be the wealth fraction invested in the SPY/SHY sleeve, $u \in \mathcal{W}_h$ its internal SPY weight, and $c \in [0,1]$ the maximum permitted cash fraction, so that $s \geq 1-c$. A cash instrument with zero risk in the model's five-session risky-sleeve metric yields the exposure-homogeneous upper-risk function

$$U_t(s,u) = s\{\rho_t(u;\widehat{Q}_t) + q_t\}, \qquad 1-c \leq s \leq 1. \tag{39}$$

The scaling applies to *both* the forecast and its buffer: applying the unchanged positive q_t to a zero-risk all-cash position would charge risk even when the modeled risky exposure is zero.

Lemma 8 (Exact minimum cash allowance for estimated feasibility)**.** *Let* $a_h(\widehat{Q}_t) = \min_{u\in\mathcal{W}_h} \rho_t(u;\widehat{Q}_t)$ *and* $T_t = q_t + a_h(\widehat{Q}_t) \geq 0$*. Under the risk definition in Equation* (39)*, a long-only cash-inclusive allocation satisfying* $U_t(s,u) \leq B$ *exists if and only if*

$$(1-c)T_t \leq B. \tag{40}$$

The smallest cash-allowance cap sufficient for a feasible allocation is

$$c_t^{\min} = \begin{cases} \left[1 - \dfrac{B}{T_t}\right]_+, & T_t > 0, \\ 0, & T_t = 0. \end{cases} \tag{41}$$

For $c = 0$ *this criterion reduces to the fully invested grid test, while* $c = 1$ *ensures existence of an estimated feasible zero-risk all-cash position.*

Proof. For each fixed $s \geq 0$, the minimum of $U_t(s,u)$ over $u \in \mathcal{W}_h$ is sT_t. Because $T_t \geq 0$, the minimum over permissible $s \in [1-c,1]$ is $(1-c)T_t$. A feasible portfolio therefore exists precisely when this quantity does not exceed B. Solving that inequality for $c \in [0,1]$ proves the threshold and its zero-denominator convention. □

Corollary 2 (One cash cap for a fixed evaluation calendar)**.** *For a finite set of decision dates* $\mathcal{T}$ *with fixed forecasts, buffers and mandate B, the smallest common cash cap ensuring at least one estimated-feasible portfolio on every date is*

$$c_{\mathcal{T}}^{\min} = \max_{t\in\mathcal{T}} c_t^{\min}.$$

On dates with $T_t > B > 0$*, the local minimum satisfies* $\partial c_t^{\min}/\partial q_t = B/T_t^2 > 0$ *and* $\partial c_t^{\min}/\partial B = -1/T_t < 0$*, with* $a_h(\widehat{Q}_t)$ *held fixed. Thus larger fitted safety margins require more permitted cash, while a larger mandate lowers the minimum required allowance.*

Proof. A common cap c is feasible on every date precisely when $c \geq c_t^{\min}$ for all $t \in \mathcal{T}$ by Lemma 8; taking their maximum proves the first claim. For $T_t > B > 0$, $c_t^{\min} = 1 - B/T_t$, so ordinary differentiation gives the stated sensitivities. □

The threshold $c_t^{\min}$ determines the smallest cash *capacity* that the mandate must permit for an estimated-feasible allocation to exist. The portfolio controller subsequently chooses an allocation within that enlarged feasible set using its tracking-error and transaction-cost preferences. Thus the mathematical result addresses the *authorization* decision, while the realized portfolio path addresses the *allocation and valuation* decisions. The zero-risk assignment to cash in Equation (39) refers specifically to the SPY/SHY sleeve's quadratic-variation contribution; dated cash carry is included in whole-portfolio net returns. Sections 5.8 and 5.9 assess both the calibrated risk bounds and the economic outcomes of the resulting choices.

The mathematical results now form a decision sequence: the critical threshold identifies the authorization state; the minimum-repair and stability formulas quantify its sensitivity; the decomposition locates the forecast and buffer contributions; the calibration implication specifies the event required for realized protection; and the cash threshold identifies a permitted alternative to relaxing the risk mandate. Section 5 evaluates those outputs using the common historical calendar and follows the selected investments into realized risk and economic value. The common cash-cap requirement in Corollary 2 is computed over a specified finite date set and can be updated as newly fitted critical budgets become available.

5. Empirical Results

A portfolio manager must answer three questions before treating an estimated risk limit as an effective investment rule: which fitted component closes the admissible set, how dependable is that component, and what happens to investment outcomes when the manager changes the admissible actions? We answer these questions in sequence using the same 437 dated decisions. The mathematical threshold from Section 4 provides the common measurement unit; the realized five-session observations then assess the risk protection and economic consequences of each choice. Figure 1 shows how the model-matched critical annual mandate evolves over time. Its two panels cover 2018–2021 and 2022–2026, respectively, while all 437 decisions, including the three from December 2017, enter every statistical comparison.

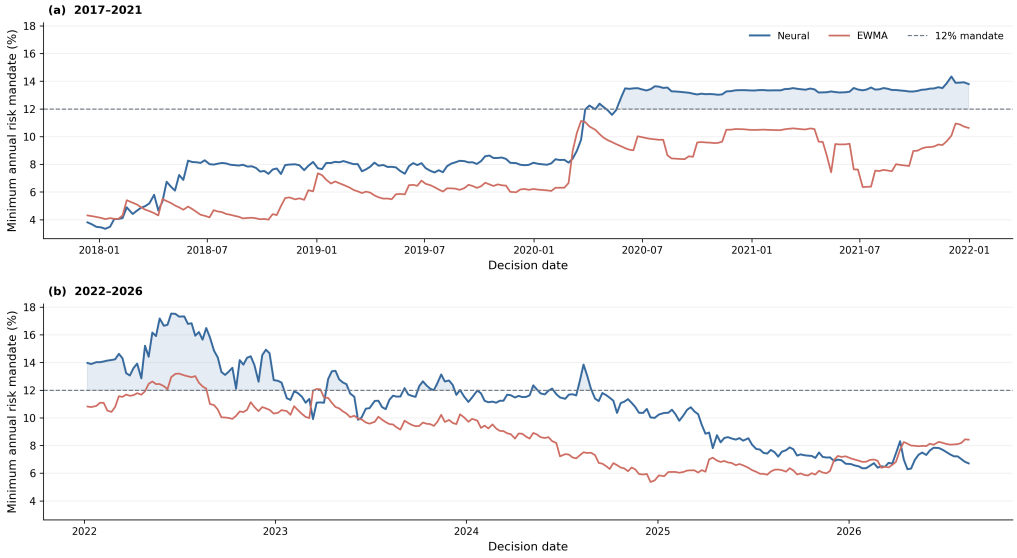


Figure 1. Annual evolution of the critical risk mandate needed to admit at least one SPY/SHY allocation. Panels (**a**,**b**) cover 2018–2026 on common scales and compare matched neural and exponentially weighted moving average (EWMA) forecasts and buffers against the 12% reference mandate. All tabulations also include the three December 2017 decisions.

5.1. Which Component Closes the Investment Set?

The first result concerns the investment decision itself. Let $a_{i,t}$ be the minimum fitted five-session point risk over the 201 candidate SPY/SHY weights, and $q_{j,t}$ the empirical margin from model j. Their sum $T_{ij,t} = a_{i,t} + q_{j,t}$ is the smallest budget that admits a candidate allocation. Hence $I_{ij,t}(B) = \mathbf{1}\{T_{ij,t} > B\}$ records a rejected investment set when it equals one. When it equals zero, at least one weight satisfies the fitted constraint $\rho_t(w; \widehat{Q}_t^i) + q_{j,t} \leq B$ and can be passed to the portfolio optimizer. Whether that chosen weight remains within the risk budget after the next five trading sessions is then measured directly in Sections 5.2 and 5.9.

This distinction separates the authorization decision from the performance assessment rather than leaving the meaning of a zero indicator implicit.

At the principal 12% annual mandate, the neural forecast paired with its own empirical margin excludes all candidate allocations on 161 of 437 dates (36.84%); the matched EWMA combination does so on 18 dates (4.12%). The contrast becomes more informative when the fitted components are exchanged without changing the date, risk limit, or point-risk grid. Substituting the EWMA margin for the neural margin reduces the neural count to nine, whereas applying the neural margin to EWMA point risk produces 165 excluded sets. Both models have a nonempty estimated set on all 437 dates when the additive margin is removed. The difference is therefore concentrated in how the margin enters the threshold, rather than in point-risk minima crossing the mandate on their own.

Table 1 summarizes the estimated investment-set exclusions and signed component accounting across the three annual risk mandates for the 437 common decisions.

Table 1. Estimated investment-set exclusions and signed component accounting at three annual risk mandates (437 common five-session decisions).

