Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Consecutive Failure Risk and Volume Frontiers of Multivariate Prediction Regions at Fixed Joint Coverage

Yixiang Wang

Wangyixiang0712@outlook.com

Jiangxi University of Technology

Haoyu Jiang

Jiangxi University of Technology

Research Article

Keywords: multivariate prediction regions; joint coverage; consecutive failures; minimumvolume sets; global optimization

Posted Date: November 26th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11850696/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Consecutive Failure Risk and Volume Frontiers of Multivariate Prediction Regions at Fixed Joint Coverage

Yixiang Wang* Haoyu Jiang

School of Information Engineering, Jiangxi University of Technology, Nanchang 330098, China

*Corresponding author: Yixiang Wang (Wangyixiang0712@outlook.com).

Abstract

Multivariate prediction regions that attain identical single-period joint coverage can produce sharply different probabilities of two consecutive failures. Existing multivariate constructions target single-period coverage and volume, leaving consecutive-failure risk outside the region design objective. We formulate the minimum consecutive-failure probability at fixed average coverage and a volume cap as a risk–volume frontier. For any two-period joint error law, allocating mass across progressively refined cells with minimum-volume cell subsets (treating point atoms as indivisible) converges monotonically to the all-shape infimum, accompanied by a two-sided finite-partition error bound. Under a common marginal whose boundary has an odd-order normal density crossing of order k, spatial variation in forward and backward return probabilities yields an optimal small-budget risk drop of order $\delta^{1/(k+1)}$ ($\sqrt{\delta}$ for $k = 1$ and $\delta^{1/4}$ for $k = 3$). The frontier directly prices geometric restrictions: at 90% coverage, disconnected sets hit 1% risk with 1% extra length versus 18.78% for a single interval in a piecewise-density model, while elliptical stretching along a persistent Gaussian axis (lag-one correlation 0.722) captures 99.65% of the all-shape $\sqrt{\delta}$ gain. We also derive global rectangular algorithms for non-Gaussian and contemporaneously dependent errors. Across six real multivariate benchmarks with one-step ridge forecasts, equal-volume per-origin width reallocation cuts consecutive-failure frequency by 18.39% (0.02150 to 0.01754) at zero volume cost and reduces risk by 9.77% at matched panel miss counts.

Keywords: multivariate prediction regions; joint coverage; consecutive failures; minimum-volume sets; global optimization.

1 Introduction

Two simultaneously predicted outputs show why forecast regions with identical coverage and area can still carry different consecutive-failure risk. Suppose the two error processes are mutually independent, and each standardized coordinate follows a standard normal distribution at every single time point. The first output follows a stationary Gaussian first-order autoregressive process, while the second output is independent across time. A joint forecast issues a rectangle, and the forecast fails if at least one coordinate falls outside that rectangle. Swapping the two half-widths leaves both area and joint coverage unchanged, yet it changes the boundary-crossing probability of the temporally persistent coordinate.

Near the minimum area, swapping widths already creates a noticeable gap. Fix joint coverage at 90%, set the lag-one error correlation of the first coordinate to 0.722, and set the rectangle area to 1.01 times the area of the minimum-area square achieving that coverage. Assigning the wider side to the first coordinate yields a consecutive-failure probability near 0.01979. Reversing the two sides raises that probability to 0.02836. Both rectangles have single-period failure probability 0.10.

This paper investigates how to reduce consecutive-failure probability under fixed coverage and a volume cap when two adjacent periods share one deterministic region. When both periods have the same marginal error law, each period meets the same coverage target. When marginals differ, one can fix either the average coverage across the two periods or each period's coverage

separately. We start from the minimum volume needed for the target coverage and specify a permitted proportional volume increase. When an application restricts region geometry, such as requiring separate upper and lower bounds for each output, both quantities are evaluated within that family. Varying the budget traces a frontier giving the infimum of risk at each volume cap.

1.1 Region size and failure ordering

Existing methods determine region size and shape from probability mass. Scott and Nowak (2006) study minimum-volume set estimation at a prescribed probability mass. Di and Kolaczyk (2010) extend this estimation to dependent samples using complexity penalties. DCP-MT (Schlembach et al. 2025) allocates error budgets across coordinates of multi-output rectangles, balancing joint coverage and volume under exchangeability. Copula methods construct joint regions from cross-output dependence (Messoudi et al. 2021), and MultiDimSPCI represents multivariate time-series errors with ellipsoids (Xu et al. 2024). These methods provide our starting point. Once a region meets its coverage target, we examine the remaining shape degrees of freedom and their effect on two consecutive joint failures.

Forecast evaluation separates failure frequency from failure ordering. Christoffersen (1998) tests unconditional coverage and first-order failure independence separately. Bröcker (2022) constructs calibration tests from conditional relations, and Retzlaff et al. (2025) apply backtests to conformal forecasts of non-stationary series. Wang (2026) investigates whether interval widths select estimators with lower mean squared error, while Min et al. (2026) compare interval stability across repeated runs of randomized algorithms. Here we hold the point forecaster fixed and turn consecutive-failure probability from a post-hoc diagnostic into a design objective for region choice.

Intertemporal return probabilities of fixed-mass sets also have independent roots. Froyland (2005) identifies almost-invariant sets through one-step retention probabilities. Gaussian noise stability studies the probability that two correlated vectors both fall inside a given set (Heilman 2021; Heilman et al. 2016). At fixed single-period coverage, that joint hit probability differs from consecutive-failure probability only by a constant. Starting from the minimum-volume set, we compute how each increment of volume lowers the smallest achievable risk. Doing so combines the boundary slope of the marginal density with the two-period return probabilities.

We begin by comparing risk across all measurable shapes. For any given two-period joint law, preserving cell masses while refining the partition approximates the population frontier. A three-piece density example yields the exact global frontier and measures the extra length demanded by a single-interval restriction. Next, we determine which part of the minimum-volume boundary should receive the first increment of volume. The crossing order of the marginal density and the spatial variation of return probabilities jointly determine the rate and direction of risk reduction. For applications requiring per-coordinate intervals, we develop global solvers for centered rectangles under non-Gaussian and contemporaneously dependent models. Finally, we derive total-variation stability bounds for frontier estimation from finite history and propose a per-origin equal-volume width adjustment rule, evaluating both on factor simulations and six real forecasting benchmarks.

2 Finite-budget risk over all measurable regions

We first define coverage, volume, and consecutive-failure probability for an arbitrary fixed region. Let $Z_t \in \mathbb{R}^M$ denote the standardized forecast error vector issued at forecast origin t for a fixed lead time, indexed in origin order. Assume initially that two adjacent periods share a common single-period distribution and fix their joint law. Write B_θ for a prediction region, where θ indexes either an arbitrary set or a parameter vector such as rectangle half-widths. Let $I_t^\theta = \mathbf{1}\{Z_t \notin B_\theta\}$ record a coverage failure, and define

$$C(\theta) = \mathbb{P}(Z_t \in B_\theta), \qquad V(\theta) = \mathrm{Vol}(B_\theta), \qquad R_2(\theta) = \mathbb{E}[I_{t-1}^\theta I_t^\theta]. \tag{1}$$

Both probabilities and the standardized volume are dimensionless. In physical units, volume carries the product of the coordinate units. Writing $\pi = 1 - C(\theta)$ gives $R_2(\theta) = \pi^2 + \mathrm{Cov}(I_{t-1}^{\theta}, I_t^{\theta})$. Fixing C fixes the independent-failure baseline π^2, so changes in R_2 come solely from the covariance between adjacent failures.

The minimum volume achieving the target coverage within a shape class sets our baseline. Given $c \in (0,1)$, define $\mathcal{M}_c = \{\theta : C(\theta) = c\}$. When the minimum volume is attained and positive, denote it by V_0. Allowing a proportional volume increase $\delta \geq 0$ leads to the population problem

$$\inf_{\theta \in \mathcal{M}_c} R_2(\theta) \quad \text{subject to} \quad V(\theta) \leq (1+\delta)V_0. \tag{2}$$

Rectangles and ellipsoids each have their own baseline V_0. Across shape classes, absolute volume V records the actual space used by each region.

2.1 Approximating all shapes at any budget

We formulate a population target valid for any two-period error law. Let P be the joint law of (Z_-, Z_+) on $\mathbb{R}^M \times \mathbb{R}^M$, with marginals μ_-, μ_+ and average marginal $\mu = (\mu_- + \mu_+)/2$. Using one deterministic region B at both origins yields consecutive-failure risk $Q(B) = P(B^c \times B^c)$. Given target coverage $0 < c < 1$ and an absolute volume cap $H < \infty$, define

$$\mathcal{R}(H, c) = \inf\{Q(B) : B \text{ measurable}, \ \mu(B) = c, \ |B| \leq H\}. \tag{3}$$

We consider only feasible pairs (H, c). When both marginals match, each period achieves coverage c. When marginals differ, (3) fixes their average coverage. If the minimum volume V_0 is positive, a relative budget δ corresponds to $H = (1+\delta)V_0$, while an absolute budget also handles laws with $V_0 = 0$. Appendix A of Online Resource 1 characterizes exact feasibility and states the counterpart when both period coverages are fixed individually.

Under unequal marginals, fixing average coverage still allows a set to shift coverage between periods. Write $c_- = \mu_-(B)$ and $c_+ = \mu_+(B)$. Even when the two periods are independent, $(c_- + c_+)/2 = c$ gives

$$Q(B) = (1 - c_-)(1 - c_+) = (1-c)^2 - \tfrac{1}{4}(c_- - c_+)^2.$$

Suppose, for instance, that the two errors are independent and uniform on $[0,1]$ and $[1,2]$. At length cap 1 and average coverage $1/2$, choosing $[0,1]$ covers the first period completely and reduces consecutive-failure probability to zero. Requiring coverage $1/2$ in each period instead makes $[1/2, 3/2]$ feasible, and every feasible set then has consecutive-failure probability $1/4$. Separate period constraints therefore isolate temporal failure dependence when marginals differ.

Within each partition cell, we retain a prescribed mass at minimal volume. If μ has density f, select points inside the cell in descending order of f until the accumulated mass reaches m, breaking ties with a fixed global ordering. This construction produces nested subsets $D_i(m)$ whose volumes $v_i(m)$ are minimal among all cell subsets of mass m, mirroring the minimum-volume set construction inside every cell (Scott and Nowak 2006). Any mass supported on curves or other zero-volume sets is selected first. We call a single point carrying positive probability an atom; a deterministic region either retains or drops an atom as a whole.

Theorem 1 (All-shape frontier for a general two-period law)**.** *Choose a nested sequence of finite Borel partitions whose union generates the Borel σ-algebra. If μ is atomless, select a cell mass m_i in each cell at level n by solving*

$$U_n = \min_{m} Q\left(\bigcup_i D_i(m_i)\right), \qquad 0 \leq m_i \leq \mu_i, \quad \sum_i m_i = c, \quad \sum_i v_i(m_i) \leq H, \tag{4}$$

where μ_i denotes the marginal mass of cell i. Under a space-wide consistent tie-breaking order, $U_n \downarrow \mathcal{R}(H, c)$. Every level yields an exact-coverage, volume-feasible deterministic region, and the joint law may concentrate on lower-dimensional sets.

If μ has atoms, apply the same construction to its atomless part while keeping each atom binary. With finitely many atoms, this adds one binary variable per atom. With countably many atoms, include progressively more atoms at level n while retaining the attainable mass set of remaining atoms in the coverage constraint, as detailed in Appendix A. The resulting feasible upper bounds still decrease to $\mathcal{R}(H, c)$, covering purely discrete joint laws as well.

The argument compares an arbitrary feasible set B with a cell-wise replacement that keeps the exact mass of B in every cell without increasing volume. As the partition refines, the total misclassified mass in cells that split B and B^c vanishes. Replacing B by another set A changes the consecutive-failure indicator only when at least one error falls in the symmetric difference $A \triangle B$, which yields

$$|Q(A) - Q(B)| \le \mu_-(A \triangle B) + \mu_+(A \triangle B) = 2\mu(A \triangle B).$$

This bound controls both the direction and the limit of approximation at every feasible budget, including the minimum volume. When μ is atomless, the frontier is continuous across the full budget interval; atoms can introduce jumps, as shown with an explicit counterexample in Appendix A.1. We next ask when a finite computation produces a certified error bound.

2.2 Finite-partition accuracy for global risk

Averaging the two-period law over partition cells yields a finite-dimensional optimization problem with an explicit gap to the true frontier. Replace P by its time-symmetrized average across the two periods. Symmetrization leaves $Q(B)$ unchanged while giving both coordinates the common marginal μ. At fixed $\mu(B) = c$, we have $Q(B) = 1 - 2c + P(B \times B)$ under the exact coverage and volume constraints of (3).

