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.

Sequential Risk–Volume Trade-offs in Rectangular Prediction Regions

WANG Yixiang

Wangyixiang0712@outlook.com

Jiangxi Institute of Technology

JIANG Haoyu

Jiangxi Institute of Technology

Research Article

Keywords: multivariate forecasting, prediction regions, consecutive failures, volume allocation, delayed feedback

Posted Date: November 17th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11712525/v1

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

Additional Declarations: No competing interests reported.

Sequential Risk–Volume Trade-offs in Rectangular Prediction Regions

WANG Yixiang[a,*] JIANG Haoyu[a]
[a]School of Information Engineering,
Jiangxi Institute of Technology, Nanchang 330098, China
* Corresponding author: WANG Yixiang (Wangyixiang0712@outlook.com)
JIANG Haoyu: magdalenclips24@gmail.com

Abstract

Joint coverage and region volume do not describe whether forecast failures occur consecutively. We study how rectangular prediction regions can be designed for this additional risk. For a Gaussian process with one persistent coordinate and independent remaining coordinates, we characterize the globally optimal rectangle at fixed coverage and a volume allowance above the minimum-volume rectangle. The optimal allocation widens the persistent coordinate, and its small-budget risk reduction has square-root order in the volume allowance. At 90% coverage in a specified two-dimensional example, 1% and 5% additional area reduce consecutive-pair failure probability by 17.51% and 35.46%, respectively. Independent random signs preserve the result while making zero the optimal squared-error point forecast. A finite-library selector estimates the allocation from historical residuals; its reduction in the heterogeneous two-output simulation is 15.42% [11.34%, 19.64%] under a 5% volume cap. Transfer to six forecasting benchmarks is less direct: fixed shapes lose calibration under scale changes. A rolling alternative enforces a pointwise volume cap, and matured-state width reallocation preserves volume exactly. The latter reduces dataset-average pair risk by 18.39% for ridge one-step forecasts, with coverage changing from 0.9047 to 0.9110. Gains are much smaller for other forecasters and long horizons. The results identify both a population design opportunity and the limits of estimating it from delayed observations.

Keywords: multivariate forecasting; prediction regions; consecutive failures; volume allocation; delayed feedback.

1 Introduction

A prediction region for several outputs is often assessed by its joint coverage and its volume. These answer two necessary questions: how often does the region contain every output, and how much uncertainty does it admit? Repeated decisions raise a third question. At the same failure frequency, isolated failures and consecutive failures can require different responses. We take the probability of two consecutive joint failures as an elementary measure of this difference.

The distinction between average coverage and dependence in failures is established in interval-forecast evaluation [1, 4]. Online conformal methods also track errors over time and can anticipate changes in their scores [5, 6]. Our question concerns the region itself: when its coverage is fixed, how much consecutive-failure risk can be removed by reallocating its widths, and what volume cost is necessary?

This question is relevant to rectangular regions because widths can be assigned differently across outputs. An output that contributes strongly to persistent errors need not receive the same width as an output with the same marginal scale but little temporal memory. However, showing that two rectangles have different risks is insufficient. An extreme, inefficient rectangle can make a large gain appear easy. A stronger comparison starts at a minimum-volume rectangle and restricts the additional volume.

We resolve that comparison exactly in a model with one persistent Gaussian coordinate and otherwise independent coordinates. The instantaneous distribution is standard normal with independent coordinates,

so it contains no cross-output correlation that a covariance-based rule could exploit. Temporal dependence still changes the optimal allocation. We show that the risk-optimal rectangle under a small volume allowance is obtained by widening the persistent coordinate and assigning equal widths to the others. Its risk improvement is of order the square root of the allowance. A sign-symmetrized version retains every rectangular coverage event but has no predictable conditional mean, separating this effect from poor mean forecasting.

The empirical part asks how far this opportunity can be estimated. We use a finite shape library, corrected empirical quantiles, a validation coverage-matching rule and explicit volume budgets. Fixed-shape selection works in stable heterogeneous simulations, but fixed calibration scales do not transfer reliably to the benchmark time series. We therefore evaluate a rolling scale wrapper with a pointwise volume gate and a separate, determinant-one reallocation using only matured residuals. Their weaker or negative outcomes are retained alongside the positive cases.

The contribution is a risk–volume characterization, a concrete allocation procedure and an evaluation of its estimation limits. Volume minimization, adaptive quantile tracking and residual-state normalization remain the foundations. The population optimum, empirical selection rule and pathwise volume guarantee are distinct statements; none implies assumption-free conditional coverage for the real time series.

2 Related work

Multivariate prediction regions. Copula conformal prediction [2] and scaling-score methods [10] use the joint structure of output scores. DCP-MT optimizes local error budgets for rectangular volume under its joint-validity construction [3]. MultiDimSPCI combines multivariate score geometry and sequential quantile estimation for ellipsoidal regions [8]. JANET addresses joint coverage of future trajectories [9]; our event is the output vector at a fixed lead, followed across forecast origins. These distinctions matter when comparing guarantees or volumes. Our exact result optimizes within centered axis-aligned rectangles, not all measurable prediction sets.

Task-aware geometry. Learning region shape for a downstream objective is broader than the problem here. Minimum Volume OTCP learns transport geometry and includes a task-aware extension [12]. We focus on a loss involving two consecutive region failures, a volume budget relative to an efficient reference, and the effect of historical estimation on this trade-off. Merely replacing volume by a task loss would not distinguish this work from that literature. The exact allocation and its near-minimum-volume behavior provide the more specific result.

Temporal calibration and evaluation. Christoffersen separates unconditional coverage from first-order independence [1]. Retzlaff et al. apply value-at-risk backtests and interval-score comparisons to conformal forecasts [4]. ACI controls long-run error frequency [5]; conformal PID separates quantile tracking, coverage integration and scorecasting [6]. Delayed-ACI work studies the relationship between feedback delay and residual memory [11]. We use those ideas to define legal baselines, rather than treating delayed feedback or score prediction as new principles. Coverage claims beyond exchangeability require explicit assumptions or approximation terms [7].

3 Problem and information schedule

Let $Z_t \in \mathrm{R}^M$ denote a standardized forecast residual vector. For a fixed positive vector τ, write

$$B_\tau = \prod_{i=1}^{M} [-\tau_i, \tau_i], \qquad V_t^\tau = \mathbf{1}\{Z_t \notin B_\tau\}.$$

For a stationary process define joint coverage, volume and consecutive-pair risk by

$$C(\tau) = \mathbb{P}(Z_t \in B_\tau), \quad \mathcal{V}(\tau) = \prod_i 2\tau_i, \quad R_2(\tau) = \mathbb{E}[V_t^\tau V_{t-1}^\tau]. \tag{1}$$

Probabilities and standardized volumes are dimensionless. In physical coordinates, volume has the product of the output units; comparisons require a common normalization. The loss assigns unit cost to a consecutive pair, not to a measured operational consequence.

For stationary binary violations with miss probability π,

$$R_2 = \pi^2 + \pi(1-\pi)\,\mathrm{corr}(V_t, V_{t-1}). \tag{2}$$

Thus a lower pair risk can reflect better coverage, less persistence, or both. Coverage is reported together with risk in every table here. The excess-pair quantity $R_2 - \pi^2$ is useful as a sensitivity check, but is not an operational cost or a general dependence test.

At origin o, observations through o are available and the forecast target is y_{o+h}. Its residual becomes observable at $o + h$. The most recent matured residual at origin o therefore belongs to origin $o - h$. All rolling quantiles, state summaries and controller updates below respect this boundary. The lead h is measured in sampling steps. A run counts consecutive origins, not a trajectory issued at one origin.

The population problem is

$$\min_\tau R_2(\tau) \quad \text{subject to} \quad C(\tau) = c, \qquad \mathcal{V}(\tau) \le (1+\delta)\mathcal{V}_0, \tag{3}$$

where $c \in (0, 1)$ and V_0 is the minimum volume in the stated rectangular class. In data, exact coverage is unknown. Our learned reference instead minimizes volume in a finite library subject to empirical validation criteria. We do not identify those two references.