Panel A. Empty estimated feasible-set counts and rates

Mandate	B_5	n	NN	NE	EN	EE	NN (%)	EE (%)
10%	0.014086	437	247	85	241	105	56.52	24.03
12%	0.016903	437	161	9	165	18	36.84	4.12
15%	0.021129	437	17	0	20	0	3.89	0.00

Panel B. Signed symmetric two-factor decomposition and individual budget components

Mandate	Gap (pp)	Point (pp)	Buffer (pp)	$q_N \geq B$	$q_E \geq B$	$a_N > B$	$a_E > B$
10%	+32.49	−1.60	+34.10	153	34	0	0
12%	+32.72	−1.49	+34.21	126	0	0	0
15%	+3.89	−0.34	+4.23	14	0	0	0

Notes: NN, NE, EN, and EE denote the neural/EWMA point-risk–buffer pairings, in that order. a_N, a_E are minimum point risks; q_N, q_E are the fitted empirical margins. The matched gap and the signed point and buffer terms are computed by Equations (32) and (33). The $q \geq B$ columns count equality as well as exceedance; the exact empty-set rule is $a + q > B$. The same 437 dates are used at each budget, yielding 1311 verified budget–date classifications.

Figures 2 and 3 show the mechanism both through time and in the two-component risk plane. The neural-margin configurations repeatedly place $a + q$ above the 12% boundary, whereas the configurations using the EWMA margin cross it much less often. The exact symmetric decomposition at this mandate is

$$\phi_Q = -0.01487, \qquad \phi_q = +0.34211, \qquad \phi_Q + \phi_q = 0.32723.$$

In percentage-point terms, the buffer adds 34.21 points to the neural–EWMA gap and the point-risk component offsets 1.49 points, giving the observed 32.72-point difference. The signed contributions can exceed the net gap because one component offsets another. The neural margin reaches or exceeds the entire five-session risk budget on 126 dates; whether a date is actually excluded is still determined by the exact $a + q > B$ threshold, including the boundary case $q = B$.

These substitutions are designed to answer a specific accounting question: *which fitted input moves the authorization threshold across the mandate when the other input is held constant?* They isolate that threshold mechanism directly. A cross-paired margin was originally calibrated to a different forecast-error process, however; choosing such a pair for live trading would require assessing its own subsequent residuals. We therefore turn from the size of the margin to the protection it delivers.

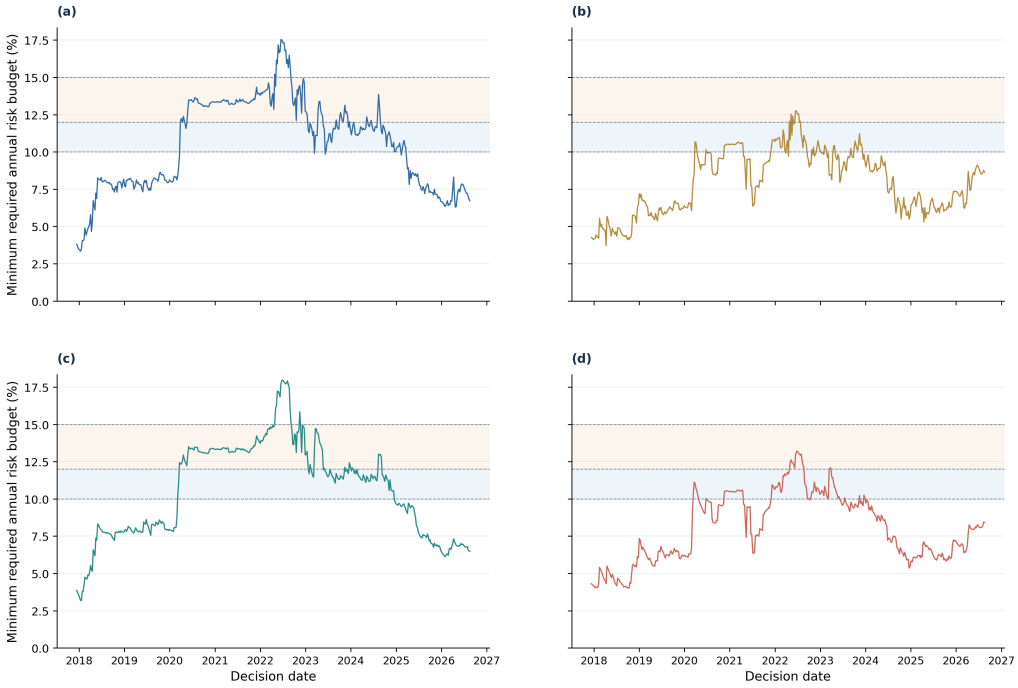


Figure 2. Critical annual risk mandates for the four fitted point-risk–margin pairings: (**a**) neural/neural, (**b**) neural/EWMA, (**c**) EWMA/neural, and (**d**) EWMA/EWMA. The blue, gold, teal and red curves identify the neural/neural, neural/EWMA, EWMA/neural and EWMA/EWMA combinations, respectively. Horizontal dashed lines mark hypothetical 10%, 12% and 15% annual risk mandates; the lightly shaded bands distinguish the mandate intervals. These references show where a fitted component exchange changes estimated admissibility.

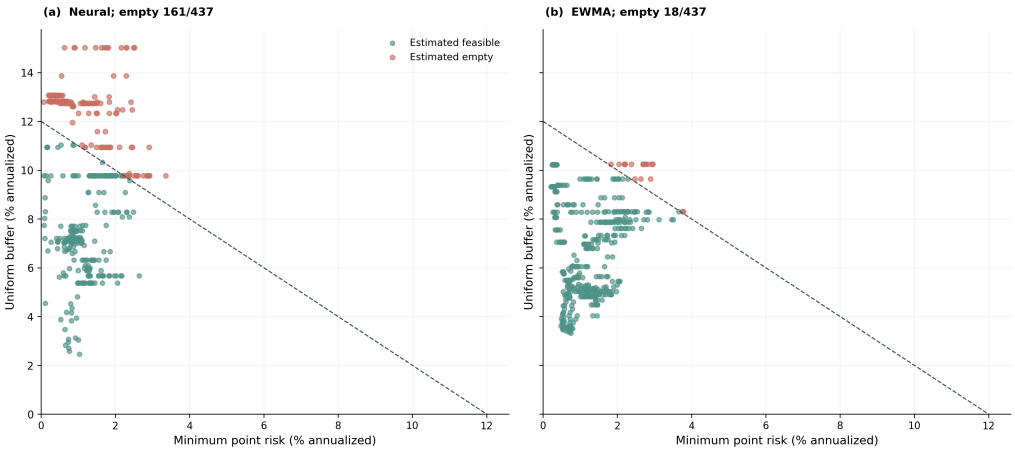


Figure 3. Joint distributions of minimum fitted point risk and empirical calibration margin for the matched neural and EWMA specifications. The descending dark dashed line is the 12% annual risk-budget boundary: points above it correspond to empty estimated feasible sets. Red and teal dots indicate estimated infeasible and feasible decisions, respectively. The horizontal and vertical axes report the fitted minimum point risk and calibration margin, respectively.

5.2. Is the Empirical Margin Informative About Subsequent Risk?

A buffer is economically useful only if its effect on authorized choices can be related to the risk it is intended to cover. The margin in Equation (3) protects the maximum positive five-session prediction residual over all 201 candidate weights. For a weight chosen from the estimated feasible set, the score event $S_t \leq q_t$ implies that its realized quadratic-variation risk stays within B, as shown in Lemma 7. The relevant empirical question is therefore

how often this upper-risk event fails after the portfolio decision, and whether those failures cluster precisely when protection is needed.

The full-grid score exceeds its matched margin in 49 of 437 neural decisions (11.21%) and 52 EWMA decisions (11.90%). Against the 10% nominal upper-tail design target, these whole-period frequencies provide an initial calibration diagnostic. Supplementary Table S19 separately reports the size and calendar clustering of realized upper-risk-bound breaches for the actually selected reference policies, complementing the full-grid score frequencies shown here. More revealing are the market-state contrasts in Table 2: on dates with lagged VIX at least 30, neural score exceedances occur in 10 of 29 decisions (34.48%); in the February–June 2020 stress window they occur in 11 of 21 decisions (52.38%). The corresponding EWMA counts are 3 of 29 and 5 of 21. An authorization rule can therefore exclude many portfolios while its empirical margin still leaves unusually large future residuals uncovered in turbulent episodes. The table makes both effects measurable instead of interpreting conservatism as protection by itself.

Table 2. Upper-risk calibration outcomes for the protected portfolio weights and stressed market subsets.

Calibration Target and Evaluation Subset	Decisions	Neural Exceed	EWMA Exceed
Uniform buffer; entire 201-weight grid	437	49 (11.21%)	52 (11.90%)
Uniform buffer; lagged VIX $\geq$ 30	29	10 (34.48%)	3 (10.34%)
Uniform buffer; February–June 2020	21	11 (52.38%)	5 (23.81%)
Min-risk buffer; chosen weight only	437	59 (13.50%)	53 (12.13%)
Min-risk buffer; all 201 weights	437	172 (39.36%)	173 (39.59%)

Notes: Exceedances compare matured positive realized-minus-predicted five-session risk residuals with the corresponding historical empirical margin. The first three rows use the *uniform* grid-wide score and are overlapping subsets of the same calendar. The last two rows instead apply a margin fitted only to the contemporaneously predicted minimum-risk action and evaluate it on the stated target. Both constructions use a nominal upper-tail rate of 10%; their protected action sets differ, so the columns should be interpreted within each row's target.

