A SHARP FINITE-WINDOW NO-FREE-LUNCH FOR COVERAGE-SUFFICIENT DEPENDENCE: EXACT BLOCK-CONFORMAL COVERAGE UNDER STRUCTURAL ENVELOPES

Anonymous authors
Paper under double-blind review

ABSTRACT

Can a finite calibration window certify that an exact conformal rank is safe to release? We impose the explicit information contract that every reported bound and release decision is measurable with respect to $O_B = (S_1, \ldots, S_B)$ and independent auxiliary randomness; future scores, transition data, or a dependence envelope enter only when declared as additional evidence. Under this contract, two stationary ergodic processes can have identical laws for O_B while their strict rank-miss probabilities lie on opposite sides of the nominal level. At $B = 9$ they are $17/162$ and $1/162$; the construction extends to every finite conformal rank and forces a rule that releases only when its pointwise-honest upper bound is nominal to refuse on the benign member. A related pair gives exact minimax absolute risk $(1 - 2^{-B})/2$ for estimating the score-block discrepancy over the stationary ergodic binary class with known uniform marginal. "Sharp" refers only to these stated restricted experiments, not to all lags, process classes, or release procedures. We then target the weaker score-block discrepancy sufficient for coverage: miscoverage is bounded by the exact rank term plus this discrepancy, at most $B\beta_S(h)$ for a stationary score process. A conservative positive route requires separately supplied finite-state Markov structure, fixed emissions, transition evidence, and Dobrushin contraction. Controlled experiments test that premise-matched route; named-data diagnostics certify no premise.

1 INTRODUCTION

Conformal prediction (CP) provides distribution-free uncertainty quantification with finite-sample coverage guarantees (Vovk et al., 2005; Angelopoulos & Bates, 2021), but rests on exchangeability of calibration and test data. Exchangeability can fail once the calibration and test indices are ordered in time: a stationary process whose finite-dimensional laws are not permutation invariant, for example one with a phase-dependent or lag-asymmetric joint law, violates it, and a nonstationary trend or regime change violates it as well. Correlation alone is not sufficient for this failure, since an exchangeable sequence may still be correlated; we therefore state the temporal condition directly and make no claim about how often it holds in any application domain.

Two routes and what each leaves open. The reactive route adapts quantiles online: ACI (Gibbs & Candes, 2021; Gibbs & Candès, 2024) and SPCI (Xu & Xie, 2023) target asymptotic or long-run coverage, while EnbPI (Xu & Xie, 2021) gives finite-sample approximate marginal coverage under strongly mixing errors. These results do not provide a finite-data certificate, computed from the calibration trajectory alone, for the exact score-block dependence quantity used here. The structural route assembles separated-block calibration with Berbee coupling (Bradley, 2007; Oliveira et al., 2024; Lee & Matni, 2024) and conformal risk control (Angelopoulos et al., 2024b; Farinhas et al., 2024). Generic versions require a valid mixing envelope, block-to-point slack, and any recalibration null as inputs. This does not make data-adaptive choices intrinsically invalid: under Markov or dedicated selection structure, thinning, dependence estimation, and model/set selection can carry their own validity analyses (Zheng & Proutiere, 2024; Grunewalder & Khaleghi, 2026; Liang et al.,

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

Table 1: Claim–evidence ledger, ordered with the primary no-free-lunch result first. Named datasets appear only in the last row and certify no premise.

Claim	Direct evidence	Premises not supplied by that evidence	Status
Finite windows do not certify $d_{2^n} < 1/2$	Theorem 2.4; exact de Bruijn enumeration	None beyond stationary ergodic binary scores	Exact finite-n boundary
Score-block dependence suffices for coverage	Theorem A.4; indexed Berbee construction	Externally valid score-block or score-mixing envelope	Exact conditional theorem
Finite-state envelope and fail-closed coverage	Public FrozenLake kernel; exact rank; 5,000 paths; 260,000 authored-chain trials	Observed stationary finite-state Markov state; fixed memoryless emission; independent path or charged historical-prefix separation	Executed conditional theorem

2024; Yang & Kuchibhotla, 2021). Those are richer information contracts than the score-only O_B contract studied here.

Our contribution: an identification boundary with a coverage consequence. Theorems 2.6–2.7 are the center of the paper: under the finite-window information contract above, identical calibration laws can hide strict rank-miss probabilities on opposite sides of the release threshold. The remaining results locate that obstruction. A finite score window does not identify the declared future-lag pairwise TV; the complete score law identifies score dependence but not full-process mixing; the exact block-score discrepancy is coverage-sufficient; and score-process mixing upper-bounds that discrepancy. Observed finite-state evidence yields one conservative structural upper bound. Block-average translation and monitoring remain assumption-conditional.

Contributions. The primary contribution is the non-identifiability construction and the weaker dependence target it reveals. The constructive finite-state route composes established transition concentration, Dobrushin contraction, coupling, and conformal rank tools; no novelty is claimed for those ingredients or for deployment reach.

Primary theorem/counterexample pair.

- **P1. Strict-rank release obstruction** (Theorems 2.6–2.7). At $B = 9, h = 1$, two processes with identical full calibration-window laws have strict miss probabilities $17/162$ and $1/162$, giving estimation lower bound $4/81$ and an honest-release refusal bound. The construction covers every $1 \le k \le B$, with exact separation $1/(B+1)$ across the nominal rank level, including nonmaximum ranks.
- **P2. Functional boundary** (Theorem 2.5). Under the stated uniform-marginal binary restriction, the block discrepancy has exact minimax value $(1 - 2^{-B})/2$. Theorem 2.8 separately gives mixture-only pairwise bounds for every $h \ge n$; neither result is a universal sharpness claim over arbitrary marginals and lags.
- **P3. Full-process non-identifiability** (Proposition A.6.7). The complete score law still cannot identify full-process mixing; a refresh family makes the boundary arbitrarily sharp.
- **P4. Coverage-sufficient dependence functional** (Theorem A.4; Theorems A.1–A.2). Score-block TV yields the rank bound, while observed finite-state evidence supplies an envelope only under explicit structural premises.

Bridge to practice. The assumption audit in Section 2.2 distinguishes theorem mode from empirical-only mode. Ordinary score diagnostics cannot discharge an external dependence premise.

Consequences. Under a supplied valid envelope, Theorem A.4 uses the exact finite-B conformal rank; there is no DKW penalty. Its only additions to α_{rank} are the B explicit coupling links and, when relevant, a separately justified translation event.

For a declared score-process envelope $\beta_S(h) \le C_\beta e^{-\lambda h}$, theorem-side block miscoverage is

$$\alpha_{\mathrm{rank}} + BC_\beta e^{-\lambda h}. \tag{1}$$

The finite-state route justifies such an envelope structurally and adds confidence-set failure δ_{env}. A fitted one-trajectory score analogue remains diagnostic only.

Theorem-level novelty comparison. Table 2 is a selected, non-exhaustive comparison of targets, assumptions, and guarantees rather than evidence of priority. The exact minimax statement is confined

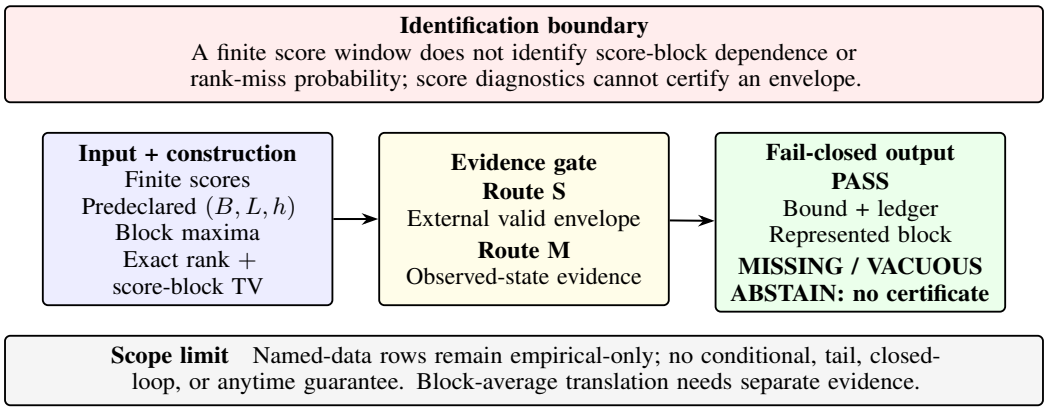

Figure 1: Whole-paper map. Route S requires an externally valid score-block or score-mixing envelope. Route M requires the complete G1–G4 audit in Section 2.2: observed stationary finite first-order Markov state with fixed memoryless emission; independent transition evidence or a separated historical prefix with predeclared cap; simultaneous nonempty transition row sets with $d_U < 1$; and an indexed $B + 1$ geometry within the full budget. Missing or vacuous evidence forces $\widehat{q} = +\infty$: abstain with no certificate.

to the displayed uniform-marginal binary experiment at lag 2^n; Theorems 2.6–2.7 instead establish the finite-window strict-rank release obstruction. We claim neither a universal sharp bound, priority for general ergodic impossibility, nor a stronger generic dependent-CP rate.

Table 2: Information contracts and guarantee targets. "Observed information" names what may enter the procedure; it is not a claim that assumptions are empirically certified.

Work	Observed information	Structure / guarantee target	Relation to this paper
Barber & Pananjady (2026)	Time series and predictor history	Switch coefficient; sharp coverage characterization over stationary β-mixing class	Population coverage quantity, not an O_B certificate
Zheng & Proutiere (2024)	Markovian calibration data	Mixing-time bound; adaptive K-split thinning	Positive structured route, not unrestricted score-only inference
Grunewalder & Khaleghi (2026)	One Markov sample path	Geometric ergodicity plus smoothness, or finite state; finite-n $\beta(m)$ estimation	Structure makes estimation possible
Liang et al. (2024); Yang & Kuchibhotla (2021)	Holdout data and candidate models/sets	Selection-specific finite-sample coverage procedures	Adaptation can be valid; not temporal lag selection
This paper	O_B and declared auxiliary evidence	Score-only strict-rank obstruction; conditional Markov escape hatch	Separates what O_B identifies from what added structure supplies

Related work. McDonald et al. and Khaleghi–Lugosi prove single-path consistency for mixing targets (McDonald et al., 2015; Khaleghi & Lugosi, 2023); Grunewalder–Khaleghi give high-probability rates under geometric Markov and smoothness/finite-state structure (Grunewalder & Khaleghi, 2026). Adams–Nobel target ergodic density estimation and Ryabko target process discrimination (Adams & Nobel, 1998; Ryabko, 2010). We claim only an exact finite-n two-point boundary for d_{2^n} and its honest-UCB corollary, not priority for general ergodic impossibility or mixing estimation. The coverage proof is in Appendix A.8.

What this paper is *not*. The conditional certificate is not a strict β-mixing test, a conditional/tail-coverage guarantee, a closed-loop control certificate, or a general fallback for unknown shift. It computes a finite-sample bound only when independently justified envelopes are supplied and otherwise refuses certification. Repeated releases receive separate per-period statements; a family-wise claim requires explicit error-budget allocation, and no anytime guarantee is implied.

2 BLOCK-WISE CONFORMAL CALIBRATION

2.1 SETUP AND ASSUMPTIONS

Let $\{(X_t, Y_t)\}_{t=1}^{T}$ be a stationary time series with predictor $\hat{f}$ and nonconformity scores $s_t = |Y_t - \hat{f}(X_t)|$.

We use $\mathrm{TV}(P, Q) = \sup_A |P(A) - Q(A)|$, equal to $\frac{1}{2}\|P - Q\|_1$ on a discrete space. Absolute regularity of a process W is $\beta_W(k) = \sup_t \mathrm{TV}(\mathcal{L}(W_{:t}, W_{t+k:}), \mathcal{L}(W_{:t}) \otimes \mathcal{L}(W_{t+k:}))$, equivalently the expected conditional-TV form. This is β-mixing (absolute regularity), not Rosenblatt α-mixing. There is no factor two: a Berbee replacement fails with probability at most $\beta_W(k)$.

Assumption 2.1 (Score stationarity and mixing)**.** For the generic score-envelope route, the score process (S_t) induced by a fixed predictor is strictly stationary, bounded by a known M, and obeys an externally justified envelope $\beta_S(k) \le C_\beta e^{-\lambda k}$. The constants are assumptions; an ACF or finite-class fit on the same trajectory does not certify them.

Remark 2.2 (Complete law versus finite trajectory). The complete score law determines β_S and the block-law discrepancy in Theorem A.4; full-process β_Z is unnecessary. Theorem 2.4 proves only the stated $1/2$ lower threshold for an honest UCB targeting d_{2^n}. In the unrestricted class, ACF, Ljung–Box, bootstrap, and Definition 2.3 remain plausibility screens unless additional assumptions justify their conversion to an envelope.

2.2 ASSUMPTION AUDIT AND FAIL-CLOSED WORKFLOW

The theorem has two evidence routes. Route S supplies an external score-block discrepancy or score-mixing envelope and records its scope. Route M derives a conservative score envelope by recording: (G1) observed stationary finite first-order Markov state and fixed memoryless emission; (G2) independent transition evidence or a separated historical prefix with predeclared cap; (G3) simultaneous nonempty transition row sets and $d_U < 1$; and (G4) an indexed $B + 1$ geometry whose full budget is at most τ. Missing evidence, a fitted ACF, or nominal empirical coverage cannot pass either route. Named datasets establish neither route and remain empirical only.

2.3 FINITE-CLASS DEPENDENCE SURROGATE

The coverage theorem depends on a declared upper envelope (C_β, λ) for the strict mixing coefficient. A data trajectory can still provide a finite-class descriptive diagnostic, but not an envelope-free finite-sample certificate. We keep these roles distinct.

Let $\mathcal{F}_M$ be a finite-VC class of $[0, M]$-valued bounded test functions on the score space (e.g., indicator functions of a finite partition with $V < \infty$ atoms). Define the empirical $\mathcal{F}_M$-restricted β-surrogate at lag k:

$$\hat{\beta}_{\mathcal{F}_M}(k) := \sup_{f,g \in \mathcal{F}_M} \left| \frac{1}{N_k} \sum_{t=1}^{N_k} f(s_t) g(s_{t+k}) - \left(\frac{1}{N_k} \sum_{t=1}^{N_k} f(s_t) \right) \left(\frac{1}{N_k} \sum_{t=1}^{N_k} g(s_{t+k}) \right) \right|. \tag{2}$$

Here $N_k = N - k$, so every average has its actual sample-count normalization. Under factorization the corresponding *population* covariance is zero; the absolute empirical supremum need not have zero finite-sample expectation and can be upward biased. It remains descriptive only.

Definition 2.3 (Finite-Class Score Diagnostic)**.** For a declared finite class $\mathcal{F}_M$, Eq. (2) is a descriptive statistic of dependence visible in the score sequence. We write the corresponding population functional as

$$\beta_{\mathcal{F}_M}(k) = \sup_{f,g \in \mathcal{F}_M} |\mathbb{E}[f(s_t) g(s_{t+k})] - \mathbb{E}[f]\mathbb{E}[g]| \,.$$

It obeys the lower-direction relation $\beta_{\mathcal{F}_M}(k) \le M^2 \beta_S(k)$, so small diagnostic values cannot upper-bound the score-process coefficient $\beta_S(k)$. We make no envelope-free root-N concentration claim. What is proved is narrower than a universal principle: for the coverage-sufficient functional targeted here, an honest upper bound for d_{2^n} based on a length-n trajectory must be at least $1/2$ with probability at least $1 - \delta$ on the common witness observation law, so this paper's own finite-sample statements carry an externally supplied envelope. We do not assert that every finite-sample statement about dependent observations must do so.