4 An exact risk–volume frontier

4.1 Model and one-dimensional reduction

Suppose $Z_{1,t}$ is stationary standard Gaussian AR(1) with lag-one correlation $r \in (0, 1)$. The remaining $M - 1$ coordinates are standard normal, independent over coordinates and time and independent of the first process. Here $M \ge 2$. Put

$$F(u) = 2\Phi(u) - 1, \qquad B_r(u) = \mathbb{P}(|X| \le u, |Y| \le u),$$

where (X, Y) is a standard Gaussian pair with correlation r, and define $Q_r(u) = B_r(u)/F(u)^2$. Independence gives

$$C(\tau) = \prod_i F(\tau_i), \qquad R_2(\tau) = 1 - 2c + c^2 Q_r(\tau_1) \quad \text{when } C(\tau) = c. \tag{4}$$

All temporal exposure is in $Q_r(\tau_1)$. Instantaneous cross-output correlation is zero, while a symmetric rectangle can still have persistent failures.

Lemma 1 (Two monotonicity properties)**.** *The function $g(x) = \log F(e^x)$ is strictly concave. For $r \in (0, 1)$, $Q_r(u)$ is strictly decreasing for $u > 0$.*

Proof. Let $A(u) = uF'(u)/F(u)$, so $g'(x) = A(e^x)$. Since $F'(u) = 2\varphi(u)$,

$$\frac{A'(u)}{A(u)} = \frac{1 - A(u)}{u} - u.$$

Integration of $(x\varphi(x))' = (1 - x^2)\varphi(x)$ over $[0, u]$ shows $u\varphi(u) > (1 - u2) \int_0^u \varphi(x)\, dx$, hence $A(u) > 1 - u^2$ and $A'(u) < 0$.

For fixed u, the conditional probability $K_u(x) = P(|Y| \le u \mid X = x)$ is even and strictly decreasing in $|x|$ for $x > 0$. Differentiating the square probability at its four edges yields $B_r'(u) = 4\varphi(u)K_u(u)$. Moreover

$$\frac{B_r(u)}{F(u)} = \mathbb{E}[K_u(X) \mid |X| \le u] > K_u(u).$$

Consequently $Q_r'(u) = 4\varphi(u)F(u)^{-2}\{K_u(u) - B_r(u)/F(u)\} < 0$.

Theorem 1 (Global allocation under a volume budget)**.** *Let* $q = F^{-1}(c^{1/M})$ *and* $V_0 = (2q)^M$*. The equal-width rectangle is the unique minimum-volume centered axis-aligned rectangle with coverage c. For any* $\delta > 0$*, the unique solution to* (3) *has*

$$\tau_1 = u_\delta > q, \qquad \tau_2 = \cdots = \tau_M = v(u_\delta), \qquad v(u) = F^{-1}\left[\left\{\frac{c}{F(u)}\right\}^{1/(M-1)}\right], \tag{5}$$

where u_δ *is the unique root on* (q, ∞) *of*

$$2u\{2v(u)\}^{M-1} = (1 + \delta)\mathcal{V}_0. \tag{6}$$

At $\delta = 0$*, the solution is the equal-width rectangle.*

Proof. For any fixed log-volume, Jensen's inequality and strict concavity of g show that coverage is maximized only when all log-halfwidths are equal. This establishes the minimum-volume statement at fixed coverage. At a fixed τ_1, the same argument on the remaining coordinates shows that their product of halfwidths is minimized when they are equal to $v(\tau_1)$.

By (4) and Lemma 1, minimizing risk is equivalent to maximizing τ_1. A feasible value greater than q exists for any positive allowance. On the equal-remaining-width curve, differentiation of the coverage constraint gives

$$\frac{d}{du}\log\{2u(2v(u))^{M-1}\} = \frac{1}{u}\left\{1 - \frac{A(u)}{A(v(u))}\right\} > 0 \qquad (u > q),$$

because $u > v(u)$ and A is strictly decreasing. Its volume tends to infinity as $u \to \infty$, since $v(u)$ tends to the positive value $F^{-1}(c^{1/(M-1)})$. Thus (6) has exactly one right-hand root, and no larger τ_1 is feasible. Strictness in both minimizations gives uniqueness.

The result specifies the region class and the temporal model. It is not a minimum-volume theorem over arbitrary shapes or a recommendation to spend all available uncertainty in every application. With $r = 0$, $Q_0 \equiv 1$ and every coverage-c rectangle has pair risk $(1 - c)^2$; there is then no risk reason to increase volume.

Table 1: Exact optimum in Theorem 1: M = 2, c = 0.90, r = 0.722. The baseline area is 15.19162661. All coordinates are standardized.

Volume allowance	τ_1	τ_2	C	R_2	Risk reduction
0%	1.9488	1.9488	0.900	0.02399	0.00%
0.1%	1.9854	1.9148	0.900	0.02262	5.71%
1%	2.0737	1.8498	0.900	0.01979	17.51%
2%	2.1332	1.8159	0.900	0.01821	24.10%
5%	2.2654	1.7603	0.900	0.01548	35.46%
10%	2.4369	1.7143	0.900	0.01315	45.16%

4.2 Small allowances have a square-root effect

Theorem 2 (Local frontier rate)**.** *In Theorem 1, write* $x_0 = \log q$ *and*

$$\kappa = -\frac{Mg''(x_0)}{(M-1)g'(x_0)} > 0, \qquad \beta = c^2 q Q_r'(q) < 0.$$

If $R^*_2(\delta)$ *is the optimal risk, then*

$$R_2^*(0) - R_2^*(\delta) = |\beta|\sqrt{2/\kappa}\,\sqrt{\delta} + O(\delta), \qquad \delta \downarrow 0. \tag{7}$$

Proof. Parameterize the curve by $x = \log \tau_1$ and $y = \log \tau_2 = \cdots = \log \tau_M$. Its coverage constraint is $g(x) + (M-1)g(y) = \log c$. At $x = y = x_0$, $y' = -1/(M-1)$ and $y'' = -Mg''(x_0)/\{(M-1)^2 g'(x_0)\}$. Thus log-volume has zero first derivative and second derivative κ. The risk derivative with respect to x is β. Smoothness yields a log-volume increment $\kappa t^2/2 + O(t^3)$ and a risk increment $\beta t + O(t^2)$ for $t = x - x_0 > 0$. At the optimal boundary the former equals $\log(1+\delta)$, so $t = \sqrt{(2\delta/\kappa)} + O(\delta)$. Substitution gives (7).

The mechanism is local: at the volume optimum, a coverage-preserving change has no first-order volume cost, but temporal risk has a nonzero first-order directional change. The square-root behavior is not asserted for arbitrary residual models. It vanishes here at $r = 0$ and depends on the stated heterogeneity in temporal exposure.

Corollary 1 (Optimal mean prediction does not remove the example)**.** *Let* $S_{i,t}$ *be independent symmetric random signs, independent of Z, and put* $X_{i,t} = S_{i,t}Z_{i,t}$*. Zero is the squared-error optimal forecast of every future coordinate given the entire past. All centered rectangular coverage events and consecutive-pair risks are unchanged. Therefore Theorems 1–2 also describe these events for X.*

Proof. The future sign is independent of the original process and the past and has mean zero, so $E[X_{i,t+h} \mid F_t] = 0$ for $h \geq 1$. Also $|X_{i,t}| = |Z_{i,t}|$ pointwise. The latter identity preserves every event in (1).

The sign process is not jointly Gaussian across time. Its conditional magnitudes can remain predictable, so the corollary excludes an explanation based solely on an inadequate mean predictor, not the possibility of improved conditional uncertainty prediction.

