VALID AND EFFICIENT SPLIT CONFORMAL REGRESSION FOR TIME SERIES

Anonymous authors
Paper under double-blind review

ABSTRACT

We study conformalized quantile regression and conformalized median regression that fit a model on one block of a time series and calibrate the conformal interval on the adjacent block. The existing theory of conformal prediction for time series rests largely on mixing conditions, which are hard to verify from a time-series model and fail for many standard processes, including simple ones with short memory. We replace this theoretical toolbox with the functional dependence measure, which in principle accommodates long-memory observations. The accuracy of the conformal interval length for time series has been understudied. To the best of our knowledge, this paper is the first work that establishes non-asymptotic coverage guarantees and accuracy of interval length simultaneously for split conformal regression on time series. Furthermore, for Gaussian linear processes with long memory, where both the estimation of the center and its calibration converge slowly, we establish a sharper rate for the length error. We show that the calibrated length converges faster than the estimated center itself, and provide a matching lower bound for the usual centers when the calibration block is sufficiently large relative to the training block. To our knowledge, this is the first theoretical analysis of conformal interval length dedicated to long memory.

1 INTRODUCTION

Split conformal prediction turns a fitted regression model into a prediction interval: the model is fitted on a training block, and the interval for a new observation is the fitted interval enlarged by an empirical quantile of scores computed on a separate calibration block (Papadopoulos et al., 2002; Vovk et al., 2005; Lei et al., 2018). For conformalized quantile regression (CQR; Romano et al., 2019) the model consists of two fitted conditional quantile functions, and for conformalized median regression (CMR) of a single fitted center whose absolute residuals are calibrated. When the calibration and test observations are exchangeable, the interval covers the new response with probability at least the nominal level whatever the fitted model is. For a time series the two blocks and the test point are adjacent segments of one trajectory, exchangeability fails, and two questions arise: whether the interval still covers at the nominal level, and whether its length approaches the length of the oracle interval formed by the true conditional quantiles. The second question is the question of efficiency, and the two questions together are the subject of this paper.

For non-exchangeable data, the coverage gap of conformal prediction can be bounded through total variation distances (Barber et al., 2023); for a stationary series, the gap of split, block, jackknife and ensemble procedures has been quantified through mixing conditions (Chernozhukov et al., 2018; Xu & Xie, 2021; Oliveira et al., 2024; Barber & Pananjady, 2026; Jiang et al., 2026), the standard toolbox for learning from dependent data (Yu, 1994; Mohri & Rostamizadeh, 2010). This toolbox has two limitations: mixing coefficients are rarely computable from the model that generates a series (Wu, 2005b; 2011; Han & Wu, 2023), and mixing conditions fail for many simple time series (Andrews, 1984; Bradley, 2005). Long memory, the property that the autocorrelations of a series are not summable (Beran, 1994), is documented for asset returns and network traffic (Ding et al., 1993; Leland et al., 1994) and in language and music (Greaves-Tunnell & Harchaoui, 2019), and standard recurrent networks do not capture it (Zhao et al., 2020). The functional dependence measure of Wu (2005b) is computed from the model that generates the series and underlies a theory of quantiles and of high-dimensional time series (Wu, 2005a; 2011; Chen et al., 2013; Zhang & Wu, 2017); for a linear process it is determined by the coefficients of the process, and the empirical processes and

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.

sample covariance matrices of linear processes are understood also under long memory (Wu, 2003; Chen & Wu, 2018; Zhai et al., 2026).

The length of conformal intervals has been studied for exchangeable data: split conformal and CQR intervals approach the oracle length when the fitted model is consistent (Lei et al., 2018; Sesia & Candès, 2020), and Yao et al. (2026) give non-asymptotic rates for the length error of CQR and CMR. For dependent data, the bound of Xu & Xie (2023, Theorem 3) on the discrepancy between their ensemble interval and the oracle interval is proved for i.i.d. errors; the efficiency analysis of adaptive conformal inference (Gibbs & Candès, 2021) by Zaffran et al. (2022) is asymptotic and assumes a known quantile function of the scores; and distributional conformal prediction, a different construction, attains an optimal interval only asymptotically (Chernozhukov et al., 2021). None of these compares the calibrated length with the oracle length at a non-asymptotic rate under temporal dependence. Zheng & Proutiere (2024) bound the effect of Markov dependence on the size of the split conformal set, but only for Markov chains with a finite uniform mixing time, and show that the set converges to the oracle set without a rate.

Sesia & Romano (2021), Kiyani et al. (2024) and Le Bars & Humbert (2025) study efficiency and length optimization for other oracle targets; we compare with the marginal equal-tailed oracle. Halkiewicz (2026) analyzes the coverage of rolling-origin conformal prediction under strong mixing and, in an extension, under a summability condition on the physical dependence measure; neither condition allows Gaussian long memory.

Our contributions can be summarized as follows.

1. The mixing conditions on which the existing guarantees largely rest fail for many standard processes, including simple ones with short memory. We replace this theoretical toolbox with the functional dependence measure, under which our guarantees hold for such processes and, in principle, allow for long memory.
2. For CQR and CMR with adjacent training and calibration blocks, we bound the coverage error of the calibrated interval and the deviation of its length from that of the oracle interval (Theorems 1–3). To our knowledge, these are the first non-asymptotic results that establish coverage and accuracy of interval length simultaneously for split conformal regression on time series.
3. For Gaussian linear processes with long memory, we establish a sharper rate for the length error of CMR (Theorem 4): with blocks of comparable size, the calibrated length converges faster than the estimated center itself. A matching lower bound shows that this rate is exact when the calibration block is large relative to the training block (Proposition 5). To our knowledge, Theorem 4 and Proposition 5 give the first theoretical analysis of conformal interval length dedicated to long memory and the first exact rate for the length of a conformal interval under dependence.

Section 2 states the two procedures and introduces the functional dependence measure, Section 3 states the three conditions, the coverage and length theorems and the examples, Section 4 contains the results on estimated centers under Gaussian long memory, Section 5 presents the numerical studies, and Section 6 discusses the scope of the results; all proofs are deferred to the appendices.

2 Problem Setup and Prerequisites

Let $Z_t = (X_t, Y_t) = G(\ldots, \varepsilon_{t-1}, \varepsilon_t)$ be a strictly stationary process driven by i.i.d. innovations ε_t through a measurable map G, and let $\mathcal{F}_t = \sigma(\varepsilon_s : s \le t)$ be the information available at time t. Causal linear processes and many nonlinear time series models have this form (Wu, 2005b). The level $\alpha \in (0, 1)$ is fixed. The predictor is trained on $Z_1, \ldots, Z_n$, calibrated on $Z_{n+1}, \ldots, Z_{n+m}$, and used to predict Y_{n+m+1} from X_{n+m+1}; no observations are discarded between the blocks, and the training algorithm may use independent randomness.

2.1 Conformalized Quantile Regression

Let $\ell_\alpha(x)$ and $u_\alpha(x)$ be the $\alpha/2$- and $(1-\alpha/2)$-quantiles of the conditional law of Y_t given $X_t = x$ under the stationary distribution, assumed continuous at these two quantiles, so that the oracle inter-

val $C_\alpha^Q(x) = [\ell_\alpha(x), u_\alpha(x)]$ has conditional coverage $1-\alpha$. This oracle conditions on the covariate, which may include lagged responses, and not on the whole past of the series. On the training block, any algorithm produces two functions $\widehat{\ell}$ and $\widehat{u}$ that estimate ℓ_α and u_α; on the calibration block, each observation receives the nonconformity score $\widehat{S}_t^Q = \max\{\widehat{\ell}(X_t) - Y_t,\ Y_t - \widehat{u}(X_t)\}$, $t = n+1, \ldots, n+m$. The calibrated correction is the k-th smallest score $\widehat{S}_{(k)}^Q$ with $k = \lceil (m+1)(1-\alpha) \rceil$, which is at most m when $m \geq (1-\alpha)/\alpha$, as assumed throughout, and the prediction interval for Y_{n+m+1} is $\widehat{C}_\alpha^Q(x) = [\widehat{\ell}(x) - \widehat{S}_{(k)}^Q,\ \widehat{u}(x) + \widehat{S}_{(k)}^Q]$ at $x = X_{n+m+1}$, an inverted interval being empty. The same steps with the true quantiles produce the oracle scores $S_t^Q = \max\{\ell_\alpha(X_t) - Y_t,\ Y_t - u_\alpha(X_t)\}$, whose distribution function F^Q satisfies $F^Q(0) = 1-\alpha$, and the k-th smallest oracle score $S_{(k)}^Q$ of the calibration block, an empirical $(1-\alpha)$-quantile of m dependent scores that estimates zero. The training step enters the results only through the uniform error of the fitted endpoints, $E_n^Q = \max\{\|\widehat{\ell} - \ell_\alpha\|_\infty, \|\widehat{u} - u_\alpha\|_\infty\}$, the norms being taken over a fixed domain that contains every realized covariate almost surely.

For CQR we ask whether the marginal coverage $\mathbb{P}\{Y_{n+m+1} \in \widehat{C}_\alpha^Q(X_{n+m+1})\}$ is close to $1-\alpha$, and whether the length of $\widehat{C}_\alpha^Q(x)$ is close to the length of the oracle interval $C_\alpha^Q(x)$, uniformly in x. On every realization,

$$|\widehat{S}_{(k)}^Q - S_{(k)}^Q| \leq E_n^Q, \qquad \sup_x |\operatorname{len} \widehat{C}_\alpha^Q(x) - \operatorname{len} C_\alpha^Q(x)| \leq 4E_n^Q + 2|S_{(k)}^Q|. \tag{1}$$

Inequality (1), proved in Appendix A.1, uses no independence between the blocks and no gap between them, and it reduces both questions to a training question, whether E_n^Q is small, and a calibration question, whether the empirical $(1-\alpha)$-quantile of m dependent oracle scores is close to the population quantile.

2.2 Conformalized Median Regression

For CMR, the training block produces one function $\widehat{\mu}$ that estimates the conditional median $\mu(x)$ of Y_t given $X_t = x$, the calibration scores are the absolute residuals $\widehat{S}_t^M = |Y_t - \widehat{\mu}(X_t)|$, the calibrated radius is $\widehat{S}_{(k)}^M$, and the prediction interval is $\widehat{C}_\alpha^M(x) = [\widehat{\mu}(x) - \widehat{S}_{(k)}^M,\ \widehat{\mu}(x) + \widehat{S}_{(k)}^M]$. Its oracle counterpart is $C_\alpha^M(x) = [\mu(x) - q_\alpha,\ \mu(x) + q_\alpha]$, where q_α is the $(1-\alpha)$-quantile of the oracle score $S_t^M = |Y_t - \mu(X_t)|$ under the stationary law, and the training error is $E_n^M = \|\widehat{\mu} - \mu\|_\infty$. The two oracles differ: $C_\alpha^Q(x)$ has coverage $1-\alpha$ at every x, whereas $C_\alpha^M(x)$ uses one radius for all x and has coverage $1-\alpha$ on average over x.

The oracle calibration error is now the distance of $S_{(k)}^M$ from q_α in place of the distance of $S_{(k)}^Q$ from zero, and a proof similar to that of (1) gives, on every realization,

$$|\widehat{S}_{(k)}^M - S_{(k)}^M| \leq E_n^M, \quad \sup_x |\operatorname{len} \widehat{C}_\alpha^M(x) - \operatorname{len} C_\alpha^M(x)| = 2|\widehat{S}_{(k)}^M - q_\alpha| \leq 2E_n^M + 2|S_{(k)}^M - q_\alpha|. \tag{2}$$

2.3 Functional Dependence Measure

The calibration question requires a measure of the dependence of the series, and we use the functional dependence measure of Wu (2005b). Let ε_0' be an independent copy of ε_0, and let $Z_t^* = G(\ldots, \varepsilon_{-1}, \varepsilon_0', \varepsilon_1, \ldots, \varepsilon_t)$ be the trajectory obtained by replacing the single innovation at time 0 by its copy, all other innovations being retained; S_t^{Q*} and S_t^{M*} denote the oracle scores computed from $Z_t^* = (X_t^*, Y_t^*)$ with the oracle endpoints evaluated at X_t^*. The functional dependence measures of the observations and of the two oracle scores at lag j are

$$\delta_q^Z(j) = \|Z_j - Z_j^*\|_q, \qquad \delta_q^Q(j) = \|S_j^Q - S_j^{Q*}\|_q, \qquad \delta_q^M(j) = \|S_j^M - S_j^{M*}\|_q, \qquad q \geq 1,$$

where $\|V\|_q = (\mathbb{E}|V|^q)^{1/q}$ with $|\cdot|$ the Euclidean norm: the L^q distance by which the series, or its score, j steps ahead responds to one innovation. For a linear process $Y_t = \sum_{j \geq 0} a_j \varepsilon_{t-j}$ the measure equals $|a_j|\,\|\varepsilon_0 - \varepsilon_0'\|_q$, for a recursion $Z_t = H(Z_{t-1}, \varepsilon_t)$ that contracts in L^q it decays geometrically (Wu & Shao, 2004), and when $a_j \asymp j^{-\beta}$ with $1/2 < \beta < 1$, so that the autocorrelation at lag h decays like $h^{1-2\beta}$ and the series has long memory, it decays like $j^{-\beta}$: the exponent of its polynomial decay is the memory exponent of the series.

The oracle calibration error is controlled through the events $\{S_t^Q \le s\}$ for thresholds s near the population quantile, and the relevant quantity is therefore the response of the conditional probability of such an event to one innovation, after the intervening innovations have been averaged out:

$$\omega^Q(j) = \sup_{s\in\mathbb{R}} \left\| \mathbb{P}(S_j^Q \le s \mid \mathcal{F}_0) - \mathbb{P}(S_j^{Q*} \le s \mid \mathcal{F}_0^*) \right\|_2, \qquad j \ge 0, \tag{3}$$

where $\mathcal{F}_0^* = \sigma(\varepsilon_0', \varepsilon_s : s \le -1)$, and $\omega^M(j)$ is defined in the same way with S_j^M and S_j^{M*}. The second probability is the first one computed from the coupled innovations, so $\omega^Q(j)$ is the supremum over s of the functional dependence measure, in L^2, of the conditional probability $\mathbb{P}(S_j^Q \le s \mid \mathcal{F}_0)$, which is the predictive dependence measure of the indicator of $\{S_j^Q \le s\}$ in the sense of Wu (2005b).

3 Main Results

This section provides the marginal coverage and the interval length guarantees of the two procedures simultaneously, for the same calibrated interval. Our theoretical guarantee relies on three assumptions. The first assumption concerns the training step and is the only requirement on the learner. Throughout, the training error E_n denotes E_n^Q for CQR and E_n^M for CMR.

Assumption 1 (Consistent endpoints)**.** *There are deterministic sequences $e_n \to 0$ and $\xi_n \to 0$ such that the training error satisfies $\mathbb{P}(E_n > e_n) \le \xi_n$ for every n.*

Assumption 1 is equivalent to $E_n \to 0$ in probability, with e_n and ξ_n recording its rate. It is stated for a general learner on purpose and is easy to check: by ergodicity, linear quantile regression satisfies it under mild conditions on every stationary series of the form of Section 2 whose conditional quantiles are linear in bounded covariates, whatever its dependence (Appendix B.1), and in the Gaussian model of Section 4 the sample mean and the empirical quantiles of the training block satisfy it with an explicit rate.

The second assumption requires that the oracle score have probability mass on both sides of its $(1-\alpha)$-quantile at a rate bounded above and below. Let F denote the distribution function of the oracle score centered at that quantile, so that $F(0) = 1-\alpha$: $F = F^Q$ for CQR, and $F = F^M$ for CMR, where $F^M(s) = \mathbb{P}(S_t^M - q_\alpha \le s)$.

Assumption 2 (Regular oracle score distribution)**.** *There are constants $r > 0$ and $0 < f_\alpha \le L_\alpha < \infty$ such that F is continuous on $[-r, r]$ and, for $0 \le t \le r$,*

(a) $F(t) - (1-\alpha) \ge f_\alpha t$ and $(1-\alpha) - F(-t) \ge f_\alpha t$;

(b) $F(t) - (1-\alpha) \le L_\alpha t$ and $(1-\alpha) - F(-t) \le L_\alpha t$.

For CQR, Assumption 2 holds when the conditional density of Y_t given X_t is bounded above and away from zero near the two oracle endpoints (Appendix B.3), a condition of the kind used for efficiency under exchangeability (Lei et al., 2018; Yao et al., 2026). The length results need only part (a); the coverage results need both parts.

The third assumption is the dependence condition. Consider the predictive dependence measure $\omega(j)$ of the thresholded oracle scores defined in (3), which is $\omega^Q(j)$ for CQR and $\omega^M(j)$ for CMR. The following assumption controls the contribution of the oracle calibration error to the coverage and length errors.

Assumption 3 (Functional dependence of the oracle scores)**.** *There are constants $A > 0$ and $\gamma > 1/2$ such that $\omega(j) \le A(1 \vee j)^{-\gamma}$ for all $j \ge 0$.*

Assumption 3 requires no mixing, can be verified from the model through the functional dependence measure of the scores, and in principle allows for long memory (Appendix B.4). It holds for the autoregression $W_t = W_{t-1}/2 + B_t$ with Bernoulli innovations and for Gaussian long memory, which violate the mixing conditions of the coverage literature, and for long memory in the scale of a volatility model and contractive nonlinear recursions (Appendices B.5 and B.6). For the Bernoulli autoregression, the coverage bounds of Oliveira et al. (2024) and Barber & Pananjady (2026) are vacuous, because its β-mixing coefficients, on which the first bound rests, and its switch coefficients, on which the second rests, equal one at every lag (Appendix B.6).

Under Assumption 3, the empirical distribution function of the oracle scores concentrates near the threshold at the rate $\rho_m = m^{-1/2}$ if $\gamma > 1$, $\rho_m = m^{-1/2}\log(m+1)$ if $\gamma = 1$ and $\rho_m = m^{1/2-\gamma}$ if $1/2 < \gamma < 1$, with a constant C_σ depending only on A and γ (Lemma 8 in Appendix A.2); apart from the training error, this is the only way in which the dependence of the series enters the results. For brevity, we first state the marginal coverage result and the interval length guarantee for CQR.

Theorem 1 (CQR marginal coverage)**.** *Suppose that Assumptions 1–3 hold. There exist $m_0 \le m_1$, depending only on α, γ, f_α, L_α, r and C_σ, such that for all n with $e_n \le r/4$ and all $m \ge m_1$,*

$$\left|\mathbb{P}\{Y_{n+m+1} \in \widehat{C}^Q_\alpha(X_{n+m+1})\} - (1-\alpha)\right| \le 2L_\alpha e_n + \xi_n + C_1\rho_m^{2/3}, \tag{4}$$

where C_1 depends only on L_α, f_α and C_σ. If in addition the conditional law of S^Q_{n+m+1} given $\mathcal{F}_{n+m}$ has a density at most L_{pred} almost surely and $\mathbb{E}(S^Q_0)^2 < \infty$, then for all n and all $m \ge m_0$,

$$\left|\mathbb{P}\{Y_{n+m+1} \in \widehat{C}^Q_\alpha(X_{n+m+1})\} - (1-\alpha)\right| \le L_{\mathrm{pred}}\{2\mathbb{E}E^Q_n + C_2\rho_m/2\}, \tag{5}$$

where C_2 depends only on α, f_α, r, $\mathbb{E}(S^Q_0)^2$ and C_σ.

Theorem 1 requires no mixing condition and, in principle, allows for long memory. Bound (4) holds without a density for the score at the test time; under the additional density condition, bound (5) replaces $\rho_m^{2/3}$ by ρ_m (Appendix A.4).

Theorem 2 (CQR accuracy of interval length)**.** *Suppose that Assumptions 2(a) and 3 hold. For every $e \ge 0$ and every t with $4/(mf_\alpha) \le t \le r$, a range that is nonempty when $m \ge 4/(f_\alpha r)$,*

$$\mathbb{P}\Big\{\sup_x |\operatorname{len}\widehat{C}^Q_\alpha(x) - \operatorname{len}C^Q_\alpha(x)| > 4e + 2t\Big\} \le \mathbb{P}(E^Q_n > e) + \frac{8C_\sigma^2\rho_m^2}{f_\alpha^2 t^2}. \tag{6}$$

If $\mathbb{E}(S^Q_0)^2 < \infty$, then for all $m \ge m_0$,

$$\mathbb{E}\sup_x |\operatorname{len}\widehat{C}^Q_\alpha(x) - \operatorname{len}C^Q_\alpha(x)| \le 4\mathbb{E}E^Q_n + C_2\rho_m, \tag{7}$$

with the constant C_2 of Theorem 1.

The proofs of Theorems 1 and 2 are given in Appendix A.4. To our knowledge, Theorem 2 is the first non-asymptotic comparison of the calibrated length with the oracle length under temporal dependence. Together, the two theorems show that under Assumptions 1–3, as $n \to \infty$ and $m \to \infty$ in any manner, the same calibrated interval is both asymptotically valid and asymptotically as short as the oracle interval.

The CMR versions are as follows.

Theorem 3 (CMR marginal coverage and accuracy of interval length)**.** *Suppose that Assumptions 1–3 hold for CMR, that is, for E^M_n, F^M and ω^M, and let m_0, m_1, C_1 and C_2 be as in Theorems 1–2, with $\mathbb{E}(S^M_0 - q_\alpha)^2$ in place of $\mathbb{E}(S^Q_0)^2$. For all n with $e_n \le r/4$ and all $m \ge m_1$,*

$$\left|\mathbb{P}\{Y_{n+m+1} \in \widehat{C}^M_\alpha(X_{n+m+1})\} - (1-\alpha)\right| \le 2L_\alpha e_n + \xi_n + C_1\rho_m^{2/3}, \tag{8}$$

and if the conditional law of S^M_{n+m+1} given $\mathcal{F}_{n+m}$ has a density at most L_{pred} almost surely and $\mathbb{E}(S^M_0)^2 < \infty$, then for all n and all $m \ge m_0$,

$$\left|\mathbb{P}\{Y_{n+m+1} \in \widehat{C}^M_\alpha(X_{n+m+1})\} - (1-\alpha)\right| \le L_{\mathrm{pred}}\{2\mathbb{E}E^M_n + C_2\rho_m/2\}. \tag{9}$$

For every $e \ge 0$ and every t with $4/(mf_\alpha) \le t \le r$,

$$\mathbb{P}\Big\{\sup_x |\operatorname{len}\widehat{C}^M_\alpha(x) - \operatorname{len}C^M_\alpha(x)| > 2e + 2t\Big\} \le \mathbb{P}(E^M_n > e) + \frac{8C_\sigma^2\rho_m^2}{f_\alpha^2 t^2}, \tag{10}$$

and if $\mathbb{E}(S^M_0)^2 < \infty$, then for all $m \ge m_0$,

$$\mathbb{E}\sup_x |\operatorname{len}\widehat{C}^M_\alpha(x) - \operatorname{len}C^M_\alpha(x)| = 2\,\mathbb{E}|\widehat{S}^M_{(k)} - q_\alpha| \le 2\mathbb{E}E^M_n + C_2\rho_m. \tag{11}$$

Theorem 3, proved in Appendix A.4, gives the same guarantees for CMR, so the CMR interval is also asymptotically valid and asymptotically as short as its oracle; for Gaussian long memory, Section 4 sharpens its length bound.