Theorem 2.4 (Sharp finite-window no-free-lunch). *Fix $n \geq 2$ and set $L = 2^n$. There exist stationary ergodic binary score processes P_0, P_1 such that*

$$\mathcal{L}_{P_0}(S_1, \ldots, S_n) = \mathcal{L}_{P_1}(S_1, \ldots, S_n) = \text{Bernoulli}(1/2)^{\otimes n},$$

but, for $d_L(P) := \text{TV}(\mathcal{L}_P(S_0, S_L), \mathcal{L}_P(S_0) \otimes \mathcal{L}_P(S_L))$,

$$d_L(P_0) = 0, \qquad d_L(P_1) = \frac{1}{2}.$$

For this result, the observation experiment is only $O_n = (S_1, \ldots, S_n)$, while the two singleton blocks are $A_1 = S_0, A_2 = S_L$; hence $B = 1$, block length one, gap L, and $\Delta^{\text{score}}_{1,2^n} = d_L$. Let $\mathcal{P}$ be all stationary ergodic binary score processes. Over all measurable real-valued estimators of O_n,

$$\inf_{\widehat{\Delta}_n} \sup_{P \in \mathcal{P}} \mathbb{E}_P |\widehat{\Delta}_n - d_L(P)| = \frac{1}{4}.$$

For all tests allowed independent auxiliary randomization, of $H_0 = \{P : d_L(P) = 0\}$ against $H_1 = \{P : d_L(P) = 1/2\}$,

$$\inf_{\phi_n} \max\left\{ \sup_{P \in H_0} P(\phi_n = 1), \sup_{P \in H_1} P(\phi_n = 0) \right\} = \frac{1}{2}.$$

Consequently, if $U_n(O_n)$ has pointwise marginal honesty, $P\{U_n \geq d_L(P)\} \geq 1 - \delta$ for every stationary ergodic binary process P, then $P_0\{U_n \geq 1/2\} \geq 1 - \delta$. Equivalently, on the common i.i.d.-looking data law, a uniformly honest bound can return below $1/2$ with probability at most δ.

Proof. Let P_0 be i.i.d. Bernoulli$(1/2)$. For P_1, take a binary de Bruijn cycle of order n, choose its phase uniformly, and repeat its L symbols periodically. The shift acts transitively on the finite orbit, so the random-phase law is stationary and ergodic. Every binary n-word occurs at exactly one phase, proving equality of the observed-window laws. Periodicity gives $S_L = S_0$; both marginals are uniform, so the diagonal joint law has joint-to-product TV $1/2$. With length-one blocks, $\Delta^{\text{score}}_{1,L} = d_L$. Identical observations and $|x| + |x - 1/2| \geq 1/2$ give the estimation lower bound; the constant estimator $1/4$ matches it because binary pairwise joint-to-product TV lies in $[0, 1/2]$. Identical hypotheses give the testing lower bound, matched by a fair-coin randomized test. The confidence statement follows by observational equality, while the constant bound $U_n = 1/2$ is uniformly valid. □

Exact witnesses. The artifact `simulation/results/fems21_finite_trajectory_impossibility_v4_20260831.json` verifies all 2^n phases for $n = 2, \ldots, 10$: every n-window occurs once, each cyclic orbit is stationary ergodic, and $\Delta^{\text{score}}_{1,2^n} = 1/2$.

Theorem 2.5 (No-free-lunch at the exact B+1-block functional). *Fix $n \geq 2$ and $B \geq 1$, and take the geometry of Theorem A.4 with block length $L = 1$ and gap $h = 2^n$, so $a_{b+1} = a_b + h$. Let P_0 be i.i.d. Bernoulli$(1/2)$ and let P_1 be a uniform random phase into a binary de Bruijn cycle of order n. Their observed n-window laws are both $\text{Bernoulli}(1/2)^{\otimes n}$, yet*

$$\Delta^{\text{score}}_{B,2^n}(P_0) = 0, \qquad \Delta^{\text{score}}_{B,2^n}(P_1) = 1 - 2^{-B}.$$

Let $\mathcal{P}_{1/2}$ be the stationary ergodic binary score processes with $P(S_t = 1) = 1/2$. For every $P \in \mathcal{P}_{1/2}$, $0 \leq \Delta^{\text{score}}_{B,2^n}(P) \leq 1 - 2^{-B}$, and over all estimators of $O_n = (S_1, \ldots, S_n)$,

$$\inf_{\widehat{\Delta}_n} \sup_{P \in \mathcal{P}_{1/2}} \mathbb{E}_P \left| \widehat{\Delta}_n - \Delta^{\text{score}}_{B,2^n}(P) \right| = \frac{1 - 2^{-B}}{2}.$$

Testing $\{\Delta^{\text{score}}_{B,2^n} = 0\}$ against $\{\Delta^{\text{score}}_{B,2^n} = 1 - 2^{-B}\}$ has minimax error $1/2$, and any pointwise $(1 - \delta)$-honest U_n satisfies $P_0\{U_n \geq 1 - 2^{-B}\} \geq 1 - \delta$. At the nonvacuous rank geometry $B = 9$ with $\alpha_{\text{rank}} = .10$ it gives $511/512$ and $511/1024$.

The proof is in Appendix A.6.

The uniform-marginal restriction is substantive. Proposition A.6.3 gives a common-q six-block law whose discrepancy exceeds $1 - q^6 - (1 - q)^6$; no arbitrary-q closed form is claimed.

Theorem 2.6 (Operational full-span strict-miss boundary). *Fix $B = 9$, $L = h = 1$, $\alpha_{\mathrm{rank}} = 1/10$, and hence $k_B = 9$. Let $P_\uparrow$ and $P_\downarrow$ be uniform random phases into the period-162 cycles $1\,0^8\,2\,0^8 \cdots 18\,0^8$ and $18\,0^8\,17\,0^8 \cdots 1\,0^8$. Their complete calibration observations $O_9 = (S_1, \ldots, S_9)$ have identical laws, but for the strict rank miss $\mu(P) = P\{S_{10} > \max_{j \le 9} S_j\}$,*

$$\mu(P_\uparrow) = \frac{17}{162} = \frac{1}{10} + \frac{2}{405}, \qquad \mu(P_\downarrow) = \frac{1}{162}.$$

Thus every estimator of μ from O_9 has minimax absolute risk at least $4/81$, and randomized testing error is $1/2$. Any pointwise $(1-\delta)$-honest upper bound must be at least $17/162$ with probability $1 - \delta$ on the common observation law; a rule releasing only when $U \le 1/10$ therefore refuses with probability at least $1 - \delta$ under $P_\downarrow$, although its true miss probability is $1/162$.

For $L \le n$ with spacing $L + h - 1 = 2^n$ the blocks are exact translates, so every A_b equals A_1, zero with probability $u = 2^{-L}$; the joint puts mass u and $1 - u$ on the two constant strings against product masses u^{B+1} and $(1-u)^{B+1}$, giving $\Delta^{\mathrm{score}}_{B,h}(P_1) = 1 - u^{B+1} - (1-u)^{B+1}$, which returns $1 - 2^{-B}$ at $L = 1$. `simulation/block_discrepancy_witness.py` recomputes every value above in exact rational arithmetic for $n \le 5$ and $B \in \{1, 2, 3, 5, 9\}$, so the obstruction holds for the functional Theorem A.4 charges, not only the singleton case.

Theorem 2.7 (Operational strict-miss boundary at every conformal rank). *Fix $B \ge 1$, $1 \le k \le B$, $L = h = 1$, and $\alpha^{\mathrm{rank}}_B = (B + 1 - k)/(B + 1)$. Put $s = B - k$ and $q = 2(B + 1)$. For each amplitude $v \in \{1, \ldots, q\}$ form the length-B block $v\,0^{k-1}(q+1)^s$. Concatenate these blocks in increasing or decreasing amplitude order and choose a uniform random phase; divide scores by $q + 1$ to place them in $[0, 1]$. Both processes are stationary ergodic and have identical full calibration laws $O_B = (S_1, \ldots, S_B)$, but their strict rank-miss probabilities $\mu(P) = P\{S_{B+1} > S_{(k)}\}$ are*

$$\mu_+ = \frac{s}{B} + \frac{q-1}{qB}, \qquad \mu_- = \frac{s}{B} + \frac{1}{qB}, \qquad \mu_+ - \mu_- = \frac{1}{B+1}.$$

Moreover $\mu_- < \alpha^{\mathrm{rank}}_B < \mu_+$. Any estimator of μ from O_B has maximum absolute risk over this pair at least $1/[2(B+1)]$; randomized minimax testing error is $1/2$. Any pointwise $(1-\delta)$-honest upper bound satisfies $U \ge \mu_+$ with probability at least $1 - \delta$ on the common law. A rule releasing only if $U \le \alpha^{\mathrm{rank}}_B$ therefore refuses on the benign process with probability at least $1 - \delta$. Appendix A.6 proves the claim; `simulation/all_rank_operational_witness.py` verifies 328 members by exact enumeration, including $(B, k) = (99, 90)$.

This covers every finite conformal rank, not only the calibration maximum. The nominal levels are the rank-grid values displayed above; the construction is stationary ergodic but not mixing, and Theorem A.6.8 quantifies when a supplied mixing envelope does and does not defeat it. For $(B, k) = (19, 18)$ the miss probabilities are $79/760$ and $41/760$, on opposite sides of the nominal level 0.1.

Theorem 2.8 (The n-versus-h boundary). *Fix $h \ge n$ and let Q_x, $x \in \{0,1\}^h$, repeat x cyclically with a uniform random phase; each Q_x is stationary ergodic. The uniform mixture of $\{Q_x\}$ has $\mathcal{L}(S_1, \ldots, S_n) = \mathrm{Bernoulli}(1/2)^{\otimes n}$ exactly, while $\Delta^{\mathrm{score}}_{1,h}(Q_x) = 2p_x(1-p_x)$ with p_x the fraction of ones in x and mean $\frac{1}{2} - \frac{1}{2h}$. For $0 < t \le 1/2$ with $\pi_t := 2^{-h}|S_t| > 0$, restrict the prior to $S_t = \{x : d_h(Q_x) \ge t\}$, the split $\mathcal{L}_{\mathrm{iid}} = \pi_t M_{S_t} + (1 - \pi_t) M_{S_t^c}$ gives $\mathrm{TV}(M_{S_t}, \mathcal{L}_{\mathrm{iid}}) \le 1 - \pi_t =: \kappa_{h,t}$, and therefore, at* every *lag $h \ge n$ and not only at 2^n, for the pairwise quantity $d_h = \Delta^{\mathrm{score}}_{1,h}$ only,*

$$\inf_{\phi} \max\{\textit{type I}, \textit{type II}\} \ge \tfrac{1}{2}(1 - \kappa_{h,t}), \quad \inf_{\widehat{\Delta}} \sup_{P} \mathbb{E}_P|\widehat{\Delta} - d_h| \ge \tfrac{t}{2}(1 - \kappa_{h,t}),$$

and any $(1-\delta)$-honest U_n valued in $[0, 1/2]$ satisfies $\mathcal{L}_{\mathrm{iid}}\{U_n \ge t\} \ge 1 - \delta - \kappa_{h,t}$. At $h = 14$, $t = 2/5$, $\delta = .05$ these are .4713, .1885, and .8926. Proposition A.6.5 identifies $h < n$ exactly; for $h \ge n$ the displayed result is mixture-only, and exact individual-process minimax characterization in the intermediate regime remains open.

The averaging is not an artifact: Theorem A.6.4 rules out any *single* ergodic process with an i.i.d. n-window and maximal lag-h dependence below $h = 2^{n-1}$, and forces $h \ge 2^n$ only when the maximal pair is required to satisfy $S_h = S_0$ almost surely. Non-maximal pairs do exist in that regime: Proposition A.6.6 exhibits one at $(n, h) = (5, 8)$ with identical window laws and gap $3/49$ (Appendix A.6).

Complete-law/full-process boundary. Even the complete score law cannot identify full-process mixing: Proposition A.6.7 in Appendix A.5 gives an arbitrarily persistent finite-state nuisance with the same i.i.d. scores. This secondary boundary is distinct from the finite-window score-functional lower bound above.

Remark 2.9 (Why there is no score-process analogue). The complete score-process law determines $\beta_S(k)$ by definition: every past–future event probability needed by $\beta_S(k)$ is a functional of that law. Therefore two processes with the same complete score law cannot have different score-process mixing coefficients. Proposition A.6.7 is the strongest possible full-process non-identifiability statement at this information level. Theorem A.4 supplies the weaker coverage-sufficient functional; this paper supplies no finite-data upper confidence bound for it or for β_S from one trajectory. The nuisance coordinate is irrelevant to score dependence and coverage.

Executable scope check. The former binary example is the special case $K = 2$, $a = 1 - 2q$ and therefore reaches only $1/2$ as $q \downarrow 0$; by itself it cannot prove the proposition above. The strengthened refresh family closes that gap. A local matrix-power check independently verifies $a^k(1 - 1/K)$. The result concerns only full-process β_Z. Score-process and coverage-sufficient quantities are identified by the complete score law and enter Theorem A.4; fitted $(\hat{C}_\beta, \hat{\lambda})$ from one trajectory may guide sensitivity analysis but never enter a theorem bound.

Relation to prior work. Prior dependent-CP results (Oliveira et al., 2024; Lee & Matni, 2024; Farinhas et al., 2024) likewise treat valid dependence envelopes as inputs. Finite-state identification results such as Zheng & Proutiere (2024) rely on additional state-space structure. Our boundary isolates the score-block object actually used by rank coverage; the next result gives one restrictive mechanism for certifying it.

2.4 Closest Theorems in This Notation, and What Is Not a Corollary

Four recent results are close enough that a reader should be able to check, from the submission alone, which statements here already follow from them. Write O_n for the observed calibration window, $\Delta^{\mathrm{score}}_{B,h}$ for the $B{+}1$-block joint-to-product TV charged by Theorem A.4, and β_S for score-process absolute regularity, so that $\Delta^{\mathrm{score}}_{B,h} \le B\beta_S(h)$ for a stationary score process.

- **Split CP under dependence** (Oliveira et al., 2024) and **single-trajectory CP** (Lee & Matni, 2024). Oliveira et al. bound miscoverage by a rank term plus a Berbee-coupling penalty that is linear in a *declared* mixing coefficient at the separation lag, times the number of coupled segments, in the form of Eq. (1); they calibrate on points inside blocks and pay an additional concentration term. Lee and Matni analyze risk-controlling prediction sets from one trajectory by blocking and decoupling dependent data, with an additive mixing-error term; we do not attribute our exact finite-B rank or joint-to-product-TV theorem to them. *Corollary status:* Theorem A.4 states the joint-to-product-TV intermediate comparison used here as the theorem, with the exact finite-B rank and no concentration term; we claim no priority for it. *Not implied:* both results take the envelope as an input and contain no statement about whether O_n can supply it, so Theorems 2.4, 2.5, and 2.6 do not follow from them; conversely, they are not contradicted, because the pairwise discrepancy d_{2^n} equals 0 and $1/2$ for the witness pair, not their full past–future mixing coefficients. The periodic de Bruijn process has $\beta_S(k) = 1 - 2^{-n}$ at every lag: both half-infinite sequences identify its uniformly random phase. It therefore fails any decaying mixing-envelope premise, while the i.i.d. witness has $\beta_S(k) = 0$.
- **CP under Markovian data** (Zheng & Proutiere, 2024). The coverage gap is controlled through the mixing time of the data chain itself, with a split scheme for unknown mixing time. *Corollary status:* Route M is the same idea in a more restrictive form (observed finite state, fixed memoryless emission, Dobrushin gate from an anytime-valid transition confidence set); we claim no priority for exploiting Markov structure. *Not implied:* the state there is the observed datum; here it must be observed *in addition to* the scores, and Proposition A.6.7 shows why: two processes with the same complete score law can have arbitrarily different full-process mixing, so scores alone cannot identify the full-process mixing time without additional structure.
- **Split CP for time series** (Barber & Pananjady, 2026). For stationary time series, Barber and Pananjady characterize coverage loss through a switch coefficient that measures temporal violations of exchangeability, with a sharp characterization over stationary β-mixing processes. This is not an empirical calibration-distribution-deviation bound. *Corollary status:* none of our statements

follows from it and it does not follow from ours. *Not implied, and consistent:* its bound is a process-dependent quantity, not a certificate computed from O_n; Theorem 2.6 exhibits two stationary ergodic processes with identical complete nine-score calibration laws whose strict miss probabilities differ by $16/162$, so no function of the calibration scores alone can be simultaneously honest and tight for the miss probability, whatever bound one attaches to the gap. The two messages are compatible: the gap may well be small, but its smallness is not identifiable from the window.