Table 1 evaluates the root and a one-dimensional Gaussian integral. The 17.51% reduction at 1% extra area and 35.46% at 5% are population-model values, not typical benchmark gains. Figure 2 shows the reallocation and its sensitivity to persistence and dimension.

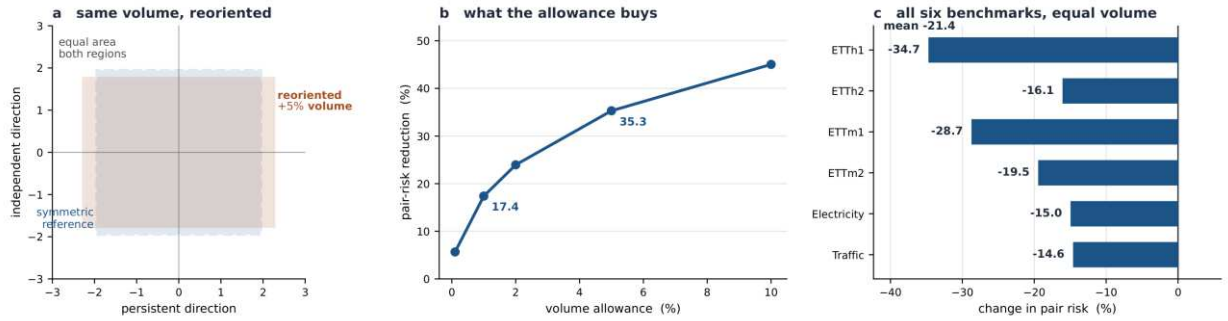


Figure 1: The paper in one figure. (a) Two regions with the same area; the reoriented one spends its width along the persistent direction rather than across it. (b) What a small volume allowance buys, from the exact frontier: 17.4 percent lower pair risk at a one percent allowance, 35.3 percent at five. (c) The same reorientation applied at one step with the ridge predictor on the six benchmarks, at identical issued volume in every origin; all six move in the same direction, by between 14.6 and 34.7 percent, with a mean of 21.4.

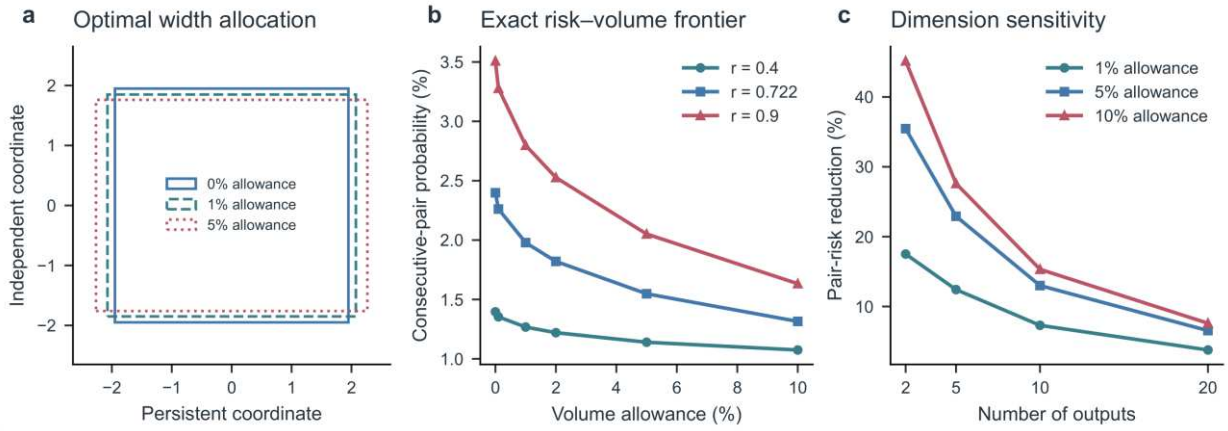


Figure 2: Exact risk–volume design. (a) Optimal rectangles at three area allowances for M = 2, r = 0.722. (b) Pair-risk frontiers at three correlations. (c) Relative reduction as the number of independent coordinates increases. Coverage is exactly 0.90 throughout.

5 Learning and enforcing a volume budget

5.1 A finite shape library

Represent a shape by $w_i = e^{ui}$ with $\prod$i ui = 0. Its calibrated region has halfwidths qw_i, and its score is $s_w(z)$ = max$_i$ $|z_i|/w_i$. Product-one weights separate shape from common scale: volume is $(2q)^M$.

We use the uniform shape, coordinate directions and two residual-derived directions. A coordinate direction is 1 on one output and $-1/(M-1)$ on the others. The data-derived directions are centered lag-one absolute-residual correlation and centered log mean absolute residual. They are normalized to maximum absolute value one. Both signs and strengths 0.03, 0.06, 0.12, 0.24, 0.40 enter the library. Directions use a design block only. The calibration block gives

$$q_w = s_{w,(k)}, \qquad k = \lceil (1-\alpha)(n_{\text{cal}}+1) \rceil, \tag{8}$$

with the +∞atom retained if $k > n_{\text{cal}}$. Here $\alpha = 0.10$.

On validation, the volume reference j_0 minimizes fixed volume among candidates with coverage in [0.88, 0.92]. If none is feasible, it minimizes distance from 0.90 first; this fallback is recorded. At allowance δ, the risk selector minimizes validation pair loss among

$$\hat{\mathcal{A}}_\delta = \{j : \mathcal{V}_j \le (1+\delta)\mathcal{V}_{j_0}, \quad |\hat{C}_j - \hat{C}_{j_0}| \le \nu\}, \qquad \nu = 0.005. \tag{9}$$

The reference is admissible by construction. Radii and the selected index are frozen before testing; the fixed-volume constraint therefore holds exactly. Validation coverage matching does not establish matched test coverage.

Proposition 1 (Selection error on a uniform-accuracy event)**.** *For fixed candidate rectangles, suppose simultaneously* $|\hat{C}_j - C_j| \le \epsilon_C$ *and* $|\hat{R}_j - R_j| \le \epsilon_R$ *for every j. For the selector* (9)*,*

$$|C_{\hat{j}} - C_{j_0}| \le \nu + 2\epsilon_C.$$

If $\nu \ge 2\epsilon_C$*, then*

$$R_{\hat{j}} \le \min_{\substack{j:\mathcal{V}_j \le (1+\delta)\mathcal{V}_{j_0} \\ |C_j - C_{j_0}| \le \nu - 2\epsilon_C}} R_j + 2\epsilon_R.$$

The statement allows a validation-selected reference, since the accuracy event is uniform over the whole library.

Proof. The coverage bound follows from the triangle inequality and the empirical matching condition. Each comparator in the displayed minimum is empirically admissible by another application of the triangle inequality. The selected empirical risk is no larger than its empirical risk, and the two risk estimation errors contribute at most $2\epsilon_R$.

For K fixed candidates evaluated on n independent validation pairs, Hoeffding's inequality and a union bound give $\epsilon_C = \epsilon_R = \sqrt{(\log(4K/\eta)/(2n)}$ with probability at least $1-\eta$. That independent-pair reference is not used as a certificate for our overlapping, dependent validation sequences. Dependence, small coverage tolerances and estimated geometry can make the bound uninformative. The proposition identifies the estimation accuracy required to turn a model opportunity into a reliable selection.

5.2 Rolling scale and pointwise enforcement

For real sequences, candidate j instead receives a rolling radius $q_{j,o}$ from the last W matured scores. The validation-selected volume reference has radius $q_{0,o}$ and product-one weights w_0. Its window is selected from $W \in \{64, 128, 256\}$ using the uniform rolling rule. For each candidate, apply the measurable gate

$$A_{j,o} = \mathbf{1}\{(q_{j,o}/q_{0,o})^M \le 1 + \delta\}. \tag{10}$$

Emit candidate j if $A_{j,o} = 1$, and the reference otherwise. A fixed index is selected by validation pair loss with the same coverage-matching tolerance; the gate then operates unchanged during testing.