Proposition 1 (Two-sided error bound for the all-shape frontier)**.** *Assume the average marginal μ is atomless and P decomposes into two parts: a diagonal component along $x = y$ assigning mass $d(x)\,\mu(dx)$ to (x, x), and an off-diagonal component with nonnegative integrable density $K(x, y)$ relative to $\mu \otimes \mu$. Take a finite partition $E_1, \ldots, E_J$ omitting zero-mass cells. Write d_i for the μ-average of d on E_i, K_{ij} for the $(\mu \otimes \mu)$-average of K on $E_i \times E_j$, and $K_{\mathcal{P}}$ for the resulting piecewise-constant kernel.*

Let $v_i(m_i)$ be the minimum volume needed for mass m_i in cell i. Denote by $p_{\mathcal{P}}$ the minimum value of the finite program (5)*.*

$$\begin{aligned} \min_{m} \quad & 1 - 2c + \sum_i d_i m_i + \sum_{i,j} K_{ij} m_i m_j, \\ \text{subject to} \quad & 0 \le m_i \le \mu(E_i), \qquad \sum_i m_i = c, \qquad \sum_i v_i(m_i) \le H. \end{aligned} \tag{5}$$

This finite value satisfies the two-sided error bound (6)*.*

$$|\mathcal{R}(H, c) - p_{\mathcal{P}}| \le e_{\mathcal{P}}, \qquad e_{\mathcal{P}} = \tfrac{1}{4} \int |K - K_{\mathcal{P}}|\, d(\mu \otimes \mu). \tag{6}$$

From an optimal solution of (5)*, assembling the lowest-volume-cost subset of mass m_i in each cell yields a feasible deterministic region whose risk is at most $\mathcal{R}(H, c) + 2e_{\mathcal{P}}$. Along nested partitions generating the Borel σ-algebra, $e_{\mathcal{P}} \to 0$, and the same bound holds uniformly across all feasible budgets.*

Appendix A.2 gives the proof of (6) and a box-subdivision solver. Partitioning the mass simplex into boxes whose relative edge lengths are at most η solves the finite program within error η, provided each box permits exact minimum-volume assembly and feasibility checks. Solver error η and cell-averaging error $e_{\mathcal{P}}$ can then be reported separately.

When the oscillation of K within cell pairs is known, a single run yields certified numeric bounds. If the essential supremum minus essential infimum of K on $E_i \times E_j$ is at most s_{ij}, then $e_{\mathcal{P}} \le \sum_{i,j} \mu(E_i)\mu(E_j)s_{ij}/4$. If the finite solver brackets $p_{\mathcal{P}}$ inside $[l, u]$, then $\mathcal{R}(H, c)$ lies in $[l - e_{\mathcal{P}}, u + e_{\mathcal{P}}]$. Cell-averaged kernels with fixed-mass quadratic objectives also appear in kernel energy representations (Borgs et al. 2012). Here mass allocation is coupled with a volume constraint, and diagonal mass uses the identity $\mathbf{1}_B(x)^2 = \mathbf{1}_B(x)$. When atoms are present, each atom enters or exits whole, restricting cell masses to sums of binary choices and atomless parts; relaxing atoms to continuous intervals changes the decision problem.

One construction shows that a single transition table does not fix the approximation error and that the constant $1/4$ in (6) does not admit universal tightening. Let μ be uniform on $[0, 1]$, split into N equal subintervals, and set $b(x) = \sin(2\pi N x)$ and $K' = 1 - ab(x)b(y)$ with $0 < a < 1$. Because b integrates to zero on each cell, K' shares every cell-pair mass with the independent law. At $c = H = 1/2$, the independent law has optimal risk $1/4$, whereas K' achieves $1/4 - a/\pi^2$ by selecting $\{b > 0\}$, where $\left|\int_B b\,dx\right|$ reaches $1/\pi$. Meanwhile $\|K' - 1\|_1 = a(\int |b|)^2 = 4a/\pi^2$, so the frontier gap equals $e_{\mathcal{P}}$ exactly.

Even under a smooth joint density, the optimal value may be approached only by a sequence of feasible sets. In the example of Appendix A.2, both coverage and length cap equal $1/2$. Every deterministic region has risk strictly above $1/4$, yet a sequence of disconnected regions drives risk to $1/4$. That is why (3) uses an infimum. The next example attains its frontier explicitly.

2.3 When a region needs to disconnect at the same budget

Three adjacent intervals let us compute the exact global frontier at every budget and compare it with single intervals. Arrange three intervals rightward from the origin with lengths 1, $1/5$, and $1/10$, carrying marginal densities f equal to $9/10$, $1/10$, and $4/5$, and zero elsewhere. Their marginal masses are 0.90, 0.02, and 0.08. Define h equal to $2/23$, $2/23$, and -1 on the three segments, and set the two-period density to

$$p(x, y) = f(x)f(y)\{1 + h(x)h(y)\}.$$

Because $\int fh = 0$ and $1 + h(x)h(y) > 0$, this defines a valid joint density with both marginals equal to f. Partitioning along the three intervals makes the density ratio constant on every cell pair, so $e_{\mathcal{P}} = 0$ in Proposition 1. Fix coverage at $c = 0.90$. The first interval alone reaches mass 0.90 at the highest density, so the minimum length is 1.

For any measurable set B of coverage c, consecutive-failure risk depends only on the mass r taken from the third interval. Integrating over $B^c \times B^c$ yields

$$R_2(B) = (1 - c)^2 + \left(\int_B fh\right)^2 = 0.01 + \left(\frac{9}{115} - \frac{25}{23}r\right)^2.$$

Transferring mass r from the first segment to the third requires at least $5/4 - 10/9 = 5/36$ extra length per unit mass, and including the low-density middle segment costs even more. When disconnected sets are permitted, taking subsets of only the first and third segments attains this lower bound. At length cap $1 + \delta$, moving r as close as possible to the root $9/125$ of the squared term solves the problem for every $\delta \ge 0$.

Restricting B to a single interval changes the geometry. Because the last two segments together hold only 0.10 mass, any single interval touching the third segment has to span the entire second segment, incurring extra length $1/5 - (0.02)/(0.90) = 8/45$. Extending continuously from

the right part of the first segment into the third attains this bound for each $r > 0$. The optimal third-segment masses for the two families are

$$r_{\text{all}} = \min\left\{\frac{9}{125}, \frac{36}{5}\delta\right\}, \qquad r_{\text{int}} = \min\left\{\frac{9}{125}, \frac{36}{5}[\delta - 8/45]_+\right\},$$

where $[z]_+ = \max(z, 0)$. At 1% extra length, the disconnected set $[0, 0.92] \cup [1.20, 1.29]$ keeps coverage 0.90 and lowers risk from 0.016125 to 0.010000. Every single interval at the same budget and coverage remains stuck at risk 0.016125. A single interval reaches risk 0.010000 only at extra length 169/900, or about 18.78%. Here 0.01 is the exact floor implied by the nonnegative square. An exact rational solver enumerating polytope vertices and integrating failure probability directly confirms both frontiers across 124 budgets.

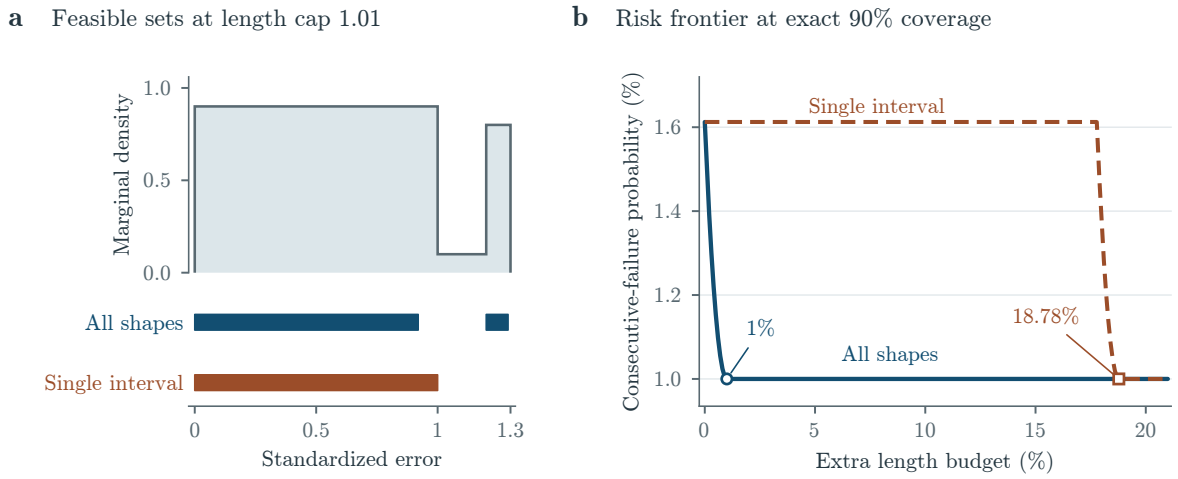

Fig. 1 All-shape and single-interval optimal frontiers for the three-block density at 90% joint coverage. **a** Representative sets at a 1% extra length cap: the disconnected set uses length 1.01, while the single interval uses length 1.00. **b** Population consecutive-failure probability against extra length budget. The two classes first reach risk 0.01 at 1% and about 18.78%. Curves are evaluated from closed-form expressions and checked point by point via rational enumeration

Figure 1 plots both frontiers on the same budget axis. All-shape regions skip the low-density gap and spend their extra length inside the third block, whereas a single interval first pays to cross the middle block before reducing risk. Appendix B adds a smooth logistic example in which a single interval already achieves the all-shape optimum at every budget.

3 Moving the boundary near minimum volume

3.1 Optimal leading term from the density boundary

We first allow any measurable prediction region. Let f denote the single-period density of the standardized error Z_t, and pick level $\lambda > 0$ so that $A = \{x : f(x) \geq \lambda\}$ has coverage c. Collecting the highest-density points makes A the classical minimum-volume set (Scott and Nowak 2006). Write $V_A = |A|$ for its Lebesgue volume; relative budgets δ in this subsection use V_A as denominator.

To quantify how boundary perturbations alter risk, let $p(x, y)$ be the two-period joint density with common marginal f. For almost every x, define

$$w(x) = \int_A [p(x, y) + p(y, x)]\, dy.$$

Where $f(x) > 0$, the ratio $w(x)/f(x)$ equals the sum of the two conditional return probabilities $\mathbb{P}(Z_+ \in A \mid Z_- = x)$ and $\mathbb{P}(Z_- \in A \mid Z_+ = x)$ for the evaluated pair of times, which need not coincide. On the level set $\Gamma = \{f = \lambda\}$, write $g(x) = \|\nabla f(x)\|$ and determine boundary values of w via a continuous representative in a neighborhood of Γ.

Theorem 2 (Small-budget frontier over all measurable sets)**.** *Suppose f is twice continuously differentiable and vanishes at infinity, and Γ is nonempty with $g > 0$ on Γ. Suppose also that p is bounded and that w admits a Lipschitz-continuous representative on an open neighborhood of Γ (unique there and restricting to boundary values on Γ). With surface element dS, define*

$$\bar{w} = \frac{\int_\Gamma (w/g)\, dS}{\int_\Gamma (1/g)\, dS}, \qquad K = \int_\Gamma \frac{(w - \bar{w})^2}{g}\, dS.$$

Let $R_{\mathrm{all}}(\delta)$ be the infimum of $R_2(B)$ over all measurable sets B satisfying $\int_B f = c$ and $|B| \le (1+\delta)V_A$. As $\delta \downarrow 0$,

$$R_2(A) - R_{\mathrm{all}}(\delta) = \sqrt{2\lambda V_A K}\,\sqrt{\delta} + O(\delta). \tag{7}$$

When $K > 0$, moving Γ outward at normal speed $h(x) = (\bar{w} - w(x))/g(x)$ and applying a second-order constant shift to restore exact coverage attains this leading term. When $K = 0$, the gain over all feasible shapes is at most $O(\delta)$.

Theorem 2 gives a local rule along the boundary. Segments with $w < \bar{w}$ move outward, while segments with $w > \bar{w}$ move inward. Steeper density gradients make normal displacements more expensive in volume, so the optimal speed divides by g. The upper bound in Appendix B.1 covers perforated, disconnected, and nonsmooth competitors alike. Computation depends solely on the two-period joint distribution and permits contemporaneous dependence among coordinates.

When the two periods have unequal marginal densities f_- and f_+, fixing average coverage sets $f = (f_- + f_+)/2$. Under $\int_B f = c$, we still have $R_2(B) = 1 - 2c + \int_{B\times B} p$, and w retains the same two-term integral definition. If f and w satisfy the regularity assumptions of Theorem 2, the leading coefficient formula remains intact.

A flat-density plateau clarifies the role of a regular level set. Take a uniform marginal on $[0, 1]$ where the next period stays in the current half-interval with probability 0.8, switches halves with probability 0.2, and draws uniformly inside the chosen half. At coverage $1/2$, both the full left half and two quarter-length pieces (one in each half) attain minimum length. Their consecutive-failure risks are 0.4 and 0.25. When multiple sets tie for minimum volume, even zero extra budget permits distinct failure risks. By contrast, under equal marginals and independent periods, every set of coverage c shares risk $(1 - c)^2$.

