The Two Prices of Dependence:
a Vanishing Coverage Tax and a Fixed Sharpness Rent

Peter Cotton
peter.cotton@microprediction.com

July 2026

Abstract

Temporal dependence charges split conformal prediction two different prices, and they scale differently. The coverage price is a tax that vanishes with the calibration size: for a stationary β-mixing score process the loss is at most $\min_\tau\{\tau/(n+1) + 2\beta(\tau)\}$ (Barber and Pananjady, 2026). The sharpness price is a rent that does not: the per-step log-score regret of the marginally calibrated predictive against the past-conditional oracle equals the entropy-rate gap $\Delta H = H(S_0) - h$, the mutual information between the present score and its infinite past, a constant independent of n. For Gaussian score processes the rent exponentiates into width — oracle intervals are narrower by the factor $e^{-\Delta H}$ at matched coverage, which for an AR(1) score process with parameter ϕ is $\sqrt{1-\phi^2}$. Dependence is therefore asymptotically free for the certificate and permanently valuable for the forecast.

1 Setting

Let $(S_t)_{t\in\mathbb{Z}}$ be a stationary ergodic sequence of conformity scores with continuous marginal distribution $\bar{F}$ and marginal density $\bar{f}$. Write $\mathcal{P}_t = \sigma(S_{t-1}, S_{t-2}, \dots)$ for the past, $F_t = \mathcal{L}(S_t \mid \mathcal{P}_t)$ for the past-conditional law with density $f(\cdot \mid \mathcal{P}_t)$, $H(S_0)$ for the differential entropy of the marginal, and $h = \lim_{m\to\infty} H(S_0 \mid S_{-1}, \dots, S_{-m}) = H(S_0 \mid \mathcal{P}_0)$ for the entropy rate (Cover and Thomas, 2006). All entropies are assumed finite. Two predictive systems are compared at each step:

- the *marginal* system, which issues the stationary law $\bar{F}$ (this is what split conformal calibrates: its quantile is an estimate of a quantile of $\bar{F}$, and the predictive density it implies is the one residual shape $\bar{f}$, re-levelled);
- the *past-conditional oracle*, which issues F_t.

The comparison isolates what marginality itself forfeits; estimation error in either system is not the subject.

2 The coverage tax

Dependence costs the marginal system coverage, and the cost is quantified by Barber and Pananjady (2026): for a stationary β-mixing score sequence, split conformal with n calibration points and a memoryless pretrained score satisfies

$$\mathbb{P}(\text{cover}) \;\ge\; 1 - \alpha - \min_{\tau\ge 0}\Big\{\tfrac{\tau}{n+1} + 2\beta(\tau)\Big\}, \tag{1}$$

with a matching upper bound, and the bound is tight up to a constant factor over this class. For fixed dependence the tax vanishes: for a process whose dependence dies at lag t (e.g. MA(t)), the loss is at most $(t+1)/(n+1)$, and for geometrically β-mixing processes (e.g. stationary Gaussian AR(1)) it is $O(\log n/n)$. The earlier rate of Oliveira et al. (2024), of order $\sqrt{\tau/n}$, is likewise vanishing. This section quotes Barber and Pananjady (2026) to fix one side of the comparison.

3 The sharpness rent

Proposition 1 (Rent identity)**.** *The per-step expected log-score regret of the marginal system against the past-conditional oracle is*

$$\mathbb{E}\Big[\log \frac{f(S_0 \mid \mathcal{P}_0)}{\bar{f}(S_0)}\Big] \;=\; \mathbb{E}\,\mathrm{KL}\big(F_0 \,\big\|\, \bar{F}\big) \;=\; I\big(S_0; \mathcal{P}_0\big) \;=\; H(S_0) - h \;=:\; \Delta H.$$