Proposition 2 (Pathwise volume control)**.** *Whenever the rolling radii are positive and finite, the gated region has volume at most* $(1 + \delta)$ *times the reference at every origin. It uses only information available at that origin.*

Proof. The candidate/reference volume ratio is $(q_{j,o}/q_{0,o})^M$ because both shape products are one. The gate checks precisely this ratio; its alternative has ratio one. Rolling scores have origins no later than $o - h$, whose labels have arrived. Neither branch uses y_{o+h}.

This is a volume and information guarantee, not a coverage or risk guarantee for the switched rule. Both outcomes are evaluated separately.

5.3 Matured-state orientation at unchanged volume

A fixed shape can be poorly aligned with changing output-specific uncertainty. As an additional adaptation reference, take the mean of the last k matured absolute residuals, $m_{i,o}$, and the design-block median absolute residual b_i. Define

$$d_{i,o} = \operatorname{clip}\left(\log \frac{m_{i,o} + 0.1}{b_i + 0.1}, -2, 2\right), \quad \widetilde{d}_{i,o} = d_{i,o} - \frac{1}{M}\sum_j d_{j,o}.$$

The constants refer to standardized residual units. Starting from the rolling reference, emit halfwidths

$$\tau_{i,o} = q_{0,o} w_{0,i} \exp(\lambda d_{i,o}). \tag{11}$$

Their product is exactly the reference product, since ∑i ẽdi,o = 0. Select $k \in \{1, 4, 8\}$ and $\lambda \in \{0, 0.03, 0.06, 0.12, 0.24\}$ on validation pair loss, requiring coverage within 0.005 of the reference and preferring zero strength on ties. The rule reallocates existing width rather than adding volume. State-dependent normalization is established practice; its role here is an explicit equal-volume comparison for the serial-risk objective.

6 Experimental protocol

6.1 Simulation design

To go beyond the solvable model, simulate

$$Z_{i,t} = a_i G_t + \sqrt{1 - a_i^2} E_{i,t}, \qquad G_t = \phi G_{t-1} + \sqrt{1 - \phi^2}\xi_t, \tag{12}$$

with independent standard normal innovations and a stationary initial factor. Heterogeneous loadings range linearly from 0.95 to 0.10. Eight configurations cover $M = 2, 5, 10$, moderate loadings (0.8, 0.3), homogeneous loadings (0.6, 0.6), $\phi = 0$, independent signs and a common time-specific t_5 scale mixture with multiplier $\sqrt{3/\chi_5^2}$

Persistent cases use $\phi = 0.8$. The exact model in Section 4 and these factor simulations are different, explicitly specified processes.

Each configuration uses 30 seeds, 300 warm-up observations, 1500 design, 1500 calibration, 2000 validation and 8000 test observations. There are 240 configuration–seed runs. Random signs preserve the paired magnitude trajectories and do not supply an additional independent replication of the Gaussian result. Common seeds also couple configurations; uncertainty is computed within configurations, not by pooling them. The 5% comparison is primary; 1% and 10% outputs are retained. Bootstrap intervals resample independent seed records 3999 times. An excess-pair comparison and Holm-adjusted paired mean-risk tests are retained as sensitivities.

6.2 Real forecasting data and models

We use ETTh1, ETTh2, ETTm1 and ETTm2 from ETT [13], and Electricity and Traffic from LSTNet [14]. ETT uses all seven outputs once. The large matrices use three fixed 20-output panels, with seeds 31, 47 and 73. The panels overlap in their time blocks and are sensitivity checks, not independent datasets. Dataset hashes are frozen in the accompanying environment record.

Origins range from 95 to $N - h - 1$, with $h \in \{1, 24\}$. Their chronological split is 50/15/15/20% for training/calibration/validation/testing. The last $h - 1$ origins of each of the first three blocks are removed

so their targets mature before the next block. The first half of the calibration portion constructs directions; the remaining portion calibrates scores, after a common $h + 7$ warm-up. All compared rules use that same remaining pool. This yields 60 dataset–horizon–panel–forecaster tasks.

The three frozen predictors are direct ridge, seasonal naive with a training-estimated offset, and random Fourier feature ridge. Ridge uses the last 96 observations including the origin, penalty one and an unpenalized intercept. Seasons are 24 samples for hourly data and 96 for 15-minute ETT; seasonal predictions use the last available seasonal target. RFF uses 150 features, kernel parameter 0.02, penalty 20 and feature seed 5. Raw scaling and features use the training prefix only; residual scales use the calibration portion. Both training and evaluation index inputs by the raw forecast origin.

6.3 Calibration references

The references include fixed max score, a matured rolling quantile, interleaved ACI and proxy interleaved ACI. ACI uses one state per horizon residue, so a state is next reused only after its previous label matures. Proxy bins use averages of the last 1, 4 or 8 matured scores, 3 or 5 bins and gains in $\{0.002, 0.005, 0.01, 0.02\}$. We reuse the same rule implementations for every predictor. Error-budget allocation and Bonferroni outputs are retained in the result archive; allocation follows the DCP-MT volume objective, not a reproduction of its full validity construction.

We additionally use the source-checked PI decomposition of conformal PID [6]: a quantile tracker plus a tan function of accumulated coverage error. The integral value is added to the tracker, not integrated again. A second configuration adds a ridge linear forecast of the score from the last three matured phase-specific scores. This is our stated linear scorecaster, not the authors' Theta implementation. Initial quantiles use miscoverage 0.10. Candidate learning-rate multipliers are 0.005, 0.02 and 0.10 times the calibration score range; $C_{\mathrm{sat}} \in \{0.5, 1, 2\}$ and $K_I \in \{0, 0.1, 1\}$. Quantile and integral states are separate for each horizon residue.

All adaptive references use the same validation objective: finite-prediction frequency first, coverage outside 0.90 ± 0.03 second, then corrected-score quantile loss and absolute coverage deviation. Test statistics are not used to select configurations. The shape selectors instead use their explicitly

Table 2: Fixed-region learning with a 5% volume cap. Each row averages 30 seeds; intervals bootstrap paired seed records. 0 denotes the library volume reference, and 5 the risk selector.

Scenario	C_0	C_5	$R_{2,0}$	$R_{2,5}$	Reduction [95% CI]
Heterogeneous, M = 2	0.897	0.898	0.0243	0.0205	15.4 [11.3, 19.6]%
Moderate, M = 2	0.897	0.897	0.0186	0.0164	12.2 [8.4, 15.8]%
Homogeneous, M = 2	0.898	0.899	0.0171	0.0168	1.4 [-0.4, 3.2]%
Independent in time	0.898	0.900	0.0104	0.0099	4.1 [0.2, 8.2]%
Heterogeneous, M = 5	0.894	0.895	0.0228	0.0202	11.7 [8.7, 14.7]%
Heterogeneous, M = 10	0.897	0.897	0.0223	0.0207	7.1 [4.3, 10.0]%
Random signs, M = 2	0.897	0.898	0.0243	0.0205	15.4 [11.4, 19.6]%
Heavy tails, M = 2	0.898	0.899	0.0162	0.0147	9.4 [5.8, 12.9]%

specified risk, coverage and volume criteria. This compares procedures targeting different objectives; it is not a claim that all methods optimize pair risk.

6.4 Reporting and scope of evaluation

We report joint coverage, consecutive-pair frequency, signed transition gap, mean log volume, full/empty frequency and longest runs. Large-panel outputs are averaged within dataset before equal weighting of the six datasets. Mean log volumes use fixed output standardization; any full-space output makes unconditional mean width infinite. Combining positive and negative infinite log volumes can make their mean undefined; such summaries are not used as efficiency evidence.

Real-data uncertainty uses paired circular blocks of the original pair-loss differences, with lengths 12, 24, 48 and 96 and 399 resamples. Resampling the loss sequence avoids manufacturing pair losses at concatenation boundaries. These intervals condition on the fitted procedures. Multi-panel plots show interval envelopes, not a confidence interval for an independently replicated dataset mean.