The action set is central to this comparison. A margin calculated solely for the model's minimum-risk weight is much smaller and can reopen the estimated set for that restricted policy, but it does not provide comparable protection across all 201 allocations: grid-wide residual exceedances approach 40% under that narrower margin. Conversely, restricting actual trades to the fixed 80% SPY sleeve is not a free gain in feasibility. With no cash and the 12% mandate, the reference-80 action with its own margin is estimated infeasible on 394 neural and 392 EWMA dates. Because that restricted action set contains only the 80% risky-sleeve split, the simulated continuation retains that action even on those dates and records it as a mandate exception. Using each model's predicted minimum-risk action and its corresponding restricted margin removes these estimated empty sets, but it also produces much lower average SPY exposure (2.81% neural; 2.39% EWMA) and 5-basis-point, zero-cash-carry annual certainty equivalents of approximately 2.20% and 2.28%, respectively. These two policies protect and select different actions; the appropriate comparison includes the permitted action, residual reliability, and investor objective together. Supplementary Table S16 summarizes the restricted-policy contrast; the full 72-policy numerical grid remains in the archived restricted-action calibration comparison.

The implication is operational: a tighter calibration target can reduce the cost of uncertainty protection *only when the eligible trading actions are narrowed consistently*. This provides a reason to redesign an action set or its calibration procedure, rather than to reduce the margin across an unchanged investment universe without assessing the resulting protection.

Calibration-window sensitivity of the fitted margins and corresponding full-grid score-exceedance tails is reported in Supplementary Table S20. The matured residual-score distributions for the uniform, fixed-80% SPY, and predicted minimum-risk targets appear in

Supplementary Table S21. These companion diagnostics identify the errors being protected, the margins fitted to those errors, and the remaining exceedances.

5.3. Does the Authorization Diagnosis Survive Numerical and Model Variation?

Because feasibility changes discontinuously at $a + q = B$, small changes in fitting can affect the investment decision even when forecast summaries appear similar. The initial neural model stopped at its 140-iteration limit in all 37 fits and yielded 131 empty sets. Raising only the numerical limit to 1000 removed those warnings; the fits ended between 204 and 635 iterations, and the number of empty sets increased to 161. The 30-count increase is a net change in exclusions under the same mandate, demonstrating that the authorization frequency depends on completing the numerical fit; individual date-by-date switches are assessed separately.

Convergence is only the first reliability check. Across ten neural initialization seeds, the excluded-set count ranges from 74 to 245; feasibility status differs across the ten seed runs on 208 of the 437 dates. The fitted upper-risk surface is therefore a material input to the manager's reported opportunity set. We further vary neural architecture, regularization and training history, the number of matured calibration scores (35, 63, 126, or 252), EWMA decay, and sample or shrinkage covariance estimation. In particular, the less-regularized neural configuration generates warnings in 13 of 37 fits, which remain flagged in the robustness audit. The positive buffer term persists in the ten-seed decomposition, but the number and timing of excluded decisions vary. The practical message is to report both the mathematical decision boundary and the fitting sensitivity of the quantities entering it, rather than to treat one successful numerical run as sufficient evidence of a stable authorization rule.

5.4. How Sensitive Is Authorization to the Covariance Estimator and Calibration Geometry?

The fixed neural–EWMA contrast is informative only for its stated uniform buffer. We therefore apply the same 437-date chronology to two rolling shrinkage forecasts and to the normalized simultaneous envelope from Section 3.7. Table 3 separates the estimator and calibration choices. With the original common margin, Ledoit–Wolf and OAS result in 117 and 99 estimated empty sets, respectively, between the neural count of 161 and the EWMA count of 18. With the normalized simultaneous rule, the count is zero for all four fitted families. Hence authorization is substantially affected by the geometry of the calibration charge, rather than being an intrinsic ranking of neural and EWMA predictive accuracy.

Table 3. Retrospective covariance-estimator and calibration-rule sensitivity on 437 common decisions (12% annual mandate, 126 matured calibration scores).

Covariance Family	Simultaneous Buffer	Empty Sets, n (%)	Full-Grid Bound Exceedance (%)
Neural	Uniform additive	161 (36.84)	11.21
Neural	Risk-scaled	0 (0.00)	10.98
EWMA	Uniform additive	18 (4.12)	11.90
EWMA	Risk-scaled	0 (0.00)	10.76
Ledoit–Wolf	Uniform additive	117 (26.77)	11.21
Ledoit–Wolf	Risk-scaled	0 (0.00)	10.98
OAS	Uniform additive	99 (22.65)	11.44
OAS	Risk-scaled	0 (0.00)	12.13

Notes: The scaled rule uses $h_t(w) = \max\{\hat{\rho}_t(w), 0.002\}/0.01$ and a single 90th-percentile rolling simultaneous normalized score; it does not independently calibrate each weight. Full-grid exceedance means that the realized risk exceeds its estimated bound at one or more of the 201 weights. These retrospective frequencies are not finite-sample or conditional coverage guarantees. Source: author-executed retrospective covariance and calibration sensitivity analysis (8 November 2026).

A zero empty-set count is not synonymous with successful risk protection. The eight full-grid exceedance frequencies range from approximately 10.76% to 12.13%; their magnitudes, stress localization, and potential use in an actual prospective policy remain distinct research questions. Likewise, the comparisons use the previously explored calendar rather than an untouched external evaluation sample.

5.5. How Much of the Gap Depends on the Calibration-History Length?

The historical 126-score reference window is a design choice, not an estimated structural constant. Table 4 changes only the number of matured scores in the original uniform rule under the 12% mandate. Neural exclusions are nonmonotone (129–161), while EWMA exclusions decline over the four illustrated windows (42 to 2). Thus the gap survives these particular choices in sign but not in magnitude. Together with the retrospective omission diagnostics in Supplementary Table S22, these patterns support a score-history sensitivity interpretation; they do not identify a pandemic causal effect.

Table 4. Calibration-history sensitivity of original uniform risk margins at the 12% mandate (437 common decisions).

Matured Score Window L	Neural Empty Sets	EWMA Empty Sets	Neural Minus EWMA
35	132	42	90
63	129	40	89
126 (reference)	161	18	143
252	130	2	128

Notes: Window lengths change the empirical 90th percentile of fully matured score histories; the investment calendar and neural/EWMA point forecasts remain fixed. Counts represent estimated empty sets, not realized risk breaches.

5.6. Do DCC–GARCH and Daily-Return HAR-Type Comparators Change the Reference Contrast?

Table 5 extends the classical benchmark set with two dynamic families fitted under the dated dynamic-benchmark protocol. On the same 437 previously explored dates and under the 12% mandate with 126 fully matured calibration scores, uniform calibration yields 8 empty sets for DCC–GARCH and 69 for the daily-return HAR-type estimator. Both show zero estimated exclusions with the normalized simultaneous scaled rule. Nonetheless their respective scaled upper-bound exceedance frequencies are 12.36% and 12.13%, versus 11.90% and 12.13% under uniform calibration. Thus neither the neural–EWMA contrast nor the absence of estimated exclusions establishes comparative forecasting accuracy or prospective risk coverage. The DCC–GARCH uniform count varies from 96 to 91 to 8 to 6 as the calibration window increases from 35 through 63 and 126 to 252; the analogous HAR-type counts are 56, 62, 69, and 20 (Supplementary Table S28). In the regime diagnostic, the eight DCC and 69 HAR-type uniform exclusions all lie within 2020–2023. Those episode-specific findings are consistent with calibration-history sensitivity and cannot be described as the causal effect of a particular market event.

Table 5. Retrospective DCC–GARCH and daily-return HAR-type comparison (12% annual mandate, 126 matured-score window, 437 evaluated decisions).

Estimator	Calibration	Estimated Empty n	Full-Grid Bound Exceed (%)
DCC–GARCH	Uniform	8	11.90
DCC–GARCH	Risk-scaled	0	12.36
Daily-return HAR-type	Uniform	69	12.13
Daily-return HAR-type	Risk-scaled	0	12.13

5.7. How Do Risk Mandates and Market Periods Change the Diagnosis?

The critical-budget characterization also makes it possible to test how the *same fitted risk surfaces* interact with different institutional mandates. Holding all point-risk minima and margins fixed, we evaluate annualized limits of 10%, 12%, and 15%, corresponding to five-session budgets of 0.014086, 0.016903, and 0.021129. Table 1 shows that the matched neural–EWMA exclusion gap is 32.49 percentage points at 10%, 32.72 at 12%, and 3.89 at 15%. The buffer contributions are 34.10, 34.21, and 4.23 points. A change in the mandate can thus strongly alter the economic importance of the same fitted margin because individual dates cross their critical-budget thresholds, precisely the nonlinear indicator mechanism described by Lemma 5.

Figure 4 connects these counts to a wider range of hypothetical mandates. Panels (a) and (b) reveal where the fitted critical budgets lie relative to each institutional threshold; panel (d) reports the minimum budget repair when the limit is binding. Panel (c) and Table 6 show how the paired result varies when observed calendar-month blocks are resampled. At 10% and 12%, the empirical 95% month-block intervals for the buffer contribution remain positive, whereas the corresponding point-risk intervals straddle zero. The intervals become markedly smaller in level at the 15% mandate, reflecting the smaller number of dates near the loosened threshold. These paired intervals quantify how calendar-month composition affects the fitted mandate comparisons within the common historical sample.