Theorem 2.8 and the exact q-counterexample (Proposition A.6.3) have no counterpart in these four works; they are internal to the identification question and use only the two-point method, for which no priority is claimed.

Data dependence of a choice is therefore not, by itself, a validity failure. The primary abstracts of Liang et al. (2024) and Yang & Kuchibhotla (2021) report dedicated model/set-selection procedures with finite-sample coverage guarantees (and efficiency tradeoffs). We use those works only for that abstract-level point: their theorem assumptions were not used to justify adaptive temporal geometry here. Our 100-cell geometry study remains descriptive because it does not instantiate either selection design.

3 End-to-End Premise-Matched Validation

We instantiate the theorem on Gymnasium `FrozenLake-v1`'s public slippery 4-by-4 transition kernel under a fixed uniform-action policy and reset-on-terminal rule. The resulting 16-state chain and its stationary law are computed exactly; no score ACF, bootstrap, or calibration-path estimate enters the envelope. For $B = 99, L = 4$, the smallest exact gap meeting a .005 coupling budget is $h = 27$, with $B \max_s \mathrm{TV}(P^h(s,\cdot),\pi) = .004607$. The resulting analytic coverage floor is .895393. Across 5,000 untouched paths with a fixed native environment-state score $|S_t - \mathbb{E}_\pi S|/15$, observed future-block coverage is .9246 (Wilson 95% [.9170,.9316]). A pointwise Split-CP threshold evaluated on the same simultaneous four-state target covers .8516 ([.8415,.8612]); mean thresholds are .4379 and .3388. Thus the comparison demonstrates target alignment rather than interval efficiency: block-max calibration pays a wider threshold to control the represented block. This is a complete public-kernel theorem instantiation with no fitted envelope, stochastic emission, or real-data claim.

The authored two-state and independently estimated-transition variants remain secondary checks in Appendix A: their measured block coverages are 89.9805% and 89.843%, respectively. All reported transition-set replications contain the true matrix; this empirical outcome is distinct from the analytic confidence guarantee.

The conformal term is the finite-grid value $\alpha_B^{\mathrm{rank}} = (B + 1 - k_B)/(B + 1)$, bounded by but never replaced with the nominal α_{rank}; ties retain the one-sided inequality. Full aggregate summaries are in Appendix A; per-seed rows are retained in the corresponding machine-readable experiment outputs. Non-circular geometry selection is evaluated on 100 evidence-length cells within 20 seed trajectories (Appendix A.12). The multi-step envelope releases in exactly 60/100 cells and refuses in 40/100. With an infinite threshold on refusal, its refusal-inclusive represented-block coverage is .95564; among released cells only, descriptive coverage is .9260667, with no conditional-on-release guarantee. The one-step rule refuses in 100/100 cells; the minimum selects the same geometries as the multi-step rule.

An independently sourced traffic-series confirmation provides a separate empirical transport check (Appendix A.14). It releases 20 of 54 complete 100-hour windows with 19/20 observed block coverage. The prespecified joint coverage/loss criterion fails; these observations neither establish the theorem premises nor demonstrate deployment benefit. Independent transition evidence also yields nonvacuous multi-step certificates under the public 11-state closure premise, reaching a minimum certified gap of 40 at the largest evidence length; the full 16-state model refuses throughout (Appendix A.13).

4 Limitations and Conclusion

The impossibility result concerns the stated score-block functional, not every dependence functional, and the constructive guarantee applies only when its finite-state, fixed-emission, independent-evidence,

and geometry premises hold. The exact individual-process minimax characterization in the intermediate regime remains open. Named-data diagnostics establish neither those premises nor deployment benefit: the traffic-series release fails its prespecified joint coverage/loss criterion, and refusal remains a substantive outcome rather than evidence of coverage. The main conclusion is therefore a boundary and a conditional remedy: finite observed score windows cannot certify the dependence quantity used by the displayed block-conformal bound, while external structure can provide a conservative envelope when every premise is independently met.

ETHICS STATEMENT

The public Electricity artifact contains anonymized customer-meter columns. We make no household-level inference. Its retained upstream record states no explicit license, and the study must not be read as validating exchangeability or a deployment population.

REPRODUCIBILITY STATEMENT

The anonymous supplementary ZIP contains public ETT copies, exact dependency pins, deterministic replay scripts, generated outputs, tests, and a checksum verifier; it excludes the proprietary Papermaking trajectory and the locally retained Electricity copy whose redistribution license is unestablished. The score-identifiability counterexample is reproduced by `simulation/src/po02_counterexample.py`; the finite-state results are reproduced by `finite_state_coverage_experiment.py` and `same_trajectory_experiment.py`. This PDF uses the unmodified official ICLR 2027 conference style copied into the build directory.

REFERENCES

Terrence M. Adams and Andrew B. Nobel. On density estimation from ergodic processes. *The Annals of Probability*, 26(2):794–804, 1998. doi: 10.1214/aop/1022855650.

Anastasios N Angelopoulos and Stephen Bates. A gentle introduction to conformal prediction and distribution-free uncertainty quantification. *arXiv preprint arXiv:2107.07511*, 2021.

Anastasios N Angelopoulos, Rina Foygel Barber, and Stephen Bates. Online conformal prediction with decaying step sizes. In *International Conference on Machine Learning*, 2024a.

Anastasios N Angelopoulos, Stephen Bates, Adam Fisch, Lihua Lei, and Tal Schuster. Conformal risk control. In *International Conference on Learning Representations*, 2024b.

Rina Foygel Barber and Ashwin Pananjady. Predictive inference for time series: why is split conformal effective despite temporal dependence? In *Proceedings of the 37th International Conference on Algorithmic Learning Theory*, 2026. To appear; arXiv:2510.02471.

Rina Foygel Barber, Emmanuel J Candes, Aaditya Ramdas, and Ryan J Tibshirani. Conformal prediction beyond exchangeability. *The Annals of Statistics*, 51(2):816–845, 2023.

Stephen Bates, Anastasios Angelopoulos, Lihua Lei, Jitendra Malik, and Michael Jordan. Distribution-free, risk-controlling prediction sets. *Journal of the ACM*, 68(6):1–34, 2021.

Richard C Bradley. *Introduction to strong mixing conditions*. Kendrick Press, 2007.

Victor Chernozhukov, Kaspar Wüthrich, and Yinchu Zhu. Exact and robust conformal inference methods for predictive machine learning with dependent data. In *Proceedings of the 31st Conference On Learning Theory*, volume 75 of *Proceedings of Machine Learning Research*, pp. 732–749. PMLR, 2018.

António Farinhas, Chrysoula Zerva, Dennis Ulmer, and André Martins. Non-exchangeable conformal risk control. In *International Conference on Learning Representations*, 2024. URL https://proceedings.iclr.cc/paper_files/paper/2024/hash/de04896f011beff76c91e094f72727f4-Abstract-Conference.html.

Isaac Gibbs and Emmanuel Candes. Adaptive conformal inference under distribution shift. *Advances in Neural Information Processing Systems*, 34:1660–1672, 2021.

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

Steffen Grunewalder and Azadeh Khaleghi. Estimating the mixing coefficients of geometrically ergodic markov processes. *IEEE Transactions on Information Theory*, 2026. doi: 10.1109/TIT.2025.3639922. URL https://arxiv.org/abs/2402.07296.

Peter Grünwald, Rianne de Heide, and Wouter M Koolen. Safe testing. *Journal of the Royal Statistical Society: Series B*, 86(5):1091–1128, 2024.

Azadeh Khaleghi and Gábor Lugosi. Inferring the mixing properties of a stationary ergodic process from a single sample-path. *IEEE Transactions on Information Theory*, 2023. doi: 10.1109/TIT.2023.3248375. URL https://arxiv.org/abs/2106.07054.

Hans R Künsch. The jackknife and the bootstrap for general stationary observations. *The Annals of Statistics*, pp. 1217–1241, 1989.

Soumendra Nath Lahiri. *Resampling methods for dependent data*. Springer Science & Business Media, 2003.

Brian Lee and Nikolai Matni. Single trajectory conformal prediction. In *2024 IEEE 63rd Conference on Decision and Control (CDC)*, pp. 3019–3024, 2024. doi: 10.1109/CDC56724.2024.10886644.

Jonghyeok Lee, Chen Xu, and Yao Xie. Kernel-based optimally weighted conformal time-series prediction. In *International Conference on Learning Representations*, 2025. arXiv:2405.16828.

Jing Lei, Max G'Sell, Alessandro Rinaldo, Ryan J Tibshirani, and Larry Wasserman. Distribution-free predictive inference for regression. *Journal of the American Statistical Association*, 113(523):1094–1111, 2018.

Ruiting Liang, Wanrong Zhu, and Rina Foygel Barber. Conformal prediction after data-dependent model selection, 2024.

Daniel J McDonald, Cosma R Shalizi, and Mark Schervish. Estimating beta-mixing coefficients via histograms. *Electronic Journal of Statistics*, 9:2855–2883, 2015.

Roberto I. Oliveira, Paulo Orenstein, Thiago Ramos, and João Vitor Romano. Split conformal prediction and non-exchangeable data. *Journal of Machine Learning Research*, 25(225):1–38, 2024.

Dimitris N Politis and Joseph P Romano. The stationary bootstrap. *Journal of the American Statistical Association*, 89(428):1303–1313, 1994.

Dimitris N Politis and Halbert White. Automatic block-length selection for the dependent bootstrap. *Econometric Reviews*, 23(1):53–70, 2004.

Aaditya Ramdas, Peter Grünwald, Vladimir Vovk, and Glenn Shafer. Game-theoretic statistics and safe anytime-valid inference. *Statistical Science*, 38(4):576–601, 2023.

Yaniv Romano, Evan Patterson, and Emmanuel Candes. Conformalized quantile regression. *Advances in Neural Information Processing Systems*, 32, 2019.

Daniil Ryabko. Discrimination between b-processes is impossible. *Journal of Theoretical Probability*, 2010. doi: 10.1007/s10959-009-0263-1. URL `http://daniil.ryabko.net/discr_jtp.pdf`.

Ryan J Tibshirani, Rina Foygel Barber, Emmanuel J Candes, and Aaditya Ramdas. Conformal prediction under covariate shift. *Advances in Neural Information Processing Systems*, 32, 2019.

Vladimir Vovk. Conditional validity of inductive conformal predictors. In *Asian Conference on Machine Learning*, pp. 475–490. PMLR, 2012.

Vladimir Vovk, Alexander Gammerman, and Glenn Shafer. *Algorithmic Learning in a Random World*. Springer, 2005.

Ian Waudby-Smith and Aaditya Ramdas. Estimating means of bounded random variables by betting. *Journal of the Royal Statistical Society: Series B*, 86(1):1–27, 2024.

Chen Xu and Yao Xie. Conformal prediction interval for dynamic time-series. In *Proceedings of the 38th International Conference on Machine Learning*, volume 139 of *Proceedings of Machine Learning Research*, pp. 11559–11569. PMLR, 2021.

Chen Xu and Yao Xie. Sequential predictive conformal inference for time series. In *Proceedings of the 40th International Conference on Machine Learning*, volume 202 of *Proceedings of Machine Learning Research*, pp. 38707–38727. PMLR, 2023.

Yachong Yang and Arun Kumar Kuchibhotla. Selection and aggregation of conformal prediction sets, 2021.

Bin Yu. Rates of convergence for empirical processes of stationary mixing sequences. *The Annals of Probability*, pp. 94–116, 1994.

Margaux Zaffran, Olivier Feron, Yannig Goude, Julie Josse, and Aymeric Dieuleveut. Adaptive conformal predictions for time series. In *Proceedings of the 39th International Conference on Machine Learning*, volume 162 of *Proceedings of Machine Learning Research*, pp. 25834–25866. PMLR, 2022.

Frederic Zheng and Alexandre Proutiere. Conformal predictions under Markovian data. In *Proceedings of the 41st International Conference on Machine Learning*, volume 235 of *Proceedings of Machine Learning Research*, pp. 61470–61497. PMLR, 2024.

Haoyi Zhou, Shanghang Zhang, Jieqi Peng, Shuai Zhang, Jianxin Li, Hui Xiong, and Wancai Zhang. Informer: Beyond efficient transformer for long sequence time-series forecasting. In *AAAI Conference on Artificial Intelligence*, volume 35, pp. 11106–11115, 2021.

APPENDIX

A A CONDITIONAL ENVELOPE THEOREM FOR OBSERVED FINITE-STATE PROCESSES

Let X_t be the *observed process state*, not an arbitrary predictor covariate: (X_t) is a stationary, time-homogeneous, first-order Markov chain on $[K]$ with transition matrix P. Conditional on the state path, scores are independent draws from a fixed kernel $Q(\mathrm{d}s \mid X_t)$ that is not refit during calibration or deployment. The transition-estimation path is independent of predictor fitting and the entire calibration/deployment path.

Exact geometry and lag convention. Fix integers $B, L, h \geq 1$. Set $a_1 = 1$, $a_{b+1} = a_b + L + h - 1$, and $\mathcal{B}_b = \{a_b, \ldots, a_b + L - 1\}$ for $b = 1, \ldots, B+1$. Thus exactly $h-1$ indices are omitted between adjacent blocks, and the first index of $\mathcal{B}_{b+1}$ is lag h from the last index of $\mathcal{B}_b$. Blocks $1{:}B$ calibrate; $\mathcal{B}_{B+1}$ is the represented future deployment block. For any real $\alpha_{\mathrm{rank}} \in (0,1)$, define the integer

$$\begin{aligned} k_B &:= \min\{k \in \mathbb{Z} : k \geq (B+1)(1-\alpha_{\mathrm{rank}})\} \\ &= \lceil (B+1)(1-\alpha_{\mathrm{rank}}) \rceil, \\ \alpha_B^{\mathrm{rank}} &:= \begin{cases} (B+1-k_B)/(B+1), & k_B \leq B, \\ 0, & k_B > B. \end{cases} \end{aligned}$$

Set $A_b = \max_{t \in \mathcal{B}_b} S_t$. When $k_B \leq B$, $\hat{q} = A_{(k_B)}$ is the one-indexed k_Bth smallest value among $A_1, \ldots, A_B$, counting ties with multiplicity and using no interpolation; when $k_B > B$, set $\hat{q} = +\infty$. Thus $\alpha_B^{\mathrm{rank}} \leq \alpha_{\mathrm{rank}}$ exactly. There are exactly B past-to-next-block coupling links, each charged directly at $\beta_S(h)$ (or conservatively at $\beta_Z(h)$ when only a full-state envelope is available); L enters the sample geometry but not this lag because absolute regularity separates sigma-fields, not individual points.