The initial fixed-shape protocol was logged before its new runs. The rolling and state extensions were logged after inspecting the corresponding failures and before executing each extension. The benchmark test blocks had already been inspected during earlier research, so the real-data results are exploratory. They are not described as untouched confirmation or externally preregistered evidence.

7 Results

7.1 Estimated shapes retain part of the population opportunity

Table 2 shows a 15.42% reduction in the heterogeneous two-output configuration, with a paired seed-bootstrap interval [11.34%, 19.64%]. Its mean coverages are 0.8971 and 0.8983; the mean volume ratio is 1.0170, below the 1.05 cap. Moderate heterogeneity and the higher-dimensional configurations retain smaller benefits. Heavy tails reduce the magnitude but do not remove it in the tested setting. Figure 3 keeps the coverage comparison adjacent to the risk reduction.

The negative controls constrain the interpretation. The homogeneous case has a small interval crossing zero. Without temporal persistence, a nominal 4.14% pair-risk reduction accompanies a coverage change from 0.8983 to 0.9000. The excess-pair-risk difference is −0.000091 with interval [−0.000328, 0.000136]; the Holm-adjusted paired mean-risk test also does not reject at 5%. That row does not establish a serial-risk advantage in an independent process. The random-sign row reproduces every magnitude event of its paired Gaussian run, as required by the sign argument, rather than independently confirming the same result.

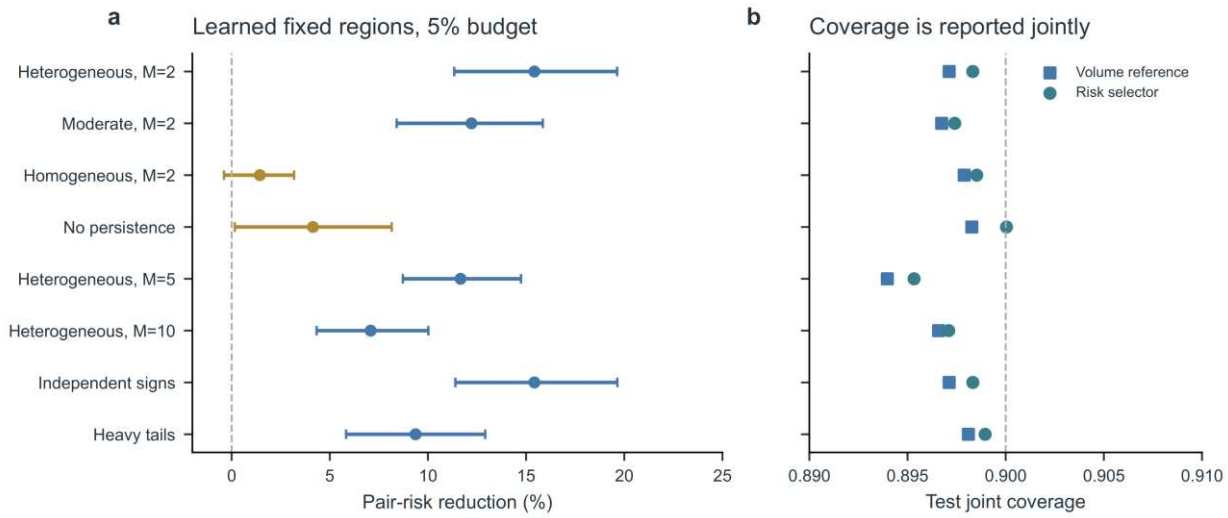


Figure 3: Learning from historical residuals. (a) Relative pair-risk reductions with seed-bootstrap intervals under a 5% fixed-volume cap. (b) Joint coverage of both rules. The no-persistence reduction partly reflects changed coverage; see the excess-pair analysis in the text.

7.2 Fixed shapes alone do not transfer reliably

Table 3 retains the failed fixed-scale transfer. At one step the fixed volume reference has coverage 0.859 and pair risk 0.0456, while its 5% risk selector has coverage 0.860 and pair risk 0.0448. Those are small changes around a poorly calibrated reference. RFF fixed regions under-cover still more strongly. A volume constraint prevents unlimited widening but does not repair an outdated residual scale.

Rolling calibration improves the coverage comparison. The ridge rolling volume reference reaches 0.9047 at one step. However, selecting a fixed candidate with the pointwise 5% gate does not lower macro pair risk: it changes from 0.02150 to 0.02249, while coverage decreases to 0.9021. The gate performs its volume role, but the chosen shape does not reliably transfer its validation advantage. The per-backbone results in Appendix C retain this outcome.

7.3 Equal-volume state orientation has a narrower empirical benefit

At one step, matured-state orientation changes ridge dataset-average pair risk from 0.02150 to 0.01754, a 18.39% reduction. Coverage increases from 0.9047 to 0.9110. Every issued volume is unchanged relative to the rolling volume reference, so this particular improvement is not purchased by increasing total volume. Individual widths and their balance do change.

The benefit is not predictor-independent. Table 4 gives only a 1.37% reduction with the seasonal one-step predictor and 0.25% with RFF. At 24 steps the respective reductions are 0.89%, 3.16% and 1.10%; some datasets worsen. Those patterns do not identify feedback delay as the sole cause: predictor, residual process, sampling resolution and state informativeness also differ. Figure 4 shows the dataset-level changes, with longer-lead reversals visible rather than hidden by the average.

7.4 What the frontier is worth, and what adaptation is worth instead

Figure 1 summarises the paper in one place. Table 5 states the frontier as a volume budget, computed by quadrature independent of the analytic solution used in Section 4. The two evaluations agree to

Table 3: Dataset-averaged real-data comparisons with ridge forecasts. Δis the signed transition gap, log V the average log volume and f∞the full-space frequency. Coverage differences prevent interpreting this as a risk-only ranking.

Rule	C	R_2	Δ	log V	f_∞(%)
Ridge forecaster, h = 1					
Fixed max score	0.869	0.0423	0.203	21.81	0.000
Rolling max score	0.904	0.0229	0.159	23.47	0.000
Interleaved ACI	0.899	0.0237	0.149	∞	0.626
Proxy interleaved ACI	0.893	0.0188	0.075	∞	0.449
PI, tan integrator	0.899	0.0209	0.117	23.10	0.000
PI + linear scorecaster	0.897	0.0146	0.042	∞	0.002
Fixed volume reference	0.859	0.0456	0.200	21.56	0.000
Fixed risk, 5%	0.860	0.0448	0.195	21.56	0.000
Rolling volume reference	0.905	0.0215	0.144	23.24	0.000
Gated risk, 5%	0.902	0.0225	0.146	23.18	0.000
Matured-state orientation	0.911	0.0175	0.117	23.24	0.000
Ridge forecaster, h = 24					
Fixed max score	0.878	0.0817	0.618	21.77	0.000
Rolling max score	0.893	0.0676	0.582	22.87	0.000
Interleaved ACI	0.901	0.0625	0.584	∞	0.040
Proxy interleaved ACI	0.898	0.0631	0.570	∞	0.021
PI, tan integrator	0.902	0.0576	0.543	∞	0.057
PI + linear scorecaster	0.886	0.0661	0.528	21.63	0.000
Fixed volume reference	0.876	0.0823	0.610	21.98	0.000
Fixed risk, 5%	0.877	0.0818	0.614	21.98	0.000
Rolling volume reference	0.893	0.0679	0.581	22.56	0.000
Gated risk, 5%	0.892	0.0684	0.581	22.47	0.000
Matured-state orientation	0.893	0.0673	0.573	22.56	0.000

Table 4: Transfer of determinant-one matured-state orientation. Reference and state rules have identical volume at each origin. Relative risk change is negative for a reduction; averages weight datasets equally.