4 Estimated Centers under Gaussian Long Memory

A common difficulty with long-memory processes is that estimation converges slowly: the sample median, for instance, converges much more slowly than under short memory (Dehling & Taqqu, 1989). The length bound of Theorem 3 inherits this slowness through both of its terms, the training error and the calibration error. In this section, devoted to long-memory processes, we show that the calibrated length can nevertheless converge faster than the estimated center. We study the Gaussian linear process. It is more convenient to analyze and is nevertheless representative: by the Wold decomposition, every stationary Gaussian process without a deterministic component can be written as a Gaussian linear process (Brockwell & Davis, 1991, Section 5.7). Specifically, we consider

$$G_t = \sum_{j\geq 0} a_j \varepsilon_{t-j}, \quad \varepsilon_t \overset{\text{i.i.d.}}{\sim} N(0,1), \quad \sum_{j\geq 0} a_j^2 = 1, \quad a_0 \neq 0, \quad |a_j| \leq A(1+j)^{-\beta}, \tag{12}$$

with $1/2 < \beta < 1$, the long-memory range of Section 2.3. Gaussian innovations are assumed for convenience; since the improvement rests on the symmetry of their law, we expect it to extend to symmetric non-Gaussian innovations.

We observe $Y_t = \mu + G_t$ with an unknown location μ and a constant covariate. The oracle intervals of CQR and CMR then coincide, since the law of G_t is symmetric: $C_\alpha^Q = C_\alpha^M = [\mu - z_\alpha,\ \mu + z_\alpha]$, where $z_\alpha = \Phi^{-1}(1-\alpha/2)$ is the radius q_α of Section 2.2. The two procedures also coincide: for any fitted endpoints, the CQR score is $\max\{\widehat{\ell} - y,\ y - \widehat{u}\} = |y - \widehat{c}| - (\widehat{u} - \widehat{\ell})/2$ with the midpoint $\widehat{c} = (\widehat{\ell} + \widehat{u})/2$; because the covariate is constant, the half-width $(\widehat{u} - \widehat{\ell})/2$ is common to all scores and cancels, and the CQR interval is the CMR interval $[\widehat{c} - \widehat{S}^M_{(k)},\ \widehat{c} + \widehat{S}^M_{(k)}]$ with center $\widehat{c}$. In this section we therefore provide the theory for CMR, with a center fitted on the training block, and it applies to CQR with the midpoint of the fitted endpoints as the center.

For the sample mean, the two terms of Theorem 3 are of order $n^{1/2-\beta}$ and $m^{1/2-\beta}$ (Appendix B.5), and in this model both can be improved. Because the centered oracle score $|G_t| - z_\alpha$ is an even function of G_t, the calibration rate improves from $m^{1/2-\beta}$ to the following rate (Appendix C.1):

$$r_m(\beta) = \begin{cases} m^{-(2\beta-1)}, & 1/2 < \beta < 3/4, \\ \{\log(m+1)/m\}^{1/2}, & \beta = 3/4, \\ m^{-1/2}, & 3/4 < \beta < 1. \end{cases} \tag{13}$$

Because the law of G_t is symmetric, an error in the center changes the population radius of the calibration residuals only at second order (Appendix C.2). The difficulty is to retain this second-order effect with a finite calibration block that is adjacent to the training block, so that the center is correlated with the calibration residuals.

Theorem 4 (CMR length under Gaussian long memory)**.** *Assume* (12)*, and let $\widehat{\mu}$ be a measurable function of $Y_1, \ldots, Y_n$ and of randomness independent of the series. There are constants m_0, depending only on α, and C, depending only on α, β, A and a_0, such that for all n and all $m \geq m_0$,*

$$\mathbb{E}|\operatorname{len} \widehat{C}^M_\alpha - 2z_\alpha| \leq C\big\{\mathbb{E}(\widehat{\mu} - \mu)^2 + r_m(\beta)\big\}. \tag{14}$$

For the sample mean, the empirical median, and the midpoint of the empirical $\alpha/2$- and $(1-\alpha/2)$-quantiles of the training block, $\mathbb{E}(\widehat{\mu} - \mu)^2 \leq Cn^{-(2\beta-1)}$, and therefore $\mathbb{E}|\operatorname{len} \widehat{C}^M_\alpha - 2z_\alpha| \leq C\{n^{-(2\beta-1)} + r_m(\beta)\}$.

Theorem 4 is proved in Appendix C.5. Under this model, the center enters the length through its mean squared error, instead of its absolute error as in Section 3, and with the midpoint as center the theorem covers CQR with the empirical training quantiles as fitted endpoints. With blocks of comparable size, $n \asymp m$, the length error is at most of the order $r_m(\beta)$, which Theorem 4 gives for a known center, although the center converges only at the rate $n^{1/2-\beta}$ when $a_j \asymp j^{-\beta}$; for $\beta > 3/4$, the rate $m^{-1/2}$ of independent scores survives both the long memory and the estimated center. To our knowledge, this is the first theoretical analysis of conformal interval length dedicated to long memory. The coverage error of the interval is at most of the same order as the right-hand side of (14) (Appendix C.7), so the interval also attains the nominal level at this rate.

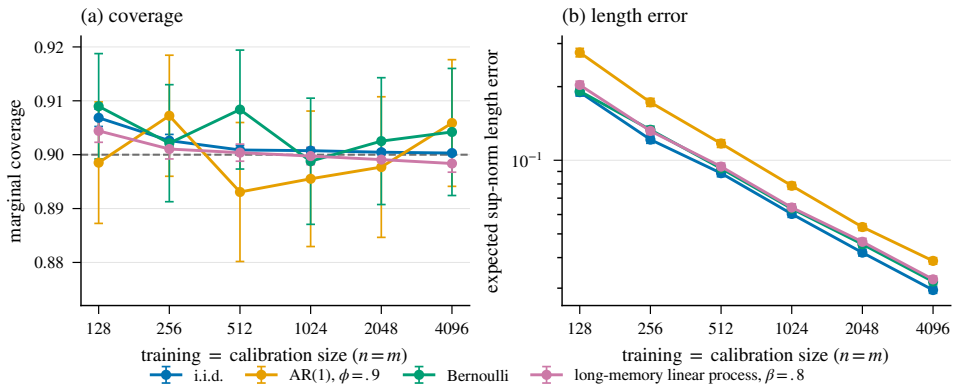

Figure 1: Study 1, CQR with quadratic quantile regression, $n = m \in \{128, \ldots, 4096\}$. (a) Marginal coverage; the dashed line is the nominal level. (b) Expected sup-norm length error, on a logarithmic scale. Pointwise 95% Monte Carlo intervals.

Proposition 5 (Lower bound). *Assume* (12)*, and let $\widehat{\mu}$ be as in Theorem 4. For all n and all $m \geq m_0$, with m_0 as in Theorem 4,*

$$\mathbb{E}|\operatorname{len} \widehat{C}_\alpha^M - 2z_\alpha| \geq c_1 \mathbb{E} \min\{(\widehat{\mu} - \mu)^2, 1\} - C\big[r_m(\beta) + \{m^{-(2\beta-1)}\mathbb{E}(\widehat{\mu} - \mu)^2\}^{1/2}\big], \tag{15}$$

where $c_1 > 0$ depends only on α and C only on α, β, A and a_0. If, in addition, $a_j \geq c(1+j)^{-\beta}$ for all $j \geq j_0$, with $c > 0$, and $\widehat{\mu} = T(Y_1, \ldots, Y_n)$, where T is nondecreasing in each coordinate, satisfies $T(y_1 + h, \ldots, y_n + h) = T(y_1, \ldots, y_n) + h$ for every real h, and has $\mathbb{E}(\widehat{\mu} - \mu)^2 \leq C_g n^{-(2\beta-1)}$, then for all $n \geq n_0$ and $m \geq K n^{\max(1, 4\beta-2)} \log(n+1)$,

$$\mathbb{E}|\operatorname{len} \widehat{C}_\alpha^M - 2z_\alpha| \geq c_3\big\{n^{-(2\beta-1)} + r_m(\beta)\big\}, \tag{16}$$

where n_0, K and $c_3 > 0$ depend only on α, β, A, a_0, c, j_0 and C_g.

Proposition 5 is proved in Appendix C.6. The sample mean, the empirical median and the midpoint satisfy its conditions on the center, so Theorem 4 and Proposition 5 together show that $n^{-(2\beta-1)} + r_m(\beta)$ is the exact rate of their length error when the calibration block is large relative to the training block. By Proposition 5, no monotone and translation-equivariant center whose mean squared error is of the same order $n^{-(2\beta-1)}$ does better. To our knowledge, Theorem 4 and Proposition 5 give the first exact rate for the length of a conformal interval under dependence.

5 Numerical Studies

We examine the theory in three simulation studies whose oracle intervals are known: the coverage and the length in four dependence setups, two of which the mixing theory does not cover (Study 1), the sharper length rate under long memory (Study 2) and the lower bound of Proposition 5 (Study 3). Throughout, $\alpha = .1$, the training and calibration blocks are adjacent, each configuration uses 1000 independent trajectories, which give Monte Carlo standard errors and pointwise 95% intervals, and each procedure is compared with its own oracle interval. Appendix D gives the designs and the complete results.

Study 1. This study illustrates that split conformal regression works well on processes beyond the existing theory. The observations are $Y_t = f(X_t) + s(X_t)h(U_t)$ with $f(x) = x + x^2$, $s(x) = 1 + .3x^2$ and $h(u) = u + u^2/2$, where the covariates X_t are i.i.d. uniform on $[-1, 1]$ and independent of (U_t). The temporal dependence is controlled by (U_t), and we consider four setups for it: independent variables, an AR(1) process, the Bernoulli autoregression of Section 3 and a long-memory linear process of the form (12) with $\beta = .8$, each transformed to the uniform law on $[0, 1]$. The exchangeable theory covers only the independent setup and the mixing theory also covers AR(1), but neither covers the Bernoulli setup or the long-memory linear process. Assumptions 1–3 hold in all four setups, so the theory of this paper covers all of them. The procedure is CQR with quadratic quantile regression

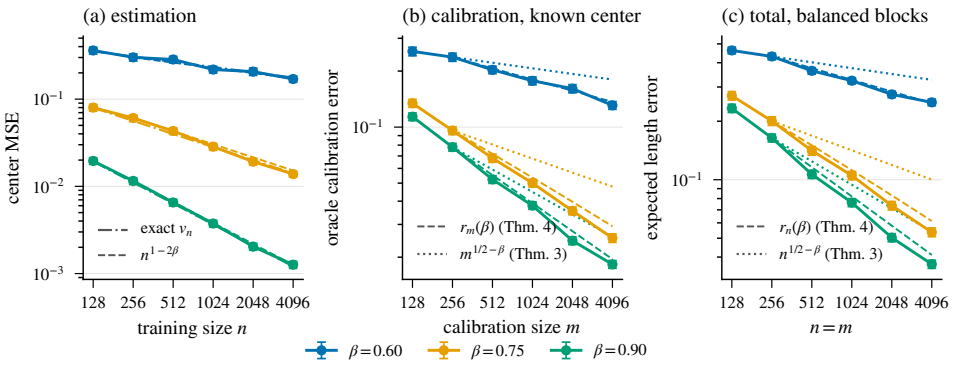

Figure 2: Study 2, CMR with the sample-mean center, Gaussian linear process (12) with $\beta \in \{.6, .75, .9\}$. (a) Mean squared error of the center, with its exact value v_n (dash-dotted) and $n^{-(2\beta-1)}$ (dashed). (b) Oracle calibration error $\mathbb{E}|S^M_{(k)} - z_\alpha|$, with $r_m(\beta)$ (dashed) and $m^{1/2-\beta}$ (dotted). (c) Length error $\mathbb{E}|\operatorname{len} \widehat{C}^M_\alpha - 2z_\alpha|$ with $n = m$, with $r_n(\beta)$ (dashed) and $n^{1/2-\beta}$ (dotted). The guides, matched to the curves at 256, compare decay, not constants. Pointwise 95% Monte Carlo intervals.

at the levels .05 and .95, which is correctly specified. For $n = m \in \{128, \ldots, 4096\}$, we estimate the marginal coverage and the expected sup-norm length error $\mathbb{E} \sup_x |\operatorname{len} \widehat{C}^Q_\alpha(x) - \operatorname{len} C^Q_\alpha(x)|$ of Theorem 2. Appendix D.2 gives the details and the paired CMR results.

Figure 1 shows that the marginal coverage stays within .01 of the nominal level in all four setups, including the two that only the theory of this paper covers, and that the length error decreases toward zero in all four. For the long-memory linear process, the calibration rate ρ_m of Theorem 2 is $m^{-.3}$, against $m^{-1/2}$ in the other three setups, but over the tested sizes its length error decreases at about the same rate as theirs, with fitted decay exponents between .51 and .53 in all four setups, whereas for this process the length error of the paired CMR interval decreases like $m^{-.3}$ (Appendix D.2). The same contrast appears in the oracle calibration step: in this setup, $U_t = \Phi(G_t)$ with G_t of the form (12), and the CQR oracle interval contains Y_t exactly when $|G_t| \le z_\alpha$, an event that is symmetric in G_t, whereas the corresponding event for the CMR oracle interval is not, because h is not linear. This agrees with the motivation of Section 4: long-memory linear processes require a separate analysis to obtain a sharper rate.

Study 2. This study examines the model of Section 4 for two purposes: to measure how slowly the center converges under long memory, and to check whether the length error nevertheless attains the rate of Theorem 4. The observations are $Y_t = \mu + G_t$, where G_t is a Gaussian linear process of the form (12) whose coefficients a_j are asymptotically proportional to $j^{-\beta}$, with $\beta \in \{.6, .75, .9\}$, one value in each case of (13). Appendix D.1 gives the coefficients and the exact simulation of the paths. The procedure is CMR with the training sample mean as the center. For $n = m \in \{128, \ldots, 4096\}$, we measure three quantities on the same trajectories: the mean squared error of the center, the oracle calibration error $\mathbb{E}|S^M_{(k)} - z_\alpha|$ and the interval length error $\mathbb{E}|\operatorname{len} \widehat{C}^M_\alpha - 2z_\alpha|$. Figure 2 compares them with the rates of Theorems 3 and 4.

There are two findings. First, the center converges slowly: its mean squared error decreases like $n^{-(2\beta-1)}$, much more slowly than under short memory, in agreement with its exact value. Second, the oracle calibration error and the length error follow the rates $r_m(\beta)$ and $r_n(\beta)$ of Theorem 4, and not the slower rates $m^{1/2-\beta}$ and $n^{1/2-\beta}$ of Theorem 3. We repeated the study with two non-Gaussian innovation laws, keeping the coefficients and taking the training median as the center. With Laplace innovations, which are symmetric, the findings are the same, as Section 4 expects. Centered exponential innovations are asymmetric by design, to test whether the improvement of Section 4 rests on symmetry: at $\beta = .75$ and .9, the length error then follows the rate of Theorem 3 more closely than that of Theorem 4. At $\beta = .6$ it still follows Theorem 4 over the tested sizes, because the stronger memory spreads each innovation over many lags and brings the law of the series closer to a Gaussian law (Appendix D.4).

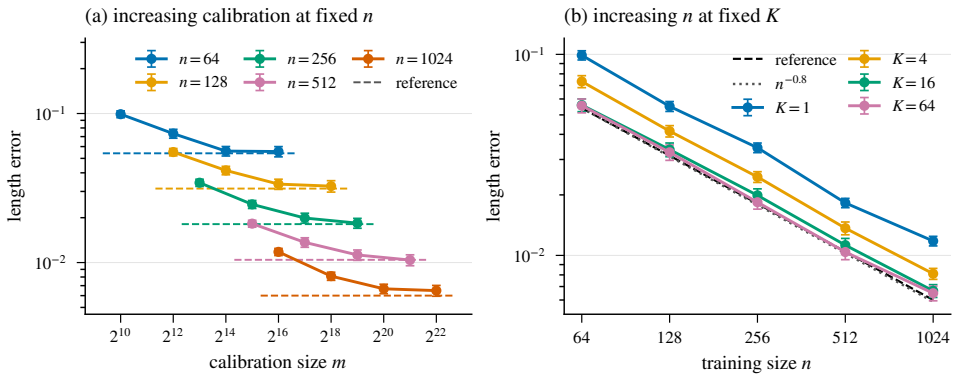

Figure 3: Study 3, CMR with the sample-mean center, $\beta = .9$ and $m \approx Kn^{1.6}$. (a) Length error against m for each training size n, with the reference of each n (dashed, same color). (b) The same errors against n for each K, with the reference (dashed) and the rate $n^{-.8}$ matched to the reference at $n = 64$ (dotted). Pointwise 95% Monte Carlo intervals.

Study 3. This study examines the lower bound of Proposition 5. Calibration does not remove the contribution from center estimation, so however long the calibration block, the length error remains at least of the order $n^{-(2\beta-1)}$, which is the order of the mean squared error of the center. We keep the model of Study 2 with $\beta = .9$ and the sample-mean center, and for each training size $n \in \{64, \dots, 1024\}$ we take the calibration sizes $m \approx Kn^{1.6}$ with $K \in \{1, 4, 16, 64\}$, which follow the threshold of Proposition 5 up to its logarithmic factor and reach thousands of times the training size. The reference is the length error of the interval calibrated on an infinitely long block, which depends only on the center and is close to z_α times its mean squared error. Figure 3 shows that once the calibration block is long enough, the length error stops decreasing and stays within 8% of this reference at every training size: more calibration data only reduces the error due to calibration, which halves with each fourfold increase of m, while the reference, set by the center, does not move. The reference decreases at the rate $n^{-.8} = n^{-(2\beta-1)}$: from $n = 64$ to 1024, the length error with the largest calibration blocks falls by a factor of 8.6, against 9.2 for $n^{-.8}$. Studies 2 and 3 thus illustrate both halves of the exact rate of Section 4: the length error attains the rate of Theorem 4, and the contribution of the center, which enters at second order, remains however much calibration data is added. Appendix D.5 gives the design, the computation of the reference and the complete results.

Real Data Analysis. We also apply CQR and CMR to day-ahead forecast errors of electricity load, which have long memory, and check both the coverage and the interval length. Both procedures cover close to the nominal level, and the mean length of CMR is within 2% of a proxy for its oracle length computed in hindsight (Appendix E).

6 Discussion

The functional dependence measure replaces mixing as the theoretical toolbox for split conformal regression on time series: it is computed from the model that generates the series, the guarantees hold for processes that are not mixing, and the same calibrated interval satisfies non-asymptotic coverage and length guarantees. Under Gaussian long memory, the length converges faster than the estimated center, at a rate that a matching lower bound shows to be exact for long calibration blocks.

The theory has two limitations. First, the coverage guarantees are marginal and concern one split of a stationary series; they extend to any fixed forecast horizon, but rolling refits and nonstationary series are not covered. Second, the sharper rate of Section 4 is proved for Gaussian linear processes with a constant covariate and rests on the symmetry of their law; Study 2 supports symmetric non-Gaussian innovations numerically, and a proof for them would replace the Hermite expansions of Appendix C.1 by the expansions of the empirical process of linear processes (Wu, 2003).

Reproducibility Statement

Complete proofs are given in Appendices A–C, with the assumptions verified for the examples of the paper in Appendix B. Appendix D specifies every study in enough detail to reproduce it: the models and their simulation, the generation of the random streams, the learners, the block sizes, the number of trajectories, and the computation of every reported quantity and of its Monte Carlo standard error. Appendix E specifies the real data analysis, which uses publicly available data of the New York Independent System Operator.

AI Use Statement

The research question and the approach of this paper are the authors' own. We used generative AI tools to help develop the theoretical models and formulate mathematical claims, to provide critical ingredients for proving these claims, to assist in the writing of the proofs, to design and provide feedback on both the simulations and the real data analysis, and to implement the methods. We have reviewed all AI-assisted work: the authors checked every mathematical claim and every proof, and checked the simulation code against the designs of Appendix D and its output against exact values where these are available. We take responsibility for the final content of this work, including text, claims or artifacts produced with the aid of generative AI.

References

Donald W. K. Andrews. Non-strong mixing autoregressive processes. *Journal of Applied Probability*, 21(4):930–934, 1984.

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*, volume 313 of *Proceedings of Machine Learning Research*, pp. 1–24, 2026.

Rina Foygel Barber, Emmanuel J. Candès, Aaditya Ramdas, and Ryan J. Tibshirani. Conformal prediction beyond exchangeability. *The Annals of Statistics*, 51(2):816–845, 2023. doi: 10.1214/23-AOS2276.

Jan Beran. *Statistics for Long-Memory Processes*. Chapman & Hall, 1994.

Jan Beran and Sucharita Ghosh. Slowly decaying correlations, testing normality, nuisance parameters. *Journal of the American Statistical Association*, 86(415):785–791, 1991. doi: 10.1080/01621459.1991.10475110.

Richard C. Bradley. Basic properties of strong mixing conditions. A survey and some open questions. *Probability Surveys*, 2:107–144, 2005. doi: 10.1214/154957805100000104.

Péter Breuer and Péter Major. Central limit theorems for non-linear functionals of Gaussian fields. *Journal of Multivariate Analysis*, 13(3):425–441, 1983. doi: 10.1016/0047-259X(83)90019-2.

Peter J. Brockwell and Richard A. Davis. *Time Series: Theory and Methods*. Springer Series in Statistics. Springer, New York, second edition, 1991. doi: 10.1007/978-1-4419-0320-4.

Ngai Hang Chan and Shiqing Ling. Residual empirical processes for long and short memory time series. *The Annals of Statistics*, 36(5):2453–2470, 2008. doi: 10.1214/07-AOS543.

Likai Chen and Wei Biao Wu. Concentration inequalities for empirical processes of linear time series. *Journal of Machine Learning Research*, 18(231):1–46, 2018.

Wei-Kuo Chen, Nikos Dafnis, and Grigoris Paouris. Improved Hölder and reverse Hölder inequalities for Gaussian random vectors. *Advances in Mathematics*, 280:643–689, 2015. doi: 10.1016/j.aim.2014.09.029.

Xiaohui Chen, Mengyu Xu, and Wei Biao Wu. Covariance and precision matrix estimation for high-dimensional time series. *The Annals of Statistics*, 41(6):2994–3021, 2013. doi: 10.1214/13-AOS1182.

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, 2018.

Victor Chernozhukov, Kaspar Wüthrich, and Yinchu Zhu. Distributional conformal prediction. *Proceedings of the National Academy of Sciences*, 118(48):e2107794118, 2021. doi: 10.1073/pnas.2107794118.

I. P. Cornfeld, S. V. Fomin, and Ya. G. Sinai. *Ergodic Theory*, volume 245 of *Grundlehren der mathematischen Wissenschaften*. Springer, New York, 1982. doi: 10.1007/978-1-4615-6927-5.

Herold Dehling and Murad S. Taqqu. The empirical process of some long-range dependent sequences with an application to U-statistics. *The Annals of Statistics*, 17(4):1767–1783, 1989. doi: 10.1214/aos/1176347394.

Zhuanxin Ding, Clive W. J. Granger, and Robert F. Engle. A long memory property of stock market returns and a new model. *Journal of Empirical Finance*, 1(1):83–106, 1993.

Isaac Gibbs and Emmanuel J. Candès. Adaptive conformal inference under distribution shift. In *Advances in Neural Information Processing Systems*, volume 34, pp. 1660–1672, 2021.

Alexander Greaves-Tunnell and Zaid Harchaoui. A statistical investigation of long memory in language and music. In *Proceedings of the 36th International Conference on Machine Learning*, volume 97 of *Proceedings of Machine Learning Research*, pp. 2394–2403, 2019.