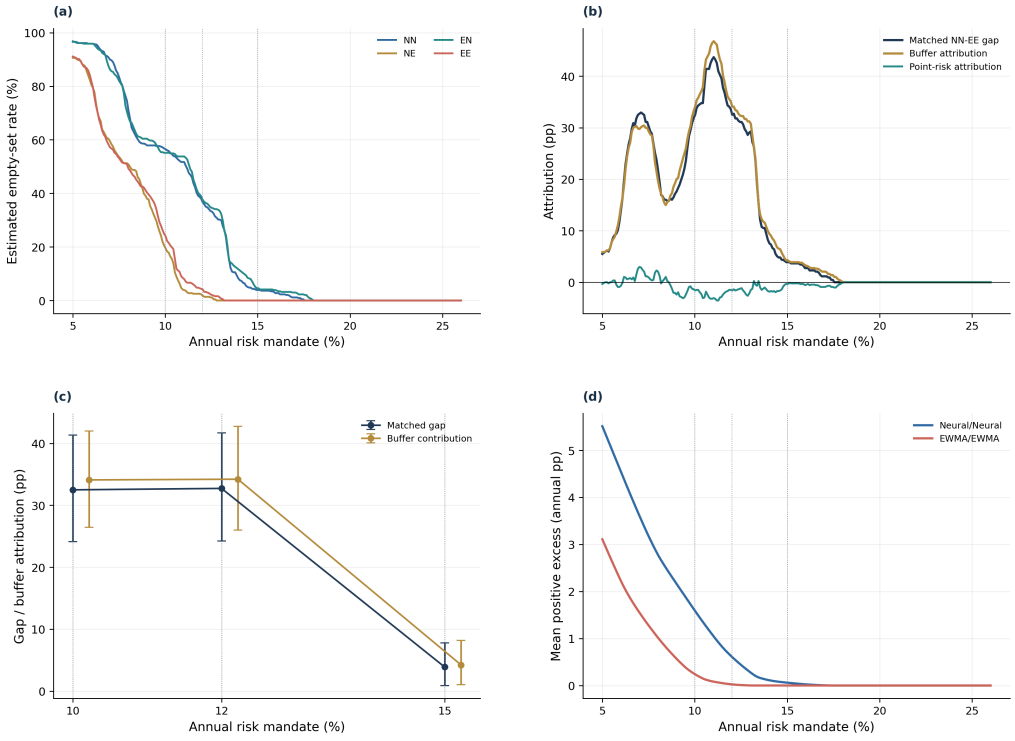


Figure 4. Risk-mandate sensitivity of the fixed historical decision surfaces: (**a**) the four empty-set profiles over annual risk budgets; (**b**) signed point-risk and buffer contributions; (**c**) paired month-block resampling intervals at 10%, 12%, and 15%; and (**d**) the additional annual risk budget needed to reopen the estimated feasible set. In panel (**b**), the dark line is the total matched neural-minus-EWMA exclusion-rate gap, the gold line is the buffer attribution, and the teal line is the point-risk attribution; the horizontal black line marks zero attribution. The dotted vertical reference lines in panels (**a**–**d**) mark the 10%, 12%, and 15% annual mandates. All panels reuse the same 437 fitted decision dates under different thresholds.

Table 6. Month-block sensitivity of the feasibility gap and calendar concentration at the primary 12% risk mandate.

Panel A. Paired historical month-block intervals (percentage points)				
Mandate	Model gap	Gap 95% interval	Point-risk 95% interval	Buffer 95% interval
10%	32.49	[24.14, 41.32]	[−4.56, 1.37]	[26.45, 41.97]
12%	**32.72**	[24.25, 41.71]	[−3.85, 0.58]	[25.97, 42.76]
15%	3.89	[0.91, 7.80]	[−0.79, 0.00]	[1.03, 8.18]
Panel B. Concentration of empty feasible sets under the primary 12% mandate				
Period	Decisions	Neural (NN), *n* (%)	EWMA (EE), *n* (%)	Gap (pp)
2017–2019	104	0 (0.0)	0 (0.0)	0.00
2020	50	34 (68.0)	0 (0.0)	68.00
2021	51	**51 (100.0)**	0 (0.0)	100.00
2022	50	**50 (100.0)**	16 (32.0)	68.00
2023	50	19 (38.0)	2 (4.0)	34.00
2024	50	7 (14.0)	0 (0.0)	14.00
2025–2026	82	0 (0.0)	0 (0.0)	0.00
Total	**437**	**161 (36.84)**	**18 (4.12)**	**32.72**

Notes: NN combines the neural point-risk estimate with its own buffer; EE combines EWMA point risk with its own buffer. The gap equals the NN minus EE exclusion rate in percentage points. Panel A uses the same 105 observed calendar months and 2000 paired resamples per mandate. Panel B groups years with no exclusions for either model at the primary 12% mandate to display the periods driving the gap; these zero counts are observed outcomes, not missing values. Unaggregated year-by-year results for both models at all three mandates (2622 model–mandate–date classifications) are retained in Supplementary Table S8. The 10% and 15% comparisons change the threshold while holding fitted forecasts and buffers fixed.

The timing of the constraint matters for implementation. At the 12% mandate, the neural matched pair has an empty set on all 51 decisions in 2021 and all 50 in 2022, compared with zero and 16 EWMA dates, respectively (the remaining two EWMA exclusions occur in 2023). This is a sustained absence of estimated admissible fully invested SPY/SHY allocations, not merely a large average risk statistic. Under 10%, the neural exclusions extend into 2025; at 15%, its remaining 17 exclusions all occur in 2022. Table 6 emphasizes the primary 12% decision rule; Supplementary Table S8 retains the complete annual counts for 10%, 12%, and 15%.

Calibration-History Concentration

The year counts become more informative when the required budget is separated into the fitted minimum point risk a_t and the empirical margin q_t. At the 12% mandate, the neural margin by itself reaches or exceeds the five-session budget on all 51 evaluated dates in 2021 and on 45 of 50 dates in 2022. Its median margin-to-budget ratio remains 106.85% and 106.53% in those years, compared with median point-risk-to-budget ratios of 4.22% and 12.61%. Thus, a high historical calibration margin, rather than the point-risk minimum alone, determines the investment-authorization boundary through much of this period.

A dated calibration-score memory analysis reproduces the original rolling 90th-percentile margins from fully matured prior scores, then holds each fitted point-risk estimate and original last-126-score window fixed while retrospectively omitting scores dated February–June 2020. Under that diagnostic the neural empty-set count changes from 51 to zero in 2021 and from 50 to four in 2022. The same audit finds a second episode: omitting 2022-dated scores changes neural exclusions from 19 to zero in 2023 and from seven to zero in 2024. These within-window sensitivity results identify how large dated forecast errors persist in the calibration record and shift subsequent investment-authorization decisions. They measure the role of historical score composition, not the causal impact of a macroeconomic event or the performance of a deployable alternative policy. Supplementary

Table S22 provides the two-model threshold decomposition, omission comparisons and corresponding selection rule. The evidence strengthens the operational case for monitoring the history that drives q_t when evaluating cash-capacity repair.

Having identified both the budget boundary and its persistence in the historical calibration record, we next assess how much cash capacity can reopen the investment set without changing the mandate.

Figure 5 combines the six fitted estimator families with a common calibration comparison. Although the scaled envelope yields zero estimated empty sets on these historical dates, the empirical any-weight exceedance rates remain nonzero for every family. The two outcomes must be read together rather than treating an increase in authorization as evidence of improved coverage.

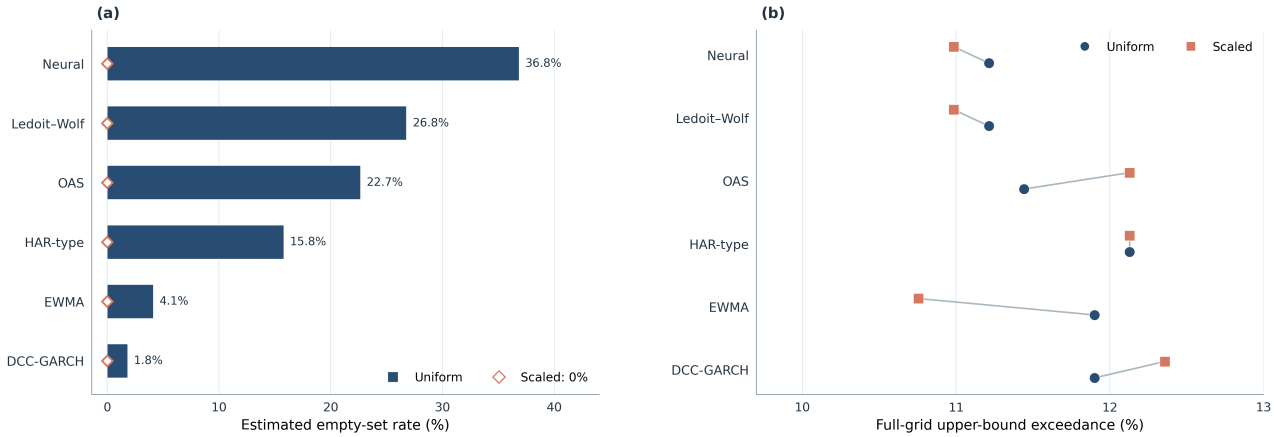


Figure 5. Cross-model authorization and historical upper-risk-bound diagnostics on the same 437 decisions (12% annual mandate; 126 matured calibration scores). (**a**) Estimated empty-set counts under the original uniform additive margin; all six risk-scaled simultaneous alternatives have zero estimated empty sets. (**b**) Empirical any-weight, 201-grid upper-bound exceedance rates under uniform and scaled margins. A lower exclusion count does not certify improved protection or prospective coverage. DCC–GARCH and daily-return HAR-type are historical comparison estimators.