An anytime-valid visit-adaptive transition set. On an independent path $\widetilde{X}_0, \ldots, \widetilde{X}_n$, let $N_i = \sum_{t<n} \mathbf{1}\{\widetilde{X}_t = i\}$ and $C_{ij} = \sum_{t<n} \mathbf{1}\{\widetilde{X}_t = i, \widetilde{X}_{t+1} = j\}$. For candidate $p \in [0,1]$, define the Jeffreys-mixture e-value

$$M_{ij}(p) = \frac{\mathrm{B}(C_{ij} + 1/2, N_i - C_{ij} + 1/2)}{\mathrm{B}(1/2, 1/2) p^{C_{ij}} (1-p)^{N_i - C_{ij}}}.$$

For orientation and normalization, let $\tau_{i,r}$ be the rth visit to state i, $Y_{i,r} = \mathbf{1}\{X_{\tau_{i,r}+1} = j\}$, and $\mathcal{G}_{i,r} = \sigma(X_0, \ldots, X_{\tau_{i,r}+1})$. For fixed alternative q,

$$L_N(q,p) = \prod_{r \leq N} (q/p)^{Y_{i,r}} \{(1-q)/(1-p)\}^{1-Y_{i,r}}$$

is a null-p $\mathcal{G}_{i,r}$-martingale for $0 < p < 1$, because each new factor has conditional mean $p(q/p) + (1-p)(1-q)/(1-p) = 1$. At $p = 0$ or $p = 1$, the null-supported factor has conditional mean $1-q$ or q, respectively, so the endpoint process is a nonnegative supermartingale. Integrating $L_N(q,p)$ against $q \sim \mathrm{Beta}(1/2, 1/2)$ gives exactly $M_{ij}(p)$ by the beta integral. There is no binomial coefficient: the likelihood is for the ordered exit sequence. Thus $M_{ij}(P_{ij})$ is the correctly oriented nonnegative $\mathcal{G}_r$ e-process. We use the continuous conventions at $p = 0, 1$: an impossible observed exit gives $M = \infty$, while $0^0 = 1$ on the null-supported path.

Set $M_{ij,0} = 1$ and $C_{ij,0} = [0,1]$. At visit count $r \geq 1$, take the closed non-strict sublevel set $C_{ij,r} = \{p \in [0,1] : M_{ij,r}(p) \leq K^2/\delta_{\mathrm{env}}\}$, and $[0,1]$ before any visit. Ville gives

$$\Pr_{P_{ij}}\{\exists r \geq 0 : M_{ij,r}(P_{ij}) \geq K^2/\delta_{\mathrm{env}}\} \leq \delta_{\mathrm{env}}/K^2.$$

If the true p is excluded from $C_{ij,r}$, then $M_{ij,r}(p)$ is *strictly* above the threshold, hence inside this Ville-controlled event; equality is retained. Therefore coverage holds simultaneously for every visit index. In particular, for every almost-surely finite $(\mathcal{G}_{i,r})$-stopping time T, $\Pr\{P_{ij} \in C_{ij,T}\} \geq$

$1 - \delta_{\text{env}}/K^2$. A union bound gives all entries. Intersect each row box with the simplex; an empty row set forces refusal. For $A \subseteq [K]$, define

$$U_i(A) = \min\left\{\sum_{j\in A} u_{ij}, 1 - \sum_{j\notin A} \ell_{ij}\right\}, \quad L_i(A) = \max\left\{\sum_{j\in A} \ell_{ij}, 1 - \sum_{j\notin A} u_{ij}\right\},$$

and $d_U = \max_{i\neq i',A}\{U_i(A) - L_{i'}(A)\}$ for $K \geq 2$, with $d_U = 0$ for $K = 1$. Distinct rows may be optimized independently in this product of row box–simplex sets; a row must not be compared with a second, independently chosen version of itself. Additional cross-row constraints require joint optimization. The implementation takes the full state count from the declared model before observing the path, retaining every unvisited row rather than inferring the state space from observed labels.

Theorem A.1 (Unconditional finite-state block coverage)**.** *Under the observed-state, stationarity, fixed-emission, independent-evidence, and geometry conditions above, the fail-closed rule releases $\hat{q}$ only if all row sets are nonempty, $d_U < 1$, and*

$$\alpha_B^{\text{rank}} + \delta_{\text{env}} + Bd_U^h \leq \tau;$$

otherwise it returns $+\infty$. Unconditionally over transition estimation and the calibration/deployment path,

$$\Pr\{\exists t \in \mathcal{B}_{B+1} : S_t > \hat{q}_{\text{released}}\} \leq \tau.$$

Consequently every predesignated point in $\mathcal{B}_{B+1}$, and all L points simultaneously, have coverage at least $1 - \tau$. The statement gives no guarantee for a point outside $\mathcal{B}_{B+1}$.

Proof. Let $E = \{P \text{ lies in every row confidence set}\}$, so $\Pr(E^c) \leq \delta_{\text{env}}$. Independence of the transition path and the complete conformal path is used here: conditioning on its realized confidence set does not change the stationary law of (Z_t) used for calibration and deployment. On E, the variational identity for total variation gives $d(P) \leq d_U$; stationarity, Markov contraction, and emission data processing give $\beta_S(k) \leq \beta_Z(k) \leq d(P)^k \leq d_U^k$. Sequential Berbee coupling of the score blocks $\mathcal{B}_1, \ldots, \mathcal{B}_{B+1}$ to independent copies has failure probability at most $\sum_{b=1}^{B} \beta_S(h) \leq Bd_U^h$. The independent-copy maxima are exchangeable unconditionally; the original miss probability is at most their strict rank-miss probability plus the coupling-failure probability. For the stated order statistic (including ties via the strict event), $\Pr\{A_{B+1} > A_{(k_B)}\} \leq (B+1-k_B)/(B+1) \leq \alpha_{\text{rank}}$; no DKW term appears. Finally,

$$\Pr\{\text{release and block miss}\} \leq \Pr(E^c) + \mathbb{E}[\mathbf{1}_E\{\alpha_B^{\text{rank}} + Bd_U^h\}\mathbf{1}_{\text{release}}] \leq \tau.$$

Refusal has infinite radius and zero miss probability. Since each future score is bounded by A_{B+1}, the simultaneous block event implies every within-block point event. This proves the claim entirely unconditionally. □

Theorem A.2 (Historical-prefix evidence on the same trajectory)**.** *Retain the observed-state, stationarity, fixed-emission, and block-geometry conditions of Theorem A.1, but estimate transitions from a historical prefix of the same trajectory. Let the first calibration-block index be lag h_0 from the prefix's last state. Predeclare $d_{\text{cap}} < 1$, independently of all observations, and release only if the transition row sets are nonempty, $d_U \leq d_{\text{cap}}$, and*

$$\alpha_B^{\text{rank}} + \delta_{\text{env}} + d_{\text{cap}}^{h_0} + Bd_U^h \leq \tau.$$

Then the unconditional simultaneous miss probability in the represented future block is at most τ.

Proof. Let E be simultaneous transition-set coverage, so $\Pr(E^c) \leq \delta_{\text{env}}$. Couple the complete calibration/deployment suffix to a stationary copy independent of the transition-estimation prefix. The mismatch probability is at most $\beta_Z(h_0)$. On E and release, data processing and Markov contraction give $\beta_Z(h_0) \leq d(P)^{h_0} \leq d_{\text{cap}}^{h_0}$. Conditional on the transition prefix, the copied suffix is stationary and independent of the random confidence set; its B internal score-block links and exact rank therefore cost at most $Bd_U^h + \alpha_B^{\text{rank}}$. Union with E^c and the prefix-to-suffix mismatch proves the displayed bound. Refusal again returns an infinite threshold. □

The cap is essential: without a predeclared deterministic d_{cap}, selecting a small random d_U from the same prefix would not by itself control the prefix-to-suffix coupling event. The theorem removes the independent-path requirement but not observed state, stationarity, Markov order, fixed emission, or the need for historical transition labels.

Table 3: Premise-matched two-state simulation of theorem accounting. Intervals are seed-cluster intervals and the bound is analytic; the result checks coverage when state, emissions, and transition evidence satisfy the declared contract, not applicability to the named datasets.

Evidence	Seeds / trials	Releases	Coverage (cluster-t 95%)	Rigorous lower bound
Entire future block ($L = 4$)	20 / 200,000	20/20	89.9805% [89.8532, 90.1078]	88.9996%
Designated first point	20 / 200,000	20/20	97.3640% [97.2891, 97.4389]	88.9996%
Transition confidence event	20 independent paths	–	20/20 contained P	$\geq 99\%$ per path

Table 4: Premise-matched observed-mode equipment simulation (20 seed clusters, 60,000 trials).

Method	Block cov. (cluster-t 95%)	Point cov.	Width	Refuse	Extra cost
Certified block	89.843% [89.624,90.063]	97.263%	3.129	0/20	1,485 skip + 40k transitions
Oracle-envelope block	89.843% [89.624,90.063]	97.263%	3.129	0/20	1,485 skip + known P
Pointwise Split CP	66.677% [66.234,67.120]	89.980%	2.068	0/20	none

Theorem-to-practice bridge. The retained record contains state definition, stationarity scope, Q, evidence source, n, δ_{env}, $[\ell, u]$, d_U, any predeclared (d_{cap}, h_0), B, L, h, and k_B. Any missing object is `NOT ESTABLISHED`. An optional process-population extension is stated but not proved or exercised here: it would require separately certified process units to be exchangeable, and even then would deliver population-marginal rather than process-conditional coverage. We report no such extension and no evidence that the exchangeability premise holds anywhere in this paper.

The construction uses $P = \left(\begin{smallmatrix} .65 & .35 \\ .35 & .65 \end{smallmatrix}\right)$, $S_t \mid X_t = x \sim \text{Unif}[0,1] + x/4$, $n = 40{,}000$, $B = 99$, $h = 16$, $\alpha_{\text{rank}} = 0.10$, $\delta_{\text{env}} = 0.01$, and $\tau = 0.12$. Mean d_U was 0.331007; the largest bound was 0.110003 miscoverage. The upper cluster-t 95% Monte Carlo limit for block miscoverage (10.1468%) remained below 11.0003%. Trials within a seed share its transition-estimation path; only the 20 seeds are treated as independent clusters. Pooled Wilson intervals are retained in the JSON as conditional descriptive summaries and are not the reported uncertainty.

Theorem A.2 was tested separately with the same P, folded-normal emissions, $n = 40{,}000$, $d_{\text{cap}} = 0.4$, and $h_0 = h = 16$. It released 20/20 seed clusters; across 60,000 continuations, block coverage was 90.0800% (cluster-t 95% [89.8752,90.2848]), mean $d_U = 0.331069$, the added prefix-link charge was $0.4^{16} = 4.295 \times 10^{-7}$, and the maximum total bound was 0.110003. This validates the extra-link accounting where premises hold; neither controlled experiment establishes them for a named dataset.

The observed steady/disturbance states, stationary P, fixed folded-normal emissions, and separate transition log satisfy the premises by construction. This is stylized synthetic evidence, not named-plant validation. With the anytime visit-adaptive set, the transition-cost sweep released 0/20, 1/20, 14/20, 20/20, 20/20, and 20/20 certificates at $n = 100, 250, 500, 1000, 5000, 40000$. On the same deterministic paths, the superseded global-n Azuma set released 0/20, 0/20, 2/20, 14/20, 20/20, and 20/20: the smallest tested cost with universal release falls fivefold, from 5,000 to 1,000 transitions. This is a controlled synthetic frontier, not a universal sample-complexity claim. The narrower pointwise baseline misses the simultaneous target; certification still pays explicit width, 1,485 skipped observations, and transition evidence.

A.1 Block Decomposition and Adaptive Block-Length Selection

We use the exact geometry already defined: B calibration blocks and one future block, each of length L, with adjacent-block lag h and exactly $h-1$ skipped indices. The block statistic is $A_b = \max_{t \in \mathcal{B}_b} s_t$ in theorem mode; high quantiles and averages are diagnostic unless their translation event is separately bounded. For *fixed* B, the relation $h \geq \lceil \lambda^{-1} \log(BC_\beta/\epsilon) \rceil$ is valid, but it cannot define h while B is simultaneously defined from h.

Proposition A.3 (Non-circular selection over predeclared geometries)**.** *Fix $N, L, \alpha_{\text{rank}}, \epsilon$, a declared envelope (C_β, λ), and a finite lag set $\mathcal{H} \subset \mathbb{N}$ before observing scores. For each $h \in \mathcal{H}$, define*

$$B_h^{\text{span}} = \left\lfloor \frac{N - L}{L + h - 1} \right\rfloor, \quad B_h^{\text{mix}} = \left\lfloor \frac{\epsilon}{C_\beta e^{-\lambda h}} \right\rfloor, \quad B_h = \min\{B_h^{\text{span}}, B_h^{\text{mix}}\}.$$

Discard h if $B_h < 1$ or $\lceil (B_h+1)(1-\alpha_{\mathrm{rank}}) \rceil > B_h$. Among the remainder, select deterministically by maximum B_h, then minimum used span $L + B_h(L+h-1)$, then minimum h; refuse if none remains. The selected geometry satisfies the exact sample-span, rank, and complete-link coupling constraints, without a fixed-point definition.

Proof. Every retained candidate obeys $L + B_h(L+h-1) \le N$, $B_h C_\beta e^{-\lambda h} \le \epsilon$, and finite rank by construction. The deterministic optimization over a finite predeclared set preserves these properties. □

An executable oracle-envelope comparison at the same $N = 1{,}885$, $L = 4$ selected $(B, h) = (156, 9)$ instead of $(99, 16)$: 57.6% more calibration blocks and 237 fewer skipped observations, with coupling charge 0.003071 versus 4.26×10^{-7}. Coverage was 90.52% versus 89.86% and width 3.153 versus 3.130 over 20 seed clusters; this is a utilization tradeoff on a controlled stationary model, not dominance or named-data evidence.

Block-to-point boundary. Block maxima need no within-block translation. The efficient block-average path uses the shifted point threshold $\hat{q} + \Delta$ and adds the externally justified slack $\bar{\eta}_{\mathrm{bt}}(\Delta)$ from Assumption A.4.2; no fitted CDF, ACF, bootstrap, or bounded-score argument supplies it. Appendix A.4 gives the exact event and the withdrawn shortcut.

A.2 Finite-Sample Coverage Theorem

Theorem A.4 (Coverage-sufficient score-block boundary)**.** *For stationary score blocks with common marginal law, let*

$$\Delta^{\mathrm{score}}_{B,h} := \mathrm{TV}\left(\mathcal{L}(A_1, \ldots, A_{B+1}),\ \bigotimes_{b=1}^{B+1} \mathcal{L}(A_b) \right).$$

With the exact integer rank, left order statistic, and infinity convention above, block maxima satisfy

$$\Pr\{\exists t \in \mathcal{B}_{B+1} : s_t > \hat{q}\} \le \alpha^{\mathrm{rank}}_B + \Delta^{\mathrm{score}}_{B,h} \le \alpha^{\mathrm{rank}}_B + B\beta_S(h) \le \alpha_{\mathrm{rank}} + B\beta_S(h). \tag{3}$$

Under Assumption 2.1, the sharper bound is $\alpha^{\mathrm{rank}}_B + BC_\beta e^{-\lambda h}$ (and hence at most $\alpha_{\mathrm{rank}} + BC_\beta e^{-\lambda h}$). The guarantee is simultaneous only inside the represented future block. For another statistic and predesignated point, add its externally justified translation event as in Assumption A.4.2.