Stanisław M. S. Halkiewicz. Rolling-origin conformal prediction under local stationarity and weak dependence, 2026. arXiv:2605.08422.

Fang Han and Wei Biao Wu. Probability inequalities for high-dimensional time series under a triangular array framework. In Hoang Pham (ed.), *Springer Handbook of Engineering Statistics*, pp. 849–863. Springer, London, 2nd edition, 2023. doi: 10.1007/978-1-4471-7503-2_41.

Tao Hong and Shu Fan. Probabilistic electric load forecasting: A tutorial review. *International Journal of Forecasting*, 32(3):914–938, 2016. doi: 10.1016/j.ijforecast.2015.11.011.

Hanyang Jiang, Rina Foygel Barber, Ashwin Pananjady, and Yao Xie. Leave a window out: Modifying the jackknife for predictive inference in time series, 2026. arXiv:2605.30292.

Shayan Kiyani, George Pappas, and Hamed Hassani. Length optimization in conformal prediction. In *Advances in Neural Information Processing Systems*, volume 37, 2024.

Hira L. Koul and Donatas Surgailis. Goodness-of-fit testing under long memory. *Journal of Statistical Planning and Inference*, 140(12):3742–3753, 2010. doi: 10.1016/j.jspi.2010.04.039.

Batiste Le Bars and Pierre Humbert. On volume minimization in conformal regression. In *Proceedings of the 42nd International Conference on Machine Learning*, volume 267 of *Proceedings of Machine Learning Research*, pp. 3030–3053, 2025.

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. doi: 10.1080/01621459.2017.1307116.

Will E. Leland, Murad S. Taqqu, Walter Willinger, and Daniel V. Wilson. On the self-similar nature of Ethernet traffic (extended version). *IEEE/ACM Transactions on Networking*, 2(1):1–15, 1994.

Paweł Lorek and Rafał Kulik. Empirical process of residuals for regression models with long memory errors. *Statistics & Probability Letters*, 86:7–16, 2014. doi: 10.1016/j.spl.2013.11.018.

Mehryar Mohri and Afshin Rostamizadeh. Stability bounds for stationary φ-mixing and β-mixing processes. *Journal of Machine Learning Research*, 11(26):789–814, 2010.

New York Independent System Operator. Integrated Real-Time Actual Load. NYISO Market Information System, data set P-58C, 2026a. Retrieved 25 September 2026.

New York Independent System Operator. ISO Load Forecast. NYISO Market Information System, data set P-7, 2026b. Retrieved 25 September 2026.

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.

Harris Papadopoulos, Kostas Proedrou, Volodya Vovk, and Alex Gammerman. Inductive confidence machines for regression. In *Machine Learning: ECML 2002*, volume 2430 of *Lecture Notes in Computer Science*, pp. 345–356. Springer, 2002. doi: 10.1007/3-540-36755-1_29.

C.-K. Peng, S. V. Buldyrev, S. Havlin, M. Simons, H. E. Stanley, and A. L. Goldberger. Mosaic organization of DNA nucleotides. *Physical Review E*, 49(2):1685–1689, 1994. doi: 10.1103/PhysRevE.49.1685.

Yaniv Romano, Evan Patterson, and Emmanuel J. Candès. Conformalized quantile regression. In *Advances in Neural Information Processing Systems*, volume 32, pp. 3543–3553, 2019.

Donald Sarason. Functions of vanishing mean oscillation. *Transactions of the American Mathematical Society*, 207:391–405, 1975. doi: 10.1090/S0002-9947-1975-0377518-3.

Matteo Sesia and Emmanuel J. Candès. A comparison of some conformal quantile regression methods. *Stat*, 9(1):e261, 2020. doi: 10.1002/sta4.261.

Matteo Sesia and Yaniv Romano. Conformal prediction using conditional histograms. In *Advances in Neural Information Processing Systems*, volume 34, pp. 6304–6315, 2021.

Vladimir Vovk, Alexander Gammerman, and Glenn Shafer. *Algorithmic Learning in a Random World*. Springer, New York, 2005. doi: 10.1007/b106715.

Wei Biao Wu. Empirical processes of long-memory sequences. *Bernoulli*, 9(5):809–831, 2003. doi: 10.3150/bj/1066418879.

Wei Biao Wu. On the Bahadur representation of sample quantiles for dependent sequences. *The Annals of Statistics*, 33(4):1934–1963, 2005a. doi: 10.1214/009053605000000291.

Wei Biao Wu. Nonlinear system theory: Another look at dependence. *Proceedings of the National Academy of Sciences*, 102(40):14150–14154, 2005b. doi: 10.1073/pnas.0506715102.

Wei Biao Wu. Asymptotic theory for stationary processes. *Statistics and Its Interface*, 4(2):207–226, 2011.

Wei Biao Wu and Xiaofeng Shao. Limit theorems for iterated random functions. *Journal of Applied Probability*, 41(2):425–436, 2004. doi: 10.1239/jap/1082999076.

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, 2021.

Chen Xu and Yao Xie. Conformal prediction for time series. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(10):11575–11587, 2023. doi: 10.1109/TPAMI.2023.3272339.

Yunzhen Yao, Lie He, and Michael Gastpar. Non-asymptotic analysis of efficiency in conformalized regression. In *International Conference on Learning Representations*, 2026.

Bin Yu. Rates of convergence for empirical processes of stationary mixing sequences. *The Annals of Probability*, 22(1):94–116, 1994. doi: 10.1214/aop/1176988849.

Margaux Zaffran, Olivier Féron, 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, 2022.

Percy S. Zhai, Mladen Kolar, and Wei Biao Wu. Simultaneous inference for covariance and precision matrices of long-range dependent time series. *IEEE Transactions on Information Theory*, 72(6): 4246–4296, 2026. doi: 10.1109/TIT.2026.3685246.

Danna Zhang and Wei Biao Wu. Gaussian approximation for high dimensional time series. *The Annals of Statistics*, 45(5):1895–1919, 2017. doi: 10.1214/16-AOS1512.

Jingyu Zhao, Feiqing Huang, Jia Lv, Yanjie Duan, Zhen Qin, Guodong Li, and Guangjian Tian. Do RNN and LSTM have long memory? In *Proceedings of the 37th International Conference on Machine Learning*, volume 119 of *Proceedings of Machine Learning Research*, pp. 11365–11375, 2020.

Frédéric 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, 2024.

A PROOFS FOR SECTIONS 2 AND 3

This appendix proves the pathwise comparison (1)–(2), the bound on the calibration variance stated before Theorem 1, and Theorems 1–3 together with the convergence statements after Theorem 2. Appendix B verifies the three assumptions for the learners and models of the paper.

The analysis of the calibration step combines known tools. The variance bound for the empirical distribution function of the oracle scores (Lemma 8) uses the method of Wu (2005b), who bounds partial sums of a causal process, its empirical distribution function included, through their projections onto individual innovations and the predictive dependence measure; when this measure is summable, the method gives $\rho_m = m^{-1/2}$, and Proposition 7 extends the bound to polynomial decay, which gives the two slower regimes of ρ_m. The pathwise comparison (1)–(2) is the deterministic comparison also used by Yao et al. (2026) for independent observations, and the two-threshold bound of Appendix A.3 is the Chebyshev bound for an empirical quantile. What is new is the formulation through the predictive dependence measure of the thresholded scores, with the transfer conditions of Proposition 12; the non-asymptotic combination of the two steps, including the window argument of Appendix A.4, which gives coverage without a density for the score at the test time; and the length statements.

A.1 PROOF OF EQUATIONS (1) AND (2)

Proof of (1). For any four real numbers, $|\max(a,b)-\max(c,d)| \le \max(|a-c|,|b-d|)$. Applying this inequality to the two endpoint residuals gives $|\widehat{S}_t^Q - S_t^Q| \le E_n^Q$ at every calibration time and at the test time. The k-th smallest values then differ by at most E_n^Q, also when ties occur: at least k of the oracle scores $S_{n+1}^Q, \dots, S_{n+m}^Q$ are at most $S_{(k)}^Q$, the corresponding fitted scores are at most $S_{(k)}^Q + E_n^Q$, and therefore $\widehat{S}_{(k)}^Q \le S_{(k)}^Q + E_n^Q$; interchanging the fitted and the oracle scores gives the reverse inequality, so $|\widehat{S}_{(k)}^Q - S_{(k)}^Q| \le E_n^Q$.

For each x, write $w(x) = u_\alpha(x) - \ell_\alpha(x) \ge 0$ and $\widehat{w}(x) = \widehat{u}(x) - \widehat{\ell}(x)$. Then $\operatorname{len} C_\alpha^Q(x) = w(x)$ and $\operatorname{len} \widehat{C}_\alpha^Q(x) = [\widehat{w}(x) + 2\widehat{S}_{(k)}^Q]_+$. Since the positive-part map is 1-Lipschitz,

$$\begin{aligned} |[\widehat{w}(x) + 2\widehat{S}_{(k)}^Q]_+ - w(x)| &\le |\widehat{w}(x) - w(x)| + 2|\widehat{S}_{(k)}^Q| \\ &\le 2E_n^Q + 2|\widehat{S}_{(k)}^Q - S_{(k)}^Q| + 2|S_{(k)}^Q| \le 4E_n^Q + 2|S_{(k)}^Q|. \end{aligned}$$

Taking the supremum over x proves the length bound in (1), including the cases of quantile crossing and empty fitted intervals. Since the bound holds simultaneously for all x in the stated domain, it applies to an adjacent, dependent test covariate without any train–test decoupling argument. □

Proof of (2). The proof follows that of (1): $|\widehat{S}_t^M - S_t^M| \le |\widehat{\mu}(X_t) - \mu(X_t)| \le E_n^M$ at every time, so the order-statistic step gives $|\widehat{S}_{(k)}^M - S_{(k)}^M| \le E_n^M$; moreover $\operatorname{len} \widehat{C}_\alpha^M(x) = 2\widehat{S}_{(k)}^M$ and $\operatorname{len} C_\alpha^M(x) = 2q_\alpha$ for every x, and the triangle inequality completes the proof. □

A.2 THE CALIBRATION VARIANCE UNDER FUNCTIONAL DEPENDENCE

This subsection fixes the notation of the remaining proofs and bounds the variance of the empirical distribution function of the oracle scores near the threshold under Assumption 3.

In the remaining appendices, S_t denotes the centered oracle score of the procedure under study, S_t^Q for CQR and $S_t^M - q_\alpha$ for CMR; $\widehat{S}_t$ denotes the corresponding fitted score, $\widehat{S}_t^Q$ or $\widehat{S}_t^M - q_\alpha$; $S_{(k)}$ and $\widehat{S}_{(k)}$ are their k-th smallest values over the calibration block; F is the distribution function of S_t, so that $F(0) = 1-\alpha$ in both cases; and E_n is the training error, E_n^Q or E_n^M. For CMR the centering cancels in the algorithm, whose radius is $\widehat{S}_{(k)}^M = q_\alpha + \widehat{S}_{(k)}$, and $S_{(k)} = S_{(k)}^M - q_\alpha$ is the calibration error of the oracle radius; the predictive dependence measure $\omega(j)$ of (3) is unchanged by the centering. With this notation, the proofs of (1) and (2) give, for both procedures,

$$|\widehat{S}_t - S_t| \le E_n \quad (n < t \le n+m+1), \qquad |\widehat{S}_{(k)} - S_{(k)}| \le E_n, \tag{17}$$

and the length error is at most $4E_n + 2|S_{(k)}|$ for CQR and equal to $2|\widehat{S}_{(k)}| \le 2E_n + 2|S_{(k)}|$ for CMR. The coverage argument of Appendix A.4 involves $2E_n + |S_{(k)}|$ for both procedures, because both the fitted test score and the fitted calibration threshold are perturbed.

Let

$$\widehat{F}_m(s) = \frac{1}{m}\sum_{i=1}^{m} \mathbf{1}\{S_{n+i} \le s\}, \qquad \sigma_m^2 = \sup_{|s|\le r} \mathrm{Var}\widehat{F}_m(s),$$

be the empirical distribution function of the oracle scores on the calibration block and its largest variance near the threshold; stationarity gives $\mathbb{E}\widehat{F}_m(s) = F(s)$. The bound to be proved is $\sigma_m \le C_\sigma \rho_m$ (Lemma 8). Let $\mathcal{P}_j W = \mathbb{E}(W \mid \mathcal{F}_j) - \mathbb{E}(W \mid \mathcal{F}_{j-1})$ denote the projection onto the innovation at time j, and let

$$\pi_j(s) = \|\mathcal{P}_0 \mathbf{1}\{S_j \le s\}\|_2, \qquad j \ge 0,$$

be the projection norm of the thresholded score, which measures how much the innovation ε_0 predicts about the event $\{S_j \le s\}$. Let $J_j(s) = \mathbf{1}\{S_j \le s\}$ and $J_j^*(s) = \mathbf{1}\{S_j^* \le s\}$, the latter computed from the coupled trajectory of Section 2.3, and let $W_j(s) = \mathbb{P}(S_j \le s \mid \mathcal{F}_0)$ and $W_j^*(s) = \mathbb{P}(S_j^* \le s \mid \mathcal{F}_0^*)$ be the two conditional probabilities compared in (3), so that $\omega(j) = \sup_s \|W_j(s) - W_j^*(s)\|_2$. Independence of the replaced innovation, of its copy and of the future innovations gives $W_j(s) = \mathbb{E}\{J_j(s) \mid \mathcal{F}_0, \varepsilon_0'\}$, $W_j^*(s) = \mathbb{E}\{J_j^*(s) \mid \mathcal{F}_0, \varepsilon_0'\}$ and $\mathbb{E}\{W_j^*(s) \mid \mathcal{F}_0\} = \mathbb{E}\{W_j(s) \mid \mathcal{F}_{-1}\}$, and therefore

$$\mathcal{P}_0 J_j = \mathbb{E}(W_j - W_j^* \mid \mathcal{F}_0) = \mathbb{E}(J_j - J_j^* \mid \mathcal{F}_0), \qquad \pi_j(s) = 2^{-1/2}\|W_j(s) - W_j^*(s)\|_2; \quad (18)$$

the second identity holds because $W_j(s)$ and $W_j^*(s)$ are the same function of $\mathcal{F}_{-1}$ and of ε_0 or ε_0', respectively, and are therefore independent and identically distributed given $\mathcal{F}_{-1}$. Assumption 3 thus gives $\sup_s \pi_j(s) \le 2^{-1/2} A(1 \vee j)^{-\gamma}$.

Proposition 6 (Projection envelope)**.** *For every s,*

$$\mathrm{Var}\Big(\sum_{i=1}^{m} \mathbf{1}\{S_{n+i} \le s\}\Big) \le \sum_{j=-\infty}^{m} \Big(\sum_{i=\max(1,j)}^{m} \pi_{i-j}(s)\Big)^2 \le m\Big(\sum_{j\ge 0} \pi_j(s)\Big)^2. \quad (19)$$

If the marginal density of S_t is at most L, then

$$\|W_j(s) - W_j^*(s)\|_2 \le \min\Big\{\{2\min(F(s), 1-F(s))\}^{1/2},\ \sqrt{3}\, L^{q/[2(q+1)]} \delta_q(j)^{q/[2(q+1)]}\Big\}, \quad (20)$$

where $\delta_q(j) = \|S_j - S_j^*\|_q$ is the functional dependence measure of the score.

Proof. Fix s and set $I_t = \mathbf{1}\{S_t \le s\} - F(s)$. As a square-integrable causal function of i.i.d. innovations, I_t has the orthogonal decomposition $I_t = \sum_{j=-\infty}^{t} \mathcal{P}_j I_t$ in L^2; the backward limit of the conditional expectations is zero by triviality of the tail sigma-field of the i.i.d. past. Summing over $t = 1, \dots, m$ and using orthogonality across j gives

$$\Big\|\sum_{t=1}^{m} I_t\Big\|_2^2 = \sum_{j=-\infty}^{m} \Big\|\sum_{t=\max(1,j)}^{m} \mathcal{P}_j I_t\Big\|_2^2 \le \sum_{j=-\infty}^{m} \Big(\sum_{t=\max(1,j)}^{m} \pi_{t-j}(s)\Big)^2,$$

which is the first inequality in (19) after a shift of the time index by stationarity. Regrouping the projections by lag and applying Minkowski's inequality,

$$\Big\|\sum_{t=1}^{m} I_t\Big\|_2 \le \sum_{k\ge 0} \Big\|\sum_{t=1}^{m} \mathcal{P}_{t-k} I_t\Big\|_2 = \sqrt{m} \sum_{k\ge 0} \pi_k(s),$$

where the equality uses orthogonality over the distinct projection times $t - k$; this is the second inequality.

For (20), the representation $W_j(s) - W_j^*(s) = \mathbb{E}\{J_j(s) - J_j^*(s) \mid \mathcal{F}_0, \varepsilon_0'\}$ and Jensen's inequality give $\|W_j(s) - W_j^*(s)\|_2 \le \|J_j - J_j^*\|_2$. The two indicators have identical marginal laws, so their

disagreement probability is at most both $2F(s)$ and $2(1-F(s))$. For every bandwidth $h>0$, disagreement also implies either $|S_j - s| \le h$ or $|S_j - S_j^*| > h$, hence

$$\|J_j - J_j^*\|_2^2 \le 2Lh + \delta_q(j)^q/h^q.$$

For $\delta_q(j)>0$, the choice $h = \{\delta_q(j)^q/L\}^{1/(q+1)}$ makes the right-hand side $3L^{q/(q+1)}\delta_q(j)^{q/(q+1)}$; if the dependence coefficient is zero, let $h \downarrow 0$. Taking square roots and combining the two bounds proves (20). □

Proposition 7. *If* $\sup_s \pi_j(s) \le B(1\vee j)^{-\gamma}$ *for all* $j \ge 0$ *and some* $\gamma > 1/2$*, then* $\sigma_m \le C_\gamma B \rho_m$ *with*

$$\rho_m = \begin{cases} m^{-1/2}, & \gamma > 1, \\ m^{-1/2}\log(m+1), & \gamma = 1, \\ m^{1/2-\gamma}, & 1/2 < \gamma < 1, \end{cases} \tag{21}$$

and C_γ *depending only on* γ*.*

Proof. Since $(1\vee j)^{-\gamma} \le 2^\gamma (j+1)^{-\gamma}$, the envelope is at most $2^\gamma B(j+1)^{-\gamma}$. For projection indices $-m < j \le m$, each inner sum in (19) is at most $2^\gamma B\sum_{k=0}^{2m}(k+1)^{-\gamma}$, and there are at most $2m$ such indices; their total contribution is bounded by CB^2m, $CB^2m\log^2(m+1)$, or $CB^2m^{3-2\gamma}$ in the three regimes $\gamma>1$, $\gamma=1$ and $1/2<\gamma<1$. For $j \le -m$, put $v=-j \ge m$ and bound the inner sum by $2^\gamma Bm(v+1)^{-\gamma}$; this contributes at most $4^\gamma B^2m^2\sum_{v\ge m}(v+1)^{-2\gamma} \le C_\gamma B^2 m^{3-2\gamma}$, which is at most a constant times m for $\gamma>1$ and is absorbed by $m\log^2(m+1)$ for $\gamma=1$. Dividing by m^2 and taking the square root gives the claim. □

Lemma 8. *Under Assumption 3,* $\sigma_m \le C_\sigma \rho_m$ *with* ρ_m *as in* (21) *and* $C_\sigma = C_\gamma A$*, where* C_γ *is the constant of Proposition 7.*

Proof. By (18), Assumption 3 gives $\sup_s \pi_j(s) \le 2^{-1/2}A(1\vee j)^{-\gamma} \le A(1\vee j)^{-\gamma}$ for all $j \ge 0$, and Proposition 7 applies with $B = A$. □

A.3 The Empirical Quantile of the Oracle Scores

This subsection turns the variance bound of Appendix A.2 into bounds on the tail and the mean of the oracle calibration error $S_{(k)}$ (Proposition 9), and fixes the thresholds m_0 and m_1 and the constants C_1 and C_2 of Theorems 1–3.

We write $d_m = k/m - (1-\alpha)$ for the offset of the rank $k = \lceil (m+1)(1-\alpha)\rceil$ from its target, so that $0 \le d_m < 2/m$, and we use the thresholds

$$m_0 = \max\Big\{\frac{4}{f_\alpha r},\ \frac{2(1-\alpha)}{\alpha}\Big\}, \tag{22}$$

for which $m \ge m_0$ implies $d_m/f_\alpha \le r/2$ and $\min(k, m-k+1) \ge m\min(1-\alpha, \alpha/2)$, and

$$m_1 = \min\Big\{m' \ge m_0 :\ \sup_{m\ge m'} t_m \le \frac{r}{2}\Big\}, \qquad t_m = \max\Big\{\frac{4}{mf_\alpha},\ \Big(\frac{16C_\sigma^2\rho_m^2}{L_\alpha f_\alpha^2}\Big)^{1/3}\Big\}, \tag{23}$$

which is finite because $\rho_m \to 0$ under Assumption 3. Indeed $k \ge (m+1)(1-\alpha) \ge m(1-\alpha)$ and $m-k+1 \ge \alpha m - (1-\alpha) \ge \alpha m/2$ when $m \ge 2(1-\alpha)/\alpha$. The constants of Theorems 1–3 are

$$C_1 = \frac{3}{2}\Big(\frac{4L_\alpha C_\sigma}{f_\alpha}\Big)^{2/3} + \frac{4L_\alpha}{f_\alpha}, \tag{24}$$

$$C_2 = \frac{4}{f_\alpha} + \frac{4\sqrt{2}\,C_\sigma}{f_\alpha}\Big(1+\frac{M}{r}\Big), \qquad M = \Big\{\frac{\mathbb{E}(S_0)^2}{\min(1-\alpha,\alpha/2)}\Big\}^{1/2}, \tag{25}$$

where $\mathbb{E}(S_0)^2$ is $\mathbb{E}(S_0^Q)^2$ for CQR and $\mathbb{E}(S_0^M - q_\alpha)^2$ for CMR.

Write $v_m(s) = \mathrm{Var}(\sum_{i=1}^m \mathbf{1}\{S_{n+i} \le s\}) = m^2\mathrm{Var}\{\widehat{F}_m(s)\}$, so that $\sigma_m^2 = \sup_{|s|\le r} v_m(s)/m^2$, and put $F_+(t) = F(t) - (1-\alpha)$ and $F_-(t) = (1-\alpha) - F(-t)$. If $F_+(t) > d_m$ and $F_-(t) + d_m > 0$, then

$$\mathbb{P}(|S_{(k)}| > t) \le \min\Big\{1,\ \frac{v_m(t)}{m^2\{F_+(t)-d_m\}^2} + \frac{v_m(-t)}{m^2\{F_-(t)+d_m\}^2}\Big\}. \tag{26}$$

Indeed, $S_{(k)} > t$ implies $\widehat{F}_m(t) < k/m$, hence $F(t) - \widehat{F}_m(t) > F_+(t) - d_m$; and $S_{(k)} < -t$ implies $\widehat{F}_m(-t) \geq k/m$, hence $\widehat{F}_m(-t) - F(-t) \geq F_-(t) + d_m$. Applying Chebyshev's inequality to each event and adding the two bounds proves (26).