5.8. What Cash Capacity Restores Admissible Investments?

The results so far identify a rejection threshold but also show why increasing the formal risk budget is not an automatic response. Lemma 8 instead holds the ETF-sleeve mandate fixed and expands the permissible asset domain to include cash. Under the exposure-homogeneous risk formulation, the minimum cash fraction is computed exactly for each fitted date. Taking the largest value over the 437 dates gives a common cash-cap requirement of 31.54% for the neural reference policy and 9.06% for EWMA at the 12% mandate. The required maxima are 42.95% and 24.22% at 10%, and 14.42% and 0% at 15%. A 50% allowance is consequently sufficient to restore a nonempty estimated feasible set on every date for both fixed reference models.

The required cash cap is specific to the fitted covariance and margin sequence, the risk definition and the chosen mandate. Under the same 12% mandate and 50% cash cap, 40 of 100 alternative model–calibration specifications still have at least one infeasible date (32/68 neural variants and 8/32 EWMA or alternative covariance variants). The comparison reveals what an implementable policy needs: a cash cap sized to the particular risk–margin process, accompanied by a defined response when the cap is insufficient. The mathematical scaling is also essential. Charging the original additive margin even when the risky-sleeve allocation approaches zero would leave 126 neural dates falsely excluded under a full-cash

option; scaling both forecast risk and margin by actual risky exposure restores the correct zero-risk limit for that modeled sleeve.

5.9. What Is the Economic Value and Realized Risk of Repair?

Expanding the feasible set changes the controller's opportunity set; its economic value is assessed through the investments actually selected. The zero-cash references include minimum-estimated-risk fallback allocations on 161 neural and 18 EWMA dates, whereas the matched 50%-cash reference paths have no estimated empty sets. We therefore report those mandate exceptions alongside certainty equivalent (CE), rather than treat the zero-cash references as continuously compliant portfolios. We retain the original selection rule and five-basis-point turnover-cost convention, apply the lagged historical three-month Treasury yield as a *cash-carry proxy*, and follow the selected holdings over the same 2185 daily return sessions. Holding these positions fixed when replacing the zero-carry assumption by historical carry separates the effect of a wider investment domain from the additional cash-return credit.

The CE differences in Table 7 can be separated into two observed changes under the stated controller and fallback convention. First, with zero assumed cash carry, increasing the cash cap from zero to 50% raises annual CE by 1.47 percentage points for neural control and 1.08 points for EWMA as the admissible domain and chosen positions change. Second, crediting those *same cash-inclusive positions* with the lagged historical Treasury-yield proxy adds 0.75 and 0.57 points, respectively. The resulting annual CEs rise from 3.77% to 5.98% for neural and from 4.37% to 6.03% for EWMA. Because the cash-cap increase also changes mean SPY exposure (23.15% to 43.41% and 37.89% to 49.59%), these CE gains describe more active investment opportunities rather than the yield on idle cash alone.

The original cash-cap month-block resampling of the historical cash-cap comparison reported neural and EWMA annual certainty-equivalent changes of approximately 2.21 and 1.65 percentage points, with descriptive 95% intervals $[-0.54, 4.86]$ and $[0.17, 3.13]$. An additional block-resampling analysis on the same dated historical policies, reported in Supplementary Table S26, gives descriptive intervals $[-0.702, 4.649]$ and $[0.221, 3.270]$. Both sets are retrospective resampling summaries of the same underlying history, not independent confirmations. Their interval endpoints differ, but both neural intervals include zero. Neither a stable welfare improvement nor a robust neural–EWMA economic ranking is established.

Table 7. Authorization, portfolio exposure and cost-adjusted welfare after permitting cash at the 12% annual ETF-sleeve risk mandate.

Model	Cash Cap	Empty Sets	Mean SPY (%)	Mean Cash (%)	CE (%)	Vol. (%)	Qv Exceed (%)
Neural	0%	161	23.15	0.00	3.77	4.82	2.75
Neural	50%	0	43.41	30.69	5.98	7.20	5.95
EWMA	0%	18	37.89	0.00	4.37	6.45	3.89
EWMA	50%	0	49.59	23.36	6.03	7.85	7.32

Notes: All policies use the same 437 five-session decisions (2185 daily return sessions) and a five-basis-point charge on the specified two-leg turnover. Annual CE uses risk aversion $\gamma = 3$; volatility is the annualized standard deviation of net daily returns. Cash earns a two-ETF-session-lagged historical three-month Treasury-yield-based return proxy. The QV exceedance column counts *realized* five-session ETF-sleeve risk contributions above the specified budget; the empty-set column measures the *predicted* allocation constraint before trading. At zero cash, the 161 neural and 18 EWMA empty-set dates trigger the baseline minimum-estimated-risk fallback, rather than an authorized feasible choice.

The additional opportunity has a measurable risk dimension. With cash capacity raised from zero to 50%, the frequency of realized five-session ETF-sleeve quadratic-variation exceedance rises from 2.75% to 5.95% for neural and from 3.89% to 7.32% for EWMA. During the February–June 2020 window, the cash-inclusive strategies record 7/21 and 5/21 ex-

ceedance dates. These observations evaluate the realized short-horizon ETF-sleeve risk of investments authorized using information available at the decision date; the estimated-feasibility indicator separately records the pre-trade constraint. In contrast, the zero-cash reference policy falls back to a minimum-estimated-risk allocation when the fitted feasible set is empty, and those fallback dates are displayed separately rather than counted as successful mandate compliance. The CE and realized-risk metrics document increased investment opportunities and equity exposure, with a potentially higher historical mean–variance CE at the cost of increased risk participation; neither an all-period welfare gain nor a compliant policy dominance claim follows.

Supplementary Table S17 reports realized risk, mean turnover and welfare at all three risk mandates and both zero- and 50%-cash caps. Tables S18 and S19 summarize robustness by model family and the magnitudes and temporal clustering of selected-action upper-risk breaches. Further cash-cap and 0-, 5- and 10-basis-point transaction-cost sensitivity paths remain available in the archived policy-sensitivity output. Together these results show which properties of the authorization remedy persist across risk budgets, fitted models and historical market conditions.

5.10. Cash Capacity Compared Directly with Mandate Relaxation

The original 12%/50%-cash comparison does not identify a unique economic benefit of cash because the fully invested 12% fallback path has many mandate exceptions. Table 8 therefore compares the cash allowance to a 15% mandate with no cash, under the same historical return and lagged-yield assumptions. Neural annual certainty-equivalent return at risk-aversion $\gamma = 3$ is 5.981% with 12%/50%-cash and 5.446% with 15%/no-cash, a descriptive +0.535 percentage-point contrast; the corresponding EWMA values are 6.026% and 6.060%, a −0.034-point contrast. The supplementary policy analysis computes a direct paired calendar-month comparison of the *two alternatives* using 2000 resamples. Both descriptive 95% intervals include zero (neural [−0.648, 1.762]; EWMA [−0.975, 0.995] percentage points), so neither alternative demonstrates robust economic superiority. The comparison is also repeated for $\gamma = 1, 3, 5, 10$, avoiding reliance on one risk-aversion coefficient.

Table 8. Historical policy sensitivity: cash allowance versus risk-mandate relaxation, uniform 126-score calibration (437 decisions; 5-bp turnover cost).

Model	Mandate/Cash Cap	Empty	CE (%)	QV Exceed (%)	Mean SPY (%)	Mean Cash (%)
Neural	12%/0%	161	3.766	2.746	23.154	0.000
Neural	12%/50%	0	5.981	5.950	43.408	30.687
Neural	15%/0%	17	5.446	4.348	41.586	0.000
Neural	15%/50%	0	7.066	5.950	54.883	20.813
EWMA	12%/0%	18	4.372	3.890	37.893	0.000
EWMA	12%/50%	0	6.026	7.323	49.592	23.364
EWMA	15%/0%	0	6.060	6.407	55.474	0.000
EWMA	15%/50%	0	6.993	7.551	61.481	13.673

Notes: CE is the annualized mean–variance certainty equivalent with $\gamma = 3$, net of the specified costs and using the original lagged three-month Treasury-yield proxy. "QV exceed" measures realized five-session risky-sleeve budget excess, not upper-bound score exceedance; there is no guaranteed whole-portfolio coverage inference. Zero-cash policies may include minimum-risk fallback allocations counted as mandate exceptions. The direct paired CE contrast and sensitivity to γ are reported in the direct paired policy analysis; see Table 9.

The accompanying descriptive month-block 95% intervals for CE changes relative to the 12%/no-cash policy are [−0.702, 4.649] points (neural, 12%/50%-cash), [−1.001, 4.161] (neural, 15%/no-cash), [0.221, 3.270] (EWMA, 12%/50%-cash), and [−0.269, 3.614] (EWMA, 15%/no-cash). In particular, the neural intervals include zero. These are retrospective block-resampling summaries, not evidence that the revised policies are prospectively

welfare-improving. The larger realized risk exceedance rates and average equity weights show why ex-ante opportunity recovery must be evaluated jointly with realized exposure.