Interpretation and proof. Under the product of the common score-block marginals, the $B+1$ statistics are i.i.d.; the strict rank event costs at most $\alpha^{\mathrm{rank}}_B \le \alpha_{\mathrm{rank}}$. This remains valid under ties because the miss event is strict; ties are counted with multiplicity in the left order statistic. Total variation changes this event probability by at most $\Delta^{\mathrm{score}}_{B,h}$. Sequential Berbee coupling applied directly to score sigma-fields has exactly B links at lag h, hence $\Delta^{\mathrm{score}}_{B,h} \le B\beta_S(h)$. For maxima, every score in $\mathcal{B}_{B+1}$ is bounded by A_{B+1}. No augmented-process mixing is used; Appendix A.8 records the sigma-fields.

Nuisance stress test. With i.i.d. Uniform scores and an independent $K = 1000$, $a = 0.999$ refresh nuisance, $\beta_Z(16) = 0.9831$ while $\beta_S(16) = 0$; at $B = 99$ the full-process surrogate refuses (97.43) but Theorem A.4 releases at 0.10, and 200,000 trials give 89.9715% block coverage (cluster-t 95% [89.8359,90.1071]).

Monitoring, risk, and recalibration boundaries. Anytime extension needs an almost-sure predictable conditional error budget, which marginal mixing does not supply (Appendix A.3); the claim is represented-block coverage, not bounded-loss risk control (Proposition A.2.1). Parameter selection is in Corollary A.6.1; named-data diagnostics are omitted because they do not instantiate this theorem.

A.2 Risk Control and Recalibration Details

Risk control. The retained bounded-loss results are empirical only. A previous theorem mixed rank calibration with a uniform concentration argument and is withdrawn; the formal claim is the represented-block coverage event in Theorems A.1 and A.4.

Recalibration. Predetermined times $\{T_1, T_2, \ldots\}$ independent of observations preserve Theorem A.4 separately in each period. Data-dependent drift triggering has no such coverage guarantee; alarms generated by Theorem A.3.2 retain only their stated anytime-valid Type-I control.

Proposition A.2.1 (Validity under fixed schedule)**.** *If recalibration occurs at predetermined times $\{T_1, T_2, \ldots\}$ independent of observations, Theorem A.4 applies to each calibration period.*

A.3 Predictable-Budget Monitoring

Assumption A.3.1 (Predictable Monitoring Budget)**.** Fix $\alpha \in (0, 1)$. Let $\bar{I}_b$ be an $\mathcal{F}_b$-measurable monitoring loss in $[0, 1]$, for example the fraction of observed misses in a completed block under its predeclared threshold. For the monitoring filtration $\mathcal{F}_{b-1}$, declared predictable values $\varepsilon_b \geq 0$ satisfy $\mathbb{E}[\bar{I}_b \mid \mathcal{F}_{b-1}] \leq \alpha + \varepsilon_b$ almost surely. Ordinary marginal β-mixing does not by itself establish this conditional inequality.

Theorem A.3.2 (Anytime Validity under the Predictable Budget)**.** *Let $E_t = \prod_{b \leq t}\{1 + \lambda_b(\bar{I}_b - \alpha)\}$ with predictable $\lambda_b \in [0, 1]$, and set $c_b = \log(1 + \lambda_b \varepsilon_b)$ and $\tilde{E}_t = E_t \exp(-\sum_{b \leq t} c_b)$. Under Assumption A.3.1, $\tilde{E}_t$ is a nonnegative supermartingale and $\mathbb{P}(\sup_t \tilde{E}_t \geq 1/\delta) \leq \delta$.*

Proof. Conditionally on $\mathcal{F}_{b-1}$, the expected wealth factor is at most $1 + \lambda_b \varepsilon_b = e^{c_b}$. Multiplication by e^{-c_b} gives the supermartingale step; Ville's inequality gives the stated bound. □

Remark A.3.3 (Scope). Earlier text derived ε_b from an unconditional Berbee mismatch probability. That conversion to an almost-sure predictable bound is invalid and withdrawn. The retained V3 simulation remains diagnostic only.

A.4 Block-to-Point Translation Details

Theorem A.4 uses block maxima directly. For a block average, retain the translation slack below rather than claim a data-derived concentration bound.

Assumption A.4.1 (Lipschitz block-statistic CDF)**.** The CDF F_A of the chosen block statistic A_b is L_F-Lipschitz on $[0, M]$. A bounded marginal density suffices; bounded scores alone do not.

Assumption A.4.2 (Declared Block-to-Point Envelope)**.** For $A_b = \bar{s}_b$ and chosen $\Delta > 0$, an externally justified value $\bar{\eta}_{\mathrm{bt}}(\Delta)$ satisfies

$$\mathbb{P}\{s_{T+1} > A_{B+1} + \Delta,\ A_{B+1} \leq \hat{q}\} \leq \bar{\eta}_{\mathrm{bt}}(\Delta).$$

Without such evidence, retain the slack symbolically or use block maxima.

Remark A.4.3 (Why the envelope is declared). A previous draft treated within-block observations as i.i.d. after one Berbee term. Ordinary β-mixing at lag L does not justify that step, so the claimed closed form is withdrawn.

A.5 Complete-Law versus Full-Process Boundary

A.6 Proof and Exact Enumeration for the n-versus-h Boundary

Proof of Theorem 2.8. Fix $h \geq n$. Under Q_x with phase φ uniform on $\mathbb{Z}_h$, the window is $(x_{\varphi+1}, \ldots, x_{\varphi+n})$ with indices modulo h; since $n \leq h$ these are n distinct coordinates, so averaging over a uniform x makes the window exactly i.i.d. uniform, which is claim 1. Periodicity gives $S_h = S_0$, so the lag-h joint law is the diagonal with marginal $\mathrm{Bernoulli}(p_x)$ and joint-to-product TV $2p_x(1 - p_x)$; averaging the binomial law of hp_x gives $\frac{1}{2} - \frac{1}{2h}$, which is claim 2. For the consequence, honesty gives $U_n \geq d_h(Q_x)$ with probability at least $1 - \delta$ under each Q_x, hence $\mathbb{E}\, U_n \geq \mathbb{E}_x[d_h(Q_x)] - \delta/2$ under the mixture because $d_h \leq 1/2$; the observed law of the mixture is i.i.d. $\mathrm{Bernoulli}(1/2)$ by claim 1, so the same holds under P_0. □

Corollary A.6.1 (Parameter selection)**.** *For a coupling allowance ϵ, choose any B, L with $k_B \leq B$ and $h \geq \lceil \lambda^{-1} \log(BC_\beta/\epsilon) \rceil$. Then simultaneous future-block miscoverage is at most $\alpha_B^{\mathrm{rank}} + \epsilon \leq \alpha_{\mathrm{rank}} + \epsilon$; larger B changes rank granularity, not validity.*

Proof of Proposition A.6.5. By stationarity the lag-h score pair is distributed as (S_0, S_h). For $h \leq n-1$ both indices lie in $\{0, \ldots, n-1\}$, so that law is the $(0, h)$ coordinate-pair marginal of the n-window law, and d_h is a functional of the pair law alone. □

Proposition A.6.2 (Computational nonattainment at the converse threshold for $3 \leq n \leq 5$)**.** *Let P be stationary ergodic binary with n-window law* $\mathrm{Bernoulli}(1/2)^{\otimes n}$ *and $d_h(P) = 1/2$. Uniform marginals put the lag-h joint on two atoms, so either $S_{t+h} = S_t$ or $S_{t+h} = 1 - S_t$ almost surely. In the second case P is $2h$-periodic, and the window law forces $2^n \mid 2h$ with every n-word equally frequent; for $2^{n-1} \leq h < 2^n$ this leaves only $h = 2^{n-1}$, where P must be an antiperiodic de Bruijn cycle of order n. The exact search enumerates each binary prefix $x \in \{0,1\}^{2^{n-1}}$, forms the cycle $x\bar{x}$, and accepts iff its 2^n cyclic n-windows are all distinct. It checks respectively* $4, 16, 256, 65{,}536$ *candidates for $n = 2, 3, 4, 5$, finding a cycle only at $n = 2$. Hence for $n \in \{3, 4, 5\}$ the smallest lag attaining the maximum is 2^n, not the converse threshold 2^{n-1}. We state this for the pairwise quantity only; whether attaining $\Delta^{\mathrm{score}}_{B,h} = 1 - 2^{-B}$ forces the same two-atom block support for general B is not proved here. The search is* `simulation/attainment_threshold.py`.

Proposition A.6.3 (The arbitrary-q closed form fails)**.** *For $m = 6$ and common marginal $q = 1/3$, let P put mass $1/3$ on each of* $110000, 001100, 000011$. *Then every coordinate is* $\mathrm{Bernoulli}(1/3)$*, but, for $Q = \mathrm{Bernoulli}(1/3)^{\otimes 6}$,*

$$\mathrm{TV}(P, Q) = \frac{227}{243} > \frac{664}{729} = 1 - q^6 - (1-q)^6,$$

with exact gap $17/729$.

Proof. Each coordinate appears in exactly one support atom, proving the marginal. Every support atom has Q-mass $(1/3)^2(2/3)^4 = 16/729 < 1/3$. Hence the overlap is $3(16/729) = 16/243$ and the TV is $1 - 16/243 = 227/243 = 681/729$. The displayed conjectured value is $1 - (1+64)/729 = 664/729$. The arithmetic is reproduced by `simulation/extremal_tv_range.py`. □

Proof of Theorem 2.6. Each cycle is primitive: every positive amplitude occurs exactly once, so a proper repeated subword cannot generate it. A uniform random phase into either finite orbit is therefore stationary and ergodic. Consecutive positive amplitudes are exactly nine positions apart. Hence every length-nine window contains exactly one amplitude, and each of the 18 amplitudes occurs once in each of the nine window positions. The two complete calibration laws are thus identical, each uniform on the same 162 vectors.

A strict miss can occur only when the calibration window begins at an amplitude; its future score is the next amplitude in the cyclic ordering, and the miss occurs exactly at a cyclic ascent. The increasing ordering has 17 ascents and the decreasing ordering has only the wraparound ascent, giving $17/162$ and $1/162$. Identical observations and $|z - 17/162| + |z - 1/162| \geq 16/162$ give risk at least $4/81$; identical hypotheses give testing error $1/2$. Honesty under $P_\uparrow$ forces $U \geq 17/162$ with probability $1 - \delta$ on the common observation law, which is the stated refusal event under $P_\downarrow$. Exact enumeration is in `simulation/operational_rank_miss_witness.py`.

Proof of Theorem 2.7. Every amplitude occurs once per cycle, so its period is exactly qB and uniform phase gives stationarity and ergodicity. Every length-B window contains one amplitude, $k-1$ zeros and s high sentinels. For each offset within a block, every amplitude occurs once; the full ordered window law is therefore independent of the ordering of the blocks. The k-th order statistic is precisely the window's amplitude. A future zero never exceeds it, whereas a future sentinel always does, contributing s/B to the miss probability. A future amplitude occurs once per block and exceeds the window's amplitude exactly at a cyclic ascent. The two orders have $q-1$ and one ascents respectively, giving the displayed probabilities. Directly,

$$\mu_+ - \alpha^{\mathrm{rank}}_B = \frac{2s+1}{2B(B+1)} > 0, \qquad \alpha^{\mathrm{rank}}_B - \mu_- = \frac{2k-1}{2B(B+1)} > 0.$$

Their gap is $1/(B+1)$. Under the common observation law, the pointwise triangle inequality $|z - \mu_+| + |z - \mu_-| \geq 1/(B+1)$ gives the risk bound, including randomized estimators. Identical observations give minimax testing error $1/2$. Honesty on the upper-miss process transfers unchanged to the benign one, proving refusal. All arithmetic in the 328 enumerated cases is rational, and the code compares complete window-frequency tables, not only marginal distributions or empirical approximations.

Proof of Theorem 2.5. Window-law equality and stationary ergodicity are as in Theorem 2.4. Under P_1 the period is 2^n, so $A_b = S_{a_b} = S_{a_1}$ almost surely for every b: the joint law of $(A_1, \dots, A_{B+1})$ puts mass $1/2$ on each of the two constant strings, while its product of marginals is uniform on $\{0,1\}^{B+1}$. Hence $\Delta = 1 - 2 \cdot 2^{-(B+1)} = 1 - 2^{-B}$. Under P_0 the $B+1$ blocks are disjoint functions of independent coordinates, so the joint equals the product exactly. For the matching upper bound set $m = B+1$, $N = 2^m$, and let p_x be any joint law with uniform one-coordinate marginals. Its product law is uniform, with mass $1/N$ per atom. For every x, $p_x \le P(A_1 = x_1) = 1/2$, and therefore

$$\min\{p_x, 1/N\} \ge (2/N)p_x.$$

Summing gives overlap at least $2/N = 2^{-B}$ and hence $\Delta \le 1 - 2^{-B}$. The constant estimator $(1-2^{-B})/2$ has worst-case absolute error at most $(1-2^{-B})/2$, while the indistinguishable endpoint pair gives the reverse inequality. The same pair gives testing error $1/2$ and the honest-bound consequence. Proposition A.6.3 shows why this proof is stated for the known uniform marginal rather than via the false arbitrary-q formula.

Proposition A.6.6 (restated). Let $n \le h < 2^{n-1}$. There exist stationary ergodic binary processes with the same n-window law and unequal d_h; at $(n,h) = (5,8)$ the gap is exactly $3/49$.

Proof of Proposition A.6.6. **General construction for every admissible intermediate lag.** For any integers $n \le h < 2^{n-1}$, necessarily $h \ge 3$. Put $m = h+2$. Let P_0 be uniform phase on a period-m binary word with one 1, and P_1 uniform phase on a period-$2m$ word with 1s at positions $0, h$. Both are stationary ergodic factors of finite cyclic shifts. The circular spacings in P_1 are $h, h+4$, both at least n; therefore neither process has two 1s in an n-window. Each of the n singleton-1 words has probability $1/m$, and the all-zero word has probability $1 - n/m$, under either process. Their common marginal success probability is $1/m$. At lag h, the $(1,1)$ joint masses are zero and $1/(2m)$, respectively. Thus

$$d_h(P_0) = \frac{2}{m^2}, \qquad d_h(P_1) = \frac{m-2}{m^2}, \qquad d_h(P_1) - d_h(P_0) = \frac{h-2}{(h+2)^2} > 0.$$

This proves the general existence statement. The common window law is sparse, not i.i.d.; it does not bypass the i.i.d.-window converse. `simulation/intermediate_general.py` checks 60 admissible (n,h) pairs in exact rational arithmetic with zero window-law or gap mismatches. The following older examples give different, sometimes larger, gaps.