Flat boundaries change the budget exponent. The rate at which density crosses its threshold governs the volume cost of moving the boundary. Theorem 2 covers first-order crossings. Suppose Γ remains a compact C^2 hypersurface with outer unit normal $n(s)$ at $s \in \Gamma$, and the density crosses λ uniformly at an odd order k:

$$f(s + vn(s)) = \lambda - a(s)v^k + O(|v|^{k+1}), \qquad \min_\Gamma a > 0.$$

Set $q = (k+1)/k$ and $D_k = \min_b \int_\Gamma a^{-1/k}|w - b|^q\, dS$. When $D_k > 0$, the optimal risk drop across all measurable sets obeys

$$R_2(A) - R_{\mathrm{all}}(\delta) = \left[(k+1)\lambda V_A \delta\right]^{1/(k+1)} D_k^{1/q} + O\left(\delta^{2/(k+1)}\right). \tag{8}$$

When $D_k = 0$, only the remainder order survives. The statement retains boundedness of p, local Lipschitz continuity of w near Γ, continuity of f vanishing at infinity, $\{f = \lambda\} = \Gamma$, and continuity of a. Setting $k = 1$ recovers Theorem 2, while $k = 3$ produces a fourth-root rate. Flatness around a density level also governs convergence in level-set estimation (Rigollet and Vert 2009); here it determines the volume-budget exponent at fixed coverage. Appendix B.2 gives the proof and a verifiable third-order example.

3.2 Shape directions inside a parametric family

We next express coverage-preserving shape changes in local coordinates. Let $\theta_\star$ be an interior volume minimizer on $\mathcal{M}_c$, possibly local. Assume C and V are three times continuously differentiable near $\theta_\star$, with $V(\theta_\star) > 0$ and $\nabla C(\theta_\star) \neq 0$. By the implicit function theorem, a coordinate $u \in \mathbb{R}^{d-1}$ parameterizes $\mathcal{M}_c$ near $\theta_\star$ via $\theta = \psi(u)$ with $\psi(0) = \theta_\star$, where $d \geq 2$ is the ambient parameter dimension. Define

$$L(u) = \log \frac{V(\psi(u))}{V(\theta_\star)}, \qquad S(u) = R_2(\psi(u)), \qquad H = \nabla^2 L(0), \qquad b = \nabla S(0).$$

Because $\theta_\star$ minimizes volume on $\mathcal{M}_c$, we have $\nabla L(0) = 0$. The Hessian H measures second-order log-volume cost, while b captures first-order risk sensitivity.

Theorem 3 (Local parametric frontier)**.** *Assume H is positive definite and R_2 is twice continuously differentiable near $\theta_\star$. Fix a small closed coordinate ball U containing 0 in its interior, and let $S_U^\star(\delta)$ be the minimum of $S(u)$ over $u \in U$ with $L(u) \leq \log(1+\delta)$. If $b \neq 0$, then as $\delta \downarrow 0$,*

$$S(0) - S_U^\star(\delta) = \gamma\sqrt{\delta} + O(\delta), \qquad \gamma = \sqrt{2b^\top H^{-1} b}. \tag{9}$$

Every optimal solution u_δ satisfies

$$\frac{u_\delta}{\sqrt{2\delta}} \longrightarrow -\frac{H^{-1}b}{\sqrt{b^\top H^{-1} b}}.$$

If $b = 0$ and $J = \nabla^2 S(0)$, then

$$S(0) - S_U^\star(\delta) = \delta \max\{0, -\lambda_{\min}(H^{-1/2} J H^{-1/2})\} + o(\delta), \tag{10}$$

where $\lambda_{\min}$ denotes the smallest eigenvalue.

At a volume minimizer, equal-coverage perturbations cost volume only at second order while changing risk at first order. Projecting the steepest-descent direction of the quadratic trust region back onto the exact coverage manifold yields (9); Appendix C supplies the proof.

Theorem 3 translates directly into an explicit calculation. Choose a matrix E whose columns form an orthonormal basis for the orthogonal complement of $\nabla C(\theta_\star)$, and take $D\psi(0) = E$. By Lagrange multipliers, $\nabla \log V = \lambda \nabla C$ at $\theta_\star$. Differentiating $C(\psi(u)) = c$ twice gives

$$b = E^\top \nabla R_2, \qquad H = E^\top \big(\nabla^2 \log V - \lambda \nabla^2 C\big) E,$$

with all derivatives evaluated at $\theta_\star$. When $b \neq 0$, solving $Hv = -b$ and finding the budget root along the exact coverage curve attains the leading gain in (9). When all sufficiently small-budget feasible sets lie inside the chart, the local frontier agrees with the global frontier in (2).

3.3 Computing local directions from probability functions

Numerical evaluation needs only two oracle functions: single-period coverage and two-period joint hit probability, since risk equals $1-2C$ plus the joint hit probability. For families with an explicit scale parameter, solving a one-dimensional root for scale at each shape keeps coverage exact at c. Risk gradients and log-volume curvature then live directly in equal-coverage coordinates.

Central finite differences approximate derivatives of smooth probability functions. With $p = d-1$ free shape parameters, the gradient uses $2p$ risk evaluations and the symmetric Hessian uses $1 + 2p + 4\binom{p}{2}$ log-volume evaluations. When H is positive definite and $b \neq 0$, we solve $Hv = -b$, normalize to $v^\top H v = 1$, and solve $L(tv) = \log(1+\delta)$. Using population derivatives recovers the optimal leading rate. Empirical coverage functions are step functions, so finite-sample selection in Section 6 replaces differentiation with validation-set screening.

When symmetry forces $b = 0$, equation (10) governs the next order. Let w be a unit eigenvector for the smallest eigenvalue of $H^{-1/2}JH^{-1/2}$. If that eigenvalue is negative, moving along $v = H^{-1/2}w$ attains the leading δ-order gain; if it is nonnegative, keeping the original shape is already optimal to order δ. Small finite-difference gradients alone do not distinguish an exact zero gradient from a weak slope.

4 Rectangular allocation and elliptical efficiency in Gaussian examples

An independent-coordinate Gaussian model makes risk gradients and volume curvature explicit. All coordinates share the same single-period marginal distribution while differing in temporal memory. Let coordinate i follow a stationary standard Gaussian AR(1) process with lag-one correlation $r_i \in (-1, 1)$, independent across coordinates $i = 1, \dots, M$ ($M \geq 2$). Each period vector is $N(0, I_M)$, and we consider centered axis-aligned rectangles $B_\tau = \prod_{i=1}^{M}[-\tau_i, \tau_i]$ with $\tau_i > 0$. Any width asymmetry therefore arises solely from temporal dependence.

4.1 Marginal coordinate effects from joint probabilities

Let Φ and ϕ denote the standard normal distribution and density functions. For a standard bivariate normal pair (X, Y) with correlation r, write

$$F(a) = 2\Phi(a) - 1, \qquad B_r(a) = \mathbb{P}\{|X| \leq a,\ |Y| \leq a\}, \qquad Q_r(a) = \frac{B_r(a)}{F(a)^2}.$$

Coordinate independence yields $C(\tau) = \prod_i F(\tau_i)$. Under $C(\tau) = c$, consecutive-failure risk factorizes as

$$R_2(\tau) = 1 - 2c + c^2 \prod_{i=1}^{M} Q_{r_i}(\tau_i). \tag{11}$$

Lemma 1. *The function $g(x) = \log F(e^x)$ satisfies $g' > 0$ and $g'' < 0$. Also $Q_{-r}(a) = Q_r(a)$, $Q_r'(a) < 0$ for all $a > 0$ when $0 < |r| < 1$, and $Q_0(a) = 1$.*

Every persistence configuration shares the same minimum-volume benchmark. Put $x_i = \log \tau_i$, $\bar{x} = M^{-1}\sum_i x_i$, and $q = F^{-1}(c^{1/M})$. Strict concavity gives $\log c = \sum_i g(x_i) \leq M g(\bar{x})$, with equality only when all half-widths coincide. Monotonicity of g implies $\bar{x} \geq \log q$, so the equal-width cube with half-widths q uniquely attains $V_0 = (2q)^M$.

Corollary 1 (Local allocation with multiple persistent coordinates)**.** *In the model above, set $x_0 = \log q$ and define*

$$h_0 = -\frac{g''(x_0)}{g'(x_0)}, \qquad P_0 = c^2 \prod_i Q_{r_i}(q), \qquad d_i = \frac{qQ_{r_i}'(q)}{Q_{r_i}(q)}, \qquad \bar{d} = \frac{1}{M}\sum_i d_i.$$

If $d^\circ = (d_i - \bar{d})_{i=1}^{M}$ is nonzero, the optimal first-order log-half-width direction is $-d^\circ$, and the square-root frontier coefficient is $\gamma = P_0\sqrt{2/h_0}\,\|d^\circ\|_2$. If all $|r_i|$ are equal, $\gamma = 0$.

Three coordinates with unequal persistence verify the numerical tangent computation. At $c = 0.90$ and $(r_1, r_2, r_3) = (0, 0.35, 0.722)$, central differences of step 10^{-4} in orthogonal tangent coordinates give $\gamma = 0.029895$. Along this direction, reducing δ through 10^{-2}, 10^{-4}, and 10^{-6} while solving for exact coverage yields risk gains divided by $\sqrt{\delta}$ equal to 0.027580, 0.029694, and 0.029875. The ratios converge toward the theoretical coefficient and validate the projection back to the coverage manifold.

Each component of $-d^\circ$ has a clear reading. The value d_i records the logarithmic slope of the two-period overlap factor with respect to log-half-width i at the cube. Coordinates below the mean $\bar{d}$ gain width, while coordinates above $\bar{d}$ give up width, and common shifts cancel under the coverage constraint. Two coordinates with equal $|r_i|$ receive the same first-order adjustment.

4.2 One-dimensional root finding for a single persistent coordinate

We now solve the global problem at any budget $\delta > 0$ when $|r_1| = r \in (0, 1)$ and $r_i = 0$ for $i \geq 2$. Although the sign of r flips signed error trajectories, symmetric intervals depend only on absolute values, so r and $-r$ share the same joint hit probability. Equation (11) reduces to a function of τ_1 alone. By Lemma 1, widening τ_1 at fixed coverage lowers risk monotonically, so the global solution pushes τ_1 as far as the volume budget allows.

Theorem 4 (Global solution for one persistent coordinate)**.** *At $\delta = 0$, the unique solution to* (2) *is the equal-width cube. For any $\delta > 0$, the unique global solution is $\tau_1 = a_\delta$ and $\tau_2 = \cdots = \tau_M = v(a_\delta)$, where*

$$v(a) = F^{-1}\left[\{c/F(a)\}^{1/(M-1)}\right], \qquad 2^M a_\delta v(a_\delta)^{M-1} = (1+\delta)(2q)^M, \qquad a_\delta > q. \tag{12}$$

Given the first half-width a, setting the remaining $M-1$ half-widths equal achieves the residual coverage at minimal volume. Global search collapses to locating one root along this monotone curve; Appendix C supplies the monotonicity and globality proof. Figure 2 compares this exact solution with the local expansion around the square.

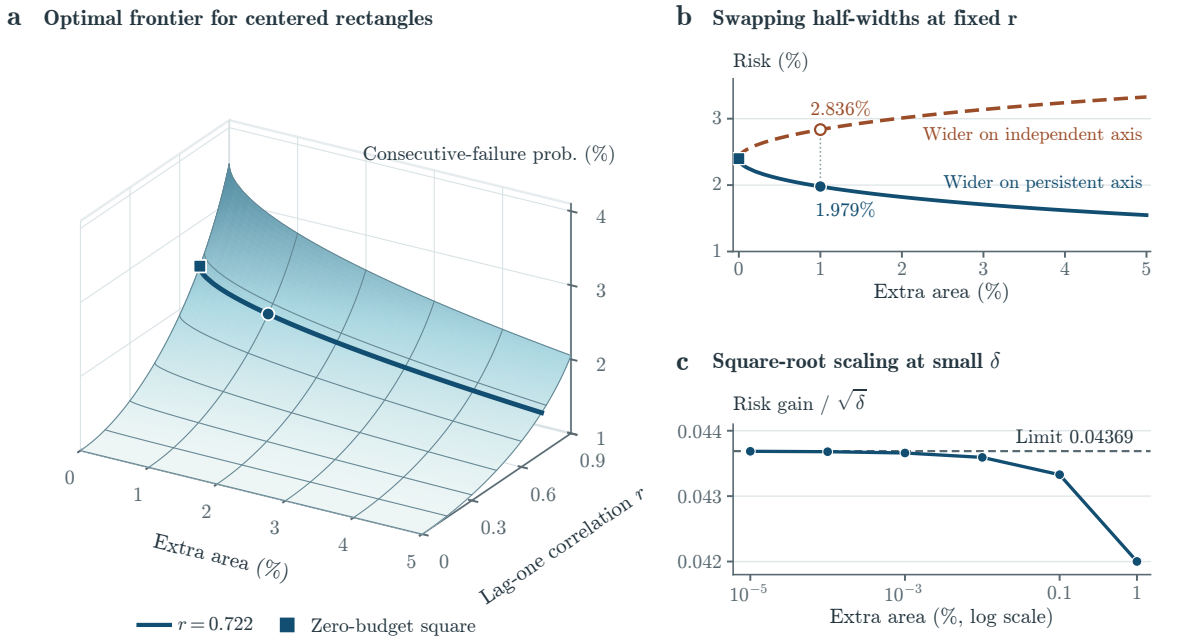

Fig. 2 Global rectangular frontier for the two-dimensional Gaussian model at 90% joint coverage, with lag-one correlation r on coordinate 1 and temporal independence on coordinate 2. **a** Minimum consecutive-failure probability across budgets and correlations, with the dark curve marking $r = 0.722$. **b** Swapping half-widths at fixed area for $r = 0.722$: solid widens coordinate 1, dashed widens coordinate 2, and the dotted segment links both designs at 1% extra area. **c** Absolute risk reduction divided by $\sqrt{\delta}$ from numerical integration, converging to the theoretical limit

Theorem 4 accounts for the swapped-rectangle contrast in the Introduction. Exchanging the first two half-widths preserves coverage and volume, yet moves the persistent half-width from a_δ down to $v(a_\delta) < a_\delta$. Combining (11) with $Q'_r < 0$ raises consecutive-failure risk strictly.