Table 9. Direct paired comparison of cash allowance and risk-mandate relaxation: 12% mandate with a 50% cash cap minus 15% mandate with no cash; annual CE percentage points, $\gamma = 3$ (2000 paired calendar-month resamples; 105 months).

Model	Difference (pp)	Descriptive 2.5%	Descriptive 97.5%
Neural	+0.535	−0.648	1.762
EWMA	−0.034	−0.975	0.995

The economic comparison in Figure 6 makes the intervention trade-off visible: adding cash or relaxing the mandate changes the observed CE and realized risky-sleeve QV exceedance rate. Because the mandates differ in the 12%-versus-15% comparison, the displayed exceedance rates have different denominating risk thresholds; neither is a common-budget superiority test.

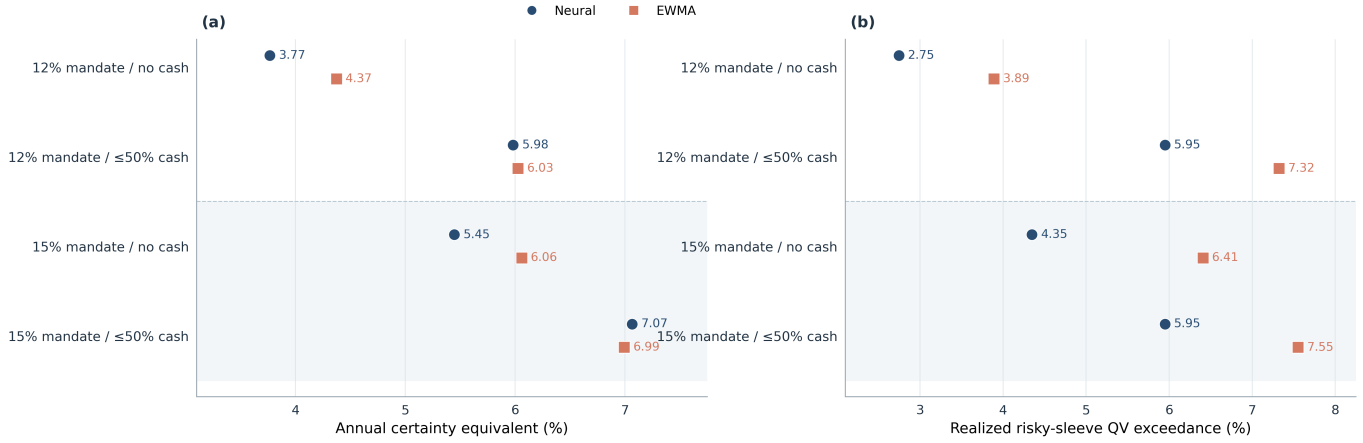


Figure 6. Historical cash-authorization and risk-mandate policy comparisons for Neural and EWMA. (**a**) Annual certainty-equivalent returns for risk aversion $\gamma = 3$ with five-basis-point transaction costs and lagged Treasury-yield cash-carry proxy. (**b**) Realized five-session risky-sleeve quadratic-variation mandate exceedance rates. Blue and orange markers denote neural and EWMA policies, respectively; the figure contains no shaded confidence region. The 12% and 15% policy rows are measured against different authorized risk thresholds and are not a common-budget coverage comparison; zero-cash reference policies permit explicitly recorded mandate-exception fallback actions.

5.11. Direct Paired Economic Uncertainty and Risk-Aversion Sensitivity

The baseline-relative retrospective intervals compare each policy with the 12%/no-cash baseline, whereas the direct paired analysis compares the two feasible-set remedies. Table 9 shows the direct alternatives-at-$\gamma = 3$ paired calendar-month result, computed on identical dated net-return histories. Neither interval excludes zero. In the wider risk-aversion audit, the neural 12%/50%-cash minus 15%/no-cash CE advantage stays near +0.53 percentage points for $\gamma = 1, 3, 5, 10$, whereas the EWMA sign changes from −0.204 at $\gamma = 1$ to +0.562 at $\gamma = 10$. Both rules are evaluated using the same dated five-basis-point trading-cost and lagged cash-carry convention.

Realized risky-sleeve quadratic-variation exceedance frequencies are 5.95% versus 4.35% for neural and 7.32% versus 6.41% for EWMA across the same two remedies. Their raw difference is not a comparison at a *common* risk threshold, because the mandates are 12% and 15%, respectively. The paired exceedance differences are therefore reported as a

descriptive policy trade-off, not as a statistical proof that either model delivers uniformly superior risk control (Supplementary Table S30).

Figure 7 displays the direct paired CE contrast over $\gamma \in \{1, 3, 5, 10\}$. In both estimators, every descriptive resampling interval crosses zero. The observed sign is not sufficient to assert a universal welfare ranking.

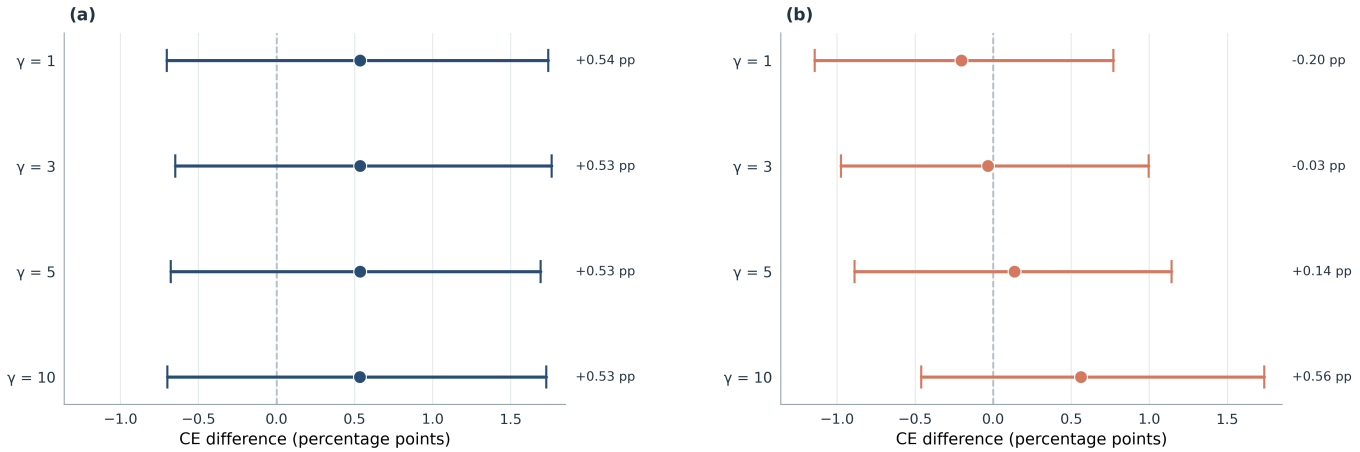


Figure 7. Direct historical certainty-equivalent comparison: 12% annual mandate with at most 50% cash minus 15% mandate with no cash. (**a**) Neural and (**b**) EWMA point differences with descriptive paired calendar-month 95% resampling intervals at $\gamma \in \{1, 3, 5, 10\}$ (105 calendar months; 2000 paired draws). The dashed vertical line is zero; every interval crosses it. The display does not establish economic dominance or untouched out-of-sample significance.

5.12. Chronological Extension Beyond the Original Source Archive

The frozen original decision sample ends on 18 August 2026 and the archived ETF vendor series ends on 15 September 2026. The chronological post-archive assessment extends matched ETF prices through 6 November 2026, with 38 additional ETF sessions. Source price returns agree exactly on overlapping trading dates; the independently earlier-ending archived market-state panel is joined at its own dated anchor without changing existing features. Neural forecasts are re-fitted under the original protocol, and all 473 prior historical forecast records, 437 original decisions, and neural/EWMA 161/18 exclusions are reproduced. Eight post-source-archive signals (16 September–4 November 2026) are observed for each reference model, seven with matured five-session targets. Neither model registers a predicted empty set or a realized all-weight risk-upper-bound exceedance on those seven matured signals. This is a very small *post-archive retrospective chronological extension*, not evidence of guaranteed future coverage. The initial prediction registry was written on 9 November 2026, after the execution dates of those 2026 historical signals; thus *zero* matured signals presently qualify as first logged before execution. The frozen future-logging protocol remains available to accumulate genuine prospective evidence separately. The chronological-extension count must not be pooled with the original 437 observations as though it were an independent precommitted test (Supplementary Table S31).

6. Discussion and Implications

The central managerial problem can now be resolved as a sequence of verifiable decisions. First identify the boundary at which the stated mandate excludes every permissible allocation. Lemma 1 characterizes the feasible set, and Lemma 3 gives the minimum risk-budget repair and conditions under which a rejection survives a change in fitted inputs. Lemmas 5 and 6 connect the same switching threshold to point-risk estimates and empirical buffers. Lemma 8 then expresses an alternative, asset-domain repair as the minimum

permissible cash fraction. Together these results turn an empty-set alert into a quantified authorization problem.

A mathematically correct decision boundary is useful only when the uncertainty margin entering it is informative. This motivates the calibration audit. A common buffer protects every admissible risky-sleeve weight on the specified score event, but its economic cost is paid across the entire action grid. We therefore evaluate matured upper-risk residuals, exceedance magnitudes and crisis clustering alongside restricted-action margins. The empirical distinction is consequential: the same high buffer that removes investment opportunities can leave sizable realized residuals uncovered in volatile episodes. The appropriate response is not automatically to reduce the buffer; it is to identify the action set to protect and assess the margin against those actions and the subsequent returns.