Take P_0 to be a uniform random phase into the period-7 block $(0,0,0,0,0,0,1)$ and P_1 a uniform random phase into the period-14 block $(0,0,0,0,0,0,0,1,0,0,0,0,0,1)$. Both are stationary ergodic binary processes. At $n = 5$ each exhibits exactly the six windows $00000, 00001, 00010, 00100, 01000, 10000$, the first with probability $2/7$ and the rest with probability $1/7$, so their 5-window laws are identical; in particular their observed law total variation is 0. At $h = 8$ the lag-8 joint law of P_0 is a cyclic shift by one because $8 \equiv 1 \pmod 7$, so $S_{t+8} = S_{t+1}$, giving $\{(0,0),(0,1),(1,0)\}$ with probabilities $5/7, 1/7, 1/7$, while P_1 is 14-periodic, so the joint-to-product TV is $2/49$ for P_0 and $5/49$ for P_1, both exactly computable in rational arithmetic (`simulation/intermediate_witness.py`). The gap is $3/49$, and $5 \le 8 < 2^4$, so this sits in the intermediate regime. By Le Cam's two-point argument, the observed laws are indistinguishable (total variation 0) while d_h differs by $3/49$, so any estimator incurs absolute error at least $3/98$ on one of the two, randomized testing has error $1/2$ from a single window, and any $[0,1/2]$-valued $(1-\delta)$-honest upper bound has expectation at least $(1-\delta)\,5/49$ on P_1. □

The gap need not be small. At $(n,h) = (5,6)$ a uniform phase into the period-12 block $(0,0,1,0,1,0,1,0,1,1,0,1)$ has lag-6 joint law with total diagonal mass $1/3$ and $d_h = 1/6$, while a uniform phase into $(0,0,1,0,1,0,1,1,0,1,0,1)$ is 6-antiperiodic, so $S_{t+6} = 1 - S_t$ almost surely and its lag-6 joint law is $\{(0,1),(1,0)\}$ with equal mass, giving the maximal $d_h = 1/2$. Both exhibit the same ten 5-windows with the same probabilities, so their 5-window laws are identical (total variation 0) and the hidden separation is $1/2 - 1/6 = 1/3$, also exactly computable in rational arithmetic (`simulation/intermediate_witness.py`). The intermediate regime therefore contains a small-gap witness at $(5,8)$ and a larger one at $(5,6)$, whose separation is $1/2 - 1/6 = 1/3$. We exhibit no indistinguishable pair whose discrepancy separation attains $1/2$ in this regime; Proposition A.6.2 shows in particular why a process with the i.i.d. n-window law cannot attain $d_h = 1/2$ at $h = 2^{n-1}$ for $3 \le n \le 5$.

Proposition A.6.2 (Appendix A.6) shows the threshold is necessary but not attained for $3 \le n \le 5$, where the smallest attaining lag is 2^n.

Theorem A.6.4 (Individual-process converse: the mixture is necessary)**.** *Let P be a stationary ergodic binary process whose n-window law is exactly $\mathrm{Bernoulli}(1/2)^{\otimes n}$ and whose lag-h score pair attains the maximal value $d_h(P) = 1/2$. Then $h \ge 2^{n-1}$; if in addition $S_h = S_0$ almost surely, then $h \ge 2^n$.*

Proof. With uniform marginals, $d_h(P) = 1/2$ forces the lag-h joint law onto two cells, so either $S_h = S_0$ a.s. or $S_h = 1 - S_0$ a.s. In the first case stationarity gives $S_{t+h} = S_t$ for all t, so P is h-periodic; in the second, $S_{t+2h} = S_t$, so P is $2h$-periodic. A periodic stationary ergodic process is supported on one cyclic orbit, and a period-p orbit exhibits at most p distinct n-windows. Matching $\mathrm{Bernoulli}(1/2)^{\otimes n}$ requires all 2^n words to carry positive equal mass, hence $p \ge 2^n$. This gives $h \ge 2^n$ in the first case and $2h \ge 2^n$ in the second. □

Proposition A.6.5 (Short-lag identification: the boundary is sharp; proof in Appendix A.6)**.** *Let P, P' be stationary binary processes with the same n-window law. Then $d_h(P) = d_h(P')$ for every $1 \le h \le n - 1$.*

So the hypothesis $h \ge n$ in Theorem 2.8 is sharp: the largest d_h gap below it is exactly 0.

Proposition A.6.6 (Individual separation in the intermediate regime; statement and proof in Appendix A.6)**.** *Let $n \le h < 2^{n-1}$. There exist stationary ergodic binary processes with the same n-window law and unequal d_h; at $(n, h) = (5, 8)$ the gap is exactly $3/49$.*

The witness and the propagated constants are given in Appendix A.6. The regime $n \le h < 2^{n-1}$ is therefore not a converse like $h < n$: it contains individual pairs whose identical observed windows hide a nonzero lag dependence gap.

Proof of the decision bounds. Fix t and let $S_t = \{x : d_h(Q_x) \ge t\}$ with mass π_t under the uniform block law. Since the full mixture's observed n-window law is exactly $\mathcal{L}_{\mathrm{iid}}$, the split $\mathcal{L}_{\mathrm{iid}} = \pi_t M_{S_t} + (1-\pi_t) M_{S_t^c}$ gives $M_{S_t} - \mathcal{L}_{\mathrm{iid}} = (1-\pi_t)(M_{S_t} - M_{S_t^c})$, so $\mathrm{TV}(M_{S_t}, \mathcal{L}_{\mathrm{iid}}) \le 1 - \pi_t = \kappa_{h,t}$ because total variation is at most one. Every member of S_t is a stationary ergodic process with $d_h \ge t$ and every draw from $\mathcal{L}_{\mathrm{iid}}$ comes from a process with $d_h = 0$, so testing $\{d_h = 0\}$ against $\{d_h \ge t\}$ is at least as hard as testing $\mathcal{L}_{\mathrm{iid}}$ against M_{S_t}; Le Cam's two-point bound gives the stated testing and estimation displays, and the confidence statement follows because an honest U_n must exceed t with probability $1 - \delta$ under M_{S_t}. Because $\kappa_{h,t}$ is computed from the exact binomial tail, all three constants are rational and reproduced by `decision_lower_bounds` in the same module.

Converse witnesses. The analytic converse above, rather than an unreported enumeration, establishes that no single orbit attains the required uniform n-window law below the exponential lag. The accompanying module reproduces finite checks for $n \le 4$ but those checks support no additional claim here. The mixture in Theorem 2.8 is therefore not a weakening chosen for convenience: in the regime $n \le h < 2^{n-1}$ no individual process can attain *maximal* lag-h dependence while matching an i.i.d. n-window, so a two-point pair of the kind used in Theorem 2.4 is unavailable there. An individual pair with strictly smaller separation does exist in that regime: Proposition A.6.6 exhibits one at $(n, h) = (5, 8)$ with gap $3/49$.

Exact short-lag witnesses. `simulation/src/short_lag_impossibility.py` enumerates every block for $h = 2, \ldots, 9$ and every window $n \le h$ in exact rational arithmetic: each orbit is a single cycle, so every Q_x is ergodic; the mixture window law equals 2^{-n} on every word with maximum error exactly 0; and the mean lag-h dependence equals the closed form $\frac{1}{2} - \frac{1}{2h}$. The tail is also concentrated: a uniform block has $d_h \ge 2/5$ with probability $91/128$ at $h = 8$ and $1749/2048$ at $h = 12$, so the obstruction is typical rather than carried by a rare block.

Proposition A.6.7 (No Nontrivial Full-Process Envelope from Scores)**.** *Let (S_t) be i.i.d. Fix any lag $k \ge 1$ and any $c < 1$. There exist two strictly stationary, geometrically absolutely regular processes $Z_t^{(0)} = (S_t, H_t^{(0)})$ and $Z_t^{(K,a)} = (S_t, H_t^{(K,a)})$ with exactly the same complete score law such that*

$$\beta_{Z^{(0)}}(k) = 0, \qquad \beta_{Z^{(K,a)}}(k) = a^k(1 - 1/K) > c.$$

Thus the constant one is the only distribution-free full-process upper envelope identified from scores alone.

Proof. Take $H_t^{(0)}$ i.i.d. For the second process, use the stationary Markov chain on $[K]$ with uniform law π and transition $P = aI + (1-a)\mathbf{1}\pi^\top$. Then $P^k = a^k I + (1-a^k)\mathbf{1}\pi^\top$ and

$$\beta_H(k) = \sum_i \pi_i \, \mathrm{TV}(P^k(i,\cdot),\pi) = a^k\left(1 - \sum_i \pi_i^2\right) = a^k(1-1/K).$$

The independent score coordinate changes neither coefficient. Choose K large and then a near one. □

A.6.1 A GEOMETRICALLY MIXING PERTURBATION OF THE ALL-RANK WITNESS

Theorem 2.7 uses a periodic phase cycle, so a practitioner who supplies a mixing envelope is entitled to ask whether the obstruction is an artefact of non-mixing. It is not, up to an explicit threshold on the mixing rate.

Theorem A.6.8 (Mixing-robust operational strict-miss boundary)**.** *Fix $B \geq 2$ and $1 \leq k \leq B$, let $P_\pm$ be the pair of Theorem 2.7 with phase cycle length $N = 2(B+1)B$, and for $\varepsilon \in (0,1)$ let $P_\pm^\varepsilon$ advance the phase deterministically with probability $1-\varepsilon$ and redraw it uniformly with probability ε.* (a) *Each $P_\pm^\varepsilon$ is a stationary, irreducible, aperiodic finite-state chain with uniform phase law and Doeblin minorisation ε, so $\sup_x \|P^t(x,\cdot) - \pi\|_{\mathrm{TV}} \leq (1-\varepsilon)^t$: the perturbed processes are uniformly geometrically ergodic while the unperturbed cycle is not mixing.* (b) *The two length-B calibration laws remain* exactly *equal,* $\mathrm{TV}\big(\mathcal{L}_+^\varepsilon(O_B), \mathcal{L}_-^\varepsilon(O_B)\big) = 0$. (c) *The strict rank misses satisfy $|\mu(P_\pm^\varepsilon) - \mu_\pm| \leq 1-(1-\varepsilon)^B \leq B\varepsilon$, hence $\mu(P_+^\varepsilon) - \mu(P_-^\varepsilon) \geq \frac{1}{B+1} - 2B\varepsilon$, and the nominal level still separates the pair whenever $B\varepsilon < \min\{\mu_+ - \alpha_B^{\mathrm{rank}},\ \alpha_B^{\mathrm{rank}} - \mu_-\}$, which, with $s = B-k$, holds for $\varepsilon \leq \min\{2s+1,\ 2k-1\}/(4B^2(B+1))$ because $B \cdot \min\{2s+1, 2k-1\}/(4B^2(B+1)) = \min\{2s+1, 2k-1\}/(4B(B+1))$ is strictly below $\min\{(2s+1),(2k-1)\}/(2B(B+1))$. An earlier version of this statement kept only the first margin, $\varepsilon \leq (2s+1)/(4B^2(B+1))$; that one-sided form fails the sufficient second-margin check when $6k \leq 2B+3$ and is corrected here.* (d) *Consequently every pointwise $(1-\delta)$-honest upper bound U satisfies $U \geq \mu(P_+^\varepsilon)$ with probability at least $1-\delta$ on the common calibration law, and a rule releasing only when $U \leq \alpha_B^{\mathrm{rank}}$ refuses under the benign, uniformly geometrically ergodic process P_-^ε with probability at least $1-\delta$, with no total-variation loss. Geometric mixing alone therefore does not restore honest release for this slowly mixing family. The condition is sufficient for the obstruction, not a necessary threshold for every process or release procedure.*

Proof. *(a)* The transition kernel is $(1-\varepsilon)\,\mathbf{1}\{\varphi' = \varphi + 1\} + \varepsilon/N$, which is a Doeblin minorisation with constant ε and invariant uniform law, because the deterministic shift and the uniform restart both preserve uniformity. The standard Doeblin coupling gives the displayed geometric bound and, for $\varepsilon > 0$, irreducibility and aperiodicity. *(b)* Condition on the restart pattern. Between restarts the phase advances deterministically from a uniform draw, so the trace law is a mixture, with process-independent weights, of products of contiguous sub-window laws of the cycle under a uniform phase. Every contiguous sub-window law of length $\ell \leq B$ is a marginal of the length-B law, and those coincide for P_+ and P_- by Theorem 2.7; hence every mixture component coincides and the two perturbed laws are identical. *(c)* The miss event is measurable with respect to O_{B+1}, which is determined by the phase path over B transitions. On the event that no restart occurs in those transitions, of probability $(1-\varepsilon)^B$, the perturbed and unperturbed paths agree, so the two miss probabilities differ by at most the complementary mass $1-(1-\varepsilon)^B \leq B\varepsilon$. The separation and threshold statements follow from the gap identities $\mu_+ - \alpha_B^{\mathrm{rank}} = (2s+1)/(2B(B+1))$ and $\alpha_B^{\mathrm{rank}} - \mu_- = (2B-2s-1)/(2B(B+1))$ with $s = B-k$. *(d)* By *(b)* the honest-upper-bound event has the same probability under both perturbed processes, so the argument of Theorem 2.7 applies verbatim with μ_+ replaced by $\mu(P_+^\varepsilon)$ and no total-variation correction. □

Executed algebra. The entry point `simulation/mixing_robust_operational_witness.py` produces the accompanying artifact record in the project evidence bundle. Exact rational enumeration on 48 (B,k) cells with $B \leq 19$ reproduces $\mu_\pm$, the separation $1/(B+1)$ and both gap identities with zero failures, and confirms $\mu_- < \alpha_B^{\mathrm{rank}} < \mu_+$ throughout. At the same 48 cells the corrected two-margin condition is verified: B times the stated bound is strictly below the smaller gap at every cell, while the earlier one-sided form fails at 16 edge cells satisfying $6k \leq 2B+3$;

the perturbed enumeration at $\varepsilon \in \{1/100, 1/50, 1/20\}$ is unchanged. For the smallest members the perturbed laws are enumerated exactly at $\varepsilon \in \{1/100, 1/50, 1/20\}$: the total variation is exactly zero, both miss drifts respect $1 - (1 - \varepsilon)^B$, and the nominal level still separates the pair. Nine regressions cover the cycle shape, the closed forms, the zero-total-variation claim, the drift bounds and the enumeration budget.

Closest prior work for the mixing-robust statement. Barber et al. (2023) bound the coverage gap of nonexchangeable conformal prediction by a total variation between the observed data distribution and its shifted counterpart, and Xu & Xie (2021) bound conditional and marginal coverage gaps for time series under decay-of-dependence assumptions. Both are upper bounds use dependence or a within-process comparison of observation laws. Small relevant divergence strengthens, rather than invalidates, such a bound. Our zero TV instead compares the calibration laws of two *different* processes; it does not set those within-process dependence terms to zero. Theorem A.6.8 gives indistinguishable calibration observations with rank misses on opposite sides of the nominal level and separation at least $1/(B+1) - 2B\varepsilon$. The exact $1/(B+1)$ identity belongs to the unperturbed pair. We claim neither an improvement on the prior bounds nor impossibility of coverage; the obstruction concerns an honest release decision from this calibration observation.

A.7 Related Work

Conformal prediction foundations. Standard conformal prediction (Vovk et al., 2005; Angelopoulos & Bates, 2021) requires exchangeability; split conformal (Lei et al., 2018) and CQR (Romano et al., 2019) inherit this. Weighted exchangeability (Barber et al., 2023; Tibshirani et al., 2019) handles covariate shift; risk control (Bates et al., 2021; Angelopoulos et al., 2024b) extends to general losses. Our block-wise approach targets *temporal* dependence via mixing, orthogonal to weight-based methods.

Online and adaptive CP. ACI (Gibbs & Candes, 2021; Gibbs & Candès, 2024) and AgACI/DtACI-style variants (Zaffran et al., 2022) adapt quantiles online and are strong choices for long-run tracking under unknown shift. EnbPI (Xu & Xie, 2021) uses ensemble residuals, while SPCI (Xu & Xie, 2023) re-estimates residual quantiles sequentially. TEMPORAL-CP instead spends calibration data on separated blocks and reports an assumption-conditional budget only when the external envelopes are supplied.

Recent non-exchangeable and single-trajectory theory (and what is genuinely new here). The closest theoretical neighbors analyze conformal degradation under non-exchangeability (Barber & Pananjady, 2026; Oliveira et al., 2024; Barber et al., 2023; Chernozhukov et al., 2018), risk control (Farinhas et al., 2024), single-trajectory blocking (Lee & Matni, 2024), and finite-state Markov estimation (Zheng & Proutiere, 2024). Finite-state results rely on observed Markov structure and transition evidence that are absent from this paper's named-dataset evidence. TEMPORAL-CP therefore keeps the score or structural envelope external for those datasets and contributes the score-block sufficiency boundary plus a conditional finite-state instantiation, not a continuous-state plug-in analogue.