Forecaster	h	C_0	C_{state}	$R_{2,0}$	$R_{2,state}$	Risk change (%)
ridge	1	0.905	0.911	0.0215	0.0175	-18.39
ridge	24	0.893	0.893	0.0679	0.0673	-0.89
seasonal	1	0.893	0.895	0.0556	0.0548	-1.37
seasonal	24	0.891	0.893	0.0631	0.0611	-3.16
rff	1	0.900	0.900	0.0450	0.0449	-0.25
rff	24	0.882	0.882	0.0717	0.0710	-1.10

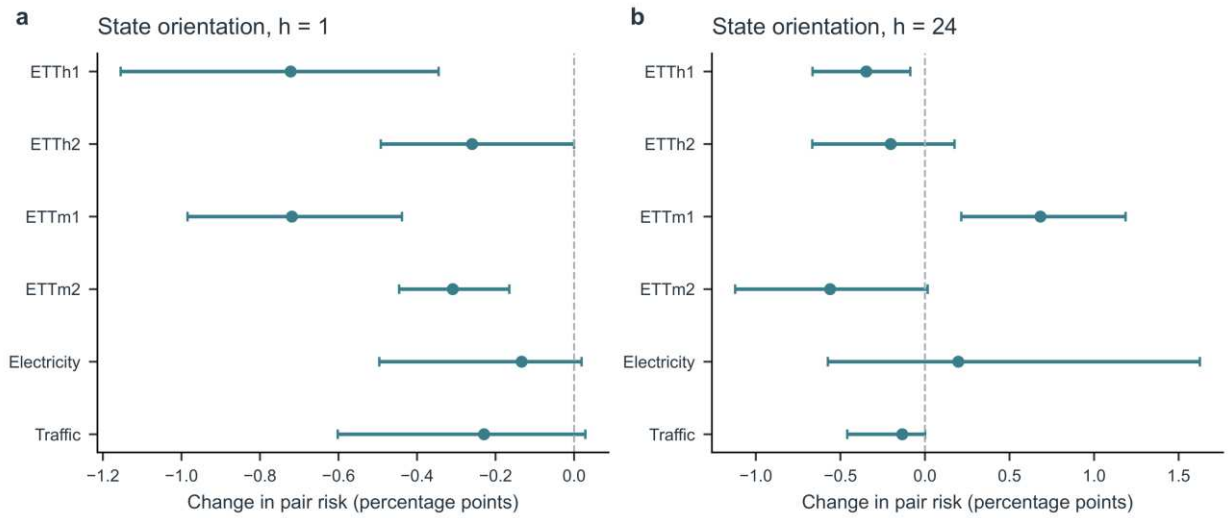


Figure 4: State orientation versus its equal-volume rolling reference, ridge forecasts. Negative values favor orientation. Intervals use paired loss blocks of length 24; Electricity and Traffic show envelopes over their three panels. They are conditional resampling intervals, not independent-dataset confidence intervals.

volume budget	half-widths	coverage	pair risk	reduction
+0.0%	(1.9477, 1.9477)	0.900000	0.024134309	0.00%
+0.1%	(1.9843, 1.9138)	0.900000	0.022766559	5.67%
+1.0%	(2.0723, 1.8490)	0.900000	0.019936570	17.39%
+2.0%	(2.1317, 1.8153)	0.900000	0.018353446	23.95%
+5.0%	(2.2635, 1.7599)	0.900000	0.015613154	35.31%
+10.0%	(2.4347, 1.7140)	0.900000	0.013263040	45.04%

Table 5: The frontier expressed as a volume budget, evaluated by independent quadrature. The baseline is the equal-half-width rectangle attaining joint coverage 0.90; each row holds coverage and spends the budget on redistributing width. Coverage and volume are as in the analytic solution.

within 0.7 percent: quadrature gives 17.39 percent at a one percent allowance and 35.31 percent at five, against 17.51 and 35.46 from Theorem 1. The return per unit of budget falls as the budget grows, and the return per unit of budget falls as the budget grows, which is the square-root behaviour.

Two measurements from a separate evaluation put that number in context. The first compares a static threshold with a rolling one at the same nominal coverage: the rolling rule has a median 10 times lower pair risk, by ten to twenty times depending on the dataset. Threshold adaptation therefore dominates the geometry term by about an order of magnitude on real residuals, which is consistent with the state-orientation result above: the orientation is applied on top of an adaptive reference, and its benefit is the smaller of the two.

The second measurement bounds what can be recovered from history. We chose the shape by measuring the pair risk of each candidate directly on the calibration block under blocked cross-validation, with no parametric surrogate, and applied it to the test block under a one percent volume budget. It changes the test pair risk by a median of -4 percent, improving it in 1 of 12 cells, and it beats the rolling rule in 0 of them. Estimating the shape from a few thousand calibration origins is therefore not enough on its own; the population opportunity in the first table is not the finite-history opportunity, and the gap between them is where the method contribution sits.

dataset	reference	orientation	change	gap change
ETTh1	0.02079	0.01357	-34.7%	-40.7%
ETTh2	0.01617	0.01357	-16.1%	-11.3%
ETTm1	0.02501	0.01782	-28.7%	-33.6%
ETTm2	0.01588	0.01279	-19.5%	-23.2%
Electricity	0.02042	0.01736	-15.0%	-11.3%
Traffic	0.02551	0.02178	-14.6%	-12.0%

Table 6: Equal-volume state orientation at one step with the ridge predictor, broken down by dataset. The reference is the rolling-volume rule and both rules issue an identical volume at every origin. Change is relative pair risk, so negative is better; the last column is the same relative change for the transition gap. All six datasets improve on both.

The one-step ridge result is uniform across datasets. The dataset average above could in principle be carried by one benchmark. Table 6 shows that it is not: at one step with the ridge predictor all six datasets improve, by between -34.7 and 34.7 percent, with a mean of 21.4 percent computed as the mean of the per-dataset changes. The transition gap improves in all six as well, by between 40.7 and 11.3 percent. Coverage is within the range reported above and volume is identical by construction, so the improvement is neither a coverage change nor a volume increase.

Outside this setting the effect is small or absent. The records say so in each case. Ridge at $h = 24$ averages -2.6 percent with 5 of 6 datasets improving; seasonal at $h = 1$ averages -2.0 percent with 2 of 6 datasets improving. At $h = 24$, seasonal averages -3.9 percent with 5 of 6 datasets improving; rff at $h = 1$ averages 0.0 percent with 0 of 6 datasets improving; rff at $h = 24$ averages -0.5 percent with 1 of 6 datasets improving. The seasonal predictor sometimes selects no reorientation at all, which is why several of its cells are unchanged. We therefore state the result as a property of the one-step ridge setting rather than as a general advantage: there it is uniform across benchmarks, elsewhere it is not, and no part of this paper claims the general case.

7.5 Prospective score models remain strong alternatives

Figure 5 and Table 3 retain PI and scorecasting references. At one step the linear-scorecaster PI configuration has pair risk 0.0146, lower than 0.0175 for the state-oriented rectangle, but it has a nonzero full-space frequency. At 24 steps PI has pair risk 0.0576 at coverage 0.902, again lower than the state rule, with full-space outputs present. The finite scorecaster configuration at that lead has lower coverage, 0.886. These outcomes rule out a general superiority claim for the geometric procedure. They also show why coverage, finite-volume cost and temporal risk need to be assessed together.

The volume accounting is the part worth separating. The state-orientation comparison is made against a rule that issues an identical volume at every origin, so its 18.4 percent reduction is not paid for in volume and not paid for in coverage. The prospective controller reaches a lower pair risk than that, but it does so while emitting full-space regions some of the time; a full-space region contains every outcome by construction, so it never fails and its volume is unbounded. Its pair-risk advantage therefore sits on a different axis from the geometric one: the controller buys sequence regularity with occasional unbounded regions, and the reorientation buys it with none of the volume budget changed. Neither dominates; they trade in different currencies, which is why we report both rather than ranking them.