Proof. The first equality is the tower property. The second is the standard identity $I(X;Y) = \mathbb{E}_Y \,\mathrm{KL}(\mathcal{L}(X \mid Y) \,\|\, \mathcal{L}(X))$ applied with $X = S_0$, $Y = \mathcal{P}_0$. The third is $I(S_0; \mathcal{P}_0) = H(S_0) - H(S_0 \mid \mathcal{P}_0)$ and the definition of the entropy rate of a stationary process. $\square$

The rent is a property of the score process alone. It does not depend on the calibration size n, and no re-levelling of the marginal shape can reduce it: conformalizing acts on $\bar{f}$ by a monotone re-levelling, which leaves every likelihood ratio $f(\cdot \mid \mathcal{P}_0)/\bar{f}(\cdot)$'s expectation unchanged (Cotton). Proposition 1 is the identity of that paper with the side information taken to be the past rather than a covariate; the transfer requires stationarity and finite entropies, nothing else.

Corollary 1 (Gaussian processes: rent as width)**.** *Let (S_t) be stationary Gaussian with marginal variance σ^2 and one-step innovation variance σ_∞^2 (the Kolmogorov–Szegő prediction variance). Then*

$$\Delta H \;=\; \tfrac{1}{2} \log \frac{\sigma^2}{\sigma_\infty^2},$$

and at any matched coverage level the past-conditional oracle's central interval is narrower than the marginal interval by the factor $\sigma_\infty/\sigma = e^{-\Delta H}$. For an AR(1) process with parameter ϕ this factor is $\sqrt{1-\phi^2}$ and $\Delta H = -\frac{1}{2}\log(1-\phi^2)$.

Proof. For Gaussian laws $H = \frac{1}{2}\log(2\pi e\,\sigma^2)$ and $h = \frac{1}{2}\log(2\pi e\,\sigma_\infty^2)$; both central intervals are z-multiples of the relevant standard deviation at matched level. For AR(1), $\sigma^2 = \sigma_\infty^2/(1-\phi^2)$. $\square$

4 The asymmetry

Fix a stationary Gaussian AR(1) score process and let n grow. The tax (1) is $O(\log n/n)$; the rent is $-\frac{1}{2}\log(1-\phi^2)$ at every n. Numerically:

ϕ	rent ΔH (nats)	oracle width / marginal width
0.3	0.047	0.954
0.6	0.223	0.800
0.9	0.830	0.436

At $\phi = 0.9$ the marginal certificate costs essentially nothing to keep, a tax of order $\log n/n$, while the forecaster who conditions on the past issues intervals less than half as wide at the same coverage, forever. The two prices answer different questions. The tax answers: how much coverage does the certificate lose because the calibration data are dependent? The rent answers: how much sharpness does the forecast lose because the calibration ignores that dependence? The literature on conformal prediction for time series has concentrated on driving the first price to zero; the second price is the one a forecaster pays at every step, and it is not reduced by any amount of calibration data.

5 Limits

Three restrictions delimit the claim. The comparison is between the marginal system and the oracle, so it prices marginality itself; a feasible past-conditional model pays estimation costs the oracle does not, and under strong dependence those costs can be material at small n. The width statement of Corollary 1 is Gaussian; for general processes the rent is the log-score statement of Proposition 1, and width comparisons depend on the shape of the conditional laws. And the rent as stated prices the score process; when scores are residuals from a fixed predictor, conditioning information in the original series that the predictor already absorbs does not appear in ΔH. Within these limits the accounting is exact, and the direction of the asymmetry does not depend on the Gaussian case: the tax vanishes in n and the rent does not.

References

R. F. Barber and A. Pananjady. Predictive inference for time series: why is split conformal effective despite temporal dependence?, 2026. Preprint, arXiv:2510.02471.

P. Cotton. Marginally useful: Formalizing the information gap in conformal prediction. Companion paper.

T. M. Cover and J. A. Thomas. *Elements of Information Theory*. Wiley, 2nd edition, 2006. Chapter 6, gambling and side information; Pinsker's inequality.

R. I. Oliveira, P. Orenstein, T. Ramos, and J. V. Romano. Split conformal prediction and non-exchangeable data. *Journal of Machine Learning Research*, 25, 2024.