Proposition 9 (Quantile form). *Under Assumption 2(a), put $a_m = d_m/f_\alpha$ and $b_m = \sqrt{2}\sigma_m/f_\alpha$. For $a_m < t \leq r$,*

$$\mathbb{P}(|S_{(k)}| > t) \leq \min\{1,\, b_m^2/(t - a_m)^2\}. \tag{27}$$

If in addition $M_m \geq \|S_{(k)}\|_2$ is a deterministic number and $a_m < r$, then

$$\mathbb{E}|S_{(k)}| \leq a_m + 2b_m + M_m \min\{1, b_m/(r - a_m)\}, \tag{28}$$

and a valid choice is

$$M_m = \left\{ \frac{m\,\mathbb{E}(S_0)^2}{\min(k, m - k + 1)} \right\}^{1/2}. \tag{29}$$

Proof. For $a_m < t \leq r$, Assumption 2(a) gives $F_+(t) - d_m \geq f_\alpha(t - a_m)$ and $F_-(t) + d_m \geq f_\alpha(t - a_m)$, and each numerator in (26) is at most $m^2\sigma_m^2$. This proves (27). For (28), the tail-integral identity gives

$$\begin{aligned}\mathbb{E}\min(|S_{(k)}|, r) = \int_0^r \mathbb{P}(|S_{(k)}| > t)\, dt &\leq a_m + \int_{a_m}^r \min\{1, b_m^2/(t - a_m)^2\}\, dt \\ &\leq a_m + \int_0^\infty \min\{1, b_m^2/u^2\}\, du = a_m + 2b_m,\end{aligned}$$

with the integral interpreted as zero when $b_m = 0$, and the Cauchy–Schwarz inequality bounds $\mathbb{E}(|S_{(k)}| - r)_+ \leq \mathbb{E}[|S_{(k)}|\mathbf{1}\{|S_{(k)}| > r\}]$ by $M_m\mathbb{P}(|S_{(k)}| > r)^{1/2} \leq M_m \min\{1, b_m/(r - a_m)\}$. For (29), if $S_{(k)} \geq 0$, then at least $m - k + 1$ scores have squares at least $(S_{(k)})^2$; if $S_{(k)} < 0$, then at least k scores do; hence $\min(k, m - k + 1)(S_{(k)})^2 \leq \sum_{i=1}^m (S_{n+i})^2$, and taking expectations proves the claim. □

At fixed α and $m \geq m_0$, (29) gives $M_m \leq M$ with the constant M of (25), a bound that does not depend on m; the constants are not uniform as $\alpha \downarrow 0$.

A.4 Proofs of Theorems 1–3

Proof of Theorem 2. For $4/(mf_\alpha) \leq t \leq r$ we have $d_m < 2/m \leq f_\alpha t/2$, so $t - a_m \geq t/2$ and (27) gives $\mathbb{P}(|S_{(k)}| > t) \leq 4b_m^2/t^2 = 8\sigma_m^2/(f_\alpha t)^2 \leq 8C_\sigma^2\rho_m^2/(f_\alpha t)^2$, by Lemma 8. On the event $\{E_n \leq e,\ |S_{(k)}| \leq t\}$, (1) bounds the length error by $4e + 2t$, and a union bound proves (6).

For the expectation, $m \geq m_0$ gives $a_m \leq r/2$, so that $\min\{1, b_m/(r - a_m)\} \leq 2b_m/r$, and (28) with (29) and $M_m \leq M$ gives

$$\mathbb{E}|S_{(k)}| \leq \frac{2}{mf_\alpha} + \frac{2\sqrt{2}\,\sigma_m}{f_\alpha}\Big(1 + \frac{M}{r}\Big) \leq \frac{C_2}{2}\rho_m, \tag{30}$$

where the second inequality uses $\sigma_m \leq C_\sigma\rho_m$ and $1/m \leq \rho_m$, which holds because $\rho_m \geq m^{-1/2}$ for $m \geq 2$ and $m_0 \geq 4$. Taking expectations in (1) proves (7). □

Proof of Theorem 1. Fix $e \geq 0$ and t with $4/(mf_\alpha) \leq t$ and $2e + t \leq r$, and put $\eta = 8C_\sigma^2\rho_m^2/(f_\alpha t)^2$, so that $\mathbb{P}(|S_{(k)}| > t) \leq \eta$ by the first step of the proof of Theorem 2. On the event $\{E_n \leq e,\ |S_{(k)}| \leq t\}$, the inequalities (17) imply

$$\{S_{n+m+1} \leq -2e - t\} \subseteq \{\widehat{S}_{n+m+1} \leq \widehat{S}_{(k)}\} \subseteq \{S_{n+m+1} \leq 2e + t\}.$$

For example, if $S_{n+m+1} \leq -2e - t$, then $\widehat{S}_{n+m+1} \leq -e - t \leq \widehat{S}_{(k)}$, since $\widehat{S}_{(k)} \geq S_{(k)} - e \geq -t - e$; the other inclusion follows from $S_{n+m+1} \leq \widehat{S}_{n+m+1} + e \leq \widehat{S}_{(k)} + e \leq t + 2e$. Removing the exceptional event at a cost of at most $\mathbb{P}(E_n > e) + \eta$ gives

$$F(-2e - t) - \mathbb{P}(E_n > e) - \eta \leq \mathbb{P}\{Y_{n+m+1} \in \widehat{C}_\alpha^Q(X_{n+m+1})\} \leq F(2e + t) + \mathbb{P}(E_n > e) + \eta,$$

and the upper bounds of Assumption 2(b), valid because $2e + t \le r$, give

$$\left|\mathbb{P}\{Y_{n+m+1} \in \widehat{C}^Q_\alpha(X_{n+m+1})\} - (1-\alpha)\right| \le L_\alpha(2e+t) + \mathbb{P}(E_n > e) + \frac{8C_\sigma^2\rho_m^2}{f_\alpha^2 t^2}.$$

Now take $e = e_n \le r/4$ and $t = t_m$ as in (23); for $m \ge m_1$ we have $t_m \le r/2$, so that $2e_n + t_m \le r$, and $t_m \ge 4/(mf_\alpha)$ by construction. Writing t^* for the second term in the maximum defining t_m, $8C_\sigma^2\rho_m^2/(f_\alpha^2 t^{*2}) = L_\alpha t^*/2$, and since $t_m \ge t^*$,

$$L_\alpha t_m + \frac{8C_\sigma^2\rho_m^2}{f_\alpha^2 t_m^2} \le L_\alpha t^* + \frac{4L_\alpha}{mf_\alpha} + \frac{L_\alpha t^*}{2} = \frac{3}{2}\Big(\frac{4L_\alpha C_\sigma \rho_m}{f_\alpha}\Big)^{2/3} + \frac{4L_\alpha}{mf_\alpha}.$$

With $1/m \le \rho_m^{2/3}$, valid for $m \ge 2$, this proves (4) with the constant C_1 of (24). The exponent $2/3$ comes from the window: the coverage error is at most $L_\alpha t$ on a window of half-width t around the threshold, plus the probability of order ρ_m^2/t^2 that the calibrated threshold leaves the window, and $t \asymp \rho_m^{2/3}$ balances the two terms; when the score at the test time has a bounded density given the past, no window is needed and (5) retains the full rate.

For (5), let $D = 2E_n + |S_{(k)}|$, a nonnegative random variable measurable with respect to $\mathcal{F}_{n+m}$ and the training randomness, which is independent of the series, so that the conditional law of S_{n+m+1} given both has the same density bound L_{pred}. The same pathwise inclusions hold with D in place of $2e + t$ on every realization. Denote $A = \{\widehat{S}_{n+m+1} \le \widehat{S}_{(k)}\}$ and $B = \{S_{n+m+1} \le 0\}$. Then $A \setminus B \subseteq \{0 < S_{n+m+1} \le D\}$ and $B \setminus A \subseteq \{-D < S_{n+m+1} \le 0\}$, and the conditional density bound gives probability at most $L_{\mathrm{pred}}\mathbb{E}D$ for each. Since $|\mathbb{P}(A) - \mathbb{P}(B)| = |\mathbb{P}(A \setminus B) - \mathbb{P}(B \setminus A)|$ is at most the larger of the two nonnegative terms, it is at most $L_{\mathrm{pred}}\mathbb{E}D$, and $\mathbb{P}(B) = F(0) = 1 - \alpha$. The bound (30) completes the proof. □

The predictive-density condition of (5) holds, for example, when X_{n+m+1} is $\mathcal{F}_{n+m}$-measurable and $Y_{n+m+1} = a + \sigma\epsilon$, where a and $\sigma \ge \sigma_0 > 0$ are $\mathcal{F}_{n+m}$-measurable and ϵ is independent of the past with density at most M_ϵ: conditional on the past, both oracle CQR endpoints are fixed and the score density is at most $2M_\epsilon/\sigma_0$, so one can take $L_{\mathrm{pred}} = 2M_\epsilon/\sigma_0$. The condition concerns the predictive law and is stronger than a density bound for the stationary marginal law; a process with a positive probability of repeating its preceding value does not satisfy it, and for such a process only (4) applies.

Proof of Theorem 3. With the centered scores of Appendix A.2, (17) gives $|\widehat{S}_{(k)} - S_{(k)}| \le E_n$ and $|\widehat{S}^M_{(k)} - q_\alpha| = |\widehat{S}_{(k)}| \le E_n + |S_{(k)}|$, and Assumptions 2 and 3 for CMR are the assumptions used above for the centered score $S_t = S_t^M - q_\alpha$. The proofs of Theorems 1 and 2 then apply verbatim: the coverage argument with the event $\{Y_{n+m+1} \in \widehat{C}^M_\alpha(X_{n+m+1})\} = \{\widehat{S}_{n+m+1} \le \widehat{S}_{(k)}\}$ and $D = 2E_n + |S_{(k)}|$ as before, which proves (8) and (9), and the length argument with $2E_n + 2|S_{(k)}|$ in place of $4E_n + 2|S_{(k)}|$, which proves (10) and (11) with (30). □

The proofs use the test observation only through its oracle score: through the marginal distribution function F, which by stationarity is the same at every time, and, for (5) and (9), through the density bound of its conditional law given $\mathcal{F}_{n+m}$. For a fixed horizon $h \ge 1$, the conditional law of S_{n+m+h} given $\mathcal{F}_{n+m}$ has the density $\mathbb{E}\{p_{n+m+h}(\cdot) \mid \mathcal{F}_{n+m}\}$, where p_{n+m+h} is the conditional density of S_{n+m+h} given $\mathcal{F}_{n+m+h-1}$ and is at most L_{pred} by stationarity. The coverage bounds of Theorems 1 and 3 therefore hold, with the same constants, for Y_{n+m+h} and the interval evaluated at X_{n+m+h}.

The convergence statements after Theorem 2 are the following: under Assumptions 1–3, as $n \to \infty$ and $m \to \infty$ in any manner,

$$\mathbb{P}\{Y_{n+m+1} \in \widehat{C}^Q_\alpha(X_{n+m+1})\} \to 1 - \alpha \quad \text{and} \quad \sup_x |\operatorname{len}\widehat{C}^Q_\alpha(x) - \operatorname{len} C^Q_\alpha(x)| \xrightarrow{p} 0, \tag{31}$$

and for CMR the coverage tends to $1 - \alpha$ and $\sup_x |\operatorname{len}\widehat{C}^M_\alpha(x) - \operatorname{len} C^M_\alpha(x)| = 2|\widehat{S}^M_{(k)} - q_\alpha| \to 0$ in probability. For (31), since $\gamma > 1/2$, $\rho_m \to 0$, and $e_n \to 0$, so that $e_n \le r/4$ and $m \ge m_1$ for all large n and m, and the right-hand side of (4) tends to zero; for the length, (6) with $e = e_n$ and any fixed $t \in (0, r]$ has a right-hand side that tends to zero and a threshold $4e_n + 2t \to 2t$, and t is arbitrary. The same argument applied to (8) and (10) gives the statements for CMR. Neither statement relates the growth of n to that of m.

B THE ASSUMPTIONS: VERIFICATION AND EXAMPLES

This appendix verifies Assumption 1 for linear quantile regression and extends Theorems 1–3 to learners whose fitted endpoints converge to a limit other than the oracle endpoints, verifies Assumption 2 from a density condition on the response and Assumption 3 from the functional dependence measure of the scores, then verifies the four examples of Section 3 and compares the resulting guarantees with those formulated through mixing.

B.1 ASSUMPTION 1 AND THE LEARNER

Assumption 1 concerns the learner only. We verify it for the parametric quantile-regression estimator computed from all n training observations, without any rate of decay of the dependence: the only property of the series that is used is ergodicity, which a stationary causal functional $Z_t = G(\dots, \varepsilon_{t-1}, \varepsilon_t)$ of i.i.d. innovations always has, because the shift on the innovation sequence is a Bernoulli shift, hence ergodic, and the process is a factor of it (Cornfeld et al., 1982).

Let $\varrho_\tau(v) = v(\tau - \mathbf{1}\{v < 0\})$ be the pinball loss at level $\tau \in (0,1)$, let $R_\tau(\theta) = \mathbb{E}\varrho_\tau(Y_t - X_t^\top\theta)$ be the population risk, and let $\widehat{R}_{n,\tau}(\theta) = n^{-1}\sum_{t=1}^n \varrho_\tau(Y_t - X_t^\top\theta)$ be the training risk.

Lemma 10 (Consistency of quantile regression on an ergodic series)**.** *Suppose that $\|X_t\| \le B_x$ almost surely, $\mathbb{E}|Y_t| < \infty$, and that for $\tau \in \{\alpha/2, 1-\alpha/2\}$ the conditional τ-quantile of Y_t given $X_t = x$ is $x^\top\theta_\tau^*$ on the covariate domain, where θ_τ^* is the unique minimizer of R_τ over a compact set $\Theta \subset \mathbb{R}^p$. Let $\widehat{\theta}_\tau \in \Theta$ satisfy $\widehat{R}_{n,\tau}(\widehat{\theta}_\tau) \le \inf_{\theta\in\Theta}\widehat{R}_{n,\tau}(\theta) + o_n$ with $o_n \to 0$ in probability. Then $\widehat{\theta}_\tau \to \theta_\tau^*$ in probability, and the fitted endpoints $\widehat{\ell}(x) = x^\top\widehat{\theta}_{\alpha/2}$ and $\widehat{u}(x) = x^\top\widehat{\theta}_{1-\alpha/2}$ satisfy $E_n^Q \le B_x \max_\tau \|\widehat{\theta}_\tau - \theta_\tau^*\| \to 0$ in probability, so that Assumption 1 holds.*

Proof. Fix τ. Since $|\varrho_\tau(v) - \varrho_\tau(v')| \le |v - v'|$, the map $\theta \mapsto \varrho_\tau(Y_t - X_t^\top\theta)$ is B_x-Lipschitz on Θ for every realization, so R_τ and $\widehat{R}_{n,\tau}$ are B_x-Lipschitz, and R_τ is finite because $\mathbb{E}|Y_t| < \infty$ and Θ is bounded. For each fixed θ, the ergodic theorem gives $\widehat{R}_{n,\tau}(\theta) \to R_\tau(\theta)$ almost surely. Given $\epsilon > 0$, let $\theta_1, \dots, \theta_J$ be a finite $\epsilon/(3B_x)$-net of Θ; on the almost sure event on which the convergence holds at every θ_j, the Lipschitz property gives $\sup_{\theta\in\Theta}|\widehat{R}_{n,\tau}(\theta) - R_\tau(\theta)| \le \epsilon$ for all large n. Hence $\sup_\Theta|\widehat{R}_{n,\tau} - R_\tau| \to 0$ almost surely.

Now $R_\tau(\widehat{\theta}_\tau) \le \widehat{R}_{n,\tau}(\widehat{\theta}_\tau) + \sup_\Theta|\widehat{R}_{n,\tau} - R_\tau| \le \widehat{R}_{n,\tau}(\theta_\tau^*) + o_n + \sup_\Theta|\widehat{R}_{n,\tau} - R_\tau| \le R_\tau(\theta_\tau^*) + o_n + 2\sup_\Theta|\widehat{R}_{n,\tau} - R_\tau|$, so $R_\tau(\widehat{\theta}_\tau) - R_\tau(\theta_\tau^*) \to 0$ in probability. Since R_τ is continuous on the compact set Θ and θ_τ^* is its unique minimizer, for every $\eta > 0$ the number $\inf\{R_\tau(\theta) - R_\tau(\theta_\tau^*) : \theta \in \Theta,\ \|\theta - \theta_\tau^*\| \ge \eta\}$ is positive, and therefore $\mathbb{P}(\|\widehat{\theta}_\tau - \theta_\tau^*\| \ge \eta) \to 0$. Finally, on the covariate domain, $|x^\top\widehat{\theta}_\tau - x^\top\theta_\tau^*| \le B_x\|\widehat{\theta}_\tau - \theta_\tau^*\|$, which gives the bound on E_n^Q; the sequences e_n and ξ_n of Assumption 1 can be taken as $e_n = \eta_n$ and $\xi_n = \mathbb{P}(B_x \max_\tau \|\widehat{\theta}_\tau - \theta_\tau^*\| > \eta_n)$ for any $\eta_n \downarrow 0$ slowly enough that $\xi_n \to 0$. □

Uniqueness of the minimizer holds, for example, when the conditional density of Y_t given X_t is positive at $x^\top\theta_\tau^*$ and $\mathbb{E}X_tX_t^\top$ is positive definite.

B.2 MISSPECIFIED LEARNERS

This subsection extends Theorems 1–3 to a learner whose fitted endpoints converge to a limit other than the oracle endpoints, as those of linear quantile regression do when the conditional quantiles are not linear in the covariates. The calibrated interval then keeps its coverage and approaches the limit interval adjusted to marginal coverage $1-\alpha$, whereas the fitted interval itself misses the nominal level unless its limit already has marginal coverage $1-\alpha$.

Let ℓ_∞ and u_∞ be measurable functions on the covariate domain, let $s_\infty(x,y) = \max\{\ell_\infty(x) - y,\ y - u_\infty(x)\}$, let q_∞ be the $(1-\alpha)$-quantile of $s_\infty(X_t, Y_t)$ under the stationary law, and put

$$C_\infty(x) = [\ell_\infty(x) - q_\infty,\ u_\infty(x) + q_\infty], \qquad E_n^\infty = \max\{\|\widehat{\ell} - \ell_\infty\|_\infty, \|\widehat{u} - u_\infty\|_\infty\}.$$

Since Y_t lies in $[\ell_\infty(X_t) - c,\ u_\infty(X_t) + c]$ exactly when $s_\infty(X_t, Y_t) \le c$, the interval $C_\infty(x)$ is the shortest interval of this form with marginal coverage at least $1-\alpha$; for the oracle endpoints, $q_\infty = 0$ and $C_\infty = C_\alpha^Q$.

Proposition 11 (Misspecified learners)**.** *Suppose that $u_\infty(x) - \ell_\infty(x) + 2q_\infty \ge 0$ for every x in the covariate domain, that Assumption 1 holds with E_n^∞ in place of E_n, and that Assumptions 2 and 3 hold for the centered limit score $S_t = s_\infty(X_t, Y_t) - q_\infty$. Then Theorems 1 and 2 hold with S_t, E_n^∞ and C_∞ in place of S_t^Q, E_n^Q and C_α^Q. Moreover, if $q_\infty \neq 0$, then for all n with $e_n \le \min(|q_\infty|, r)/2$ the coverage of the fitted interval satisfies*

$$\left|\mathbb{P}\{\widehat{\ell}(X_{n+m+1}) \le Y_{n+m+1} \le \widehat{u}(X_{n+m+1})\} - (1-\alpha)\right| \ge \frac{f_\alpha \min(|q_\infty|, r)}{2} - \xi_n.$$

Proof. Put $\widehat{S}_t = \widehat{S}_t^Q - q_\infty$. The inequality for maxima in the proof of (1) gives $|\widehat{S}_t - S_t| \le E_n^\infty$ for $n < t \le n+m+1$, so that (17) holds with E_n^∞ in place of E_n. The distribution function F of S_t satisfies $F(0) = 1-\alpha$, by the definition of q_∞ and the continuity of F at 0 required by Assumption 2. The coverage event is $\{\widehat{S}_{n+m+1} \le \widehat{S}_{(k)}\}$, as before, because subtracting q_∞ from every fitted score does not change the calibrated interval. The proof of Theorem 1 uses the oracle endpoints only through these three facts, and it applies verbatim. For the length, put $\widehat{w}(x) = \widehat{u}(x) - \widehat{\ell}(x)$ and $w_\infty(x) = u_\infty(x) - \ell_\infty(x)$, so that $\operatorname{len} \widehat{C}_\alpha^Q(x) = [\widehat{w}(x) + 2q_\infty + 2\widehat{S}_{(k)}]_+$ and $\operatorname{len} C_\infty(x) = w_\infty(x) + 2q_\infty \ge 0$. Since the positive-part map is 1-Lipschitz,

$$|\operatorname{len} \widehat{C}_\alpha^Q(x) - \operatorname{len} C_\infty(x)| \le |\widehat{w}(x) - w_\infty(x)| + 2|\widehat{S}_{(k)}| \le 2E_n^\infty + 2\big(E_n^\infty + |S_{(k)}|\big),$$

which is the length bound of (1) with E_n^∞ and C_∞ in the roles of E_n^Q and C_α^Q, and the proof of Theorem 2 applies verbatim.

For the fitted interval, Y_{n+m+1} lies between $\widehat{\ell}(X_{n+m+1})$ and $\widehat{u}(X_{n+m+1})$ exactly when $\widehat{S}_{n+m+1}^Q \le 0$, that is, when $\widehat{S}_{n+m+1} \le -q_\infty$. Put $\delta = \min(|q_\infty|, r)/2$, so that $e_n \le \delta$. If $q_\infty > 0$, this event and $\{E_n^\infty \le e_n\}$ imply $S_{n+m+1} \le -q_\infty + e_n \le -\delta$, so that the coverage is at most $F(-\delta) + \xi_n \le 1-\alpha-f_\alpha\delta+\xi_n$ by Assumption 2(a). If $q_\infty < 0$, the events $\{S_{n+m+1} \le |q_\infty| - e_n\}$ and $\{E_n^\infty \le e_n\}$ together imply it, and $|q_\infty| - e_n \ge \delta$, so that the coverage is at least $F(\delta) - \xi_n \ge 1-\alpha+f_\alpha\delta-\xi_n$. □

The same argument applies to CMR: for a limit center μ_∞, the $(1-\alpha)$-quantile q_∞ of $|Y_t - \mu_\infty(X_t)|$ and $E_n^\infty = \|\widehat{\mu} - \mu_\infty\|_\infty$, Theorem 3 holds with $[\mu_\infty(x) - q_\infty,\ \mu_\infty(x) + q_\infty]$ in place of $C_\alpha^M(x)$, provided that Assumption 1 holds for E_n^∞ and Assumptions 2 and 3 hold for the centered score $|Y_t - \mu_\infty(X_t)| - q_\infty$.