The small-budget coefficient also follows directly along the one-dimensional curve. Put $t = \log(a/q)$ and $y(t) = \log v(qe^t)$. Differentiating $g(x_0 + t) + (M-1)g(y(t)) = \log c$ yields $y'(0) = -1/(M-1)$. A second derivative gives log-volume increment $\kappa t^2/2 + O(t^3)$ with $\kappa = Mh_0/(M-1)$, while risk changes by $\beta t + O(t^2)$ with $\beta = c^2 q Q'_r(q) < 0$. Solving (12) gives

$$t = \sqrt{2\delta/\kappa} + O(\delta), \qquad R_2(q, \ldots, q) - R_2(a_\delta, v(a_\delta), \ldots) = |\beta|\sqrt{2/\kappa}\,\sqrt{\delta} + O(\delta).$$

When $r = 0$, the periods are independent with common marginals, so every fixed region of coverage c has risk $(1-c)^2$ and the frontier is flat.

Figure 2 checks the global solution and local limit under $M = 2$, $c = 0.90$, and $r = 0.722$. The two markers in panel b match the swapped rectangles from the Introduction. Varying r generates the surface in panel a, where the $r = 0.722$ slice tracks the optimal branch. In panel c, shrinking the extra area from 1% to 0.00001% raises the ratio of risk reduction to $\sqrt{\delta}$ from 0.04200 toward the theoretical limit 0.04369.

A sign randomization isolates mean forecasting from failure persistence. Multiplying each $Z_{i,t}$ by an independent Rademacher sign (± 1 with equal probability) independent of the error process makes the conditional mean given the past identically zero, so zero is the optimal squared-error point forecast. Absolute coordinates remain unchanged, so centered axis-aligned rectangles produce the exact same hit sequence and retain the full allocation gain of Theorem 4.

4.3 Allocating area starting from a disk

Keeping the same two-dimensional Gaussian process while replacing centered rectangles with centered axis-aligned ellipses tests whether the allocation direction transfers across geometry classes. Coordinate 1 retains lag-one correlation $r \in (0,1)$ and coordinate 2 remains white noise. Each class has its own minimum area at coverage c, so each starts from its own baseline.

Descent direction on the disk. Parameterize centered axis-aligned ellipses E_s by half-axes $q(s)e^s$ and $q(s)e^{-s}$, adjusting the common scale $q(s) > 0$ at each s to hold coverage fixed at c. Write $q_* = q(0) = \sqrt{-2\log(1-c)}$; the disk E_0 has area πq_*^2. Because the standard bivariate normal density decreases strictly with radius, exchanging any interior subset of E_0 for an equal-area exterior subset lowers mass, making the disk the unique minimum-area measurable set at coverage c up to null sets.

Proposition 2 (Stretching the disk along the persistent axis)**.** *In the two-dimensional model above at fixed coverage c,*

$$\log \frac{|E_s|}{|E_0|} = \tfrac{1}{2} q_*^2 s^2 + O(s^4), \qquad \beta_E = \left.\frac{dR_2(E_s)}{ds}\right|_{s=0} < 0.$$

Within centered axis-aligned ellipses, taking $s > 0$ to lengthen the first axis attains the optimal small-budget risk drop $|\beta_E|\sqrt{2\delta}/q_* + O(\delta)$ near the disk.

Positive s assigns extra length to the persistent coordinate. Larger current $|Z_{1,t}|$ lowers the conditional probability that the prior error fell inside the disk, and integrating around the circle yields a negative risk derivative; Appendix C details the proof.

Because both ellipses and general measurable sets start from the same disk $A = E_0$, we can compute the exact fraction of the all-shape leading gain captured by elliptical stretching. Let $h_* = (\bar{w} - w)/g$ be the optimal normal speed from Theorem 2, and let $h_E(\theta) = q_* \cos(2\theta)$ be the normal speed of an elliptical stretch. Equip boundary speeds with the weighted inner product $\langle u, v\rangle_g = \int_{\partial A} g\, uv\, dS$ and norm $\|u\|_g$. Preserving coverage forces zero circular integral. Along any such speed h, the first-order risk reduction is $\langle h_*, h\rangle_g$ and the second-order area cost is $\|h\|_g^2/(2\lambda)$. The ratio of square-root coefficients is the cosine between h_* and h_E:

$$\frac{\gamma_E}{\gamma_{\text{all}}} = \frac{|\langle h_*, h_E\rangle_g|}{\|h_*\|_g \, \|h_E\|_g}, \qquad \gamma_{\text{all}} = \sqrt{2\lambda V_A}\, \|h_*\|_g. \tag{13}$$

At $c = 0.90$ and $r = 0.722$, write $P_0(x)$ for the conditional probability that the prior error fell inside E_0 given current first coordinate x. On the circle, $\lambda = (1-c)/(2\pi)$, $g = q_*\lambda$, and $w(\theta) = 2\lambda P_0(q_* \cos\theta)$. Numerical quadrature gives $\gamma_{\text{all}} = 0.04259980484$ and $\gamma_E = 0.04245261381$. As the area allowance tends to zero from the disk, stretching along the persistent axis captures 99.65% of the optimal all-shape leading coefficient. Figure 3**c** displays both normalized velocities around the quadrant.

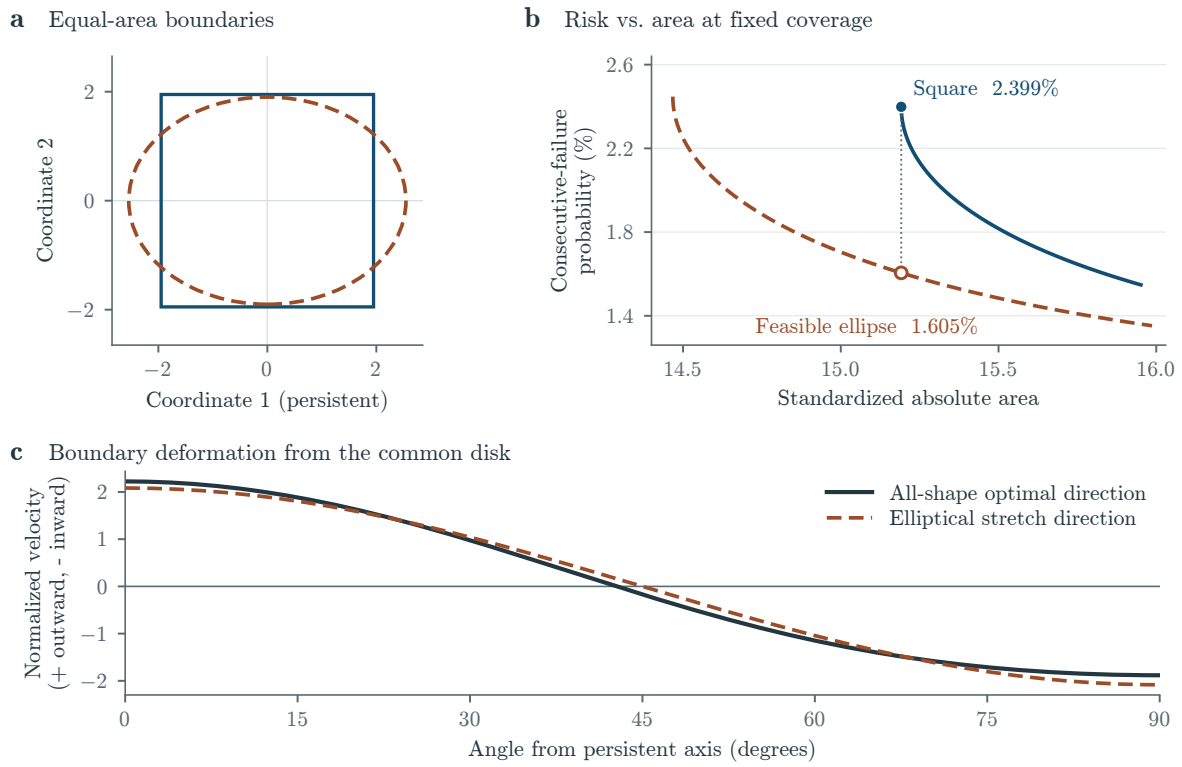

Fig. 3 Gaussian example at 90% joint coverage with lag-one correlation 0.722 on coordinate 1. **a** Minimum-area square and equal-area stretched ellipse on identical axes. **b** Optimal rectangular risk and feasible elliptical risk from their respective minimum areas; markers correspond to the two boundaries in **a**. **c** Optimal all-shape normal velocity and elliptical stretch velocity from the common disk, normalized to equal second-order area cost. Angles start from the persistent axis, with positive values moving outward and negative values inward; elliptical stretching attains 99.65% of the optimal leading coefficient

4.4 Comparing the two shape classes at equal area

Deterministic quadrature evaluates both the optimal rectangular branch and the stretched elliptical branch. Let a and d denote horizontal and vertical half-widths or half-axes, and let $m(x)$ be the conditional probability that the second coordinate falls in the vertical slice at x. For $|x| \le a$, rectangles have $m(x) = F(d)$ and ellipses have $m(x) = F(d\sqrt{1 - x^2/a^2})$, with $m(x) = 0$ outside $[-a, a]$. With standard normal density φ and bivariate normal density φ_r,

$$C = \int m(x)\varphi(x)\,dx, \qquad R_2 = 1 - 2C + \iint m(x)m(y)\varphi_r(x, y)\,dx\,dy.$$

Fixing the within-class area fixes the product ad. We solve the coverage equation along $s \ge 0$ and evaluate the double integral, substituting $x = a\sin v$ on ellipses to smooth the endpoint square root into $\cos v$. Using 96 and 144 Gauss–Legendre nodes per dimension gives risks in Table 1 that agree within 10^{-9}, and setting $r = 0$ returns $(1 - c)^2$ at every budget.

Figure 3 pairs finite-area comparisons with small-budget boundary directions. Reaching 90% coverage takes area 15.19163 for a square and 14.46757 for a disk. Spending that area difference to stretch the disk along the persistent axis yields an ellipse with the same area as the square while cutting consecutive-failure risk from 0.023985 to 0.016052. Returning to the disk, panel c shows that the outward and inward arcs selected by the all-shape optimum closely track those of elliptical stretching.

Table 1 reports the half-widths, half-axes, areas, and risks from each family's minimum area. Adding 1% within-class area lowers rectangular risk by 17.51% and elliptical risk along the stretched branch by 16.83%.

Table 1 Area budgets, half-widths or half-axes, and consecutive-failure probabilities ($c = 0.90$, $r = 0.722$). Allowances are relative to each class's minimum area; the first and last rows share the same absolute area.

Shape	Class allowance	Horizontal half-axis	Vertical half-axis	Area	R_2
Rectangle	0	1.94882	1.94882	15.19163	0.023985
Rectangle	1%	2.07368	1.84979	15.34354	0.019785
Rectangle	5%	2.26535	1.76034	15.95121	0.015479
Ellipse	0	2.14597	2.14597	14.46757	0.024472
Ellipse	1%	2.30316	2.01950	14.61224	0.020355
Ellipse	5%	2.54317	1.90134	15.19095	0.016055
Ellipse	5.0047%	2.54339	1.90126	15.19163	0.016052

Computational cost of one-dimensional root finding versus grid search. The globally optimal rectangle in Table 1 also illustrates the cost of grid search versus root finding. Fix $M = 2$, $c = 0.90$, $r = 0.722$, a 1% area allowance, and search interval $[0, 0.5]$ for $t = \log(a/q)$. Evaluating area across grids of 33, 129, and 513 points and taking the largest feasible a gives risks 0.020719, 0.019995, and 0.019819; the coarsest grid uses only 0.590% extra area. Exploiting the monotonicity in Theorem 4, Brent's method reaches risk 0.019785 in just 12 budget evaluations over the same interval. Both procedures perform one risk integration after locating a, so the savings come directly from solving for the budget boundary.