An unavailable portfolio creates several possible responses with different institutional meanings. Increasing the risk budget would require an authorized mandate change; Lemma 3 quantifies the smallest change without recommending that it always be granted. Revising calibration changes the protected error set and must be evaluated against the proposed trading policy. Allowing cash changes the investment domain while keeping the risky-sleeve mandate explicit. Our cash identity computes how much such capacity must be permitted for a given fitted path, and the 100-specification comparison shows why a cash threshold established for one model is not automatically transferable to another. These responses require different approvals and different risk evidence: mandate relaxation changes the institutional limit, calibration redesign changes the protected errors, and cash capacity changes the allowable asset mix.

The economic evidence completes the chain. In the two reference models, permitting cash eliminates estimated infeasibility and raises cost-adjusted historical CE, while the new allocations also increase average equity exposure and the frequency of realized five-session risky-sleeve QV budget exceedance. The result is a measurable authorization–protection–welfare trade-off. A manager who reports only a drop in empty-set counts will miss the change in realized market exposure; a manager who reports only realized exceedances will miss the economic opportunity created by the enlarged feasible set. Both measures are needed to evaluate the same cash-cap decision.

The measurement design follows the institutional decision being evaluated. The five-session uncentered quadratic-variation proxy measures the SPY/SHY sleeve's realized short-horizon risk contribution, and alternative volatility or tail-loss functionals can lead to different assessments (Patton, 2011). The rolling empirical 90th-percentile buffer enters the authorization threshold, while fully matured prequential residuals supply the separate evidence needed to evaluate its upper-risk reliability under serial dependence. In the cash extension, the constraint preserves this ETF-sleeve risk definition and measures the reduction in risky exposure as cash capacity increases; the lagged Treasury-yield series supplies dated interest-rate carry to the whole-portfolio welfare calculation. Accordingly, the reported risk-budget exceedances refer to the ETF-sleeve contribution, whereas CE summarizes the net return path including cash carry. An extension to a traded cash asset with explicitly modeled return innovations and cross-asset dependence would answer a different question about the risk of the entire portfolio return. This separation makes the authorization and valuation results comparable without conflating their risk targets.

The historical analysis uses one two-ETF investment universe and 437 dated decisions that also informed earlier research development. Consequently, seed variations, alternative calibrations and paired month-block intervals identify sensitivity within this history; transportability to new markets or policy regimes calls for a separately chosen chronological evaluation. Adjusted-opening prices and lagged yields provide dated execution and cash-return proxies, so executable trading-cost and interest-rate implementation should

be confirmed using venue-specific fill and cash-account data. This next validation step is important when translating the observed CE gains or crisis frequencies into an institutional performance claim. The analytical feasibility, perturbation and cash-threshold results continue to hold under their explicit risk-model assumptions.

For implementation, the study suggests a concrete reporting protocol: publish the critical-budget path and exact feasible-set indicator; separate point-risk and calibration effects; document the calibration's realized residuals, fit sensitivity and fallback handling; disclose the minimum authorized budget or cash repair; and report realized risky exposure, transaction costs and certainty-equivalent welfare. Because the protocol acts on fitted risk matrices and margins, it can be applied to neural, EWMA or alternative positive-semidefinite risk estimators. Its purpose is a decision system whose authorization, protection and economic consequences can be jointly audited.

What the additional sensitivity and temporal analyses show.

The new evidence does not invalidate the exact fixed-input threshold calculations. It limits their interpretation: the large baseline 161-versus-18 contrast is a property of a specified model–calibration–window combination, not a stable ranking of estimator quality. The scaled simultaneous envelope can remove estimated exclusions while preserving near-nominal historical full-grid exceedance frequencies, but this is not a new coverage theorem and does not certify protection of a forward-deployed trading rule. The shrinkage, DCC–GARCH, daily-return HAR-type and calibration-window comparisons make the model-design contingencies empirically visible. Direct paired month-block tests do not establish that allowing cash is economically superior to relaxing the mandate; the small post-archive extension is chronologically later, but remains retrospective rather than prospectively pre-logged. The choice of $h_t(w)$ and the consequences under regime shifts require independent validation before operational deployment.

Marginal Authorization and Managerial Implications

The decision consequence is marginal and conditional, not a universal treatment effect. Under the original uniform margin, a one-unit increase in q_t raises the fitted critical budget $B_t^{\text{crit}} = a_t + q_t$ by exactly one unit, with point-risk forecasts held fixed. Yet the observed empty-set indicator changes only at dates whose critical budget crosses the mandate. For a fixed risk-scaled score q_t^{sc}, the alternative critical budget is the minimum of the affine functions $\hat{\rho}_t(w) + q_t^{\text{sc}} h_t(w)$ across admissible weights. A locally stable, unique minimizing weight w_t^* therefore has marginal critical-budget sensitivity $h_t(w_t^*)$ to a change in that calibrated score, with fitted forecasts and scale factors held fixed; the minimizing weight can switch at a tie. These are comparative statics of the estimated authorization rule, not causal effects on market returns or statistical coverage claims.

This distinction yields three practical decisions. First, a risk officer should report the fitted minimum point risk, empirical buffer, critical budget, and amount by which the stated mandate is binding, rather than reporting only an "infeasible" flag. Second, any proposed buffer revision must be evaluated on the protected investment actions and matured future residuals, not justified by a lower exclusion count alone: the risk-scaled rule eliminates all 437 historical estimated exclusions across the six model families, while its historical full-grid upper-bound exceedance rates remain around 11–12%. Third, after an exclusion occurs, the investment committee must distinguish a formally approved mandate change from permission to hold cash; these interventions change different institutional constraints. At $\gamma = 3$, the direct historical cash-versus-mandate certainty-equivalent contrasts have descriptive paired intervals including zero for both reference estimators. The evidence does not rank these two repairs as universally superior.

The resulting audit recommendation is to record, for each decision, (i) the risk-limit authority and candidate action set; (ii) the exact fitted feasibility threshold and the historical scores supporting its margin; (iii) a permitted repair and any exception fallback; and (iv) ex-post upper-risk exceedance, risky exposure, turnover, and cost-adjusted certainty equivalent. These quantities must be reported together, because satisfying an estimated bound, improving risk protection, and increasing economic welfare are distinct outcomes. They can be monitored prospectively under the locked pre-execution prediction registry, but the current evidence remains retrospective and the registry has no matured qualifying prospective forecasts. A sufficiently large, genuinely advance-logged evaluation across additional assets and market regimes, together with independently validated execution and cash-carry data, is necessary before generalizing to live institutional performance.

7. Conclusions

A binding portfolio risk mandate gives an investment manager a precise authorization problem: when no admissible allocation appears, the manager must identify the constraint-driving input, determine the smallest institutionally permissible adjustment, and evaluate the risk and value of the resulting position. We develop an exact mathematical and historically evaluated decision procedure for those three tasks.

Exact feasible-set geometry, discretization bounds, minimum budget repair, perturbation and dual certificates, and budget-dependent component identities identify the authorization threshold and its sensitivity to estimation and calibration. An exposure-homogeneous cash extension quantifies the capacity required to reopen the permissible investment set. The 437-decision neural–EWMA contrast shows that the common additive calibration buffer is the main arithmetic source of the reference gap. The additional Ledoit–Wolf, OAS, DCC–GARCH, daily-return HAR-type, risk-scaled simultaneous envelope, and 35–252-score experiments show that estimated exclusions depend on both forecasting and calibration design. Direct paired economic intervals encompass zero, and only seven matured post-archive decisions per reference model are currently available, none prospectively logged before execution. These retrospective results are not evidence of prospective coverage or a new general multi-asset theory. Cash-inclusive policies then make the consequences observable in realized risky exposure, trading costs, and certainty-equivalent welfare.