For linear quantile regression, the proof of Lemma 10 uses correct specification only in its last step: without it, and with θ_τ^* the unique minimizer of R_τ over Θ as before, $\widehat{\theta}_\tau \to \theta_\tau^*$ in probability, and $E_n^\infty \le B_x \max_\tau \|\widehat{\theta}_\tau - \theta_\tau^*\| \to 0$ in probability for the limit pair $\ell_\infty(x) = x^\top \theta_{\alpha/2}^*$ and $u_\infty(x) = x^\top \theta_{1-\alpha/2}^*$. Assumption 1 therefore holds for E_n^∞ whether or not the conditional quantiles are linear in the covariates, and Assumptions 2 and 3 for the limit score are verified as in Appendices B.3 and B.4, with the endpoints of $C_\infty(x)$ in place of the oracle endpoints. The length of the calibrated interval thus approaches that of C_∞. When the covariates include a constant and θ_τ^* is an interior point of Θ, the first-order condition for the intercept gives $\mathbb{P}(Y_t \le X_t^\top \theta_\tau^*) = \tau$ for continuous conditional laws, so that $q_\infty = 0$ if $\ell_\infty \le u_\infty$: the fitted interval then attains the nominal level in the limit, and calibration adds the non-asymptotic coverage bound of Theorem 1. When $q_\infty \neq 0$, as can happen without a constant, calibration restores the nominal level that the fitted interval misses.

B.3 Assumption 2 from the Response Density

This subsection derives Assumption 2 for CQR from bounds on the conditional density of the response near the oracle endpoints. Assume $u_\alpha(x) - \ell_\alpha(x) \ge 2r$. For $|s| \le r$,

$$F^Q(s) = \mathbb{E}\big[F_{Y|X}(u_\alpha(X) + s \mid X) - F_{Y|X}(\ell_\alpha(X) - s \mid X)\big].$$

If the conditional density lies between f_0 and L_0 on the two radius-r endpoint neighborhoods, then for $0 \le t \le r$,

$$F^Q(t) - F^Q(0) = \mathbb{E}\bigg[\int_{u_\alpha(X)}^{u_\alpha(X)+t} f_{Y|X}(v \mid X)\, dv + \int_{\ell_\alpha(X)-t}^{\ell_\alpha(X)} f_{Y|X}(v \mid X)\, dv\bigg],$$

and each integral lies between $f_0 t$ and $L_0 t$. The same reasoning applies to $F^Q(0) - F^Q(-t)$. Hence Assumption 2 holds with $f_\alpha = 2f_0$ and $L_\alpha = 2L_0$. If the conditional density is at most L_0 everywhere, then $F^Q(s)$ is, for every s, the expectation of the positive part of $F_{Y|X}(u_\alpha(X) + s \mid X) - F_{Y|X}(\ell_\alpha(X) - s \mid X)$, a nondecreasing function of s with derivative at most $2L_0$, so that the marginal density of S_t^Q is at most $L = 2L_0$, the density bound of Proposition 12(a).

B.4 Assumption 3 from the Functional Dependence Measure

This subsection gives two sufficient conditions for Assumption 3 in terms of the functional dependence measure of the scores (Proposition 12) and bounds this measure through the dependence of the observations. In the notation of Appendix A.2, $\omega(j) = \sup_s \|W_j(s) - W_j^*(s)\|_2$ with $W_j(s) - W_j^*(s) = \mathbb{E}\{J_j(s) - J_j^*(s) \mid \mathcal{F}_0, \varepsilon_0'\}$, and $\delta_q(j) = \|S_j - S_j^*\|_q$ is the functional dependence measure of the score.

Proposition 12 (Sufficient conditions)**.** *Let $A' > 0$ and $\gamma' > 0$.*

- *(a) If the marginal density of S_t is at most L and $\delta_q(j) \le A'(1 \vee j)^{-\gamma'}$ for all $j \ge 0$ and some $q \ge 1$, then $\omega(j) \le \sqrt{3}(LA')^{q/(2q+2)}(1 \vee j)^{-\gamma' q/(2q+2)}$ for all $j \ge 0$, so that Assumption 3 holds with $\gamma = \gamma' q/(2q+2)$ whenever $\gamma' q > q + 1$.*
- *(b) Suppose $S_t = H(V_t, \varepsilon_t)$, where V_t is $\mathcal{F}_{t-1}$-measurable, and that $K_s(v) = \mathbb{P}\{H(v, \varepsilon_t) \le s\}$ is L_H-Lipschitz in v for every s. Then $\|W_j(s) - W_j^*(s)\|_2 \le L_H \|V_j - V_j^*\|_2$ for every $j \ge 1$ and every s. Consequently, if $\|V_j - V_j^*\|_2 \le A'(1 \vee j)^{-\gamma'}$ for all $j \ge 1$ and $\gamma' > 1/2$, then Assumption 3 holds with $\gamma = \gamma'$ and $A = \max(2^{-1/2}, L_H A')$.*

Proof. Part (a) is (20). For part (b), fix $j \ge 1$ and s. Conditional on $\mathcal{F}_{j-1}$ and ε_0', the common current innovation ε_j is fresh, so that $\mathbb{E}\{J_j(s) \mid \mathcal{F}_{j-1}, \varepsilon_0'\} = K_s(V_j)$ and $\mathbb{E}\{J_j^*(s) \mid \mathcal{F}_{j-1}, \varepsilon_0'\} = K_s(V_j^*)$. Since $j \ge 1$, the tower property gives

$$W_j(s) - W_j^*(s) = \mathbb{E}\{K_s(V_j) - K_s(V_j^*) \mid \mathcal{F}_0, \varepsilon_0'\},$$

and conditional Jensen's inequality together with the Lipschitz assumption gives $\|W_j(s) - W_j^*(s)\|_2 \le L_H \|V_j - V_j^*\|_2$. The requirement $j \ge 1$ ensures that the innovation being integrated out is not the innovation being replaced; at $j = 0$, $W_0(s) = J_0(s)$ and $W_0^*(s) = J_0^*(s)$ are independent and identically distributed given $\mathcal{F}_{-1}$, so that $\|W_0(s) - W_0^*(s)\|_2^2 = 2\mathbb{E}\{p(1-p)\} \le 1/2$ with $p = \mathbb{P}(S_0 \le s \mid \mathcal{F}_{-1})$, which is the reason for the constant $A = \max(2^{-1/2}, L_H A')$. □

Part (a) requires only a bounded marginal density of the score, at the price of the power $q/(2q+2)$; since $\gamma = \gamma' q/(2q+2) > 1/2$ forces $\gamma' > 1 + 1/q$, it does not admit long memory, and it is the route for series whose conditional next-step law has atoms, such as the Bernoulli autoregression of Appendix B.5. Part (b) applies when the conditional distribution function of the score given the past, $K_s(V_t)$, is a Lipschitz function of the part V_t that the past determines; the decay exponent of $\|V_j - V_j^*\|_2$ then transfers to $\omega(j)$ without loss. This is the structure of a location model $Y_t = m_t + a_0 \varepsilon_t$ and of a scale model $Y_t = \sigma_t \varepsilon_t$ with past-determined m_t or σ_t (Appendix B.5), for which $\|V_j - V_j^*\|_2$ is bounded by the functional dependence measure of m_t or σ_t plus a Lipschitz constant times that of X_t; for a linear process the exponent is the memory exponent β, so part (b) admits long memory.

When the oracle endpoints are Lipschitz, the dependence measure of the scores is controlled by that of the observations. For the CQR score, the inequality for maxima used in the proof of (1) gives

$$|S_t^Q - S_t^{Q*}| \le |Y_t - Y_t^*| + \max\{|\ell_\alpha(X_t) - \ell_\alpha(X_t^*)|, |u_\alpha(X_t) - u_\alpha(X_t^*)|\},$$

and Minkowski's inequality proves $\delta_q^Q(j) \le \|Y_j - Y_j^*\|_q + K_0 \|X_j - X_j^*\|_q$ for K_0-Lipschitz endpoints; for CMR the corresponding constant is the Lipschitz constant of μ. For a stationary recursion $Z_t = H(Z_{t-1}, \varepsilon_t)$, a sufficient contraction condition is $\|H(z, \varepsilon_t) - H(z', \varepsilon_t)\|_q \le \lambda_H \|z - z'\|$ for deterministic z, z' and some $0 < \lambda_H < 1$, with the norm inside the expectation interpreted in the state space. If a stationary L^q solution exists, iterating the coupled recursion gives $\delta_q^Z(j) \le \lambda_H^j \delta_q^Z(0)$ for the observations; this is the geometric moment-contraction route of Wu & Shao (2004), and the contraction has to be checked for the particular recursion.

B.5 Verification of the Examples of Section 3

This subsection verifies the assumptions for the four examples of Section 3.

A linear process with discrete innovations. Let B_t be i.i.d. Bernoulli$(1/2)$ and $W_t = \sum_{j\geq 0} 2^{-j} B_{t-j} = W_{t-1}/2 + B_t$, which is uniform on $[0,2)$ by the binary expansion of $W_t/2$. Changing B_0 changes W_j by exactly $2^{-j}(B_0 - B_0')$, so $\|W_j - W_j^*\|_q = 2^{-j}\|B_0 - B_0'\|_q \leq 2^{-j}$. To verify the failure of strong mixing directly, let $\alpha_{\mathrm{mix}}(h)$ be the supremum of $|\mathbb{P}(A\cap D) - \mathbb{P}(A)\mathbb{P}(D)|$ over $A \in \sigma(W_s : s \leq 0)$ and $D \in \sigma(W_s : s \geq h)$. The recursion gives $2^h W_h = 2\sum_{j=1}^{h} 2^{j-1} B_j + W_0$, so W_0 is the remainder of $2^h W_h$ modulo 2 and is a measurable function of W_h; taking $A = \{W_0 \leq 1\}$ and D the same event expressed through W_h gives $\alpha_{\mathrm{mix}}(h) \geq 1/4$ for every h, which is the argument of Andrews (1984). In the location model $Y_t = \mu + W_t$ with a constant covariate, the conditional median is $\mu + 1$, the oracle CMR score $S_t^M = |W_t - 1|$ is uniform on $[0,1]$ with $q_\alpha = 1-\alpha$, so the centered distribution function F^M has slope one on $[-r, r]$, and Assumption 2 holds with $f_\alpha = L_\alpha = 1$ and $r = \min(\alpha, 1-\alpha)$; the oracle CQR endpoints are $\mu + \alpha$ and $\mu + 2 - \alpha$, so that the CQR score $S_t^Q = |W_t - 1| - (1-\alpha)$ coincides with the centered CMR score, and what follows applies to CQR as well. The score has marginal density one, and $\delta_q^M(j) \leq \|W_j - W_j^*\|_q \leq 2^{-j}$ because the absolute value is 1-Lipschitz, so Proposition 12(a) gives $\omega^M(j) \leq \sqrt{3}\, 2^{-jq/(2q+2)}$, and Assumption 3 holds with every $\gamma > 1/2$, so that $\rho_m = m^{-1/2}$. The empirical median, and for CQR the empirical $\alpha/2$- and $(1-\alpha/2)$-quantiles, of the training block are consistent by ergodicity (Lemma 10 with a constant covariate), so Assumption 1 holds for these learners.

Gaussian long memory. Under (12), the oracle CMR score in $Y_t = \mu + G_t$ is $S_t^M = |G_t|$, with $q_\alpha = z_\alpha$; the centered score $S_t = |G_t| - z_\alpha$ has distribution function $F^M(s) = 2\Phi(z_\alpha + s) - 1$ and density $2\phi(z_\alpha + s)$, and on $|s| \leq r = z_\alpha/2$ the density lies between $2\phi(3z_\alpha/2)$ and $2\phi(z_\alpha/2)$, which gives Assumption 2. Write $S_t = H(V_t, \varepsilon_t)$ with $V_t = \sum_{j\geq 1} a_j \varepsilon_{t-j}$ and $H(v,\varepsilon) = |v + a_0\varepsilon| - z_\alpha$. Then $K_s(v) = \Phi\{(z_\alpha + s - v)/|a_0|\} - \Phi\{(-z_\alpha - s - v)/|a_0|\}$ is Lipschitz in v with constant $2\phi(0)/|a_0|$, and $\|V_j - V_j^*\|_2 = |a_j|\,\|\varepsilon_0 - \varepsilon_0'\|_2 = \sqrt{2}|a_j| \leq \sqrt{2}A(1+j)^{-\beta}$, where A is the coefficient bound of (12), so Proposition 12(b) gives Assumption 3 with $\gamma = \beta > 1/2$ and the constant $\max\{2^{-1/2}, 2\sqrt{2}\phi(0)A/|a_0|\}$ in the role of the constant of Assumption 3. The same holds for the CQR score with the constant oracle endpoints $\mu \mp z_\alpha$, which coincides with the centered CMR score. The functional dependence of the observations themselves is $\|G_j - G_j^*\|_q = |a_j|\,\|\varepsilon_0 - \varepsilon_0'\|_q$ for every finite q, and the failure of strong mixing is proved in Appendix B.6. For the sample mean, $E_n^M = |\bar{G}_n|$ with $\bar{G}_n = n^{-1}\sum_{t\leq n} G_t$ and $\mathbb{E}|\bar{G}_n| = (2v_n/\pi)^{1/2} \leq Cn^{1/2-\beta}$, where $v_n = \mathrm{Var}(\bar{G}_n)$, by (33) and (34), and Markov's inequality gives Assumption 1 with, for example, $e_n = n^{(1/2-\beta)/2}$ and $\xi_n = Cn^{(1/2-\beta)/2}$.

Long memory in the scale. Let $Y_t = \sigma_t \varepsilon_t$ with i.i.d. innovations ε_t whose law is symmetric with $\mathbb{E}\varepsilon_t^2 = 1$ and with a bounded, continuous and positive density f_ε satisfying $\sup_{x\geq 0} x f_\varepsilon(x) < \infty$, as for the normal law, and let the volatility be $\sigma_t = g(h_t)$ with $h_t = \sum_{j\geq 1} a_j \varepsilon_{t-j}$ a linear process of the past innovations whose coefficients satisfy $|a_j| \leq A(1+j)^{-\beta}$ for some $1/2 < \beta < 1$, and g an L_g-Lipschitz function with values in $[\sigma_{\min}, \sigma_{\max}]$, $\sigma_{\min} > 0$. With a constant covariate the conditional median of Y_t is 0, and the oracle CMR score is $S_t^M = \sigma_t|\varepsilon_t|$; its marginal density at s is $\mathbb{E}\{f_{|\varepsilon|}(s/\sigma_t)/\sigma_t\}$, which, for $|s - q_\alpha| \leq r$ with some $r < q_\alpha$, lies between $\sigma_{\max}^{-1}$ and $\sigma_{\min}^{-1}$ times the infimum and the supremum of $f_{|\varepsilon|}$ over the compact set of arguments s/σ with $|s - q_\alpha| \leq r$ and $\sigma \in [\sigma_{\min}, \sigma_{\max}]$, so that Assumption 2 holds. The score has the form $H(V_t, \varepsilon_t) = V_t|\varepsilon_t|$ of Proposition 12(b) with $V_t = \sigma_t$, and $K_s(v) = F_{|\varepsilon|}(s/v)$ has $|\partial_v K_s(v)| = (s/v^2) f_{|\varepsilon|}(s/v) \leq \sigma_{\min}^{-1} \sup_{x\geq 0} x f_{|\varepsilon|}(x) = L_H$ uniformly in s for $v \geq \sigma_{\min}$, while $\|V_j - V_j^*\|_2 \leq L_g \|h_j - h_j^*\|_2 = \sqrt{2} L_g |a_j| \leq \sqrt{2} L_g A(1+j)^{-\beta}$. Hence Assumption 3 holds with $\gamma = \beta$, the score has a finite second moment since $\sigma_t \leq \sigma_{\max}$, and the theorems apply with $\rho_m = m^{1/2-\beta}$: the current innovation smooths the score through the scale exactly as it does through the location in the previous example. The empirical median and the empirical quantiles of the training block satisfy Assumption 1 by ergodicity (Lemma 10 with a constant covariate). Since the law of Y_t is symmetric, the CQR score

with the constant oracle endpoints $\mp q_\alpha$ coincides with the centered CMR score, as in the previous example.

Nonlinear recursions. Under the contraction condition of Appendix B.4, the observations satisfy $\delta_q^Z(j) \le \lambda_H^j \delta_q^Z(0)$, and the scores inherit geometric decay when the oracle endpoints are Lipschitz. Proposition 12(a) then gives Assumption 3 with every $\gamma > 1/2$ whenever the score has a bounded marginal density, and part (b) applies when the recursion has the additive form $Y_t = h(Z_{t-1}) + \varepsilon_t$ with an innovation density bounded by M_ε.

B.6 Mixing versus Functional Dependence

This subsection compares mixing with the functional dependence measure on two examples of Section 3: Gaussian long memory is not strongly mixing, and for the Bernoulli autoregression both the β-mixing and the switch coefficients equal one, so that the coverage bounds formulated through them are vacuous.

Long memory is not strongly mixing. A stationary Gaussian sequence is strongly mixing if and only if its spectral density has the form $|P|^2 \exp(u + \tilde{v})$ with P a trigonometric polynomial, u and v continuous and $\tilde{v}$ the conjugate function of v (Bradley, 2005, Theorem 7.1); equivalently, $\log f - \log |P|^2$ has vanishing mean oscillation for some P, since the functions of vanishing mean oscillation are exactly the sums $u + \tilde{v}$ with u and v continuous (Sarason, 1975). Suppose $f(\lambda) = |\lambda|^{2\beta-2} L(\lambda)$ with $1/2 < \beta < 1$ and L continuous and positive at the origin, as for the linear process with the coefficients of Appendix D.1, whose spectral density is proportional to $|2\sin(\lambda/2)|^{2\beta-2}$. Then $\log f - \log |P|^2 = -(2 - 2\beta + 2m) \log |\lambda| + \psi(\lambda)$ near the origin, where $m \ge 0$ is the multiplicity of the zero of P at the origin and ψ is continuous, hence of vanishing mean oscillation; but $\log |\lambda|$ is not, because its mean oscillation over $[-h, h]$ is the same positive number for every $h > 0$, and $2 - 2\beta + 2m > 0$. Hence no such representation exists, and the sequence is not strongly mixing, in particular not β-mixing.

The switch coefficient of the Bernoulli autoregression. The coverage guarantees for split conformal prediction on a stationary series are formulated through β-mixing coefficients (Oliveira et al., 2024) or through the switch coefficient of Barber & Pananjady (2026). For a sequence $V = (V_1, \dots, V_N)$ of scores, a gap $\tau \ge 0$ and $1 \le k \le N - 1 - \tau$, the switch coefficient $\Psi_{k,\tau}(V)$ is the total variation distance between the law of $(V_1, \dots, V_{N-\tau-k}, V_{N-k+1}, \dots, V_N)$ and the law of $(V_{\tau+k+1}, \dots, V_N, V_1, \dots, V_k)$, that is, between the laws of two blocks of lengths $N - \tau - k$ and k separated by the gap τ in their two temporal orders; for the remaining values of k the two laws coincide by stationarity. The coverage bound of Barber & Pananjady (2026, Theorem 4) subtracts from $1 - \alpha$ a gap term, which is at least τ divided by the number of scores, and the average of $\Psi_{k,\tau}(V)$ over k, minimized over the gap; for stationary β-mixing sequences, the switch coefficient satisfies $\Psi_{k,\tau}(V) \le 2\beta(\tau)$ (Barber & Pananjady, 2026, Proposition 1). The switch coefficient cannot be bounded through the functional dependence measure in general, and the two routes separate on the first example of Appendix B.5. Let $W_t = W_{t-1}/2 + B_t$ be that process, for which $\delta_q^Z(j) \le 2^{-j}$. Its β-mixing coefficient equals one at every lag, because the conditional law of W_τ given $\mathcal{F}_0$ is supported on 2^τ points while the marginal law is continuous. Its switch coefficients also equal one. Put $L = N - \tau - k$. Under the first law above, the coordinates at positions L and $L + 1$ are W_L and $W_{L+\tau+1}$, and the recursion gives $2^{\tau+1} W_{L+\tau+1} - W_L = 2 \sum_{j=1}^{\tau+1} 2^{j-1} B_{L+j} \in 2\mathbb{Z}$ almost surely. Under the second law, the same coordinates are W_N and W_1, and $W_N = 2^{-(N-1)} W_1 + D_N$ with D_N a function of $B_2, \dots, B_N$, so that $2^{\tau+1} W_1 - W_N \in 2\mathbb{Z}$ requires W_1 to lie in a countable set determined by $(B_2, \dots, B_N)$; since W_1 is a function of $(B_s)_{s \le 1}$, independent of $(B_2, \dots, B_N)$, with a continuous law, this event has probability zero. The event $\{2^{\tau+1} v_{L+1} - v_L \in 2\mathbb{Z}\}$ therefore has probability one under the first law and zero under the second, and $\Psi_{k,\tau}(W) = 1$ for every τ and every $1 \le k \le N - 1 - \tau$. The same holds for the CMR score $V_t = |W_t - 1|$ of that example, with the event that $2^{\tau+1} w' - w \in 2\mathbb{Z}$ for some $w \in \{1 \pm v_L\}$ and $w' \in \{1 \pm v_{L+1}\}$, since each of the four sign combinations leads to a relation of the same kind with a nonzero coefficient of W_1. Consequently the average switch coefficient is at least $1 - (\tau + 1)/N$, the coverage bound of Barber & Pananjady (2026) is vacuous for every gap once $N > 1/\alpha$, and so is the β-mixing bound of Oliveira et al. (2024), whereas Theorem 3 with the true center, for which $E_n = 0$, gives the coverage error $C_1 m^{-1/3}$ with no gap.

C PROOFS FOR SECTION 4

This appendix proves Theorem 4 and Proposition 5 and bounds the coverage error of the interval. Write $g = \widehat{\mu} - \mu$ for the error of the center, a function of the training block and of randomness independent of the series, and let $a(g)$ be the $(1-\alpha)$-quantile of $|Z-g|$ for $Z \sim N(0,1)$, the radius that an infinitely long calibration block would return for a center that is off by g, so that $a(0) = z_\alpha$. The length of the CMR interval is $2\widehat{S}^M_{(k)}$, where $\widehat{S}^M_{(k)}$ is the k-th smallest of the calibration residuals $|Y_{n+i} - \widehat{\mu}| = |G_{n+i} - g|$. Since $a(g) \geq z_\alpha$ by Lemma 15(i), the reverse triangle inequality gives

$$\big||\widehat{S}^M_{(k)} - z_\alpha| - \{a(g) - z_\alpha\}\big| \leq |\widehat{S}^M_{(k)} - a(g)|. \tag{32}$$

The term $a(g) - z_\alpha$ is the effect of the center on the population radius, and Appendix C.2 shows that it is of second order in g. The right-hand side is the error of a finite calibration block that is adjacent to the training block and therefore dependent on the center, and Appendix C.3 bounds it. The proof of Theorem 4 uses (32) as an upper bound on $|\widehat{S}^M_{(k)} - z_\alpha|$, and the proof of Proposition 5 uses it as a lower bound (Appendices C.5 and C.6). Appendix C.1 collects the Hermite expansions used in Appendix C.3 and for the calibration rate $r_m(\beta)$ of Section 4, and Appendix C.4 bounds the mean squared error of empirical quantiles of the training block, which gives the rates of Theorem 4 for the empirical median and the midpoint. Appendix C.7 uses the bound of Appendix C.3 on $|\widehat{S}^M_{(k)} - a(g)|$ to show that the coverage error has the order of the length bound of Theorem 4.