Localized, weighted, and kernel CP. Weighted CP under covariate shift (Tibshirani et al., 2019; Barber et al., 2023) and KOWCPI/kernel-localized time-series CP (Lee et al., 2025) pursue localized efficiency through weights or kernels. They are complementary to the separated-block assumption ledger. The surrogate experiment includes a lightweight temporal-kernel baseline as a stress test, not a full KOWCPI reproduction.

E-processes, safe testing, and dependence. Safe testing and e-processes (Ramdas et al., 2023; Grünwald et al., 2024; Waudby-Smith & Ramdas, 2024) provide the right language for optional stopping, and online risk-control work (Angelopoulos et al., 2024a) studies sequential prediction under broad online protocols. Our monitoring statement is deliberately conditional: Theorem A.3.2 corrects wealth given a predictable conditional budget ε_b. It does not derive that budget from ordinary marginal β-mixing or from Berbee's unconditional coupling probability.

Table 5: Related-work map: each neighboring line solves a different problem. TEMPORAL-CP's contribution is the auditable, self-refusing finite-sample layer, not a claim of universal dominance over these methods.

Neighboring line	What it gives	What TEMPORAL-CP adds / does not claim
Split CP under dependence (Barber & Pananjady, 2026; Oliveira et al., 2024)	Coverage degradation/bounds under dependence inputs	We identify why score laws cannot supply those process inputs; blocking itself is not claimed new
Single-trajectory CP (Lee & Matni, 2024)	Blocking/decoupling guarantees for one trajectory	Explicit evidence audit and refusal boundary; no data-derived process envelope
Markovian-data CP (Zheng & Proutiere, 2024)	CP bounds exploiting finite-state Markov structure	Independent transition confidence set $\rightarrow$ Dobrushin gate, paired with a score-only impossibility result
ACI, DtACI, SPCI, EnbPI (Gibbs & Candes, 2021; Gibbs & Candès, 2024; Zaffran et al., 2022; Xu & Xie, 2021; 2023)	Online/asymptotic tracking under shift	Different target: use these when theorem premises cannot be established
Weighted/kernel CP (Tibshirani et al., 2019; Lee et al., 2025)	Localized efficiency under weighting assumptions	Orthogonal to our marginal finite-state dependence audit
E-process/safe testing (Ramdas et al., 2023; Grünwald et al., 2024; Waudby-Smith & Ramdas, 2024)	Optional-stopping-valid testing framework	Conditional wealth correction given, not derived from, a predictable budget
Mixing estimation (McDonald et al., 2015; Khaleghi & Lugosi, 2023; Grunewalder & Khaleghi, 2026)	Asymptotic consistency or structured finite-sample rates	Exact de Bruijn equivalence rules out an assumption-free finite-sample score certificate

Mixing estimation and the finite-class boundary. Single-path estimators give asymptotic consistency (McDonald et al., 2015; Khaleghi & Lugosi, 2023), while finite-sample high-probability control is available under geometrically ergodic Markov and smoothness/finite-state assumptions (Grunewalder & Khaleghi, 2026). Theorem 2.4 instead fixes one target, d_{2^n}, and proves an exact finite-n threshold: an honest UCB cannot be below $1/2$ more often than δ. Adams–Nobel address ergodic density estimation and Ryabko process discrimination (Adams & Nobel, 1998; Ryabko, 2010); we do not claim priority for those broader impossibility questions. Definition 2.3 is descriptive.

Portfolio and deployment boundary. Conformal calibration also appears in control and admission systems, but this paper is not a closed-loop control or admission-decision paper. Here the contribution is the assumption-conditional marginal UQ layer and the separation of coverage-sufficient score dependence from stronger full-process assumptions.

Block bootstrap and structural methods. Block resampling methods (Künsch, 1989; Politis & Romano, 1994; Lahiri, 2003; Politis & White, 2004) target estimator distributions, not predictive coverage; their data-driven block-length results are complementary motivation for our adaptive selection rule but do not provide a CP coverage bound. TEMPORAL-CP uses Berbee coupling and finite-sample rank accounting for computable bounds under stated premises, and records the conditional results and boundaries above.

A.7.1 Public Benchmark Details

Honest reading. On all three ETT datasets TEMPORAL-CP is wider than Split CP (by 1.2–$1.6\times$) and SPCI (by 1.2–$2.0\times$); it is narrower than ACI (by 1.9–$3.0\times$). TEMPORAL-CP delivers higher empirical coverage than Split/SPCI (94–99% vs. 89–95%) at this width premium. *We do not claim Pareto improvement over Split CP or SPCI on ETT.* We report only the frozen-lockfile replay values and make no cross-environment robustness or causal attribution claim.

Honest reading of the extension. The two additional public benchmarks split the verdict, and we report both. On Traffic the three split-type comparators undercover at the 90% target – Split CP

Table 6: Public ETT stress tests: coverage(%)/width. Numbers are produced by the released script `src/run_real_experiments.py`, which downloads ETTh1/ETTh2/ETTm1 from the Informer/ETDataset source (Zhou et al., 2021) and writes `benchmark_results_real_metrics.csv`; the values below are read directly from that CSV (single-seed deterministic LightGBM pipeline) under the exact-pin `requirements.lock` environment. These results are empirical comparability evidence, not theorem certificates.

Data	Split CP	SPCI	ACI	**TEMPORAL-CP**
ETTh1	92.67/0.293	89.63/0.262	99.71/0.851	**98.13/0.449**
ETTh2	90.78/0.238	89.14/0.235	99.86/0.878	**94.43/0.292**
ETTm1	95.04/0.155	90.17/0.122	99.81/0.489	**98.78/0.242**

Table 7: Additional public benchmarks: coverage(%)/width at a 90% target. Produced by the same released script `src/run_real_experiments.py` with `-datasets electricity traffic` over 5 seeds, reading the real LSTNet-format archives (Electricity $26{,}304 \times 321$, Traffic $17{,}544 \times 862$); values are read directly from `benchmark_results_real_metrics_public_ext.csv`. EnbPI and Rolling CP are in that CSV and omitted here only for width. These are empirical comparability evidence, not theorem certificates.

Data	Split CP	SPCI	ACI	**TEMPORAL-CP**
Electricity	95.80/1.721	90.12/0.878	99.87/4.540	**98.00/2.547**
Traffic	85.45/0.300	89.58/0.589	99.71/7.188	**91.01/0.615**

85.45%, Rolling CP 85.84% and EnbPI 85.49% – and TEMPORAL-CP is the only method other than ACI to reach the target, at 91.01% with width 0.615 against ACI's 99.71% at width 7.188, roughly a twelvefold width premium for the same coverage direction. On Electricity the ordering reverses against us: TEMPORAL-CP overcovers at 98.00% with width 2.547, while SPCI sits closest to the target at 90.12% with width 0.878, so on that series a simpler comparator is preferable and our separated-block spend buys nothing. We therefore claim no dominance on public data. In these runs, separated-block calibration reaches the target on Traffic but overcovers with greater width than SPCI on Electricity; the experiment does not isolate dependence or the causal effect of separated blocks. Coverage here is empirical over the released pipeline's seeds, not validity in the sense of the theorems. Both series are read from the archived LSTNet files decompressed to `data/electricity.csv` and `data/traffic.csv`; the pipeline's remote download path returns HTTP 404 for these two sources. The historical automatic synthetic fallback was not used for these reported rows and has been removed: public-data loaders now raise an actionable error when the named data are unavailable. The archived copies are the reproducible route; the shapes $26{,}304 \times 321$ and $17{,}544 \times 862$ are printed by the loader as a check.

Self-refusal in certified mode (operational honesty). The legacy `--certified --self-certify` flags select block-max, $G = 240$, and a score diagnostic. They preserve the measured ETT coverage and width outputs, but the generated Boolean must be read only as a legacy plug-in diagnostic. The external process envelope remains unverified whether that Boolean is true or false.

A.8 Proof of Theorem A.4 and Auxiliary Remarks

Remark A.8.1 (No implicit test point). Theorem A.4 has no floating "next point." Deployment is exactly $\mathcal{B}_{B+1}$, whose first index is lag h from the end of $\mathcal{B}_B$. A point elsewhere requires a newly specified geometry and coupling charge.

Remark A.8.2 (Multivariate outputs). For vector-valued targets, apply TEMPORAL-CP per component with Bonferroni correction (α/d) for joint coverage, or use block-max aggregation over components for conservative multivariate coverage; tighter dependence-aware joint guarantees are an open direction.

We provide the complete proof with explicit constants.

A.8.1 KEY LEMMAS

Lemma A.8.3 (Indexed block coupling). *Write $Y_b = (S_t : t \in \mathcal{B}_b)$, let $e_b = \max \mathcal{B}_b$, and let $a_{b+1} = \min \mathcal{B}_{b+1}$. By construction $a_{b+1} = e_b + h$: there are $h - 1$ omitted indices, while the separation in the definition of $\beta_S(k)$ is the index difference k. Moreover,*

$$\sigma(Y_1, \ldots, Y_b) \subseteq \sigma(S_t : t \le e_b), \qquad \sigma(Y_{b+1}) \subseteq \sigma(S_t : t \ge a_{b+1}),$$

so monotonicity of absolute regularity gives $\beta(\sigma(Y_1, \ldots, Y_b), \sigma(Y_{b+1})) \le \beta_S(h)$. The standard sequential Berbee construction for block vectors (Yu, 1994, Lemma 4.1) therefore extends the probability space with mutually independent $Y'_1, \ldots, Y'_{B+1}$, each $Y'_b \overset{d}{=} Y_b$, such that

$$\Pr\{\exists b : Y_b \neq Y'_b\} \le \sum_{b=1}^{B} \beta_S(h) \le BC_\beta e^{-\lambda h}.$$

At induction step $b + 1$, the new copy is independent of the entire vector $(Y'_1, \ldots, Y'_b)$, not merely of Y'_b; the union bound then gives the display. The sum has B, not $B - 1$, terms because it includes the calibration-to-test link. There is no factor L: $\beta_S(h)$ already compares full score-past and score-future sigma-fields.

A.8.2 MAIN PROOF, LINE BY LINE

(1) Apply Lemma A.8.3 to all $B + 1$ blocks. (2) Measurability maps each coupled block to its maximum, preserving the coupling event. (3) Stationarity makes the independent-copy maxima identically distributed, hence exchangeable. (4) With $k_B = \lceil (B+1)(1-\alpha_{\mathrm{rank}}) \rceil \le B$, symmetry of ranks and the strict miss event give $\Pr\{A'_{B+1} > A'_{(k_B)}\} \le (B+1-k_B)/(B+1) = \alpha_B^{\mathrm{rank}} \le \alpha_{\mathrm{rank}}$, also with ties. If $k_B > B$, the prescribed threshold is infinite and $\alpha_B^{\mathrm{rank}} = 0$. (5) Union with coupling failure gives $\alpha_B^{\mathrm{rank}} + BC_\beta e^{-\lambda h}$. (6) For every $t \in \mathcal{B}_{B+1}$, $s_t \le A_{B+1}$ deterministically, so one block event yields simultaneous coverage of all represented points. No such inequality is used outside that block. (7) For a non-max statistic, the point threshold is $\hat{q} + \Delta$, not $\hat{q}$. Indeed $\{s_{T+1} > \hat{q} + \Delta\} \subseteq \{A_{B+1} > \hat{q}\} \cup \{s_{T+1} > A_{B+1} + \Delta,\ A_{B+1} \le \hat{q}\}$, so the declared translation event adds $\bar{\eta}_{\mathrm{bt}}(\Delta)$ at that shifted threshold. These steps prove Theorem A.4 without DKW or an unstated test-block coupling.

A.9 LOWER-BOUND EVIDENCE BOUNDARY

No empirical calibration-rate claim is made: the earlier $0.8/\sqrt{B}$ fit had no retained script or machine-readable output and is withdrawn. The exact local witness below is analytic and executable; it concerns threshold estimation only, not conformal validity or block-to-point translation.

A.10 SAMPLE-COMPLEXITY, LOWER-BOUND, AND IMPOSSIBILITY STATEMENTS

Corollary A.10.1 (Geometry sample requirement). *For chosen B, L and coupling allowance ϵ, take $h \ge \lceil \lambda^{-1} \log(BC_\beta/\epsilon) \rceil$. The complete calibration-plus-future-block span is exactly $N = (B+1)L + B(h-1)$. If $k_B \le B$, Theorem A.4 gives simultaneous represented-block miscoverage at most $\alpha_B^{\mathrm{rank}} + \epsilon \le \alpha_{\mathrm{rank}} + \epsilon$. No additional $B^{-1/2}$ validity term or universal $128K$ cutoff is required.*

Theorem A.10.2 (Explicit local quantile-estimation lower bound). *Fix $\alpha \in (0, 1)$ and $c_0 \in (0, 1]$. For $B \ge 1$, observe B i.i.d. block statistics from either $P_0 = \mathcal{N}(0, 1)$ or $P_1 = \mathcal{N}(c_0/\sqrt{B}, 1)$. These laws define stationary processes with $\beta(k) = 0$ for every $k \ge 1$. For every estimator $\hat{q}$ of the $(1 - \alpha)$ quantile,*

$$\max_{i \in \{0,1\}} \mathbb{E}_i |\hat{q} - q_i| \ge \frac{c_0}{8\sqrt{B}}, \qquad \max_{i \in \{0,1\}} \mathbb{E}_i (\hat{q} - q_i)^2 \ge \frac{c_0^2}{16B}.$$

Thus the local threshold-estimation scales $B^{-1/2}$ in absolute risk and B^{-1} in squared risk already in an i.i.d. subclass. This is not a coverage lower bound or a block-to-point statement. In particular, the two displayed denominators are $\sqrt{B}$ and B, respectively; no B^{-1} absolute-risk or $B^{-1/2}$ squared-risk claim is made.

Proof. Both CDFs are Gaussian location CDFs, so they are valid and their stationary marginals are exactly the displayed laws. Writing $z_{1-\alpha} = \Phi^{-1}(1-\alpha)$ gives $q_i = z_{1-\alpha} + \theta_i$ and hence the exact quantile gap $\Delta = |q_1 - q_0| = c_0/\sqrt{B}$. The one-observation divergence is $\mathrm{KL}(P_0 \| P_1) = \Delta^2/2$; product additivity gives $\mathrm{KL}(P_0^{\otimes B} \| P_1^{\otimes B}) = c_0^2/2$ exactly. Pinsker's inequality therefore yields $\mathrm{TV}(P_0^{\otimes B}, P_1^{\otimes B}) \le c_0/2 \le 1/2$.

For any $\hat{q}$, classify the nearer of q_0, q_1. On a classification error the absolute loss is at least $\Delta/2$ and the squared loss at least $\Delta^2/4$. The sum of the two testing error probabilities is at least $1 - \mathrm{TV}$. Taking the larger of the two risks gives

$$\max_i \mathbb{E}_i|\hat{q} - q_i| \ge \frac{\Delta}{4}(1 - \mathrm{TV}) \ge \frac{c_0}{8\sqrt{B}},$$

and

$$\max_i \mathbb{E}_i(\hat{q} - q_i)^2 \ge \frac{\Delta^2}{8}(1 - \mathrm{TV}) \ge \frac{c_0^2}{16B}.$$

No coupling, dependent-product KL approximation, or unspecified perturbation is used. The exact quantities are checked by `simulation/src/lower_bound_witness.py` and its regression test. □

Remark A.10.3 (Implications and scope). Theorem A.10.2 concerns numerical threshold error in one explicit regular family. Exact conformal rank validity in Theorem A.4 has no $B^{-1/2}$ penalty, and this result supplies no translation term, generic constant optimality, or lower bound for dependent-process estimation.