The resulting contribution is a *model–calibration–authorization–outcome audit* that connects mathematical constraint diagnosis to real investment choices. It gives portfolio managers an auditable sequence: locate the switching threshold, quantify an institutionally available remedy, test the reliability of its uncertainty adjustment, and evaluate the net risk and value of the resulting allocation. This is the operational significance of treating portfolio feasibility as a decision outcome in its own right.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/ijfs14110327/s1. The technical Supplementary Materials contain 32 tables and four figures. Table S1: Archived input identifiers for the 437 decision epochs; Table S2: Covariance–buffer substitution and estimated empty-set episodes; Table S3: Indicator-rate decomposition of the reference exclusion gap; Table S4: Neural fitting-iteration convergence sensitivity; Table S5: Twelve covariance–buffer component-swapping comparisons; Table S6: Two-factor arithmetic attribution of the matched exclusion gap; Table S7: Exploratory paired month-block bootstrap intervals; Table S8: Annual estimated empty-set counts for the two matched estimators; Table S9: Annual outcomes under the original 140-iteration neural configuration; Table S10: Historical portfolio performance with five-basis-point trading costs; Table S11: Positioning relative to methodological and IJFS research; Table S12: Full-grid uniform upper-risk score exceedances; Table S13: Numerical and specification sensitivity of authorization outcomes; Table S14: Minimum modeled cash capacity needed for estimated feasibility; Table S15: Cash-inclusive portfolio outcomes and lagged Treasury

carry; Table S16: Action-restricted calibration and selected-portfolio results; Table S17: Realized risky-sleeve risk and welfare under alternative mandates; Table S18: Sensitivity across 100 forecast–calibration specifications; Table S19: Selected-action exceedance magnitudes and temporal clustering; Table S20: Distributions and tails of uniform calibration margins; Table S21: Positive forecast residuals by protected investment action; Table S22: Calibration-score persistence and dated omission tests; Table S23: Covariance estimator and calibration-rule outcomes at 126 scores; Table S24: Calibration-history window-length sensitivity; Table S25: Historical cash-capacity and mandate comparisons; Table S26: Paired month-block CE changes relative to the reference policy; Table S27: DCC–GARCH and daily-return HAR-type feasibility diagnostics; Table S28: Dynamic-model calibration-window dependence; Table S29: Direct cash-versus-mandate paired CE comparison; Table S30: Direct-remedy CE sensitivity to risk aversion; Table S31: Chronological post-archive evaluation and prospective-status audit; Table S32: Zero-cash-carry simple-cash and calibrated-policy comparison; Figure S1: Calendar-year estimated empty-set counts across mandates and covariance–buffer combinations; Figure S2: Neural fitting-iteration and component-comparison diagnostics; Figure S3: Calibration-window effects on estimated exclusions and full-grid exceedances for six estimators; Figure S4: Historical-regime concentration of estimated exclusions and exceedances. The supplementary captions provide the complete variable definitions, numerical settings, and interpretation limits.

Author Contributions: Conceptualization, H.T.H., J.C., S.M.H., L.P. and H.T.H.L.; methodology, H.T.H., J.C., S.M.H. and L.P.; validation, H.T.H., J.C., L.P. and H.T.H.L.; formal analysis, H.T.H., J.C., S.M.H., L.P. and H.T.H.L.; investigation, H.T.H., J.C. and L.P.; data curation, H.T.H., J.C., S.M.H., L.P. and H.T.H.L.; writing—original draft preparation, H.T.H., J.C., S.M.H., L.P. and H.T.H.L.; writing—review and editing, H.T.H. and J.C.; visualization, H.T.H., J.C., S.M.H., L.P. and H.T.H.L. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The derived data supporting the results of this study are publicly available in Zenodo at https://doi.org/10.5281/zenodo.24351764 (version 1.0.0). The original SPY and SHY market data were obtained from Yahoo Finance and remain subject to the provider's terms of use.

Acknowledgments: During the preparation of this manuscript, the authors used ChatGPT (OpenAI; https://chatgpt.com/; accessed on 19 November 2026) for language editing, formatting, and code error checking. The authors reviewed and edited the outputs and take full responsibility for the content of this publication.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

CE	Certainty equivalent
EWMA	Exponentially weighted moving average
MLP	Multilayer perceptron
NN	Neural point-risk/Neural buffer
NE	Neural point-risk/EWMA buffer
EN	EWMA point-risk/Neural buffer
EE	EWMA point-risk/EWMA buffer
PSD	Positive semidefinite
ETF	Exchange-traded fund
FRED	Federal Reserve Economic Data
OAS	Oracle-approximating shrinkage
DCC	Dynamic conditional correlation

HAR	Heterogeneous autoregressive (daily-return proxy here)
QV	Quadratic variation
SPY	SPDR S&P 500 ETF Trust ticker
SHY	iShares 1–3 Year Treasury Bond ETF ticker
VIX	Cboe Volatility Index
pp	Percentage points

References

Angelopoulos, A., Bates, S., Fisch, A., Lei, L., & Schuster, T. (2024). Conformal risk control. In *International conference on learning representations*. Available online: https://proceedings.iclr.cc/paper_files/paper/2024/hash/f3549ef9b5ff520a7e41ff3cc306ab2b-Abstract-Conference.html (accessed on 9 November 2026).

Bertsimas, D., & Sim, M. (2004). The price of robustness. *Operations Research*, *52*(1), 35–53. [CrossRef]

Chaweewanchon, A., & Chaysiri, R. (2022). Markowitz mean-variance portfolio optimization with predictive stock selection using machine learning. *International Journal of Financial Studies*, *10*(3), 64. [CrossRef]

Christensen, K., Siggaard, M. V., & Veliyev, B. (2023). A machine learning approach to volatility forecasting. *Journal of Financial Econometrics*, *21*(5), 1680–1727. [CrossRef]

Corsi, F. (2009). A simple approximate long-memory model of realized volatility. *Journal of Financial Econometrics*, *7*(2), 174–196. [CrossRef]

Delage, E., & Ye, Y. (2010). Distributionally robust optimization under moment uncertainty with application to data-driven problems. *Operations Research*, *58*(3), 595–612. [CrossRef]

DeMiguel, V., Garlappi, L., & Uppal, R. (2009). Optimal versus naive diversification: How inefficient is the 1/N portfolio strategy? *The Review of Financial Studies*, *22*(5), 1915–1953. [CrossRef]

Engle, R. (2002). Dynamic conditional correlation: A simple class of multivariate generalized autoregressive conditional heteroskedasticity models. *Journal of Business & Economic Statistics*, *20*(3), 339–350. [CrossRef]

Gibbs, I., & Candès, E. J. (2024). Conformal inference for online prediction with arbitrary distribution shifts. *Journal of Machine Learning Research*, *25*(162), 1–36.

Giunta, N., Orlando, G., Carleo, A., & Ricci, J. M. (2024). Exploring entropy-based portfolio strategies: Empirical analysis and cryptocurrency impact. *Risks*, *12*(5), 78. [CrossRef]

Goldfarb, D., & Iyengar, G. (2003). Robust portfolio selection problems. *Mathematics of Operations Research*, *28*(1), 1–38. [CrossRef]

Jagannathan, R., & Ma, T. (2003). Risk reduction in large portfolios: Why imposing the wrong constraints helps. *The Journal of Finance*, *58*(4), 1651–1683. [CrossRef]

Jensen, T. I., Kelly, B., Malamud, S., & Pedersen, L. H. (2026). Machine learning and the implementable efficient frontier. *The Review of Financial Studies*, *39*(10), 3035–3078. [CrossRef]

Ledoit, O., & Wolf, M. (2004). A well-conditioned estimator for large-dimensional covariance matrices. *Journal of Multivariate Analysis*, *88*(2), 365–411. [CrossRef]

Markowitz, H. (1952). Portfolio selection. *The Journal of Finance*, *7*(1), 77–91. [CrossRef]

Mashayamombe, E., Harley, C., & Paepae, T. (2026). A methodological framework for intrinsic explainability in portfolio allocation: Constraint-aware portfolio reasoning network. *International Journal of Financial Studies*, *14*(9), 229. [CrossRef]

Miri, S., Salavati, E., & Shamsi, M. (2025). Robust portfolio selection under model ambiguity using deep learning. *International Journal of Financial Studies*, *13*(1), 38. [CrossRef]

Mohajerin Esfahani, P., & Kuhn, D. (2018). Data-driven distributionally robust optimization using the Wasserstein metric: Performance guarantees and tractable reformulations. *Mathematical Programming*, *171*, 115–166. [CrossRef]

Moreira, A., & Muir, T. (2017). Volatility-managed portfolios. *The Journal of Finance*, *72*(4), 1611–1644. [CrossRef]

Natarajan, K., Pachamanova, D., & Sim, M. (2009). Constructing risk measures from uncertainty sets. *Operations Research*, *57*(5), 1129–1141. [CrossRef]

Orlando, G., & Bufalo, M. (2021). Empirical evidences on the interconnectedness between sampling and asset returns' distributions. *Risks*, *9*(5), 88. [CrossRef]

Patton, A. J. (2011). Volatility forecast comparison using imperfect volatility proxies. *Journal of Econometrics*, *160*(1), 246–256. [CrossRef]

Roos, E., & den Hertog, D. (2020). Reducing conservatism in robust optimization. *INFORMS Journal on Computing*, *32*(4), 1109–1127. [CrossRef]

Wang, T., & Dobriban, E. (2026). Optimal decision-making based on prediction sets. In *43rd international conference on machine learning* (Vol. 306, pp. 125896–125918). PMLR. Available online: https://proceedings.mlr.press/v306/wang26ar.html (accessed on 9 November 2026).

Wiesemann, W., Kuhn, D., & Sim, M. (2014). Distributionally robust convex optimization. *Operations Research*, *62*(6), 1358–1376. [CrossRef]

Xu, C., & Xie, Y. (2021). Conformal prediction interval for dynamic time-series. In *38th international conference on machine learning* (Vol. 139, pp. 11559–11569). PMLR. Available online: https://proceedings.mlr.press/v139/xu21h.html (accessed on 9 November 2026).

Yu, J., & Chang, K.-C. (2026). A hierarchical signal-to-policy learning framework for risk-aware portfolio optimization. *International Journal of Financial Studies*, *14*(3), 75. [CrossRef]

Zaffran, M., Feron, O., Goude, Y., Josse, J., & Dieuleveut, A. (2022). Adaptive conformal predictions for time series. In *39th international conference on machine learning* (Vol. 162, pp. 25834–25866). PMLR. Available online: https://proceedings.mlr.press/v162/zaffran22a.html (accessed on 9 November 2026).

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.