5 Finite-budget rectangular frontiers

When an application must issue separate upper and lower limits for each output, free cell-wise assemblies are not publishable. We switch to coordinate coverage probabilities to allocate widths within centered axis-aligned rectangles at finite volume budgets. Let the standardized error vector have $M \geq 2$ coordinates with a common single-period marginal across two adjacent origins and independent coordinates at each single origin. Write $F_i(a) = \mathbb{P}(|Z_{t,i}| \leq a)$, assuming each F_i is continuous and strictly increasing from 0 to 1 on $[0, \infty)$. For half-widths τ_i, introduce

$$\begin{aligned} p_i &= F_i(\tau_i), \qquad x_i = -\log p_i, \qquad s = -\log c, \\ w_i(x) &= \log\bigl[2F_i^{-1}(e^{-x})\bigr], \qquad W(x) = \sum_i w_i(x_i). \end{aligned} \tag{14}$$

Standardization keeps each logarithmic width dimensionless. Joint coverage c corresponds to $\sum_i x_i = s$, half-widths satisfy $\tau_i = F_i^{-1}(e^{-x_i})$, and $W(x)$ is the log-volume of the rectangle. Let W_0 be its minimum, achieved at x^0, and set $k = W_0 + \log(1 + \delta)$. Because each $w_i(x) \to \infty$ as $x \downarrow 0$, the minimum lies in the interior of the simplex.

Loss follows from the probability of hitting the region at both origins. Putting $U_{t,i} = F_i(|Z_{t,i}|)$ and

$$J(p) = \mathbb{P}\{U_{t-1,i} \leq p_i,\ U_{t,i} \leq p_i,\ 1 \leq i \leq M\},$$

gives $R_2 = 1 - 2c + J(p)$ on the equal-coverage simplex. When the two-period coordinate pairs $(U_{t-1,i}, U_{t,i})$ are mutually independent across i, the joint hit probability factors as $J(p) = \prod_i D_i(p_i)$, where $D_i(v) = \mathbb{P}(U_{t-1,i} \leq v,\ U_{t,i} \leq v)$ is the diagonal of coordinate i's two-period rank distribution. Every calculation below uses only two adjacent periods.

5.1 Global allocation under retain-and-redraw dynamics

We first examine a temporal dependence model admitting a direct global solution. Coordinate i retains its previous magnitude rank $U_{t-1,i}$ with probability λ_i and redraws independently from the uniform distribution on $[0,1]$ otherwise, starting from independent uniform ranks. Then $D_i(p) = \lambda_i p + (1-\lambda_i)p^2$, so

$$J(p) = c^2 \exp\left(\sum_i f_i(x_i)\right), \qquad f_i(x) = \log(1-\lambda_i+\lambda_i e^x).$$

Magnitude marginals govern the volume cost of coverage, whereas retention probabilities govern the impact on consecutive failures. Because retention places positive mass on equal adjacent magnitudes, the two-period law carries a singular diagonal part, yet using $J(p)$ directly still solves the finite-budget problem.

Theorem 5 (Finite-budget optimum for the retain-and-redraw model)**.** *Assume $0 < \lambda_i < 1$ and each w_i is differentiable and convex. Subject to $\sum_i x_i = s$ and $W(x) \le k$, the optimal allocation is unique. Put $[z]_+ = \max(z,0)$ and $\ell_i = \log(\lambda_i/(1-\lambda_i))$, and let a be the unique scalar satisfying*

$$\widehat{x}_i = [a-\ell_i]_+, \qquad \sum_i \widehat{x}_i = s. \tag{15}$$

If every $\widehat{x}_i > 0$ and $W(\widehat{x}) \le k$, then $\widehat{x}$ is optimal and further budget increases leave risk unchanged. Otherwise, for every positive budget the optimum satisfies $W(x) = k$. Multipliers $\mu > 0$ and $\nu \in \mathbb{R}$ then satisfy

$$\frac{\lambda_i e^{x_i}}{1-\lambda_i+\lambda_i e^{x_i}} + \mu w_i'(x_i) + \nu = 0. \tag{16}$$

If all F_i belong to a common scale family, $F_i(t) = F(t/b_i)$ with $b_i > 0$, then $\lambda_i > \lambda_j$ implies $x_i \le x_j$ at the optimum, so the standardized half-width τ_i/b_i is at least τ_j/b_j.

Because the transformed objective is strictly convex, first-order optimality identifies the unconstrained minimum; if that point exceeds the volume cap, the optimum lies on the budget boundary. Appendix C writes these conditions in coordinate-wise root-finding form.

Theorem 5 also tells when a budget stops helping. Take $M = 2$, $c = 0.90$, $F_i(t) = 1-e^{-t}$, and retention probabilities 0.5 and 0.505. Putting $v(x) = -\log(1-e^{-x})$ gives $w_i(x) = \log[2v(x)]$, whose second derivative shares the sign of $e^x v(x) - 1 > 0$ since $-\log(1-u) > u$ for $0 < u < 1$. Both width costs are identical and strictly convex, so $x^0 = (s/2, s/2)$. Solving (15) yields $\widehat{x} = (0.062681, 0.042680)$ and $\exp\{W(\widehat{x}) - W(x^0)\} - 1 \approx 0.008420$. Adding roughly 0.842% area already hits the unconstrained minimum risk. When every $\lambda_i = 0$, we get $J = c^2$ and every feasible rectangle has risk $(1-c)^2$.

For more general cross-period dependence with independent contemporaneous coordinates, Appendix C.3 subdivides the coverage simplex using the full joint hit probability J. When log-width costs are convex, every positive budget admits an explicit error bound that shrinks to zero under refinement.

5.2 Non-Gaussian global computation with three persistent coordinates

Two models sharing identical marginals but distinct temporal mechanics test finite-budget allocation in three dimensions. Three mutually independent coordinate processes each have magnitude distribution $F(t) = 1 - e^{-t}$ for $t \ge 0$; attaching independent symmetric signs gives unit-scale Laplace marginals, and half-widths use that scale. At joint coverage $c = 0.90$, both models share the same equal-width baseline, with half-widths 3.36649 and reference volume 305.22586.

The first model uses the retain-and-redraw mechanism of Theorem 5. Starting from uniform ranks $U_{i,t} = F(|Z_{i,t}|)$, each coordinate copies $U_{i,t-1}$ with probability λ_i and redraws uniformly

otherwise, giving $D_i(u) = \lambda_i u + (1-\lambda_i)u^2$. Setting $\lambda = (0.75, 0.35, 0.10)$ makes all three coordinates persistent while allowing exact magnitude repeats.

The second model replaces atoms on the diagonal with a smooth rank transition density. Coordinate processes remain independent across i, so J still factors into diagonal functions. Given prior rank u, let the next rank $v \in [0,1]$ have density $1 + \theta_i(1-2u)(1-2v)$ with $0 \le \theta_i \le 1$, which is nonnegative and preserves the uniform marginal. Integrating over $[0,u]^2$ yields the Farlie–Gumbel–Morgenstern diagonal $D_i(u) = u^2[1 + \theta_i(1-u)^2]$. Setting $\theta = (0.90, 0.45, 0.10)$ assigns zero probability to exact magnitude ties. Together, the two models test global allocation under identical marginals, coverage, and absolute volume budgets.

Single-axis versus joint allocation at the same budget. We compare three half-width rules at identical coverage and volume caps. The equal-width reference uses minimum volume. The single-axis rule widens the most persistent first coordinate while keeping coordinates 2 and 3 equal, solving along $x_1 \le s/3$ and $x_2 = x_3 = (s - x_1)/2$ up to the budget cap; this tests whether isolating only the most persistent coordinate suffices. The joint rule lets all three half-widths vary freely under the same volume cap.

Table 2 Population consecutive-failure probabilities for three-dimensional Laplace marginals at joint coverage 0.90, with relative volume allowance δ. Single-axis uses the full allowance on axis 1; joint gives the global minimum over all feasible centered axis-aligned rectangles, displayed to nine decimals.

Temporal structure	δ	Equal-width ref.	Single-axis	Joint allocation
Retain-and-redraw, heterogeneous	1%	.045129801	.041411211	.040985344
Retain-and-redraw, heterogeneous	5%	.045129801	.037516862	.036278096
Smooth transition, heterogeneous	5%	.011399430	.011044980	.010862990
Retain-and-redraw, $\lambda_i = 0.35$	5%	.040782325	.040825575	.040782325
Temporally independent, $\lambda_i = 0$	5%	.010000000	.010000000	.010000000

Joint allocation captures the persistence gap between the second and third coordinates. In Table 2, a 5% budget in the heterogeneous retain-and-redraw model sets volume cap 320.48715. Single-axis allocation reaches risk 0.037516862, whereas letting the three half-widths take values 3.92881, 3.43444, and 2.96895 cuts risk further to 0.036278096; the moderately persistent second coordinate also expands relative to the cube. The smooth transition model shows the same pattern. Two control rows confirm when the equal-width cube remains optimal. Under homogeneous retention ($\lambda_i = 0.35$), symmetry and strict convexity make equal x_i uniquely optimal, so spending extra budget only worsens risk for the single-axis branch while joint allocation retains the baseline volume. Under temporal independence ($\lambda_i = 0$), every fixed region at coverage c has risk $(1-c)^2$.

Convex duality bounds certify every global value in Table 2. At 5% budget, the certified interval width is 4×10^{-12} for the heterogeneous retain-and-redraw model and 10^{-12} for the smooth transition model, incorporating both optimization error and interval-arithmetic rounding. Appendix D lists the exact endpoints and verification steps.

5.3 Direct half-width computation under contemporaneous dependence

To allow contemporaneous correlation across coordinates, we operate directly on half-widths using single-period coverage and consecutive-failure probability. Assume both periods share the same marginal vector law, with each magnitude distribution F_i continuous and strictly increasing on $[0,\infty)$. For centered rectangle half-widths a, write $C(a)$ for joint coverage, $Q(a)$ for consecutive-failure probability, and $V(a) = 2^M \prod_i a_i$. Increasing any coordinate of a does not decrease C and does not increase Q, and marginal continuity makes both functions continuous. Evaluating C and Q uses the full joint law, allowing both contemporaneous and cross-period dependence.

Proposition 3 (Finite-budget rectangular frontier for a general joint law)**.** *Let H be an absolute volume cap and H_0 be the minimum volume of a centered rectangle achieving coverage c. For $H \geq H_0$, define $Q^\star(H) = \min\{Q(a) : C(a) = c,\ V(a) \leq H\}$. The box subdivision below yields bounds $L_h \leq Q^\star(H) \leq U_h$ with*

$$\lim_{h \downarrow 0} L_h = Q^\star(H).$$

If $H > H_0$ and $Q^\star$ is left-continuous at H, then $U_h \to Q^\star(H)$ as well; this holds at all $H > H_0$ outside an at most countable set of budgets.

Proof. Coordinate marginals bound the search domain. Putting $m_i = F_i^{-1}(c)$, coverage c requires $a_i \geq m_i$, and the volume cap forces $a_i \leq T_i$, where $T_i = H/\big(2^M \prod_{j \neq i} m_j\big)$. The feasible set is compact, so $Q^\star(H)$ is attained; similarly H_0 is attained by restricting search within the volume of an equal-width rectangle at coverage c.

Cover $\prod_i [m_i, T_i]$ with boxes $[l, u]$ of edge length at most h. Discard any box with $C(l) > c$, $C(u) < c$, or $V(l) > H$. On every surviving box, $Q(u)$ lower-bounds $Q(a)$, and taking the minimum over surviving boxes defines L_h. Along the diagonal from l to u, continuity yields a point y with $C(y) = c$; if $V(y) \leq H$, include $Q(y)$ as an upper-bound candidate. Since consecutive-failure risk does not exceed $1 - c$, define U_h as the minimum of $1 - c$, any known feasible rectangles, and all diagonal candidates. Every box holding an optimal $a^\star$ survives with $Q(u) \leq Q^\star(H)$, while every candidate corresponds to a feasible rectangle, yielding $L_h \leq Q^\star(H) \leq U_h$.

Take any convergent subsequence of boxes attaining L_h. As $h \downarrow 0$, both corners converge to a single point a satisfying $C(a) = c$ and $V(a) \leq H$ by continuity, so every cluster point of L_h is a feasible risk and equals $Q^\star(H)$. For any $H_0 \leq H' < H$, a box containing an optimal point for H' satisfies $V(u) < H$ once h is small enough, so its diagonal candidate yields $\limsup_{h \downarrow 0} U_h \leq Q^\star(H')$. Letting $H' \uparrow H$ yields convergence under left-continuity, and monotonicity of $Q^\star$ restricts jump points to a countable set. □

A checkable convexity condition makes both bounds converge at every $H > H_0$. Write $b_i = \log a_i$, and suppose $\{b : C(e^b) \geq c\}$ is convex while $t \mapsto C(ta)$ increases strictly for every positive a. Moving fraction $\eta \in (0, 1)$ from an optimal $\log a^\star$ at budget H toward the minimum-volume $\log a^0$ maintains coverage at least c and lowers volume to $V(a^\star)^{1-\eta} H_0^\eta < H$. Scaling all half-widths down equally to restore exact coverage c uses a factor converging to 1 as $\eta \downarrow 0$, approximating $a^\star$ by strictly sub-budget rectangles and giving left-continuity. For instance, if the log-magnitude vector $(\log |Z_{t,i}|)_i$ has an everywhere positive log-concave density, convexity of the orthant constraint $y \leq b$ and preservation of log-concavity under integration (Prékopa 1973) make $b \mapsto C(e^b)$ log-concave while allowing arbitrary coordinate correlation.