Theorem A.10.4 (Scoped impossibility of distribution-free exact conditional coverage)**.** *Without restrictions on the joint law of (X, Y), a distribution-free procedure cannot provide nontrivial finite intervals with $\Pr\{Y \in \mathcal{C}(X) \mid X = x\} \ge 1 - \alpha$ for almost every x uniformly over all laws (Vovk, 2012). This already holds in the i.i.d. subclass, for which $\beta(k) = 0$; therefore merely assuming β-mixing cannot restore exact object-conditional coverage. The statement makes no dimension-dependent rate claim. Group-, region-, or model-conditional guarantees remain possible under their corresponding restrictions.*

Corollary A.10.5 (Threshold accuracy is separate from validity)**.** *Additional density or tail assumptions may sharpen high-probability numerical error of the empirical block quantile. They do not alter the exact rank term in Theorem A.4, suppress any of the B coupling links, or convert block averages into point coverage.*

A.11 Compact Operational Boundaries

The executable calibration cost is score computation plus sorting B block statistics, $O(N + B \log B)$, with constant-time threshold lookup. The formal scope is already complete in the main-text ledger: score diagnostics do not instantiate a score-process upper envelope; general-loss experiments remain empirical; and the weighted, translation, and monitoring extensions require their separately stated premises.

Drift-triggered recalibration (heuristic). The local rule monitors $\hat{c}_t = W^{-1} \sum_{i=t-W+1}^{t} \mathbb{I}\{Y_i \in \mathcal{C}(X_i)\}$ and triggers when $|\hat{c}_t - (1-\alpha)| > \tau$. It has no coverage guarantee because the trigger depends on deployment outcomes; only pre-specified schedules fall under Proposition A.2.1, and e-process alarms are optional-stopping valid only under Assumption A.3.1. The retained proprietary-data cadence and former self-certified score-fit rows are diagnostics, not general guidance or certificates.

A.12 GEOMETRY SELECTION AFTER THE ASSUMPTION AUDIT

Algorithm 1 Fail-Closed Non-Circular Geometry Selection

Require: Passed G1–G3 record; envelope $g(h)$, δ_{env}; $N, L, \alpha_{\mathrm{rank}}, \tau$; predeclared finite $\mathcal{H}$
Ensure: A finite geometry (B, L, h) or `REFUSE`

```
if τ ≤ α_rank + δ_env then
    return REFUSE with an infinite threshold
end if
for each h ∈ H do
    B_h^span ← ⌊(N − L)/(L + h − 1)⌋
    B_h^budget ← ∞ if g(h) = 0, else ⌊(τ − α_rank − δ_env)/g(h)⌋
    B_h ← min{B_h^span, B_h^budget}
    discard if B_h < 1 or ⌈(B_h + 1)(1 − α_rank)⌉ > B_h
end for
Select maximum B_h, then minimum span, then minimum h
if no candidate remains then
    return REFUSE with an infinite threshold
end if
return (B_h, L, h) and the complete G1–G4 ledger
```

Score-ACF and bootstrap sweeps may be run separately to describe empirical sensitivity, but their outputs cannot overwrite d_U, alter the confidence budget after inspection, or turn a refused theorem-mode result into an accepted one.

Executed selected-gap audit. On the public FrozenLake chain, each certified box–simplex rectangle in 100 evidence-length cells (20 seeds $\times$ evidence lengths $\{1000, 5000, 20000, 80000, 320000\}$) feeds the non-circular selection rule of Algorithm 1 with frozen candidates $h \in \{1, \ldots, 128\}$, $\alpha = .10$, $\delta = .01$, $\tau = .12$, and conformal rank $k_B = \lceil (B+1)(1-\alpha) \rceil$ recomputed per selected B. Three envelope rules are compared on identical evidence: the multi-step submultiplicative envelope $g(h) = \min_{r \le \min(32,h)} D_r^{\lfloor h/r \rfloor}$, the lower-overlap one-step envelope D_1^h, and their minimum. This D_1 is not the exact robust box–simplex coefficient d_U. Since the multi-step minimum includes $r = 1$, its minimum with D_1^h equals $g(h)$ algebraically; this is not an independent third method. The algorithm uses the corresponding $g(h)$, rather than refusing all multi-step candidates merely because $D_1 = 1$. Evidence lengths within a seed share trajectory prefixes and are not independent replicates. At every one of the 60 longer-evidence cells the minimum coincides with the multi-step rule; the one-step rule admits no feasible geometry at any cell, and all 40 short-evidence cells refuse. Hence release frequency over the full inventory is .600, not 1.000. The original scoring loop inadvertently used one transition between blocks regardless of the selected lag, and the original summary omitted refusals and treated seed–length cells as independent. Those coverage and interval claims are superseded.

The 2026-09-12 correction retains the original selected geometries and independently initialized score streams, but advances each inter-block link with P^h, so block starts differ by $L + h - 1$. Across 60,000 stationary score paths per rule, the multi-step rule has refusal-inclusive represented-block coverage .95564 (an infinite threshold makes each refusal a covered outcome; approximate seed-t 95% interval $[.95425, .95703]$). The minimum has .95378 $[.95231, .95525]$; its geometry is identical but it uses a separate Monte Carlo stream, so this difference is not an envelope effect. First average all five lengths within each seed, then form intervals over 20 seed means (19 degrees of freedom). These are Monte Carlo diagnostics, not a proof of nominal interval coverage. Exactly 60/100 cells release and 40/100 refuse for each of these two rules. Their released-only descriptive coverages are .9260667 and .9229667, respectively; these are not release-conditional guarantees. The one-step rule's refusal-inclusive coverage is one solely because it refuses in 100/100 cells. Among released cells the mean threshold is .4512, mean span 139,589, mean blocks 2,178 and mean lag 71.67. Identical seed-level release frequencies have no assessed t interval. The runner `simulation/selected_gap_experiment.py` retains selections and aggregate score outcomes, not individual paths. The corrected artifact record binds the input results hash and execution-time scoring source (a subsequent docstring-only correction names D_1 as lower-overlap,

not exact); the original record is preserved. This is post-run repair on the same transition evidence, not a new independent confirmation.

A.13 MULTI-STEP ENVELOPES FROM INDEPENDENT TRANSITION EVIDENCE

Premises and construction. Assume an observed stationary Markov chain on a declared finite state space $\{1, \dots, K\}$, with transition matrix P, and fixed memoryless score emissions, conditionally independent given the states. Transition evidence E is independent of the entire calibration and test score path. Let C_i be nonempty row box–simplex confidence sets satisfying $\Pr\{P_{i\cdot} \in C_i \ \forall i\} \geq 1 - \delta_{\text{env}}$, where δ_{env} is the simultaneous failure budget. All declared states, including unvisited states, remain represented. This extension assumes that confidence guarantee rather than deriving a new transition concentration result.

For a positive integer power cap R, initialize $L^{(0)} = U^{(0)} = I$ and define, for $0 \leq r < R$,

$$L_{ij}^{(r+1)} = \min_{p \in C_i} \sum_k p_k L_{kj}^{(r)}, \qquad U_{ij}^{(r+1)} = \max_{p \in C_i} \sum_k p_k U_{kj}^{(r)}.$$

Each objective is linear; residual probability mass can be allocated in sorted coefficient order subject to the row bounds. For $K > 1$, use the implemented conservative row-overlap bound

$$D_r = 1 - \min_{i \neq j} \sum_k \min\{L_{ik}^{(r)}, L_{jk}^{(r)}\}.$$

Set $D_r = 0$ for $K = 1$. Since each true row dominates its lower endpoints, $\text{TV}(P_{i\cdot}^r, P_{j\cdot}^r) \leq D_r$. This is an upper bound, not exact optimization of TV over the row boxes. Propagation and budget comparisons use rational arithmetic; floating mixture-inversion endpoints are widened until 40-digit interval evaluation proves the required outside-set inequality, or revert to trivial bounds.

Envelope proposition and proof. On simultaneous confidence coverage, induction gives $L^{(r)} \leq P^r \leq U^{(r)}$ entrywise: insert the true row into each optimization and use nonnegative transition weights. Consequently $d(P^r) \leq D_r$, where $d(Q) = \max_{i,j} \text{TV}(Q_{i\cdot}, Q_{j\cdot})$ is the Dobrushin coefficient. For integer gap $h \geq 1$, define

$$g_U(h) = \min_{1 \leq r \leq \min(R,h)} D_r^{\lfloor h/r \rfloor}.$$

Submultiplicativity, with the remaining power bounded by one, and memoryless-emission data processing yield $\beta_S(h) \leq d(P^h) \leq g_U(h)$. Including $r = 1$ gives $g_U(h) \leq D_1^h$ for its own lower-overlap bound, not necessarily for the earlier exact box–simplex bound d_U^h. Separate optimizers across recursion levels enlarge uncertainty; this is not exact optimization over one fixed kernel. No additional confidence budget is required for these powers.

For example, let $K = 3$ and every row set be $0 \leq p_j \leq 1/2$, $\sum_j p_j = 1$. Its three vertices are the permutations of $(0, 1/2, 1/2)$, so the exact robust one-step coefficient is $d_U = 1/2$. All propagated lower endpoints are zero, however, giving $D_r = g_U(h) = 1$. Thus dominance over d_U^h is not established. If a certified d_U from the same row sets is available, $\min\{g_U(h), d_U^h\}$ is a valid stronger envelope on the same confidence event. The numerical frontier below did not evaluate that combined rule.

Independent-evidence coverage proposition and proof. Fix B calibration blocks, their length L, and separation h. Let α_B^{rank} denote the exact independent-block rank miss bound and τ the target miss probability. Release the rank threshold only if

$$\alpha_B^{\text{rank}} + \delta_{\text{env}} + B g_U(h) \leq \tau;$$

otherwise return infinity. Conditional on any covering evidence realization, independence preserves the stationary score-path law. The rank-plus-TV argument bounds released-threshold miss probability by $\tau - \delta_{\text{env}}$; refusal contributes zero. Integrating and charging confidence failure gives unconditional represented-block miss probability at most τ. Because the same covering event bounds every tested gap, selection of h from E alone also preserves this argument when the score process is independently sampled at that geometry. This is not coverage conditional on release and extends neither historical-prefix nor online-monitoring guarantees.

Retained numerical frontier. The runner `simulation/multistep_frozenlake_experiment.py` records 100 completed jobs, 20 per evidence length, with $R = 32$, $B = 99$, block length 4, rank 90, $\delta_{\text{env}} = .01$, $\tau = .12$, and tested gaps 1–128. The closed 11-state analysis uses extra public map/terminal-reset knowledge, never observed-label selection. The full 16-state confidence model retains five unreachable/unvisited labels and refuses.

Evidence length	D_1^h successes	$g_U(h)$ successes	Minimum certified gaps
1000	0/20	0/20	none
5000	0/20	0/20	none
20000	0/20	20/20	72–110
80000	0/20	20/20	48–54
320000	0/20	20/20	40

Success means certification at some tested gap, not empirically verified performance at the selected gap. Both implemented envelopes still refuse at $h = 27$. The first column uses the lower-overlap D_1^h, not the earlier exact box–simplex d_U^h; it establishes no superiority over that exact one-step comparator. Recorded confidence failures and numerical fallbacks are zero; these counts do not establish universal arithmetic correctness. The fixed-$h = 27$ experiment evaluates 1,000 independent score paths per job; because every estimated gate refuses there, its zero miss count is not evidence of useful finite-threshold coverage. This earlier fixed-geometry run did not measure selected-gap coverage; the corrected follow-up is in Appendix A.12. The structural premises remain unverified for the observational real-data diagnostics.

Reproduction record. The entry point is `simulation/multistep_frozenlake_experiment.py`. JSON result paths in this appendix are relative to the project evidence bundle at the repository root. Source and configuration bindings accompany the local records; this does not establish independent external replication.

A.14 Independent empirical transport and refusal-inclusive block loss

We preregistered an empirical transport study on the UCI Metro Interstate Traffic Volume dataset (dataset 492, `https://doi.org/10.24432/C5X60B`, CC BY 4.0), a single Minnesota I-94 ATR station 301 trajectory. This source was not used in earlier development. Its observational nature does not establish finite-state Markovity, stationarity, fixed memoryless emissions, an independent transition-evidence stream, or a valid mixing envelope.

From 48,204 raw rows, deterministic preprocessing removes 17 exact duplicate rows and collapses 7,612 additional same-target timestamp duplicates, leaving 40,575 unique hourly observations. No conflicting-target timestamp occurs. Missing hours are not imputed or compressed. The 34,042 observations before 2018 are separated into pre-2016 predictor development, 2016 calibration, and 2017 method/baseline selection; the 6,533 observations from 2018 are sealed until code, configuration, thresholds and selection are fixed. Timestamps retain the source's naive local-clock convention.

The released LightGBM predictor uses 48 target lags and calendar features, with 100 trees, depth 6, learning rate .1 and seed 42. These are rolling one-hour-ahead predictions using previously observed values, not a 100-step open-loop forecast made at a window's start. Lag features and calibration blocks cannot cross a missing hour or partition boundary. The frozen empirical geometry is $L = 100, G = 50$. Selection in 2017 chooses the block-maximum path over the default within-block .90-quantile path.

For a common target across policies, each contiguous run is partitioned into complete nonoverlapping 100-hour windows. A window is released only if all 100 constituent hours are released by the separated-block schedule; partial releases are conservatively classified as whole-window refusal. This full-window operational restriction is explicit and is not a new theorem instantiation. A released window misses if any of its hourly intervals misses. The primary conditional loss is $10\,\mathbf{1}\{\text{released window misses}\} + \mathbf{1}\{\text{window refused}\}$. The ratio is a declared sensitivity convention, not elicited utility, and contains no width penalty. Hourly results remain secondary.

The 2017 comparison grid includes always-release Split CP, six fixed-width multiples of its calibrated radius, and always-refuse. Always-refuse has the smallest baseline block loss and is fixed before confirmation. The 2018 analysis has 5,959 usable lagged examples and 54 complete common windows; 559 incomplete tail hours are retained in accounting but excluded from these window metrics.

Policy	Released	Refused	Coverage	10:1 loss
TEMPORAL-CP block maximum	20	34	$19/20 = .95$	$44/54 = .815$
Selected always-refuse baseline	0	54	undefined	1.000

Coverage in the table is conditional on release. All 34 refused windows are partial hourly releases; none is counted as a covered forecast. Released intervals have width 5091.93 in hourly traffic volume units. The full-window point loss difference is $-.1852$. Within each contiguous run, adjacent window pairs form prespecified 200-hour clusters; 25 clusters are retained and four unpaired windows are omitted from interval estimation. Independent recomputation reproduces 10,000 paired resamples with seed 2026091401: the cluster-weighted loss difference is $-.18$, its upper 97.5th percentile is .26, and the released-block coverage lower 2.5th percentile is $.8235$.

The prespecified joint criterion fails: the coverage lower endpoint is below .90 and the loss upper endpoint is not below zero, although the release fraction $20/54 = .3704$ exceeds the declared .25 floor. These resampling summaries are conditional on approximate stationarity, weak dependence across clusters and retained-cluster representativeness; observational diagnostics do not prove those assumptions. They are not certified confidence bounds, population guarantees or independent-trajectory replications. No threshold is retuned after confirmation. The study supplies a bounded negative external transport result, not evidence of general operational benefit. Complete window records, hashes, selection and the failed primary decision remain in `simulation/metro_interstate_v1/`.