Throughout this appendix, (12) is in force, C denotes a constant depending only on α, β, A and a_0 whose value may change from line to line, and constants that depend on fewer quantities are named. Two elementary consequences of (12) are used repeatedly. First, the autocorrelation $\rho(h) = \mathrm{Cov}(G_t, G_{t+h}) = \sum_{j\geq 0} a_j a_{j+h}$ of the series satisfies, splitting the sum at $j = h$,

$$|\rho(h)| \leq \sum_{j\geq 0} |a_j a_{j+h}| \leq C_\rho (1+h)^{1-2\beta}, \tag{33}$$

where $C_\rho = A^2/(1-\beta) + A^2/(2\beta-1)$; note that $C_\rho \geq 1$, because $1 = \sum_j a_j^2 \leq A^2 \sum_j (1+j)^{-2\beta} \leq A^2\{1 + 1/(2\beta-1)\} \leq C_\rho$. Moreover, the same bound holds for $\mathrm{Cov}(\sum_{j<i} a_j \varepsilon_{t-j}, \sum_{j<i'} a_j \varepsilon_{t'-j})$, a covariance of two truncated sums at times t and t', in terms of $h = |t - t'|$. Second, for $s \geq 1$,

$$\sum_{h=0}^{s-1} (1+h)^{1-2\beta} \leq \frac{s^{2-2\beta}}{2-2\beta}, \quad \frac{1}{s}\sum_{h=0}^{s-1} (1+h)^{2-4\beta} \leq C_\beta\, r_s(\beta)^2, \quad \sum_{j\geq i} (1+j)^{-2\beta} \leq \frac{i^{1-2\beta}}{2\beta-1}\ (i \geq 1), \tag{34}$$

where C_β equals $1/(3-4\beta)$, 2 and $1 + 1/(4\beta-3)$ in the respective regimes $\beta < 3/4$, $\beta = 3/4$ and $\beta > 3/4$, and where $r_s(\beta)$ is the rate (13) with s in place of m. We also use $s^{1-2\beta} \leq r_s(\beta)$ and $1/s \leq r_s(\beta)$ for every $s \geq 2$; the thresholds m_0 and n_0 below are at least 2.

C.1 HERMITE EXPANSIONS AND THE CALIBRATION RATE OF THE SYMMETRIC SCORE

This subsection bounds covariances of functions of a Gaussian series through their Hermite expansions. An even function has no odd Hermite coefficients, so its covariances decay like the squared autocorrelations of the series; this gives the calibration rate $r_m(\beta)$ of the symmetric score (Lemma 14), and the same expansion bounds the conditional variance in Appendix C.3. The reduction of even functionals of a Gaussian series to Hermite rank two goes back to Breuer & Major (1983) and Dehling & Taqqu (1989). Let H_k denote the probabilists' Hermite polynomials, normalized by $\mathbb{E}H_j(\xi)H_k(\xi) = j!\mathbf{1}\{j=k\}$ for $\xi \sim N(0,1)$. For a standard bivariate Gaussian pair (ξ, ξ') with correlation ϱ, comparing coefficients in

$$\mathbb{E}\, e^{u\xi - u^2/2} e^{v\xi' - v^2/2} = e^{\varrho uv} \tag{35}$$

gives $\mathbb{E}H_j(\xi)H_k(\xi') = j!\varrho^j \mathbf{1}\{j=k\}$, and therefore, for square-integrable ψ, ψ' with Hermite coefficients $\eta_k = \mathbb{E}\psi(\xi)H_k(\xi)$ and $\eta'_k = \mathbb{E}\psi'(\xi)H_k(\xi)$,

$$\mathrm{Cov}(\psi(\xi), \psi'(\xi')) = \sum_{k\geq 1} \frac{\eta_k \eta'_k}{k!} \varrho^k, \qquad \mathrm{Var}(\psi(\xi)) = \sum_{k\geq 1} \frac{\eta_k^2}{k!}. \tag{36}$$

Lemma 13. *Let (ξ, ξ') be a standard bivariate Gaussian pair with correlation ϱ, and let ψ and ψ' be even and square-integrable. Then* $|\mathrm{Cov}(\psi(\xi), \psi'(\xi'))| \le \varrho^2 \{\mathrm{Var}(\psi(\xi))\mathrm{Var}(\psi'(\xi'))\}^{1/2}$.

Proof. The odd Hermite coefficients of an even function vanish, so (36) runs over even $k \ge 2$, where $|\varrho^k| \le \varrho^2$; the Cauchy–Schwarz inequality and the Parseval identity in (36) give the bound. □

For indicators of symmetric intervals, whose variances are at most $1/4$, the covariance is at most $\varrho^2/4$, including when the two intervals differ.

Lemma 14 (Calibration rate of the symmetric score)**.** *Let $S_t = |G_t| - z_\alpha$ and $r = z_\alpha/2$, as in Appendix B.5. Then $\sigma_m \le C_\rho C_\beta^{1/2}\, r_m(\beta)$. If $\rho(h) \ge c_\rho(1+h)^{1-2\beta}$ for all $h \ge h_0$, with $c_\rho > 0$, then for every real a the standard deviation of $m^{-1}\sum_{i=1}^m \mathbf{1}\{G_{n+i} \le a\}$, the corresponding quantity for a one-sided indicator, is of exact order $m^{1/2-\beta}$.*

Proof. For $|s| \le r$ and $a = z_\alpha + s > 0$, the indicator $\mathbf{1}\{S_t \le s\} = \mathbf{1}\{|G_t| \le a\}$ is an even function of G_t, so Lemma 13 bounds each covariance $\mathrm{Cov}(\mathbf{1}\{|G_{n+i}| \le a\}, \mathbf{1}\{|G_{n+j}| \le a\})$ by $\rho(i-j)^2/4$, and (33) and (34) give

$$\mathrm{Var}\{\widehat{F}_m(s)\} \le \frac{1}{4m^2}\sum_{i,j\le m}\rho(i-j)^2 \le \frac{C_\rho^2}{2m}\sum_{h<m}(1+h)^{2-4\beta},$$

which is at most $C_\rho^2 C_\beta r_m(\beta)^2/2$. For the one-sided indicator, the first Hermite coefficient of $x \mapsto \mathbf{1}\{x \le a\}$ is $-\phi(a) \ne 0$, and by (36) and the Cauchy–Schwarz inequality the remaining terms contribute at most $\rho(h)^2/4$, so the covariance at lag h is $\phi(a)^2\rho(h) + O(\rho(h)^2)$, of exact order $\rho(h)$ for all large h; the variance of the partial sum is therefore of order $m^{3-2\beta}$, which gives the scale $m^{1/2-\beta}$. □

C.2 The Shifted Population Radius

This subsection deals with the term $a(g) - z_\alpha$ in (32), the effect of the center on the population radius. Because the law of Z is symmetric, $a(g)$ is an even function of g, so it has no first-order term, and $a(g) - z_\alpha$ is of order g^2 instead of $|g|$. Lemma 15 makes this precise: part (iii) gives the two-sided bounds, of order $\min(g^2, 1)$ and g^2, that enter the proofs of Proposition 5 and Theorem 4, and part (iv) gives the exact second-order constant; parts (i) and (ii), the position of $a(g)$ relative to $|g|$ and a lower bound on the density of $|Z - g|$ near $a(g)$, are used in Appendix C.3.

For $g \in \mathbb{R}$ and $w \ge 0$, let $H_g(w) = \mathbb{P}(|Z-g| \le w) = \Phi(w-g) + \Phi(w+g) - 1$ with $Z \sim N(0,1)$. For each g the map $w \mapsto H_g(w)$ is continuous and strictly increasing from 0 to 1, with derivative $H_g'(w) = \phi(w-g) + \phi(w+g) \le 2\phi(0) < 1$, so the $(1-\alpha)$-quantile $a(g)$ of $|Z-g|$ is the unique solution of $H_g(a(g)) = 1-\alpha$, and $a(0) = z_\alpha$. Put

$$\begin{aligned} w_* &= \frac{z_\alpha}{2}, \qquad f_* = \min\{\phi(u) : \Phi^{-1}(1-\alpha) - w_* \le u \le z_\alpha + w_*\}, \\ \kappa_\alpha &= \inf_{0<u\le 1}\frac{\phi(z_\alpha - u) - \phi(z_\alpha + u)}{u}, \qquad f_1 = \min\{\phi(u) : |u| \le \max(z_\alpha, 1)\}. \end{aligned} \tag{37}$$

The constant κ_α is positive because the ratio is continuous and positive on $(0, 1]$ with limit $2z_\alpha\phi(z_\alpha)$ at zero, and the ratio is at most $2\sup|\phi'| = 2\phi(1)$ by the mean value theorem.

Lemma 15 (Shifted radius)**.** *(i) a is even and nondecreasing in $|g|$, and $\Phi^{-1}(1-\alpha) \le a(g) - |g| \le z_\alpha$ for every g.*

(ii) $H_g'(w) \ge f_*$ for every g and every w with $|w - a(g)| \le w_*$.

(iii) $\frac{1}{2}\kappa_\alpha \min(g^2, 1) \le a(g) - z_\alpha \le C_\alpha g^2$ for every g, with $C_\alpha = \max\{\phi(1)/f_1, 1\}$.

(iv) $a(g) = z_\alpha + \frac{1}{2}z_\alpha g^2 + O(g^4)$ as $g \to 0$.

Proof. (i) $H_g(w)$ is even in g, and $\partial_g H_g(w) = \phi(w+g) - \phi(w-g) \le 0$ for $g, w \ge 0$, so $H_g(w)$ is nonincreasing in $|g|$ and $a(g)$ is nondecreasing in $|g|$; this is Anderson's inequality for the intervals

$[g-w, g+w]$. For $g \geq 0$, $1-\alpha = H_g(a(g)) \leq \Phi(a(g)-g)$ gives the lower bound on $a(g)-g$, and $H_g(g+z_\alpha) = \Phi(z_\alpha) + \Phi(2g+z_\alpha) - 1 \geq 2\Phi(z_\alpha) - 1 = 1-\alpha$ gives $a(g) \leq g + z_\alpha$; evenness covers $g < 0$.

(ii) For $g \geq 0$ and $w \geq 0$, $|w-g| \leq w+g$, so $H_g'(w) \geq \phi(w-g) = \phi(w-|g|)$, and the same holds for $g < 0$ by evenness. If $|w - a(g)| \leq w_*$, then by (i) $w - |g| \in [\Phi^{-1}(1-\alpha) - w_*, z_\alpha + w_*]$, on which $\phi \geq f_*$.

(iii) For $0 \leq g \leq 1$,

$$H_0(z_\alpha) - H_g(z_\alpha) = \int_0^g \{\phi(z_\alpha - u) - \phi(z_\alpha + u)\}\, du \in \left[\kappa_\alpha g^2/2,\ \phi(1) g^2\right].$$

Since $a(g) \geq z_\alpha$ and $H_g' < 1$, $1-\alpha = H_g(a(g)) \leq H_g(z_\alpha) + a(g) - z_\alpha$, and $1-\alpha = H_0(z_\alpha)$, so $a(g) - z_\alpha \geq H_0(z_\alpha) - H_g(z_\alpha) \geq \kappa_\alpha g^2/2$; for $|g| \geq 1$, monotonicity gives $a(g) - z_\alpha \geq a(1) - z_\alpha \geq \kappa_\alpha/2$. For the upper bound with $|g| \leq 1$, every $u \in [z_\alpha, a(g)]$ satisfies $u - |g| \in [z_\alpha - 1, z_\alpha]$ by (i), so $H_g'(u) \geq \phi(u-|g|) \geq f_1$ there, and $H_g(a(g)) - H_g(z_\alpha) = H_0(z_\alpha) - H_g(z_\alpha) \leq \phi(1) g^2$ gives $a(g) - z_\alpha \leq \phi(1) g^2 / f_1$; for $|g| > 1$, (i) gives $a(g) - z_\alpha \leq |g| \leq g^2$.

(iv) The function $(g, w) \mapsto H_g(w)$ is smooth with $\partial_w H_g(w) > 0$, so a is smooth by the implicit function theorem, and even, so $a'(0) = 0$. Differentiating $H_g(a(g)) = 1-\alpha$ twice at $g = 0$ gives $\partial_g^2 H_0(z_\alpha) + \partial_w H_0(z_\alpha) a''(0) = 0$ with $\partial_g^2 H_0(z_\alpha) = 2\phi'(z_\alpha) = -2 z_\alpha \phi(z_\alpha)$ and $\partial_w H_0(z_\alpha) = 2\phi(z_\alpha)$, whence $a''(0) = z_\alpha$, and the odd derivatives vanish. □

The second-order behavior of residual empirical processes on the estimation sample under long memory is classical (Beran & Ghosh, 1991; Koul & Surgailis, 2010; Chan & Ling, 2008; Lorek & Kulik, 2014); there it arises from the exact cancellation of the first-order term of the error process by the estimated location, which uses the same observations, whereas here the residuals are out of sample, the object is the expected error of an order statistic at finite n and m, and the mechanism is the evenness of the population radius. The mechanism rests on the symmetry of the error law and not on the absolute-value form of the score: for a skewed error the radius $a(g)$ is no longer even, its derivative at $g = 0$ is proportional to the difference of the error density at the two oracle endpoints, and the center error re-enters the length at first order.

For symmetric innovations that are not Gaussian, Lemma 15(i)–(iii) remain valid with the marginal density of the series in place of ϕ, provided that it is positive, continuously differentiable, nonincreasing on $[0, \infty)$ and has a negative derivative at the radius, with the constants redefined accordingly; the rates, however, use Gaussianity throughout, in the Hermite expansions of Lemmas 13 and 14, in the Gaussian conditioning of Lemmas 16–18, and in the Gaussian Hölder inequality of Lemma 19.

C.3 Calibration Residuals Conditional on the Training Innovations

This subsection deals with the right-hand side of (32), the error of a finite calibration block. The difficulty is that the calibration block is adjacent to the training block, so that the center and the calibration residuals are dependent. We condition on the innovations up to time n: each calibration observation splits into a part that is determined by these innovations and a Gaussian part that is independent of them, so that, conditionally, the center is fixed and each residual is Gaussian, shifted by the error of the center and by the memory of the training period, which decays with the distance from the training block. Lemmas 16 and 17 bound the conditional mean and variance of the empirical distribution function of the residuals, and Lemma 18 turns these bounds into a bound on $\mathbb{E}|\widehat{S}^M_{(k)} - a(g)|$ of order $r_m(\beta)$, up to a term that couples the error of the center with the memory of the training period.

Let $\mathcal{F}_n$ denote the sigma-field generated by the innovations up to time n and by the randomness of the training algorithm, so that $g = \widehat{\mu} - \mu$ is $\mathcal{F}_n$-measurable. For $1 \leq i \leq m+1$ decompose the future observation into its predictable and its innovative parts,

$$G_{n+i} = \Pi_i + U_i, \qquad \Pi_i = \sum_{j \geq i} a_j \varepsilon_{n+i-j}, \qquad U_i = \sum_{j=0}^{i-1} a_j \varepsilon_{n+i-j}, \tag{38}$$

so that Π_i is $\mathcal{F}_n$-measurable and $(U_1, \dots, U_{m+1})$ is a centered Gaussian vector independent of $\mathcal{F}_n$, with $s_i^2 = \mathrm{Var}(U_i) = \sum_{j<i} a_j^2 \in [a_0^2, 1]$ and $\tau_i = \mathbb{E}\Pi_i^2 = \sum_{j \geq i} a_j^2 = 1 - s_i^2 \leq A^2 i^{1-2\beta}/(2\beta - 1)$ by (34). Put

$$T_1 = \frac{1}{m}\sum_{i=1}^{m} |\Pi_i|, \quad \Pi_2 = \frac{1}{m}\sum_{i=1}^{m} \Pi_i^2, \quad \bar{\tau} = \frac{1}{m}\sum_{i=1}^{m} \tau_i \leq C_\tau m^{1-2\beta}, \quad C_\tau = \frac{A^2}{(2\beta-1)(2-2\beta)}, \tag{39}$$

where the bound on $\bar{\tau}$ follows from (34); note that $\mathbb{E}\Pi_2 = \bar{\tau}$ and $\mathbb{E}T_1^2 \leq \bar{\tau}$ by Jensen's inequality. By (33) applied to the truncated sums, $|\mathrm{Cov}(U_i, U_j)| \leq C_\rho(1 + |i-j|)^{1-2\beta}$, so the correlations $R_{ij} = \mathrm{Cov}(U_i, U_j)/(s_i s_j)$ satisfy $|R_{ij}| \leq C_\rho a_0^{-2}(1 + |i-j|)^{1-2\beta}$, and by (34)

$$\max_{i \leq m} \sum_{j=1}^{m} |R_{ij}| \leq \frac{C_\rho}{a_0^2}\,\frac{m^{2-2\beta}}{1-\beta}, \qquad \frac{1}{m^2}\sum_{i,j=1}^{m} R_{ij}^2 \leq \frac{2C_\rho^2 C_\beta}{a_0^4}\, r_m(\beta)^2. \tag{40}$$

Finally, for $w \geq 0$, $b \in \mathbb{R}$ and $s \in [|a_0|, 1]$, let $\pi_w(b, s) = \mathbb{P}(|s\xi + b| \leq w) = \Phi\{(w-b)/s\} - \Phi\{(-w-b)/s\}$ for $\xi \sim N(0,1)$; then $\pi_w(-g, 1) = H_g(w)$, and

$$\partial_b \pi_w(0, s) = 0, \qquad |\partial_b^2 \pi_w(b, s)| \leq \frac{2\phi(1)}{a_0^2}, \qquad |\partial_s \pi_w(b, s)| \leq \frac{2\phi(1)}{|a_0|}, \tag{41}$$

because $\partial_b \pi_w(b, s) = s^{-1}[\phi\{(w+b)/s\} - \phi\{(w-b)/s\}]$, $|\phi'| \leq \phi(1)$ and $|x\phi(x)| \leq \phi(1)$.

Lemma 16 (Conditional mean)**.** *Let $\widehat{H}_m(w) = m^{-1}\sum_{i=1}^{m} \mathbf{1}\{|G_{n+i} - g| \leq w\}$ for $w \geq 0$, with g $\mathcal{F}_n$-measurable. Then, for every $\mathcal{F}_n$-measurable $w \geq 0$,*

$$\big|\mathbb{E}\{\widehat{H}_m(w) \mid \mathcal{F}_n\} - H_g(w)\big| \leq B, \qquad B = K_1\big(|g|T_1 + \Pi_2 + \bar{\tau}\big), \qquad K_1 = \frac{2\phi(1)}{\min(a_0^2, |a_0|)},$$

and $\mathbb{E}B \leq K_1\{(\bar{\tau}\,\mathbb{E}g^2)^{1/2} + 2\bar{\tau}\}$.

Proof. Conditionally on $\mathcal{F}_n$, $|G_{n+i} - g| = |U_i + \Pi_i - g|$ with $U_i \sim N(0, s_i^2)$, so $\mathbb{E}\{\mathbf{1}\{|G_{n+i} - g| \leq w\} \mid \mathcal{F}_n\} = \pi_w(\Pi_i - g, s_i)$. A Taylor expansion in the first argument about $-g$ and (41) give $|\pi_w(\Pi_i - g, s_i) - \pi_w(-g, s_i)| \leq |\partial_b \pi_w(-g, s_i)||\Pi_i| + \phi(1)\Pi_i^2/a_0^2 \leq 2\phi(1)a_0^{-2}(|g||\Pi_i| + \Pi_i^2/2)$, since $\partial_b \pi_w(\cdot, s)$ vanishes at zero and is $2\phi(1)a_0^{-2}$-Lipschitz. The mean value theorem in the second argument gives $|\pi_w(-g, s_i) - \pi_w(-g, 1)| \leq 2\phi(1)|a_0|^{-1}(1 - s_i) \leq 2\phi(1)|a_0|^{-1}\tau_i$, because $1 - s_i \leq 1 - s_i^2 = \tau_i$. Averaging over i proves the first claim, and the Cauchy–Schwarz inequality with $\mathbb{E}T_1^2 \leq \bar{\tau}$ and $\mathbb{E}\Pi_2 = \bar{\tau}$ proves the second. □

Lemma 17 (Conditional variance)**.** *In the setting of Lemma 16, for every $\mathcal{F}_n$-measurable $w \geq 0$,*

$$\mathrm{Var}\{\widehat{H}_m(w) \mid \mathcal{F}_n\} \leq W, \qquad W = K_2\big\{r_m(\beta)^2 + m^{1-2\beta}(\Pi_2 + g^2)\big\},$$

where $K_2 = \max\{8\phi(1)^2 C_\rho/(a_0^4(1-\beta)),\ C_\rho^2 C_\beta/(2a_0^4)\}$.

Proof. Write $\xi_i = U_i/s_i$ and $b_i = \Pi_i - g$, so that, conditionally on $\mathcal{F}_n$, the summand $\mathbf{1}\{|G_{n+i} - g| \leq w\} = \psi_i(\xi_i)$ with $\psi_i(x) = \mathbf{1}\{|s_i x + b_i| \leq w\}$ and (ξ_i) a standard Gaussian vector with correlations R_{ij}. The first Hermite coefficient of ψ_i is $\eta_{i,1} = \mathbb{E}[\xi \mathbf{1}\{(-w - b_i)/s_i \leq \xi \leq (w - b_i)/s_i\}] = \phi\{(-w - b_i)/s_i\} - \phi\{(w - b_i)/s_i\}$, which vanishes at $b_i = 0$ and is $2\phi(1)|a_0|^{-1}$-Lipschitz in b_i, so $|\eta_{i,1}| \leq 2\phi(1)|a_0|^{-1}|b_i|$. By (36), the Cauchy–Schwarz inequality over $k \geq 2$ and $\mathrm{Var}(\psi_i(\xi_i)) \leq 1/4$,

$$|\mathrm{Cov}(\psi_i(\xi_i), \psi_j(\xi_j) \mid \mathcal{F}_n)| \leq \frac{4\phi(1)^2}{a_0^2}|b_i b_j||R_{ij}| + \frac{R_{ij}^2}{4},$$

also on the diagonal. Since $\sum_{i,j} |b_i b_j||R_{ij}| \leq (\max_i \sum_j |R_{ij}|)\sum_i b_i^2$ for a symmetric matrix with nonnegative entries, and $\sum_i b_i^2 \leq 2m(\Pi_2 + g^2)$, the bounds (40) give

$$\mathrm{Var}\{\widehat{H}_m(w) \mid \mathcal{F}_n\} \leq \frac{4\phi(1)^2}{a_0^2} \cdot \frac{C_\rho m^{2-2\beta}}{a_0^2(1-\beta)} \cdot \frac{2m(\Pi_2 + g^2)}{m^2} + \frac{C_\rho^2 C_\beta}{2a_0^4} r_m(\beta)^2,$$

which is the claim. □

Lemma 18 (Finite-calibration comparison). *Let g be $\mathcal{F}_n$-measurable with $\mathbb{E}g^2 < \infty$, and let $\widehat{S}^M_{(k)}$ be the k-th smallest of $|G_{n+i} - g|$, $1 \le i \le m$. There are m_0, depending only on α, and K_3, depending only on α, β, A and a_0, such that for $m \ge m_0$,*

$$\mathbb{E}|\widehat{S}^M_{(k)} - a(g)| \le K_3\big\{r_m(\beta) + m^{1-2\beta} + (m^{1-2\beta}\mathbb{E}g^2)^{1/2}\big\}.$$

Proof. Let $m_0 = \max\{8/(f_* w_*), 2(1-\alpha)/\alpha\}$, so that $d_m = k/m - (1-\alpha) < 2/m \le f_* w_*/4$ and $\min(k, m-k+1) \ge m\min(1-\alpha, \alpha/2)$ for $m \ge m_0$, by the argument given for (22).