Proposition 3 separates theoretical optimality from finite-run certification. The lower limit matches $Q^\star(H)$ at every budget, and observing $U_h - L_h \leq \varepsilon$ terminates with certified tolerance ε. Left-continuity is sharp for upper-bound convergence. Appendix D constructs a two-coordinate nonconvex counterexample with a jump in $Q^\star$. In contrast, under log-concave densities or the convex log-width cost of Proposition 6, the frontier is continuous everywhere and stops in finitely many steps with explicit error bounds. Interval subdivision follows global optimization methods (Hansen 1979), relying here only on monotonicity of joint probabilities and the volume constraint.

A two-state example with contemporaneous dependence. A shared latent state creates contemporaneous and temporal dependence simultaneously. Let $S_t \in \{-1, 1\}$ be a stationary symmetric two-state Markov chain with lag-one correlation $\rho = 0.8$ (persistence probability 0.9). Conditional on the full state path, observations are independent across coordinates and time. Given $S_t = s$, coordinate i's magnitude rank has density $1 + s\eta_i(1 - 2u)$ on $[0, 1]$, with $(\eta_1, \eta_2) = (0.85, 0.25)$. Unconditional ranks are uniform, and magnitude marginals are $F_i(a) =$

$1 - e^{-a}$. Writing $p_i = F_i(a_i)$, conditional coverage in state s is $C_s = \prod_{i=1}^{2}[p_i + s\eta_i p_i(1 - p_i)]$. Setting $C = (C_+ + C_-)/2$ and $D = (C_+ - C_-)/2$ yields $Q = (1 - C)^2 + \rho D^2$. Because $C = p_1 p_2[1 + \eta_1\eta_2(1 - p_1)(1 - p_2)]$, joint coverage no longer factors across coordinates.

At $c = 0.90$ and absolute area cap $H = 45$, the equal-width reference has area 35.15371 and risk 0.01208407. Evaluating Q from state transition probabilities and subdividing half-width space inspects 1,251 boxes for tolerance 10^{-5} and 2,123 boxes for 10^{-6}, returning outward-rounded double-precision bounds $[0.01060631, 0.01060727]$.

Solving along the equal-coverage curve checks this result independently. Fixing a_1 makes $C = c$ quadratic in p_2. Setting $x = 1 - p_1$, $y = 1 - p_2$, and $k = \eta_1\eta_2$, we have $x, y \leq 0.1$ along the curve, and implicit differentiation gives $|dy/dx| \leq [c(1 - k(1 - c))]^{-1} < 1.14$. Differentiating $D/c = (\eta_1 x + \eta_2 y)/(1 + kxy)$ and multiplying by $(1 + kxy)^2$ yields a lower bound $\eta_1 - k\eta_2(1 - c)^2 - 1.14\eta_2 > 0$, so widening coordinate 1 lowers risk. At fixed area, stretching from the square toward axis 1 decreases both $F_1(a_1)F_2(a_2)$ (by strict concavity of $\log F_i(e^t)$) and $1 + ke^{-a_1-a_2}$. The budget boundary with larger a_1 is therefore globally optimal. Root finding yields half-widths $(4.71566, 2.38567)$ and risk 0.01060725; swapping the two half-widths keeps coverage and area intact while raising risk to 0.01419377. Common-factor models supply a related benchmark for our simulations: equal loadings yield zero first-order slope at the square by exchange symmetry, whereas unequal loadings allow small stretches to lower risk at fixed coverage, as worked out in Appendix C.4.

6 Selecting prediction regions from finite history

Population integration identifies the allocation direction when the error law is known, whereas real data provide only a finite residual history. Selecting both a minimum-volume reference and a minimum-risk competitor from the same candidate library and evaluating them out of sample tests directly whether temporal ordering helps shape selection.

How accurately finite history determines the all-shape frontier depends on its information about the target two-period law. If the true and estimated average marginals agree and the two-period total-variation error is e, optimal risk differs by at most e at every feasible budget. When marginals are also estimated, patching coverage inside a finite-volume reserve with surplus probability mass yields a two-sided bound with an explicit volume shift. Appendix F derives this total-variation stability bound, gives estimation rates from adjacent residual pairs, and presents a minimax lower bound showing that cross-period information is needed to identify the future frontier. Below we evaluate finite-sample rules in factor simulations and six multivariate forecasting benchmarks.

6.1 Selecting a fixed rectangle on a validation sample

Separating a scalar radius from normalized coordinate weights decouples size calibration from shape choice. Standardize each coordinate by its residual scale so half-widths and volumes are dimensionless. Set half-width i to qw_i with $q > 0$, $w_i > 0$, and $\prod_i w_i = 1$, giving volume $(2q)^M$. Defining the score $s_w(z) = \max_i |z_i|/w_i$ makes the failure event $s_w(z) > q$, so one scalar quantile calibrates every candidate shape.

On an evaluation window of T origins, empirical coverage is $1 - T^{-1}\sum_{t=1}^{T} I_t$ and consecutive-failure frequency is $(T - 1)^{-1}\sum_{t=2}^{T} I_{t-1}I_t$. A run of three consecutive misses contributes two adjacent pairs.

A design split builds the shape library, and a separate calibration split sets each shape's radius. The library contains the equal-width shape, single-axis stretches and contractions, and candidates moving along two data-driven directions: coordinate-wise lag-one autocorrelation of absolute residuals and log mean absolute residual. Each direction is centered in log-half-width space and applied in both signs at five strengths $(0.03, 0.06, 0.12, 0.24, 0.40)$. Given n calibration

scores, the radius is the $\lceil 0.90(n+1) \rceil$-th order statistic (or $+\infty$ if the rank exceeds n). The library contains at most $1 + 10(M+2)$ candidates, and scoring n residuals of dimension M per candidate costs $O(nM)$.

A validation split first locks the volume reference and then screens shapes within the budget. Step 1 picks the smallest-volume candidate j_0 with validation coverage in $[0.88, 0.92]$ (or closest to 0.90 if none falls in that band). Step 2 retains candidates whose volume is at most $(1+\delta)$ times that of j_0 and whose validation coverage stays within 0.005 of j_0, selecting the one with lowest validation consecutive-failure frequency (breaking ties by smaller volume, then fixed index). Freezing shape and radius before testing holds the volume ratio exact throughout the test period while letting test coverage reflect subsequent realizations.

Two estimation tolerances bound validation selection accuracy. If every candidate's coverage error is at most e_C and risk error is at most e_R, the chosen shape's population coverage differs from the reference by at most $0.005 + 2e_C$. If $2e_C \leq 0.005$, every budgeted candidate within population coverage gap $0.005 - 2e_C$ passes screening, and the selected shape's population risk exceeds that of the best passing candidate by at most $2e_R$.

Lemma 2 in Appendix E gives uniform simultaneous bounds for e_C and e_R over all fixed centered rectangles under stationary β-mixing errors with continuous magnitude marginals. Because the bound holds uniformly across all $T-1$ adjacent validation pairs, it applies directly to data-selected shapes. Resolving gains of order $\sqrt{\delta}$ requires both a candidate library containing a strong feasible shape and estimation noise small enough to separate it from the reference.

6.2 Shape selection in common-factor simulations

Extending the common-factor model of Appendix C.4 across dimensions tests fixed-shape selection from finite samples. Heterogeneous configurations space factor loadings a_i evenly from 0.95 down to 0.10, using $\rho = 0.8$ for persistent cases and $\rho = 0$ for temporal independence. Each setting runs 30 seeds, discarding 300 burn-in steps and splitting subsequent draws chronologically into 1,500 design, 1,500 calibration, 2,000 validation, and 8,000 test observations. Moderate heterogeneity uses loadings $(0.8, 0.3)$, homogeneous uses $(0.6, 0.6)$, and the heavy-tailed setting multiplies each vector by an independent factor $\sqrt{3/\chi^2_5}$. Paired bootstrap resampling of the 30 seeds (3,999 replicates) constructs the intervals in Table 3.

Table 3 Test performance of fixed-shape selection at a 1.05× volume cap relative to the library reference. C denotes joint coverage and R_2 denotes consecutive-failure frequency. The final column gives relative risk reduction from seed-mean ratios and paired 95% bootstrap intervals.

Configuration	C_0	C_{sel}	$R_{2,0}$	$R_{2,\text{sel}}$	Reduction and 95% CI (%)
Heterogeneous, $M = 2$	.897	.898	.0243	.0205	15.4 [11.3, 19.6]
Moderate heterogeneity, $M = 2$	.897	.897	.0186	.0164	12.2 [8.4, 15.8]
Homogeneous, $M = 2$	.898	.899	.0171	.0168	1.4 [−0.4, 3.2]
Temporally independent, $M = 2$	.898	.900	.0104	.0099	4.1 [0.2, 8.2]
Heterogeneous, $M = 5$	.894	.895	.0228	.0202	11.7 [8.7, 14.7]
Heterogeneous, $M = 10$	.897	.897	.0223	.0207	7.1 [4.3, 10.0]
Random signs, $M = 2$	.897	.898	.0243	.0205	15.4 [11.4, 19.6]
Heavy-tailed, $M = 2$	.898	.899	.0162	.0147	9.4 [5.8, 12.9]

Comparing chronological data against shuffled residual vectors isolates the role of temporal ordering. In the two-dimensional heterogeneous setting, the selected rectangle spends only 1.70% extra volume on average while reducing consecutive-failure frequency from 0.0243 to 0.0205 (Table 3). Shuffling row order within the design and validation blocks while keeping calibration and test blocks intact raises test consecutive-failure frequency to 0.0236. Shuffling preserves every cross-sectional vector and destroys only the temporal information visible during selection. Gains persist in five and ten dimensions at 11.7% and 7.1%.

Two controls verify dependence on temporal structure. Under homogeneous loadings, the 95% interval crosses zero. Under temporal independence, a slight coverage uptick accounts for the raw risk change: writing $\widehat{\pi} = 1 - \widehat{C}$, the seed-level difference in excess consecutive-failure frequency $\widehat{R}_2 - \widehat{\pi}^2$ averages -0.000091 with 95% interval $[-0.000328, 0.000136]$.

Randomizing coordinate signs tests sign invariance. Because the selector uses only absolute residuals, multiplying each coordinate by independent random signs preserves every failure path of the Gaussian heterogeneous setting, so both rows share the same underlying hit events.

6.3 Per-origin region choice under a fixed volume

6.3.1 Conditional risk of an adjacent pair

Holding the previously issued region fixed, we choose how to deploy the current origin's volume. Let $\mathcal{F}_o$ denote information available before issuing the forecast at origin o, Z_o the standardized error of this h-step forecast once observed, and I_{o-1} the failure indicator of the previous origin's forecast. Although the previous region is fixed, I_{o-1} is usually still unobserved when $h > 1$. Consider the single adjacent pair $(o-1, o)$.

Proposition 4 (Optimal region for one adjacent pair)**.** *Assume Z_o has conditional density f_o given $\mathcal{F}_o$, and the current volume allowance $v_o \in (0, \infty)$ is known prior to release. Define the subdensity m_o via*

$$\int_D m_o(z)\, dz = \mathbb{P}(I_{o-1} = 1,\ Z_o \in D \mid \mathcal{F}_o) \tag{17}$$

for measurable D. Among all $\mathcal{F}_o$-measurable regions of volume v_o, any set $B_o^\star$ satisfying

$$\{m_o > \ell_o\} \subseteq B_o^\star \subseteq \{m_o \geq \ell_o\}, \qquad \mathrm{Vol}(B_o^\star) = v_o$$

minimizes the conditional consecutive-failure probability of this pair, where $\ell_o \geq 0$ is chosen to meet the volume constraint.