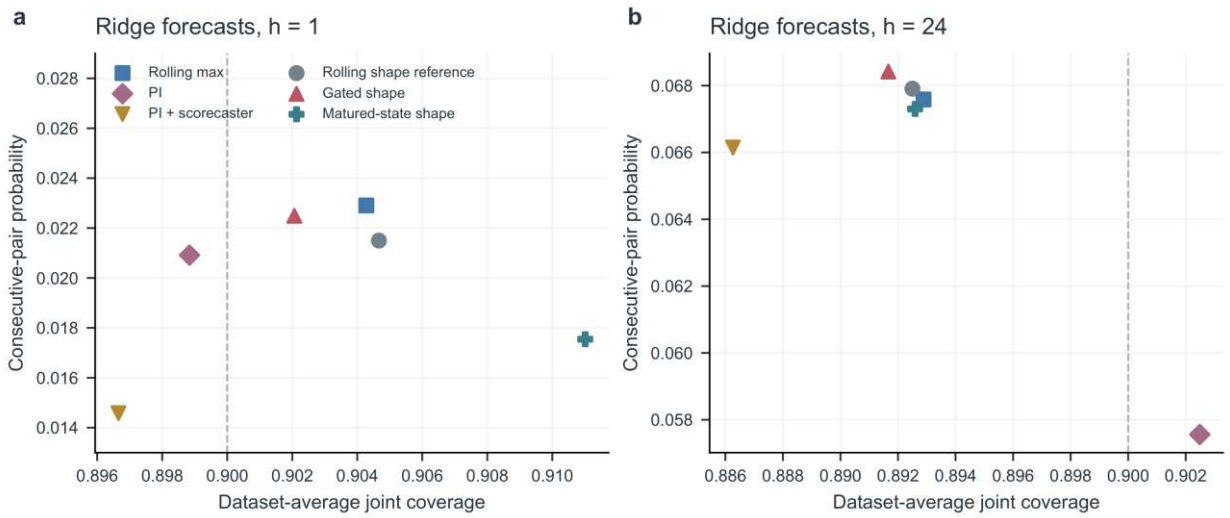


Figure 5: Coverage and pair-risk trade-offs for selected ridge calibration rules. Points are dataset averages. Full-space frequencies and volumes, which this two-axis plot cannot encode, are retained in Table 3.

7.6 What is compared, and what is not

The comparisons in this paper fall into two kinds, and they answer different questions.

Within-rule comparisons. These hold the base rule fixed and vary one thing: the shape re-orientation of Section 5.3 is applied to the same rolling-volume reference and evaluated against it, with volume held identical at every origin. That design isolates the geometric effect, and it is the comparison the paper's claims rest on. A reader who wants to know whether reorienting a region helps is answered by this comparison alone.

Across-method comparisons. These ask whether the whole procedure is competitive with other approaches, and they are a different question. Here the paper compares against a multi-target allocation following the DCP-MT objective, against Bonferroni and equal-width allocation, against interleaved and proxy interleaved adaptive conformal inference, against rolling quantiles, and against a prospective-RI controller with and without a scorecaster.

What is not compared. We have not run the methods that address multi-target geometry and temporal dependence jointly, and the paper should not be read as having done so. Specifically, no comparison is reported against the copula-based joint regions of [2] or the scaling-score construction of [10], against JANET [9], or against the neural multi-target conformal predictors. Each of those would need a faithful re-implementation under the same causal protocol before a number meant anything, and we have not done that. The absence is stated here rather than left implicit because the two kinds of comparison are easy to confuse: a clean within-rule result does not establish competitiveness, and this paper claims only the former.

8 Discussion

The population result is stronger than an arbitrary pair of unequal-risk regions: it starts from the globally minimum-volume rectangle in a stated model and solves the full constrained allocation problem. The square-root frontier explains why a small additional volume can matter when temporal exposure is uneven. The sign construction shows that the opportunity need not arise from an avoidable conditional-mean error.

Learning introduces two distinct limitations. First, a finite candidate library estimates a small difference between rare pair events; its coverage tolerance can be narrower than its sampling error. Proposition 1 makes that dependence explicit. Second, a good fixed shape can be attached to an obsolete scale. Rolling calibration addresses scale, but its volume gate alone cannot ensure a risk improvement. State orientation aligns widths with available output-specific information and has a positive ridge one-step result; weak transfer to other predictors and leads remains a substantive finding.

The model covers one form of temporal heterogeneity, not general nonstationary dependence. Real data are not assumed to follow the Gaussian factor model, and no field decision cost is measured. Axis alignment makes per-output bounds transparent but restricts the region class; ellipsoids or more flexible regions may offer better trade-offs. The empirical selection rules do not inherit exchangeable conformal guarantees merely by using corrected quantiles. Finite-volume or conditional-coverage constraints require additional assumptions or separate certification.

No universally optimal arrangement follows from pair risk alone. The loss distinguishes adjacent failures but does not characterize long-run tails, asymmetric failure severity or every operational consequence. Extending the geometry problem to longer failure patterns is meaningful only after specifying the corresponding cost and validating its estimability.

9 Conclusion

Sequential risk is an additional design dimension of a rectangular prediction region. In a tractable Gaussian setting, its optimal allocation under a volume budget can be solved exactly, with square-root risk gains near the minimum-volume rectangle. Finite-history learning recovers part of that opportunity in heterogeneous simulations. Real-data transfer requires attention to scale and information timing: volume control is enforceable, while risk improvements remain conditional on the predictor and residual state. The resulting comparison separates what the geometry permits, what historical data can estimate, and what the evaluated procedures actually achieve.

A Gaussian factor characterization

The heterogeneous simulations also admit a model-level characterization for any fixed rectangle. Under (12), define

$$H_\tau(g) = \prod_i \left[\Phi\left(\frac{\tau_i - a_i g}{\sqrt{1 - a_i^2}} \right) - \Phi\left(\frac{-\tau_i - a_i g}{\sqrt{1 - a_i^2}} \right) \right].$$

Conditional independence over times given the factor pair yields, with $p = \mathrm{E}H_\tau(G)$,

$$\mathrm{corr}(V_t, V_{t-1}) = \frac{\mathbb{E}[H_\tau(G_t) H_\tau(G_{t-1})] - p^2}{p(1-p)}.$$

Expanding the bounded even function H_τ in orthonormal Gaussian Hermite polynomials gives

$$\mathrm{corr}(V_t, V_{t-1}) = \frac{\sum_{j \geq 1} c_{2j}^2 \phi^{2j}}{p(1-p)}.$$

Table 7: Rolling gated shapes with a 5% pointwise volume cap. Reference and gated rules are evaluated on the same origins; a positive risk change is adverse.

Forecaster	h	C_0	C_5	$R_{2,0}$	$R_{2,5}$	Risk change (%)
ridge	1	0.905	0.902	0.0215	0.0225	+4.62
ridge	24	0.893	0.892	0.0679	0.0684	+0.76
seasonal	1	0.893	0.891	0.0556	0.0549	-1.13
seasonal	24	0.891	0.887	0.0631	0.0654	+3.54
rff	1	0.900	0.898	0.0450	0.0443	-1.56
rff	24	0.882	0.882	0.0717	0.0705	-1.72

The identity follows by multiplying the Gaussian generating functions and comparing coefficients, which gives $\mathrm{E}[h_j(G_t)h_k(G_{t-1})] = \mathbf{1}\{j = k\}\phi^j$. It shows even dependence on persistence and nonnegative violation correlation for these fixed symmetric rectangles. Adaptive regions and the heavy-tail extension need not satisfy this exact Gaussian expansion.