Step 1: a conditional tail bound. Let $0 < t \le w_*$ with $f_* t > B + d_m$. The event $\widehat{S}^M_{(k)} > a(g) + t$ implies $\widehat{H}_m(a(g)+t) < k/m \le 1 - \alpha + d_m$, whereas Lemma 16 and Lemma 15(ii) give $\mathbb{E}\{\widehat{H}_m(a(g)+t) \mid \mathcal{F}_n\} \ge H_g(a(g)+t) - B \ge 1 - \alpha + f_* t - B$; the event therefore requires a downward deviation of $\widehat{H}_m(a(g)+t)$ from its conditional mean by more than $f_* t - B - d_m > 0$. The event $\widehat{S}^M_{(k)} < a(g) - t$ implies $\widehat{H}_m(a(g)-t) \ge k/m \ge 1-\alpha$, whereas the conditional mean is at most $1 - \alpha - f_* t + B$, so it requires an upward deviation of at least $f_* t - B \ge f_* t - B - d_m$. Chebyshev's inequality conditionally on $\mathcal{F}_n$ and Lemma 17 give

$$\mathbb{P}\{|\widehat{S}^M_{(k)} - a(g)| > t \mid \mathcal{F}_n\} \le \min\Big\{1, \frac{2W}{(f_* t - B - d_m)^2}\Big\}. \tag{42}$$

Step 2: a conditional second moment. By Lemma 15(i), $|g| - a(g) \in [-z_\alpha, -\Phi^{-1}(1-\alpha)]$, so $\big||x - g| - a(g)\big| \le |x| + C_z$ for every real x, with $C_z = z_\alpha + |\Phi^{-1}(1-\alpha)|$. The quantity $\widehat{S}^M_{(k)} - a(g)$ is the k-th order statistic of $T_i = |G_{n+i} - g| - a(g)$, and for any real numbers, $T^2_{(k)} \le \sum_i T_i^2 / \min(k, m-k+1)$, because at least $m-k+1$ of the T_i are at least $T_{(k)}$ when $T_{(k)} \ge 0$ and at least k of them are at most $T_{(k)}$ when $T_{(k)} < 0$. Since $\mathbb{E}(G^2_{n+i} \mid \mathcal{F}_n) = s_i^2 + \Pi_i^2 \le 1 + \Pi_i^2$,

$$M^2 := \mathbb{E}\{(\widehat{S}^M_{(k)} - a(g))^2 \mid \mathcal{F}_n\} \le \frac{2\sum_{i \le m}(1 + \Pi_i^2 + C_z^2)}{m\min(1-\alpha, \alpha/2)} \le C_M(1 + \Pi_2), \tag{43}$$

with $C_M = (2 + 2C_z^2)/\min(1-\alpha, \alpha/2)$.

Step 3: the conditional expectation. Put $a' = (B + d_m)/f_*$ and $b' = (2W)^{1/2}/f_*$, so that the right-hand side of (42) is $\min\{1, b'^2/(t-a')^2\}$. If $a' < w_*$, integrating (42) over $0 < t \le w_*$ as in the proof of Proposition 9 gives $\mathbb{E}\{\min(|\widehat{S}^M_{(k)} - a(g)|, w_*) \mid \mathcal{F}_n\} \le a' + 2b'$, and the conditional Cauchy–Schwarz inequality gives $\mathbb{E}\{(|\widehat{S}^M_{(k)} - a(g)| - w_*)_+ \mid \mathcal{F}_n\} \le M\min\{1, b'/(w_* - a')\}$. Let $b_* = f_* w_*/4$. On the event $\{B \le b_*\}$ we have $a' \le w_*/2$, hence

$$\mathbb{E}\{|\widehat{S}^M_{(k)} - a(g)| \mid \mathcal{F}_n\} \le a' + 2b' + \frac{2Mb'}{w_*} \qquad \text{on } \{B \le b_*\}. \tag{44}$$

Step 4: the unconditional expectation. Taking expectations in (44) and using $\mathbb{E}b' \le (2\mathbb{E}W)^{1/2}/f_*$ and $\mathbb{E}(Mb') \le (\mathbb{E}M^2\,\mathbb{E}b'^2)^{1/2}$,

$$\mathbb{E}\big[|\widehat{S}^M_{(k)} - a(g)|\mathbf{1}\{B \le b_*\}\big] \le \frac{\mathbb{E}B + d_m}{f_*} + \frac{2(2\mathbb{E}W)^{1/2}}{f_*}\Big\{1 + \frac{(\mathbb{E}M^2)^{1/2}}{w_*}\Big\}.$$

On the complementary event, $\mathbb{E}\{|\widehat{S}^M_{(k)} - a(g)| \mid \mathcal{F}_n\} \le M \le C_M^{1/2}(1 + \Pi_2^{1/2})$ by (43), so that, by Markov's inequality and the Cauchy–Schwarz inequality,

$$\mathbb{E}\big[|\widehat{S}^M_{(k)} - a(g)|\mathbf{1}\{B > b_*\}\big] \le C_M^{1/2}\big\{\mathbb{P}(B > b_*) + (\bar\tau\mathbb{P}(B > b_*))^{1/2}\big\} \le C_M^{1/2}\Big\{\frac{3\mathbb{E}B}{2b_*} + \frac{\bar\tau}{2}\Big\}.$$

Now $\mathbb{E}M^2 \le 2C_M$ because $\bar\tau \le 1$, $\mathbb{E}W \le K_2\{r_m(\beta)^2 + m^{1-2\beta}(\bar\tau + \mathbb{E}g^2)\}$ by Lemma 17, $\mathbb{E}B \le K_1\{(\bar\tau\mathbb{E}g^2)^{1/2} + 2\bar\tau\}$ by Lemma 16, $\bar\tau \le C_\tau m^{1-2\beta}$, $d_m \le 2/m \le 2r_m(\beta)$, and $(m^{1-2\beta}\bar\tau)^{1/2} \le C_\tau^{1/2} m^{1-2\beta}$. Collecting the terms proves the lemma. □

C.4 Empirical Quantiles of a Long-Memory Gaussian Sample

The bound of Theorem 4 involves the center only through its mean squared error. For the sample mean this is the variance of $\bar{G}_n$, which is bounded directly in Appendix C.5; for the empirical median and for the midpoint of two empirical quantiles it requires the mean squared error of an empirical quantile of the training block, which this subsection bounds by a multiple of $n^{1-2\beta}$ for every n. The same bound gives the explicit rate for Assumption 1 mentioned after that assumption.

Lemma 19. *Let $\tau \in (0,1)$, $\zeta_\tau = \Phi^{-1}(\tau)$ and $j_n = \lceil n\tau \rceil$. Under* (12)*, $\mathbb{E}(G_{(j_n)} - \zeta_\tau)^2 \le C_q n^{1-2\beta}$ for all $n \ge 1$, where $G_{(j_n)}$ is the j_n-th smallest of $G_1, \dots, G_n$ and C_q depends only on τ, β and A.*

Proof. Let Γ_n be the correlation matrix of $(G_1, \dots, G_n)$ and λ_n its largest eigenvalue, so that $1 \le \lambda_n \le \max_i \sum_j |\rho(i-j)| \le 1 + 2C_\rho n^{2-2\beta}/(2-2\beta) \le C_\lambda n^{2-2\beta}$ by (33)–(34), with $C_\lambda = 1 + C_\rho/(1-\beta)$. Fix $u \in \mathbb{R}$ and $t > 0$, and let $f_i(x) = \exp[t\{\mathbf{1}\{x \le u\} - \Phi(u)\}]$. The Gaussian Hölder inequality of Chen et al. (2015, Theorem 1) states that for a centered Gaussian vector with covariance matrix T, nonnegative measurable functions f_i and exponents $p_i > 0$ such that $T \preceq \operatorname{diag}(p_1 T_{11}, \dots, p_n T_{nn})$, one has $\mathbb{E}\prod_i f_i(X_i) \le \prod_i \{\mathbb{E} f_i(X_i)^{p_i}\}^{1/p_i}$; for two standard Gaussian variables with correlation ϱ and $p_1 = p_2 = p$ the condition reads $(p-1)^2 \ge \varrho^2$, which is Nelson's hypercontractivity condition. Since Γ_n has unit diagonal and $\Gamma_n \preceq \lambda_n I$, the condition holds with all exponents equal to λ_n, and the inequality gives

$$\mathbb{E}\prod_{i=1}^n f_i(G_i) \le \prod_{i=1}^n \big\{\mathbb{E} f_i(G_i)^{\lambda_n}\big\}^{1/\lambda_n} \le \exp\Big(\frac{n\lambda_n t^2}{8}\Big),$$

where the last step is Hoeffding's lemma for the centered variable $\mathbf{1}\{G_i \le u\} - \Phi(u)$, whose range has length one. With $F_n(u) = n^{-1}\sum_{i \le n} \mathbf{1}\{G_i \le u\}$, the Chernoff bound with $t = 4x/\lambda_n$, applied also to $-\{\mathbf{1}\{x \le u\} - \Phi(u)\}$, yields

$$\mathbb{P}\{|F_n(u) - \Phi(u)| \ge x\} \le 2\exp(-2nx^2/\lambda_n), \qquad x > 0. \tag{45}$$

Let $f_\tau = \min\{\phi(v) : |v - \zeta_\tau| \le 1\}$, so that $\Phi(\zeta_\tau + t) - \tau \ge f_\tau t$ and $\tau - \Phi(\zeta_\tau - t) \ge f_\tau t$ for $0 \le t \le 1$, and note that $\tau \le j_n/n \le \tau + 1/n$. For $2/(nf_\tau) \le t \le 1$, the event $G_{(j_n)} > \zeta_\tau + t$ implies $F_n(\zeta_\tau + t) < j_n/n$, hence $F_n(\zeta_\tau + t) - \Phi(\zeta_\tau + t) < 1/n - f_\tau t \le -f_\tau t/2$, and the event $G_{(j_n)} < \zeta_\tau - t$ implies $F_n(\zeta_\tau - t) \ge \tau$, hence $F_n(\zeta_\tau - t) - \Phi(\zeta_\tau - t) \ge f_\tau t$; by (45),

$$\mathbb{P}(|G_{(j_n)} - \zeta_\tau| > t) \le 4\exp\Big(-\frac{nf_\tau^2 t^2}{2\lambda_n}\Big), \qquad \frac{2}{nf_\tau} \le t \le 1.$$

Integrating $2t$ times the tail probability, the interval $t < 2/(nf_\tau)$ contributes at most $4/(nf_\tau)^2$, the interval up to $t = 1$ contributes at most $\int_0^\infty 8t\exp\{-nf_\tau^2 t^2/(2\lambda_n)\}\,dt = 8\lambda_n/(nf_\tau^2)$, and the remainder is $\mathbb{E}[(G_{(j_n)} - \zeta_\tau)^2 \mathbf{1}\{|G_{(j_n)} - \zeta_\tau| > 1\}] \le \{\mathbb{E}(G_{(j_n)} - \zeta_\tau)^4\}^{1/2}\, 2\exp\{-nf_\tau^2/(4\lambda_n)\}$. The rank inequality of Step 2 in the proof of Lemma 18, applied to fourth powers, gives $G_{(j_n)}^4 \le \sum_i G_i^4/\min(j_n, n - j_n + 1) \le \sum_i G_i^4/\{n\min(\tau, 1-\tau)\}$, so $\mathbb{E}(G_{(j_n)} - \zeta_\tau)^4 \le 8\{3/\min(\tau, 1-\tau) + \zeta_\tau^4\}$. Since $\lambda_n/n \le C_\lambda n^{1-2\beta}$, $1/n^2 \le n^{1-2\beta}$ and $\exp(-cn^{2\beta-1}) \le n^{1-2\beta}/(ce)$ for $c > 0$, all three contributions are at most a constant multiple of $n^{1-2\beta}$; for the finitely many $n < 2/f_\tau$ the bound follows from the fourth-moment bound by enlarging the constant. □

C.5 Proof of Theorem 4

Proof of Theorem 4. Step 1: any center. Write $g = \widehat{\mu} - \mu$, which is $\mathcal{F}_n$-measurable, and assume $\mathbb{E}g^2 < \infty$, since otherwise there is nothing to prove. The calibrated radius $\widehat{S}_{(k)}^M$ is the k-th smallest of $|Y_{n+i} - \widehat{\mu}| = |G_{n+i} - g|$, so Lemma 18 applies, and since $m^{1-2\beta} \le r_m(\beta)$ it gives $\mathbb{E}|\widehat{S}_{(k)}^M - a(g)| \le 2K_3\{r_m(\beta) + (m^{1-2\beta}\mathbb{E}g^2)^{1/2}\}$. Lemma 15(iii) gives $0 \le a(g) - z_\alpha \le C_\alpha g^2$, and (32) gives

$$\begin{aligned}\mathbb{E}|\operatorname{len}\widehat{C}_\alpha^M - 2z_\alpha| = 2\mathbb{E}|\widehat{S}_{(k)}^M - z_\alpha| &\le 2\mathbb{E}|\widehat{S}_{(k)}^M - a(g)| + 2\mathbb{E}\{a(g) - z_\alpha\} \\ &\le 4K_3\big\{r_m(\beta) + (m^{1-2\beta}\mathbb{E}g^2)^{1/2}\big\} + 2C_\alpha \mathbb{E}g^2.\end{aligned}$$

Finally $(m^{1-2\beta}\mathbb{E}g^2)^{1/2} \le \{m^{1-2\beta} + \mathbb{E}g^2\}/2 \le \{r_m(\beta) + \mathbb{E}g^2\}/2$, which proves (14). The center enters the concentration of $\widehat{S}_{(k)}^M$ around $a(g)$ only through $m^{1-2\beta}\mathbb{E}g^2$: a shift g breaks the evenness

of the indicator $\mathbf{1}\{|G_{n+i}-g| \le w\}$, whose first Hermite coefficient is of order g and carries the long-memory variance $m^{1-2\beta}$ (Lemma 17). The center is never treated as independent of the calibration block: the argument conditions on the training innovations and retains the conditional means $\Pi_i - g$ of the calibration residuals, whose interaction with the center appears in the term $(m^{1-2\beta}\mathbb{E}g^2)^{1/2}$.

Step 2: the three centers. The sample mean is a function of $Y_1, \dots, Y_n$ with $\widehat{\mu} - \mu = \bar{G}_n$, and

$$v_n = \mathrm{Var}(\bar{G}_n) = \frac{1}{n^2}\sum_{s,t=1}^{n} \rho(s-t) \le \frac{1}{n}\Big\{1 + 2\sum_{h=1}^{n-1}|\rho(h)|\Big\} \le \frac{2C_\rho}{n}\,\frac{n^{2-2\beta}}{2-2\beta} = \frac{C_\rho}{1-\beta}\,n^{1-2\beta}$$

by (33)–(34) and $C_\rho \ge 1$. The empirical median is the $\lceil n/2 \rceil$-th smallest of $Y_1, \dots, Y_n$, so that $\widehat{\mu} - \mu = G_{(\lceil n/2 \rceil)}$, and Lemma 19 with $\tau = 1/2$ gives $\mathbb{E}(\widehat{\mu} - \mu)^2 \le C_q n^{1-2\beta}$. For the midpoint $\widehat{c} = (\widehat{\ell} + \widehat{u})/2$ of $\widehat{\ell} = Y_{(\lceil n\alpha/2 \rceil)} = \mu + G_{(\lceil n\alpha/2 \rceil)}$ and $\widehat{u} = Y_{(\lceil n(1-\alpha/2) \rceil)} = \mu + G_{(\lceil n(1-\alpha/2) \rceil)}$, since $\zeta_{\alpha/2} + \zeta_{1-\alpha/2} = 0$ for the normal quantiles $\zeta_\tau = \Phi^{-1}(\tau)$,

$$\widehat{c} - \mu = \tfrac{1}{2}\big\{\big(G_{(\lceil n\alpha/2 \rceil)} - \zeta_{\alpha/2}\big) + \big(G_{(\lceil n(1-\alpha/2) \rceil)} - \zeta_{1-\alpha/2}\big)\big\},$$

and Lemma 19 at the two levels gives $\mathbb{E}(\widehat{c}-\mu)^2 \le C_q n^{1-2\beta}$, with C_q the larger of the two constants. Step 1 with these three bounds proves the second claim. By the identity of Section 4, which holds for any real endpoints, also when $\widehat{\ell} > \widehat{u}$, the CQR interval with the empirical quantiles as fitted endpoints is the CMR interval with center $\widehat{c}$. The same lemma gives $\mathbb{E}(E_n^Q)^2 \le 2C_q n^{1-2\beta}$ for the training error $E_n^Q = \max\{|\widehat{\ell} - \ell_\alpha|, |\widehat{u} - u_\alpha|\}$ of these endpoints, which is the explicit rate mentioned after Assumption 1. □

The identity used in Step 2 relies on the constant covariate, and the improvement concerns an error of location. With a covariate, a width error that varies with the covariate enters the length at first order: for CQR with $\mu = 0$, fitted endpoints $\widehat{\ell}(x) = -(z_\alpha + ex)$ and $\widehat{u}(x) = z_\alpha + ex$ at a covariate $X_t = \pm 1$ independent of G_t, the common correction is of order e^2 and the length error at each covariate value is $2e + O(e^2)$; a width error common to all covariates cancels in the calibration, by the same identity.

C.6 Proof of Proposition 5

Proof of Proposition 5. Step 1: any center. Write $g = \widehat{\mu} - \mu$ and assume $\mathbb{E}g^2 < \infty$, since otherwise the right-hand side of (15) is $-\infty$. By (32), Lemma 15(iii) and Lemma 18,

$$\begin{aligned}\mathbb{E}|\operatorname{len}\widehat{C}_\alpha^M - 2z_\alpha| &= 2\mathbb{E}|\widehat{S}_{(k)}^M - z_\alpha| \ge 2\mathbb{E}\{a(g) - z_\alpha\} - 2\mathbb{E}|\widehat{S}_{(k)}^M - a(g)| \\ &\ge \kappa_\alpha \mathbb{E}\min(g^2, 1) - 2K_3\big\{r_m(\beta) + m^{1-2\beta} + (m^{1-2\beta}\mathbb{E}g^2)^{1/2}\big\},\end{aligned}$$

which is (15) with $c_1 = \kappa_\alpha$ and $C = 4K_3$, since $m^{1-2\beta} \le r_m(\beta)$.

Step 2: a common Gaussian component of the center. Let $V = n^{-1/2}\sum_{s=-2n+1}^{-n}\varepsilon_s \sim N(0,1)$, a function of the innovations preceding the training block, and let $b_t = \mathrm{Cov}(G_t, V) = n^{-1/2}\sum_{s=-2n+1}^{-n} a_{t-s}$ for $1 \le t \le n$. The indices $t-s$ range over $[t+n, t+2n-1] \subset [n+1, 3n-1]$, so for $n \ge j_0$ the lower bound on the coefficients gives

$$b_t \ge n^{-1/2} \cdot n \cdot c\,(3n)^{-\beta} = h_n, \qquad h_n = c\,3^{-\beta} n^{1/2-\beta}.$$

Write $W_t = G_t - b_t V$. The vector $(W_1, \dots, W_n)$ is jointly Gaussian with V and uncorrelated with it, hence independent of V. Since T is translation equivariant, $g = T(Y_1, \dots, Y_n) - \mu = T(G_1, \dots, G_n) = T(W + bV)$, where $b = (b_1, \dots, b_n)^\top$. For fixed W, put $f_W(v) = T(W + bv)$; for $h \ge 0$, every coordinate of $W + b(v+h)$ exceeds the corresponding coordinate of $W + bv + h_n h\mathbf{1}$, so monotonicity and equivariance give $f_W(v+h) \ge f_W(v) + h_n h$. If $f_W(0) \ge 0$, then $g \ge h_n$ on the event $\{V \ge 1\}$; if $f_W(0) < 0$, then $g \le f_W(0) - h_n|V| \le -h_n$ on the event $\{V \le -1\}$. Since V is independent of W and $\mathbb{P}(V \ge 1) = \mathbb{P}(V \le -1) = \Phi(-1)$,

$$\mathbb{P}(|g| \ge h_n) \ge \Phi(-1), \qquad \mathbb{E}\min(g^2, 1) \ge \Phi(-1)\min(h_n^2, 1) = \Phi(-1)c^2 3^{-2\beta} n^{1-2\beta}$$

for $n \ge n_0$, where $n_0 \ge \max(j_0, 2)$ is such that $h_n \le 1$. No Gaussianity or unbiasedness of g is used.

Step 3: a calibration block large relative to the training block. Let $\nu = 2\beta - 1 \in (0, 1)$ and $c_2 = \Phi(-1)c^2 3^{-2\beta}$. By Steps 1 and 2 and $\mathbb{E}g^2 \le C_g n^{-\nu}$, for $n \ge n_0$ and $m \ge m_0$,

$$\mathbb{E}|\operatorname{len} \widehat{C}_\alpha^M - 2z_\alpha| \ge c_1 c_2 n^{-\nu} - C\big\{r_m(\beta) + C_g^{1/2}(nm)^{-\nu/2}\big\}.$$

Let $m \ge Kn^{\max(1,2\nu)} \log(n+1)$ with $K \ge m_0$, and recall that $n \ge 2$, so that $\log(n+1) \ge 1$. If $\nu < 1/2$, then $m \ge Kn$, so $r_m(\beta) = m^{-\nu} \le K^{-\nu} n^{-\nu}$ and $(nm)^{-\nu/2} \le K^{-\nu/2} n^{-\nu}$. If $\nu = 1/2$, then, since $\log(m+1)/m$ is decreasing in m and $Kn\log(n+1) + 1 \le K(n+1)^2$, $r_m(\beta)^2 \le \{\log K + 2\log(n+1)\}/\{Kn\log(n+1)\} \le (\log K + 2)/(Kn)$, and $(nm)^{-1/4} \le K^{-1/4} n^{-1/2}$. If $\nu > 1/2$, then $m \ge Kn^{2\nu}$, so $r_m(\beta) = m^{-1/2} \le K^{-1/2} n^{-\nu}$ and $(nm)^{-\nu/2} \le K^{-\nu/2} n^{-\nu(1+2\nu)/2} \le K^{-\nu/2} n^{-\nu}$, since $2\nu > 1$. In each regime $r_m(\beta) + C_g^{1/2}(nm)^{-\nu/2} \le \epsilon_K n^{-\nu}$ with $\epsilon_K \to 0$ as $K \to \infty$, so for K large the right-hand side of the last display is at least $c_1 c_2 n^{-\nu}/2$ and $r_m(\beta) \le n^{-\nu}$, which gives (16) with $c_3 = c_1 c_2/4$. □

The sample mean, the empirical median and the midpoint of the empirical $\alpha/2$- and $(1-\alpha/2)$-quantiles are nondecreasing in each coordinate and translation equivariant, and Step 2 of the proof of Theorem 4 bounds their mean squared error by a multiple of $n^{1-2\beta}$, so they satisfy the conditions of Proposition 5 on the center.

C.7 Coverage under Gaussian Long Memory

This subsection bounds the coverage error of the interval of Theorem 4. The bound has the order of the length bound (14), so the interval attains the nominal level at the rate of its length, although its center converges at the slower rate $n^{1/2-\beta}$.