Proof. For any candidate B, the conditional pair risk is $\mathbb{P}(I_{o-1} = 1 \mid \mathcal{F}_o) - \int_B m_o(z)\, dz$, so volume should go to regions of high m_o. Because $m_o \geq 0$ has total integral at most 1, choose ℓ_o so $\{m_o > \ell_o\}$ has volume at most v_o and $\{m_o \geq \ell_o\}$ has volume at least v_o. If a tie-breaking slice on $\{m_o = \ell_o\}$ is needed, collect points in increasing distance from the origin; spheres have zero volume, filling v_o exactly using only $\mathcal{F}_o$. For every equal-volume competitor B,

$$\int_{B_o^\star} m_o - \int_B m_o = \int (\mathbf{1}_{B_o^\star} - \mathbf{1}_B)(m_o - \ell_o)\, dz \geq 0,$$

where the sign of the integrand follows directly from the threshold inequalities. □

Factoring $m_o(z) = f_o(z)\, \mathbb{P}(I_{o-1} = 1 \mid Z_o = z, \mathcal{F}_o)$ makes the weighting transparent. For multi-step forecasts, the factor gauges how likely a prior miss becomes when $Z_o = z$. For one-step forecasts whose previous outcome has already arrived, $m_o = I_{o-1} f_o$: after a miss, maximizing current conditional coverage minimizes this pair's risk, whereas after a hit, this pair's loss is already zero. Only the current adjacent pair's loss is counted here. At any horizon, if I_{o-1} and Z_o are conditionally independent given $\mathcal{F}_o$, the weight is likewise independent of z.

6.3.2 Rectangular half-widths in a conditional Gaussian model

We next maximize current conditional coverage over centered rectangles at fixed volume. Suppose that given $\mathcal{F}_o$, coordinates are independent with $Z_{i,o} \sim N(0, \sigma_{i,o}^2)$ and $\sigma_{i,o} > 0$. Writing the known volume as $v_o = (2q_o)^M$ with $q_o > 0$ constrains half-widths to $\prod_i \tau_{i,o} = q_o^M$. The unique optimal half-widths are

$$\tau_{i,o}^\star = q_o \frac{\sigma_{i,o}}{\left(\prod_{j=1}^M \sigma_{j,o}\right)^{1/M}}. \tag{18}$$

Comparing standardized half-widths gives uniqueness directly. Put $x_i = \log(\tau_{i,o}/\sigma_{i,o})$, whose sum is fixed by the volume constraint. With $F(t) = 2\Phi(t) - 1$, conditional log-coverage is $\sum_i \log F(e^{x_i})$. By the strict concavity in Lemma 1, this sum reaches its unique maximum when all x_i match, yielding (18). The same argument extends to any conditionally independent common-scale family whose standardized magnitude distribution F makes $x \mapsto \log F(e^x)$ strictly concave. Standard Laplace errors, for instance, have $F(t) = 1 - e^{-t}$, whose derivative $t/(e^t - 1)$ $(t = e^x)$ decreases strictly in x. When scales evolve over time, practical rules estimate their relative sizes from data.

When m_o is a positive scalar multiple of f_o, equation (18) also minimizes the pair risk among centered axis-aligned rectangles. Dropping the rectangular constraint under that condition returns the conditional Gaussian level set from Proposition 4, namely an ellipsoid scaled by $\sigma_{i,o}$. Below we estimate relative scales from matured residuals and evaluate observed consecutive failures.

6.3.3 Estimating relative scales from matured residuals

In real forecasting, coordinate scales come from residuals whose target values have already arrived, which we call matured residuals. At horizon h, the newest matured forecast at origin o was issued at $o-h$. Given a unit-product reference shape w_0, let $q_{0,o}$ be the empirical quantile of the latest W matured s_{w_0} scores up to $o-h$. Window W is chosen on validation from $\{64, 128, 256\}$ using equal-width rolling scores, and the rolling reference shape is chosen from the library for lowest mean log-volume subject to the coverage rule. Pairing a fixed candidate shape with its rolling radius $q_{j,o}$ also allows a pre-release volume gate $(q_{j,o}/q_{0,o})^M \le 1 + \delta$, reverting to the reference if the gate fails so the volume cap holds at every origin.

To adapt widths dynamically without changing volume, we tilt the rolling reference according to recent scale ratios relative to design-period scales. Let $m_{i,o}$ be the mean of the latest k matured absolute residuals up to $o - h$, and let b_i be the design-split median absolute residual. Set

$$d_{i,o} = \mathrm{clip}_{[-2,2]}\left(\log \frac{m_{i,o} + 0.1}{b_i + 0.1}\right), \qquad \bar{d}_o = M^{-1} \textstyle\sum_i d_{i,o},$$
$$\tau_{i,o} = q_{0,o} w_{0,i} \exp\{\lambda(d_{i,o} - \bar{d}_o)\}. \tag{19}$$

The constant 0.1 sits in standardized residual units, and clipping caps single-step shifts. Subtracting the cross-sectional mean $\bar{d}_o$ keeps the sum of log-half-width adjustments at zero, holding per-origin volume identically equal to the reference. On validation, we screen $(k, \lambda) \in \{1, 4, 8\} \times \{0, 0.03, 0.06, 0.12, 0.24\}$, retaining pairs within 0.005 of reference coverage and selecting the lowest consecutive-failure frequency (breaking ties by smaller λ, then smaller k). Equation (19) applies shrinkage toward estimated scales, whereas (18) uses full conditional scales to maximize current coverage directly.

6.3.4 Checking the adjustment direction under known scales

Two alternating volatility regimes give an exact controlled benchmark. At each step, the regime switches with probability 0.03 between conditional standard deviations $(2, 0.5)$ and $(0.5, 2)$ with independent standard normal innovations; a constant $(1, 1)$ process is included as control. An equal-width reference draws its radius from the latest 128 matured scores, and every rule uses that identical radius so per-origin volumes match exactly. The feasible rule reads only matured residuals and tunes (19) on validation. The oracle benchmark reads the true current scales and applies (18), maximizing conditional coverage directly without the 0.005 validation screening rule.

Holding residual paths fixed while varying feedback delay isolates staleness from process dynamics. Each setting runs 30 replications of 12,000 steps: 2,000 design, 4,000 validation

(trimming the final $h-1$ immature origins), and 6,000 test steps. At each test origin, conditional coverage $\prod_i F(\tau_{i,o}/\sigma_{i,o})$ is evaluated from the true state and averaged over origins and seeds.

Matured residuals respond distinctly across the two scale structures. Under alternating volatility at $h = 1$, recent-residual reallocation raises conditional coverage by 0.00425 on average; delaying feedback to $h = 24$ trims that gain to 0.00139. Under constant equal scales, the equal-width reference is already conditionally optimal (0.90775), and finite-sample estimation noise slightly nudges recent-residual coverage to 0.90703 ($h = 1$) and 0.90693 ($h = 24$), still staying above the 0.90 target (Table 4).

Table 4 Equal-volume conditional coverage under two volatility regimes. The final column reports Monte Carlo standard errors across seeds for recent-residual minus reference coverage; matching seeds share identical residual paths across both lags.

Scale structure	h	Equal-width ref.	Recent residuals	Known state	Gain SE
Alternating volatility	1	.90731	.91155	.99791	.00027
Alternating volatility	24	.90732	.90871	.99791	.00018
Constant equal scale	1	.90775	.90703	.90775	.00018
Constant equal scale	24	.90775	.90693	.90775	.00018

7 Equal-volume comparisons on real forecast series

7.1 Evaluation units and baselines

We evaluate shape reallocation across six multivariate time-series datasets: ETTh1, ETTh2, ETTm1, and ETTm2 with seven outputs each (Zhou et al. 2021), plus Electricity and Traffic with three fixed 20-output panels each (Lai et al. 2018). Testing horizons $h \in \{1, 24\}$ with direct ridge regression, training-offset seasonal naive forecasts, and random Fourier feature ridge regression produces 60 panel tasks in total. Both regressions look back 96 observations; ridge uses penalty 1, while random Fourier features use 150 features, kernel width 0.02, and penalty 20. Seasonal periods equal one day (24 for hourly series, 96 for 15-minute series).

Information boundaries stay identical across methods at every test origin. Chronological train, calibration, validation, and test blocks take 50%, 15%, 15%, and 20% of each series, trimming the last $h-1$ immature origins from each of the first three blocks. Residuals are standardized by calibration-split sample standard deviations plus 10^{-9}. The first half of calibration constructs shape directions and the second half calibrates scores, with all rules sharing warm-up windows, point forecasters, and historical information. For each forecaster and horizon, metrics are averaged over panels within each dataset and then averaged with equal weight across the six datasets; the three panels of Electricity or Traffic share a common timeline. Relative risk percentages below are ratios of dataset-averaged risks.

Fixed-radius and rolling-radius baselines isolate scale calibration from shape allocation. Fixed baselines comprise equal-width rectangles, Bonferroni splits, and discrete error-budget optimization following the DCP-MT volume objective (Schlembach et al. 2025). The basic rolling reference updates only the scalar score quantile. Adding interleaved adaptive conformal inference (ACI) modulo h, along with a recent-score grouped variant (Gibbs and Candès 2021), measures threshold feedback effects.

Predictive threshold baselines further test scalar adaptation. Proportional-integral (PI) control combines quantile tracking with cumulative coverage error (Angelopoulos et al. 2023), and our linear scorecast variant forecasts the next score via ridge regression on the three newest matured scores in the same residue class. Adaptive threshold baselines are tuned on validation by finite-prediction rate, coverage compliance, and pinball loss. To isolate coordinate width allocation cleanly, we match per-origin volume identically between the rolling volume reference and

recent-scale adjustment, choosing shapes via the volume, coverage, and consecutive-failure rule of Section 6.

7.2 Total-volume calibration and equal-volume width allocation

Table 5 separates total-volume calibration from cross-axis width allocation. Static fixed-threshold rectangles suffer distribution shift and drop below 86% test coverage; updating scalar quantiles over time in the rolling volume reference restores coverage to 0.9047 and lowers consecutive-failure frequency to 0.02150. Switching among static library shapes with a per-origin volume gate (5% budget) fails to track time-varying relative axis scales during testing, giving risk 0.02249. By contrast, dynamically updating aspect ratios from recent matured residuals (recent-scale equal-volume adjustment) keeps per-origin volume strictly identical to the rolling reference (both have mean log-volume 23.24 and zero full-space rate $f_\infty = 0$). At the same time it cuts consecutive-failure frequency further to 0.01754 and lifts coverage to 0.9110.

As scalar-threshold references, ACI and PI with linear scorecasting reduce consecutive-failure frequency by varying total volume over time (reaching 0.01458 for the latter). They nevertheless output the full space on 0.4489% and 0.0016% of test origins, giving mean log-volume $+\infty$. In the bottom block of Table 5, the rolling volume reference and recent-scale adjustment release finite rectangles of exact equal volume at every single origin, cleanly isolating the net gain of width allocation.

Table 5 Calibration baselines versus equal-volume width allocation on one-step ridge forecasts. Here f_∞ is the fraction of origins issuing the full space. Metrics are averaged within datasets and then equally across all six datasets; no rule issued an empty set.

Rule	Coverage	Pair risk	Mean log-vol.	f_∞ (%)
Static fixed-threshold baselines (frozen radius and shape)				
Fixed equal-width	.8695	.04232	21.81	0
Bonferroni	.8922	.03102	22.68	0
Discrete error-budget allocation	.8706	.04310	21.91	0
Fixed-library volume reference	.8592	.04555	21.56	0
Fixed-library risk selection, 5%	.8597	.04480	21.56	0
Variable-volume scalar threshold baselines (changing total size)				
Equal-width rolling threshold	.9043	.02290	23.47	0
Interleaved ACI grouped by recent score	.8930	.01881	$+\infty$	.4489
PI	.8988	.02091	23.10	0
PI with linear scorecasting	.8967	.01458	$+\infty$	.0016
Finite-volume reference and equal-volume width allocation (zero full-space)				
Rolling volume reference	.9047	.02150	23.24	0
Per-origin volume gate, 5% (static shape switch)	.9021	.02249	23.18	0
Recent-scale equal-volume adjustment (ours)	**.9110**	**.01754**	**23.24**	**0**

7.3 Equal-volume gains and coverage shifts

Fixing per-origin volume first isolates real-series allocation effects. On one-step ridge forecasts, recent-residual adjustment reduces six-dataset average pair risk from 0.02150 to 0.01754 (an 18.39% drop) while raising joint coverage from 0.9047 to 0.9110.

Higher coverage lowers consecutive misses mechanically, so we decompose each panel into $\widehat{R}_2 = \widehat{\pi}^2 + (\widehat{R}_2 - \widehat{\pi}^2)$ before averaging. The baseline $\widehat{\pi}^2$ drops from 0.00911 to 0.00794, while the excess term $\widehat{R}_2 - \widehat{\pi}^2$ falls from 0.01239 to 0.00960, accounting for 29.5% and 70.5% of the total risk reduction. Both terms represent an algebraic breakdown of observed frequencies. Under

equal dataset weighting, the mean longest consecutive-failure streak per panel shortens from 7.44 to 6.44 origins.

Paired circular block bootstrapping assesses dataset-level sampling variability. Forming paired per-origin loss differences and drawing circular blocks of length 24 across 399 replicates on one-step ridge forecasts gives a risk difference of -0.00722 (95% CI $[-0.01155, -0.00345]$) on ETTh1 and -0.00260 (95% CI $[-0.00492, 0.00000]$) on ETTh2.