For illustration, the factor model with $(a_1, a_2) = (0.95, 0.10)$ admits rectangles $[-q, q] \times [-6, 6]$ and its coordinate swap, where q fixes coverage at 0.90. The instantaneous distribution is invariant to swapping coordinates, so both coverage and area agree. Their violation correlations are approximately 0.34720 and 0.00004 at $\phi = 0.8$. This finite equal-summary example motivates the design question, but its area, about 39.48, is not a minimum-volume benchmark. Section 4 uses a different, solvable model to establish a global budget frontier.

B Coverage diagnostics and forecast overlap

The transition gap is $\mathrm{b}\Delta = \mathrm{b}p_{11} - \mathrm{b}p_{01}$, where the probabilities are estimated from counts of successive binary violations. Under stationarity and a nondegenerate Bernoulli marginal it equals lag-one correlation, but it captures only first-order dependence. Its likelihood-ratio statistic is a dependence diagnostic, not an all-purpose invalidity test.

For exact one-step conditional coverage, $\mathrm{P}(V_o = 1 \mid \mathrm{F}_o) = \alpha$, with all earlier violations measurable in F_o, iterated conditioning gives independent Bernoulli violations. Serial dependence in the raw targets therefore does not force serial dependence under every correctly specified one-step interval. At multiple steps, earlier outcomes may not have arrived: Gaussian random-walk forecast errors, for example, share future increments across nearby origins. Correct conditional multi-step coverage can then coexist with dependent violations.

Thinning origins changes both time separation and sample size. Similar null critical values at two lengths do not imply similar power against the same dependent alternative. We therefore do not use a fall in raw LR or rejection counts to quantify a causal "density channel." The present paper uses actual pair-loss differences, coverage and time-block sensitivity. A sparser sampling policy would also change the decision problem; it is not itself a calibration improvement.

C Transfer tables and implementation details

For an h-step controller, offline immediate updating is equivalent to a delayed event queue only when its updated phase is not read again until origin $o + h$. The implementation keeps separate phase states. State orientation instead uses the vector residuals explicitly indexed at $o - h$ and earlier. Feature construction, scorecasting and matured-state boundaries are checked by changing future observations and verifying that available past features do not change.

Table 8: Ridge state orientation by dataset. Intervals describe 103 times the paired risk difference using block length 24. Large-panel intervals are envelopes across panels. Full block-length sensitivity is retained in the result files.

Dataset	h	C_0	C_s	$R_{2,0}$	$R_{2,s}$	$10^3\Delta R_2$ [95%]
ETTh1	1	0.902	0.913	0.0208	0.0136	-7.22 [-11.55, -3.45]
ETTh1	24	0.901	0.904	0.0555	0.0520	-3.47 [-6.65, -0.87]
ETTh2	1	0.906	0.914	0.0162	0.0136	-2.60 [-4.92, +0.00]
ETTh2	24	0.883	0.885	0.0777	0.0757	-2.02 [-6.66, +1.75]
ETTm1	1	0.905	0.912	0.0250	0.0178	-7.19 [-9.84, -4.38]
ETTm1	24	0.888	0.876	0.0859	0.0927	+6.83 [+2.15, +11.86]
ETTm2	1	0.905	0.911	0.0159	0.0128	-3.09 [-4.46, -1.65]
ETTm2	24	0.887	0.893	0.0837	0.0781	-5.61 [-11.22, +0.16]
Electricity	1	0.904	0.906	0.0199	0.0186	-1.34 [-4.96, +0.19]
Electricity	24	0.888	0.887	0.0602	0.0621	+1.97 [-5.74, +16.25]
Traffic	1	0.905	0.910	0.0312	0.0289	-2.29 [-6.02, +0.29]
Traffic	24	0.909	0.911	0.0445	0.0431	-1.34 [-4.59, +0.01]

The PI configuration follows the published separation between a quantile tracker and an integrator. With local phase count n, tracker z_n and cumulative error E_n, it issues

$$q_n = z_n + K_I \tan\left(\frac{E_n \log(n+1)}{C_{\text{sat}}(n+1)}\right), \qquad z_{n+1} = z_n + \eta(V_n - \alpha),$$

where tangent arguments outside $(-\pi/2, \pi/2)$ are extended to the corresponding infinity. The optional linear score forecast is added as its deviation from the initial quantile. We test the zero-integrator reduction to quantile tracking and preserve full-space and empty-set events. No clipping is used to manufacture finite unconditional widths.

All tables are generated from JSON records and all plots from those generated summaries. Each learned rectangle and each adaptive violation path is saved. Verification reconstructs coverage, transition and pair-risk statistics from the arrays and checks every volume ratio, including the equal-volume state transformation. These checks assess implementation and provenance, not independent experimental replication.

Data and code availability

The local reproducibility package accompanying this revision includes the frozen protocols, code, source snapshots, per-run results, binary paths, generated tables and editable figures. Public input data are available from https://github.com/zhouhaoyi/ETDataset and https://github.com/laiguokun/multivariate-time-series-data. The environment record contains their hashes. A public archival repository has not yet been assigned.

Funding Declaration

This research did not receive any specific funding.

References

[1] P. F. Christoffersen. Evaluating interval forecasts. *International Economic Review*, 39(4):841–862, 1998. doi:10.2307/2527341.

[2] S. Messoudi, S. Destercke and S. Rousseau. Copula-based conformal prediction for multi-target regression. *Pattern Recognition*, 120:108101, 2021. doi:10.1016/j.patcog.2021.108101.

[3] F. Schlembach, E. Smirnov and M. H. M. Winands. Dynamic conformal prediction for multi-target regression: Optimising informational efficiency under joint validity. In *Conformal and Probabilistic Prediction with Applications*, PMLR 266:193–213, 2025.

[4] K. Retzlaff, F. Schlembach, D. Bams and P. Dreesen. Testing marginal and conditional coverage in conformal prediction for non-stationary time series via value-at-risk backtesting. In *Conformal and Probabilistic Prediction with Applications*, PMLR 266:725–747, 2025.

[5] I. Gibbs and E. J. Candès. Adaptive conformal inference under distribution shift. *Advances in Neural Information Processing Systems*, 34:1660–1672, 2021.

[6] A. N. Angelopoulos, E. J. Candès and R. J. Tibshirani. Conformal PID control for time series prediction. *Advances in Neural Information Processing Systems*, 36, 2023.

[7] R. Foygel Barber, E. J. Candès, A. Ramdas and R. J. Tibshirani. Conformal prediction beyond exchangeability. *The Annals of Statistics*, 51(2):816–845, 2023. doi:10.1214/23-AOS2276.

[8] C. Xu, H. Jiang and Y. Xie. Conformal prediction for multi-dimensional time series by ellipsoidal sets. In *Proceedings of ICML*, PMLR 235:55076–55099, 2024.

[9] E. English, E. Wong-Toi, M. Fontana, S. Mandt, P. Smyth and C. Lippert. JANET: Joint adaptive prediction-region estimation for time-series. *Machine Learning*, 114:177, 2025. doi:10.1007/s10994-025- 06812-2.

[10] S. Rousseau and S. Messoudi. Scaling-score conformal prediction for multi-target regression. arXiv:2609.17091, 2026.

[11] L. El Halabi and A. Brandt. Adaptive conformal inference under delayed feedback: Coverage guarantees and a delay-to-memory diagnostic. arXiv:2609.07251, 2026.

[12] C. Zhang, H. Ren, Z. Wang, Y. Bao and C. Zou. Minimum volume optimal transport for multivariate conformal prediction. arXiv:2610.14907, 2026.

[13] H. Zhou, S. Zhang, J. Peng, S. Zhang, J. Li, H. Xiong and W. Zhang. Informer: Beyond efficient transformer for long sequence time-series forecasting. *Proceedings of the AAAI Conference on Artificial Intelligence*, 35(12):11106–11115, 2021.

[14] G. Lai, W.-C. Chang, Y. Yang and H. Liu. Modeling long- and short-term temporal patterns with deep neural networks. In *Proceedings of SIGIR*, pages 95–104, 2018. doi:10.1145/3209978.3210006.