To examine failure clustering independently of coverage level, we match test failure counts panel by panel. For each panel, divide each origin's residual by its adjusted half-widths, take the coordinate maximum ratio, and set the threshold so that the count of exceedances equals that of the rolling reference. This post-hoc diagnostic scales every issued rectangle in the panel by one common constant. Macro-averaged pair risk becomes 0.01940, remaining 9.77% below the same-coverage reference (0.02150), at a mean relative volume ratio of 0.813. Figure 4 aligns both comparisons at zero. Five of the six datasets keep fewer consecutive failures after count matching; Traffic reverses sign, moving from 2.29 fewer pairs per thousand to 0.29 more. Online Resource 1 (Table S2) records exact per-dataset values.

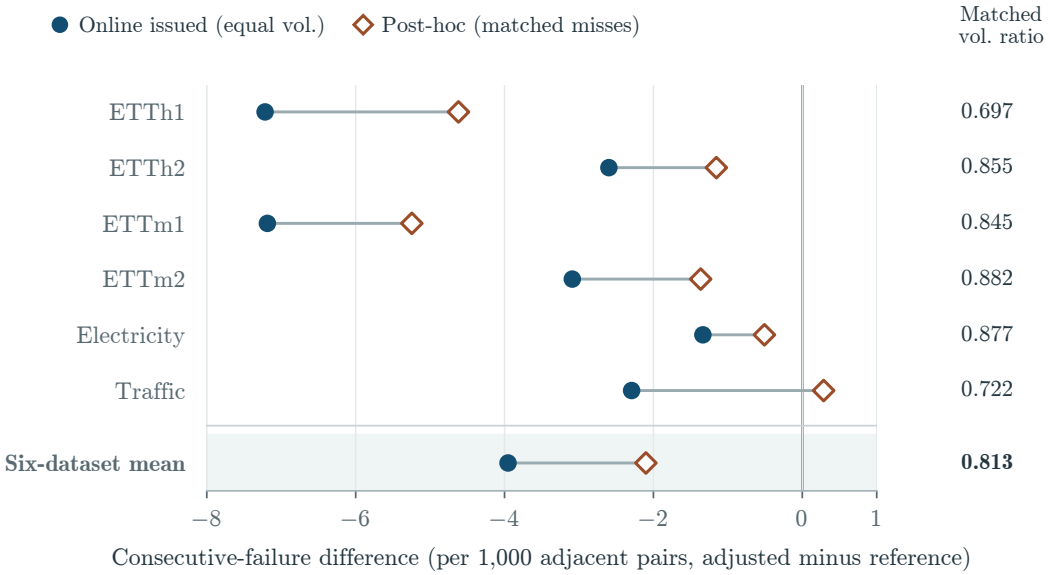

Fig. 4 Consecutive-failure frequency differences for one-step ridge forecasts relative to the rolling reference. Negative values indicate fewer consecutive failures per 1,000 adjacent forecast pairs. Circles show per-origin equal-volume online releases; diamonds show post-hoc panel-matched failure counts; the right column lists matched-to-reference volume ratios. Metrics are averaged over panels within each dataset before taking the equal-weighted six-dataset mean in the bottom row

Across the remaining forecasters and 24-step forecasts, equal-volume width adjustment reduces six-dataset average pair risk in all six (forecaster $\times$ horizon) combinations. One-step seasonal naive and random Fourier forecasts drop by 1.37% and 0.25%. The 24-step ridge, seasonal naive, and random Fourier forecasts drop by 0.89%, 3.16%, and 1.10%. The 24-step gain is smaller than the one-step gain. Feedback lag $h = 24$ weakens the ability of matured residuals to track current relative scales, consistent with the control experiment in Table 4. Multi-step forecast errors also share $h - 1$ common future shocks; for example, adjacent error correlation reaches $(h - 1)/h$ under a random walk with independent standard normal increments. The control design in Table 4 provides a reference that alters only feedback age.

8 Discussion

This paper formulates consecutive-failure risk at fixed joint coverage as a volume-budgeted frontier. Given any two-period joint error law, nested cell-wise partitioning approximates the risk of arbitrary measurable regions while keeping exact coverage and volume feasibility. The framework handles non-Gaussian marginals, contemporaneous cross-output dependence, and exact ties across adjacent periods. When single-period atoms exist, including an atom shifts coverage in discrete jumps; atomless marginals yield continuous frontiers across budgets, whereas atoms can introduce discontinuities.

Practical shape restrictions carry directly measurable volume costs. In the three-block density example, forcing a connected interval requires paying to cover a low-density gap before reaching the third block, whereas a two-piece set allocates length directly to risk reduction. In the smooth logistic example, a single interval already achieves the all-shape frontier across every budget. The Gaussian example sits in between: starting from the minimum-area disk, stretching an ellipse along the persistent axis captures 99.65% of the optimal all-shape small-budget gain. Every comparison holds coverage identical and reports explicit volume.

8.1 Tightness of theoretical bounds and online width allocation

Both the approximation error bound and the small-budget power laws developed here are tight, while the finite-history stability bound is paired with a minimax lower bound showing the necessity of cross-period information. For global numerical approximation, Proposition 1 bounds the two-sided error uniformly across all feasible budgets by $e_{\mathcal{P}}$, and the sinusoidal perturbation in Section 2.2 shows that the constant $1/4$ does not admit universal reduction. The smooth Fourier construction in Appendix A.2 further shows why defining the deterministic frontier as an infimum is necessary. Locally, Theorem 2 and equation (8) characterize the $\delta^{1/(k+1)}$ power law for odd-order boundary crossings, while the infinite-order flat-density example in Appendix B.2 shows that finite-order crossing is required for any positive power-law rate. For finite-history estimation and deployment, Proposition 7 quantifies how total-variation error in the joint law propagates to the whole frontier, and the minimax lower bound in Appendix F shows that cross-period dependence information is required to identify the future frontier. Faced with time-varying scale structures in real multivariate series, the matured-residual equal-volume width adjustment in Sections 6.3 and 7 lowers consecutive-failure risk while preserving finite volume and target coverage without adding any volume budget.

In our view, separating per-origin volume control from cross-axis width allocation is the right operational division for online multivariate prediction regions. Rolling scalar quantiles maintain target coverage under drift. Equal-volume width updates capture temporal persistence without risking infinite prediction sets. We believe that practitioners constrained to coordinate-wise intervals gain more from adaptive width reallocation than from static library switching when relative residual scales evolve over time.

Statements and Declarations

Funding. The authors did not receive support from any organization for the submitted work.

Competing interests. The authors have no competing interests to declare that are relevant to the content of this article.

Author contributions. Both authors contributed to the study conception, methodological development, theoretical analysis, and numerical evaluation. Both authors reviewed and finalized the manuscript.

Data availability. The six real multivariate benchmark series come from the public ETDataset repository (`ETTh1`, `ETTh2`, `ETTm1`, `ETTm2`) and the LSTNet archive (`Electricity`, `Traffic`) (Zhou et al. 2021; Lai et al. 2018). The electronic supplementary source package (`ESM`, Online Resource 2) contains test-origin indices (`evidence/test-origins.csv`), SHA-256 checksums (`results/environment.json`), rational enumerations, quadrature grids, and per-seed records for factor and two-state simulations. It also includes all source CSV and frozen JSON/NPZ files underlying every figure and table in the main manuscript and Online Resource 1.

Code availability. All Python programs reproducing the numerical results and figures are provided in the accompanying electronic supplementary source package (`ESM`, Online Resource 2), verified under `Python 3.9.13`, `NumPy 2.0.2`, `SciPy 1.13.1`, `Matplotlib 3.9.4`, and `mpmath 1.3.0`. For theoretical frontiers, `all_budget_example.py` and `scope_grid_solver.py` compute exact rational frontiers and finite-partition bounds, `flat_checks.py` runs 65- and 85-digit third-order quadrature, and `boundary_checks.py` and `frontier.py` evaluate leading coefficients and Gaussian elliptical efficiency. For global rectangles and empirical tasks, `global_examples.py` and `joint_box.py` produce interval-certified rectangular frontiers, while `factor.py`, `empirical_checks.py`, `pi_matched.py`, and `pi_paired.py` run the simulations and real-series diagnostics (including ACI and PI baselines). Vector figure scripts sit in `figures/`, and full reproduction commands and file manifests appear in `REPRODUCE.md` and `MANIFEST.sha256`.

Supplementary information. Online Resource 1 (`ESM_1.pdf`): Complete mathematical proofs for all-shape frontiers, boundary power laws, parametric and rectangular global algorithms, finite-history validation and transfer bounds, and per-dataset tables (Appendices A–G, Tables S1–S2). **Online Resource 2** (`ESM_2.zip`): Reproducible Python source package, frozen result records, checksums, and vector figure scripts.

References

Angelopoulos, A. N., Candès, E. J., Tibshirani, R. J.: Conformal PID control for time series prediction. Adv. Neural Inf. Process. Syst. 36, 23047–23074 (2023). https://doi.org/10.52202/075280-1000

Borgs, C., Chayes, J. T., Lovász, L., Sós, V. T., Vesztergombi, K.: Convergent sequences of dense graphs II. Multiway cuts and statistical physics. Ann. Math. 176, 151–219 (2012). https://doi.org/10.4007/annals.2012.176.1.2

Bröcker, J.: Uniform calibration tests for forecasting systems with small lead time. Stat. Comput. 32, 102 (2022). https://doi.org/10.1007/s11222-022-10144-9

Christoffersen, P. F.: Evaluating interval forecasts. Int. Econ. Rev. 39, 841–862 (1998). https://doi.org/10.2307/2527341

Di, J., Kolaczyk, E.: Complexity-penalized estimation of minimum volume sets for dependent data. J. Multivar. Anal. 101, 1910–1926 (2010). https://doi.org/10.1016/j.jmva.2010.04.009

Froyland, G.: Statistically optimal almost-invariant sets. Phys. D 200, 205–219 (2005). https://doi.org/10.1016/j.physd.2004.11.008

Gibbs, I., Candès, E. J.: Adaptive conformal inference under distribution shift. Adv. Neural Inf. Process. Syst. 34, 1660–1672 (2021)

Hansen, E. R.: Global optimization using interval analysis: The one-dimensional case. J. Optim. Theory Appl. 29, 331–344 (1979). https://doi.org/10.1007/BF00933139

Heilman, S.: Low correlation noise stability of symmetric sets. J. Theor. Probab. 34, 2192–2240 (2021). https://doi.org/10.1007/s10959-020-01031-y

Heilman, S., Mossel, E., Neeman, J.: Standard simplices and pluralities are not the most noise stable. Isr. J. Math. 213, 33–53 (2016). https://doi.org/10.1007/s11856-016-1320-y

Lai, G., Chang, W.-C., Yang, Y., Liu, H.: Modeling long- and short-term temporal patterns with deep neural networks. In: Proc. 41st Int. ACM SIGIR Conf. Res. Dev. Inf. Retr., pp. 95–104 (2018). https://doi.org/10.1145/3209978.3210006

Messoudi, S., Destercke, S., Rousseau, S.: Copula-based conformal prediction for multi-target regression. Pattern Recognit. 120, 108101 (2021). https://doi.org/10.1016/j.patcog.2021.108101

Min, Y., Lu, Y., Li, L., Zhang, Z., Teng, J.: Questioning the coverage-length metric in conformal prediction: When shorter intervals are not better. Proc. Mach. Learn. Res. 306, 88935–88963 (2026)

Prékopa, A.: On logarithmic concave measures and functions. Acta Sci. Math. 34, 335–343 (1973). https://acta.bibl.u-szeged.hu/14411/

Retzlaff, K., Schlembach, F., Bams, D., Dreesen, P.: Testing marginal and conditional coverage in conformal prediction for non-stationary time series via value-at-risk backtesting. Proc. Mach. Learn. Res. 266, 725–747 (2025)

Rigollet, P., Vert, R.: Optimal rates for plug-in estimators of density level sets. Bernoulli 15, 1154–1178 (2009). https://doi.org/10.3150/09-BEJ184

Schlembach, F., Smirnov, E., Winands, M. H. M.: Dynamic conformal prediction for multi-target regression: Optimising informational efficiency under joint validity. Proc. Mach. Learn. Res. 266, 193–213 (2025)

Scott, C. D., Nowak, R. D.: Learning minimum volume sets. J. Mach. Learn. Res. 7, 665–704 (2006)

Wang, Y.: On the alignment between marginal uncertainty measures and conditional decision risk in estimator selection. Stat. Comput. (2026). https://doi.org/10.1007/s11222-026-11045-x

Xu, C., Jiang, H., Xie, Y.: Conformal prediction for multi-dimensional time series by ellipsoidal sets. Proc. Mach. Learn. Res. 235, 55076–55099 (2024)

Zhou, H., Zhang, S., Peng, J., Zhang, S., Li, J., Xiong, H., Zhang, W.: Informer: Beyond efficient transformer for long sequence time-series forecasting. Proc. AAAI Conf. Artif. Intell. 35, 11106–11115 (2021). https://doi.org/10.1609/aaai.v35i12.17325

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- ESM1supplementen.pdf
- ESM2sourcecode.zip