Corollary 20 (Coverage)**.** *Assume* (12)*, and let $\widehat{\mu}$ be as in Theorem 4, with $\mathbb{E}(\widehat{\mu} - \mu)^2 < \infty$. For all n and all $m \ge m_0$, with m_0 as in Theorem 4,*

$$\begin{aligned}\big|\mathbb{P}\{Y_{n+m+1} \in \widehat{C}_\alpha^M\} - (1-\alpha)\big| &\le C\big[r_m(\beta) + \{m^{1-2\beta}\mathbb{E}(\widehat{\mu} - \mu)^2\}^{1/2} + m^{1-2\beta}\big] \\ &\le C\big\{\mathbb{E}(\widehat{\mu} - \mu)^2 + r_m(\beta)\big\}.\end{aligned} \tag{46}$$

For the sample mean, the empirical median and the midpoint of Theorem 4, the coverage error is therefore at most $C\{n^{-(2\beta-1)} + r_m(\beta)\}$.

The middle term of (46) is the interaction between the error of the center and the part of the test observation that the training innovations predict, to which a shifted center is sensitive at first order.

Proof of Corollary 20. Write $g = \widehat{\mu} - \mu$. The calibrated radius $\widehat{S}_{(k)}^M$ is the k-th smallest of $|G_{n+i} - g|$, $1 \le i \le m$, and $Y_{n+m+1} \in \widehat{C}_\alpha^M$ exactly when $|G_{n+m+1} - g| \le \widehat{S}_{(k)}^M$. Let $\mathcal{F}_{n+m}$ denote the sigma-field generated by the innovations up to time $n+m$ and by the randomness of the training algorithm. Conditionally on $\mathcal{F}_{n+m}$, $G_{n+m+1} = a_0 \varepsilon_{n+m+1} + \Pi'$ with $\Pi' = \sum_{j \ge 1} a_j \varepsilon_{n+m+1-j}$ measurable with respect to $\mathcal{F}_{n+m}$, so G_{n+m+1} is Gaussian with variance a_0^2, whereas g, $a(g)$ and $\widehat{S}_{(k)}^M$ are $\mathcal{F}_{n+m}$-measurable. Hence the conditional probability $\mathbb{P}(|G_{n+m+1} - g| \le q \mid \mathcal{F}_{n+m}) = \Phi\{(q + g - \Pi')/|a_0|\} - \Phi\{(-q + g - \Pi')/|a_0|\}$ is Lipschitz in q with constant $2\phi(0)/|a_0|$, and

$$\big|\mathbb{P}\{|G_{n+m+1} - g| \le \widehat{S}_{(k)}^M\} - \mathbb{P}\{|G_{n+m+1} - g| \le a(g)\}\big| \le \frac{2\phi(0)}{|a_0|}\,\mathbb{E}|\widehat{S}_{(k)}^M - a(g)|.$$

For the second probability, condition on $\mathcal{F}_n$ and use the decomposition (38) at $i = m+1$: $\mathbb{P}\{|G_{n+m+1} - g| \le a(g) \mid \mathcal{F}_n\} = \pi_{a(g)}(\Pi_{m+1} - g, s_{m+1})$, whereas $\pi_{a(g)}(-g, 1) = H_g(a(g)) = 1 - \alpha$ by the definition of $a(g)$. The bound on a single summand in the proof of Lemma 16, at $i = m+1$, and the Cauchy–Schwarz inequality give

$$\begin{aligned}\big|\mathbb{P}\{|G_{n+m+1} - g| \le a(g)\} - (1-\alpha)\big| &\le K_1 \mathbb{E}\big(|g\Pi_{m+1}| + \Pi_{m+1}^2 + \tau_{m+1}\big) \\ &\le K_1\big\{(\tau_{m+1}\mathbb{E}g^2)^{1/2} + 2\tau_{m+1}\big\},\end{aligned}$$

and $\tau_{m+1} \le A^2 m^{1-2\beta}/(2\beta - 1)$ by (34). Combining the two displays with Lemma 18 proves the first inequality in (46); the second follows from $m^{1-2\beta} \le r_m(\beta)$ and $2(xy)^{1/2} \le x + y$. For the three centers, Step 2 of the proof of Theorem 4 gives $\mathbb{E}g^2 \le Cn^{1-2\beta}$. □

D Experimental Design and Complete Results

This appendix specifies the studies of Section 5 and reports their complete results. Appendix D.1 gives the conventions shared by all studies and the Gaussian linear process, Appendices D.2 and D.3 give Studies 1 and 2, Appendix D.4 extends Study 2 to non-Gaussian innovations, and Appendix D.5 gives Study 3.

D.1 Conventions and Simulation

This subsection fixes the conventions of all studies and the simulation of the Gaussian linear process.

All studies use $\alpha = .1$ and the rank $k = \lceil (m+1)(1-\alpha) \rceil$, without interpolation. Each trajectory is one stationary path, whose first n observations form the training block and whose next m observations form the calibration block. A cell, that is, a model with fixed block sizes, consists of 1000 independent trajectories, each generated from its own random stream, a PCG64 generator seeded through NumPy's SeedSequence with a fixed base seed and the identifiers of the cell and the trajectory. A reported value is the average over the 1000 trajectories, and its Monte Carlo standard error is the sample standard deviation divided by $1000^{1/2}$. The studies were run in Python 3.12.3 with NumPy 2.5.3 and SciPy 1.18.1.

The Gaussian linear process of Studies 1–3 is (12) with the coefficients $a_0 = c$ and $a_j = a_{j-1}(1 - \beta/j)$ for $j \geq 1$, where the constant $c > 0$ makes $\sum_j a_j^2 = 1$. These coefficients are positive and asymptotically proportional to $j^{-\beta}$, so that (12) holds, and so does the coefficient lower bound of Proposition 5. The autocorrelation $\rho(h) = \sum_{j \geq 0} a_j a_{j+h}$ satisfies $\rho(0) = 1$ and $\rho(h) = \rho(h-1)(h-\beta)/(h-1+\beta)$ for $h \geq 1$, and it is asymptotically proportional to $h^{1-2\beta}$; the variance of the mean of n consecutive observations, $v_n = n^{-2}\{n + 2\sum_{j=1}^{n-1}(n-j)\rho(j)\}$, is computed exactly. A path of length N is generated exactly by circulant embedding of its covariance matrix, with an embedding whose size is the smallest power of two that is at least $2(N-1)$, so that the path has the stationary law, without truncation of (12) and without burn-in.

D.2 Study 1: Four Dependence Setups

This subsection specifies Study 1 and reports its complete results for CQR and CMR.

For every integer t, let A_t, ξ_t and B_t be independent, with A_t uniform on $[-1, 1]$, ξ_t standard normal and B_t Bernoulli$(1/2)$, and i.i.d. over t. The covariate is $X_t = A_t$ and the response is $Y_t = f(X_t) + s(X_t)h(U_t)$, with f, s and h as in Section 5, where U_t is uniform on $[0, 1]$ in each of the four setups:

(1) independent: $U_t = \Phi(\xi_t)$; (2) AR(1): $U_t = \Phi(H_t)$, $H_t = .9H_{t-1} + .19^{1/2}\xi_t$;
(3) Bernoulli: $U_t = (U_{t-1} + B_t)/2$; (4) long-memory linear process: $U_t = \Phi(G_t)$,

where $G_t = \sum_{j \geq 0} a_j \xi_{t-j}$ in (4), with the coefficients of Appendix D.1 and $\beta = .8$. The process in (3) is $W_t/2$ for the Bernoulli autoregression W_t of Section 3 and is simulated exactly on a 128-bit binary state; (2) and (3) start from their stationary laws. Setups (3) and (4) are not strongly mixing (Appendices B.5 and B.6), and neither is (X_t, Y_t), of which U_t is a function.

The learner is linear quantile regression on the basis $(1, x, x^2)$ over the coefficient box $[-5, 5]^3$, solved exactly as a linear program, at the levels .05 and .95 for CQR and .5 for CMR. Since h is increasing, the conditional τ-quantile of Y_t given $X_t = x$ is $f(x) + s(x)h(\tau)$, so the learner is correctly specified; the oracle CQR interval has the endpoints $f(x) + s(x)h(.05)$ and $f(x) + s(x)h(.95)$, and the oracle CMR interval is $\mu(x) \pm q_\alpha$, with $\mu(x) = f(x) + s(x)h(.5)$ and $q_\alpha = .752$ the $(1-\alpha)$-quantile of $|Y_t - \mu(X_t)|$. Assumptions 1–3 hold in all four setups: Assumption 1 by Lemma 10, since every setup is ergodic and the learner is correctly specified; Assumption 2 because the conditional density of Y_t given X_t lies between $1/2.6$ and 1 on its support (Appendix B.3); and Assumption 3 by Proposition 12, through part (a) with every $\gamma > 1/2$ in setups (1)–(3), whose scores have geometrically decaying functional dependence, and through part (b) with $\gamma = \beta = .8$ in setup (4). Hence $\rho_m = m^{-1/2}$ in setups (1)–(3) and $\rho_m = m^{-.3}$ in setup (4).

The marginal coverage is estimated by averaging over the trajectories the conditional coverage probability of the calibrated interval given the observed past, computed from the known conditional law

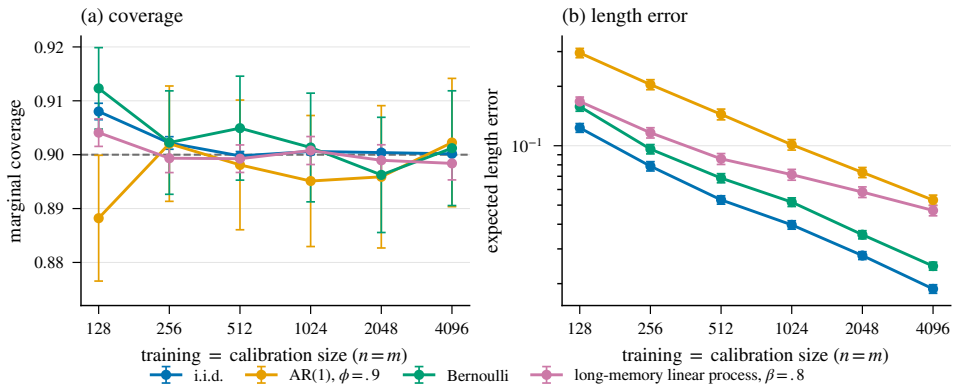

Figure 4: Study 1, CMR with quadratic median regression, on the trajectories of Figure 1. (a) Marginal coverage; the dashed line is the nominal level. (b) Length error $2\mathbb{E}|\widehat{S}^M_{(k)} - q_\alpha|$, on a logarithmic scale. Both panels show pointwise 95% Monte Carlo intervals.

of U_{n+m+1}; this has a smaller Monte Carlo variance than averaging coverage indicators. The length error is $\mathbb{E}\sup_x |\operatorname{len}\widehat{C}^Q_\alpha(x) - \operatorname{len} C^Q_\alpha(x)|$ for CQR and $2\mathbb{E}|\widehat{S}^M_{(k)} - q_\alpha|$ for CMR, and the cells are $n = m \in \{128, \dots, 4096\}$ for each setup.

Figure 1 shows the CQR results and Figure 4 the CMR results. The coverage of both procedures lies between .888 and .912 in all cells; in setup (1), where the scores are exchangeable, it agrees with the exact value $k/(m+1)$, for example $117/129 = .907$ at $m = 128$. The length errors decrease in all four setups. Least-squares fits of the logarithm of the length error on $\log n$ over the four largest sizes give the decay exponents .53, .53, .51 and .51 for CQR and .50, .48, .50 and .29 for CMR, in the order of the setups. In setup (4), over the tested sizes, the length error of CMR decreases like $m^{-.3}$, the rate ρ_m of Theorem 3, whereas that of CQR decreases like $m^{-1/2}$, as in the other setups.

The oracle calibration errors show the same contrast. On the calibration blocks of setup (4), these errors, $\mathbb{E}|S^Q_{(k)}|$ and $\mathbb{E}|S^M_{(k)} - q_\alpha|$, which use the oracle scores and involve no training, decrease with the exponents .50 and .30. At its population threshold, the CQR oracle indicator $\mathbf{1}\{S^Q_t \le 0\} = \mathbf{1}\{|G_t| \le z_\alpha\}$ is an even function of G_t, like the indicators of Section 4, so its first Hermite coefficient vanishes and its covariances decay like $\rho(h)^2$ (Lemma 13), which leads to the rate $m^{-1/2}$ at $\beta = .8$. The CMR indicator is not even in G_t, because h is not linear, and its nonzero first Hermite coefficient leads to the rate $m^{1/2-\beta} = m^{-.3}$. This explains the oracle calibration rates; the fitted CQR length can also retain a first-order training-error term, so the observed slope does not establish its asymptotic rate.

D.3 Study 2: The Sharper Length Rate

This subsection specifies Study 2 and the three numbers of Table 1.

The observations are $Y_t = \mu + G_t$, with $\mu = 0$, which is no loss of generality because the procedure is translation equivariant, and with G_t the Gaussian linear process of Appendix D.1 for $\beta \in \{.6, .75, .9\}$. On each path of length $n + m$, the center is the training sample mean $\widehat{\mu} = \bar{Y}_n$, the calibrated radius $\widehat{S}^M_{(k)}$ is the k-th smallest of $|Y_{n+i} - \widehat{\mu}|$, and the known-center radius $S^M_{(k)}$ is the k-th smallest of $|Y_{n+i} - \mu|$. The cells are $n = m \in \{128, \dots, 4096\}$. Figure 2 shows the three quantities of the study: the mean squared error $\mathbb{E}(\widehat{\mu} - \mu)^2$ of the center, which equals v_n; the oracle calibration error $\mathbb{E}|S^M_{(k)} - z_\alpha|$; and the length error $\mathbb{E}|\operatorname{len}\widehat{C}^M_\alpha - 2z_\alpha| = 2\mathbb{E}|\widehat{S}^M_{(k)} - z_\alpha|$.

The three numbers of Table 1 summarize the rates of these quantities, each computed over the five largest sizes $n = m \in \{256, \dots, 4096\}$. The estimation error exponent checks the rate of the center: it is minus the least-squares slope, with an intercept, of the logarithm of the mean squared error on $\log n$, to be compared with $2\beta - 1$. The oracle calibration error ratio checks whether

Table 1: Study 2 and its extension, rates over $n = m \in \{256, \dots, 4096\}$, with Monte Carlo standard errors in parentheses; the reference is $2\beta - 1$ for the estimation error exponent and one for the two ratios. The center is the training mean for Gaussian innovations and the training median for the non-Gaussian innovations of Appendix D.4.

Innovations	β	$2\beta - 1$	Estimation error exponent	Oracle calibration error ratio	Total length error ratio
Gaussian	.60	.20	.211 (.021)	1.029 (.055)	.986 (.046)
	.75	.50	.544 (.019)	1.118 (.028)	1.110 (.028)
	.90	.80	.806 (.021)	1.056 (.022)	1.080 (.022)
Laplace	.60	.20	.202 (.021)	.905 (.055)	1.043 (.045)
	.75	.50	.459 (.020)	1.095 (.026)	1.106 (.027)
	.90	.80	.829 (.020)	1.054 (.022)	1.057 (.023)
centered exponential	.60	.20	.206 (.019)	.973 (.056)	1.026 (.044)
	.75	.50	.487 (.021)	.720 (.026)	.686 (.026)
	.90	.80	.802 (.020)	.891 (.022)	.876 (.022)

the oracle calibration error decreases at the rate $r_m(\beta)$ of (13): it is the least-squares slope of the logarithm of this error on $\log r_m(\beta)$, which equals the ratio of the decay exponent of the error to that of $r_m(\beta)$ when both are powers of m; the value one means that the error decreases at the rate $r_m(\beta)$, and a larger value that it decreases faster. The total length error ratio is the same number for the length error, on $\log r_n(\beta)$, and checks whether the length error decreases at the rate of Theorem 4. The Monte Carlo standard error of each number is its standard deviation over 2000 bootstrap replications that resample the 1000 trajectories of each cell with replacement, keeping the quantities of a trajectory together.

The Gaussian rows of Table 1 report the three numbers. The estimated mean squared errors agree with the exact v_n within 1.7 Monte Carlo standard errors in all 18 cells, and the estimation error exponents .211, .544 and .806 are close to $2\beta - 1$: the center converges slowly. The two ratios are close to one. At $\beta = .6$, they correspond to decay exponents of about .21 and .20, against .1 for the rate $m^{1/2-\beta}$ of Theorem 3, and at $\beta = .9$ to about .53 and .54, against .5 for $r_m(\beta)$ and .4 for $m^{1/2-\beta}$. At $\beta = .75$ the ratios 1.118 and 1.110 exceed one: over this range the errors decrease faster than $\{\log(m+1)/m\}^{1/2}$. The oracle calibration error and the length error therefore follow the rates of Theorem 4 and decrease faster than the center converges.

D.4 Extension of Study 2: Non-Gaussian Innovations

Section 4 proves the sharper rate for Gaussian innovations and attributes it to the symmetry of their law. This subsection repeats Study 2 with two non-Gaussian innovation laws and the same coefficients. The Laplace law is symmetric, and tests whether the improvement extends beyond Gaussian innovations, as Section 4 expects. The centered exponential law is asymmetric by design, and tests whether the improvement rests on symmetry, in which case the length error should approach the rate of Theorem 3.

The observations are $Y_t = G_t = \sum_{j \geq 0} a_j \varepsilon_{t-j}$, with the coefficients of Appendix D.1, $\beta \in \{.6, .75, .9\}$ and i.i.d. innovations with mean zero and variance one: Laplace with scale $2^{-1/2}$, or $\mathrm{Exp}(1) - 1$, which has skewness 2. The center is the empirical median of the training block, which estimates the center c of the CMR oracle interval $c \pm q$, the median of G_t; the radius q is the $(1-\alpha)$-quantile of $|G_t - c|$, computed by numerical inversion of the characteristic function of G_t with an error below 2×10^{-6}. The cells are $n = m \in \{128, \dots, 4096\}$, and the quantities and the three numbers are those of Study 2, with the known center c in the oracle calibration error. Paths with non-Gaussian innovations cannot be generated by circulant embedding. For a path of length N and L the smallest power of two that is at least $\max(2^{20}, 64N)$, we compute $G_t = H_t + R_t$ for $1 \leq t \leq N$, where $H_t = \sum_{s=1-L}^{t} a_{t-s} \varepsilon_s$ is obtained exactly from $L + N$ simulated innovations by a fast Fourier convolution, and the remote part $R_t = \sum_{s \leq -L} a_{t-s} \varepsilon_s$ is replaced by a Gaussian vector with its exact covariance. This keeps the covariance of the path and changes only the higher

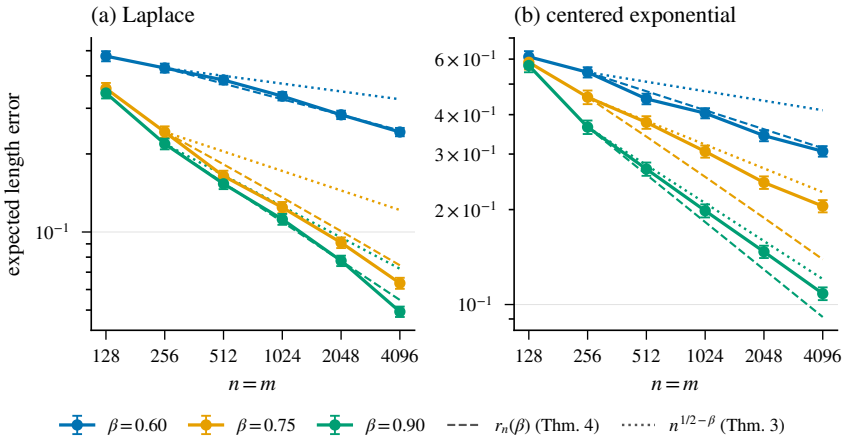

Figure 5: Extension of Study 2, CMR with the empirical median as center, for (a) Laplace and (b) centered exponential innovations and $\beta \in \{.6, .75, .9\}$: length error with $n = m$, with the rate $r_n(\beta)$ of Theorem 4 (dashed) and the rate $n^{1/2-\beta}$ of Theorem 3 (dotted), matched to the empirical curves at 256. Pointwise 95% Monte Carlo intervals.

cumulants of R_t, a weighted sum of infinitely many innovations with small weights. With Gaussian innovations, the same algorithm reproduces Study 2 within Monte Carlo error.

Figure 5 shows the length error, and Table 1 gives the three numbers. With Laplace innovations, the findings of Study 2 hold: the estimation error exponents are close to $2\beta - 1$, and the oracle calibration and total length error ratios are close to one, between .905 and 1.106. With centered exponential innovations, the total length error ratios at $\beta = .75$ and .9 are .686 and .876, well below one, whereas the same ratios computed with the rate $n^{1/2-\beta}$ of Theorem 3 in place of $r_n(\beta)$ are 1.171 and 1.095, close to one: the length error follows the rate of Theorem 3 more closely than that of Theorem 4. At $\beta = .6$, however, the total length error ratio is 1.026, and 2.051 with the rate of Theorem 3.

The exception at $\beta = .6$ has a simple source: long memory makes the law of G_t more symmetric. Under the centered exponential law, the skewness of G_t is $2\sum_{j\geq 0} a_j^3$, which equals .76, 1.60 and 1.95 at $\beta = .6$, .75 and .9, since $a_0 = .70$, .92 and .99: the stronger the memory, the more the coefficients spread each innovation over many lags, and the closer the law of G_t is to a Gaussian law. The asymmetry enters the length through the first-order term of the population radius. Let f be the density of G_t, c its median and q the oracle radius, and let $a(g)$ now denote the $(1-\alpha)$-quantile of $|G_t - c - g|$, so that $a(0) = q$; implicit differentiation gives

$$a'(0) = \frac{f(c-q) - f(c+q)}{f(c-q) + f(c+q)},$$

which vanishes for a symmetric law, as in Appendix C.2. The densities $f(c-q)$ and $f(c+q)$ at the oracle endpoints are .078 and .106, .005 and .108, and below 10^{-4} and .101 at $\beta = .6$, .75 and .9, so that $|a'(0)| = .15$, .91 and 1.00. By the reduction principle for the empirical process of long-memory linear processes (Wu, 2003), the empirical distribution function of the calibration block differs from the distribution function of G_t at x, to first order, by $-f(x)\bar{G}_m$, where $\bar{G}_m = m^{-1}\sum_{i=1}^{m} G_{n+i}$. The known-center radius, the k-th smallest of $|G_{n+i} - c|$, therefore has the first-order error $-a'(0)\bar{G}_m$, whose expected absolute value is about $|a'(0)|(2v_m/\pi)^{1/2}$.

At $m = 4096$, this first-order part is about .05, .09 and .03 at $\beta = .6$, .75 and .9, against observed oracle calibration errors of .16, .10 and .05. At $\beta = .75$ and .9 it is the larger part of the error, which then decreases at about the rate $m^{1/2-\beta}$. At $\beta = .6$ the error is dominated by the part that Gaussian innovations also produce, about .14 at $m = 4096$, which decreases like $r_m(\beta) = m^{-.2}$, whereas the first-order part decreases only like $m^{-.1}$ and would dominate only for m of the order of 10^8. Accordingly, the total length error ratios computed with the rate of Theorem 3, 2.05, 1.17 and 1.09 at $\beta = .6$, .75 and .9, decrease as the skewness grows. Hence the more symmetric the law of the

series, the closer the length error is to the rate of Theorem 4, and the more skewed, the closer it is to the rate of Theorem 3. The extension thus supports both the expectation of Section 4 for symmetric innovations and its dependence on symmetry.

D.5 Study 3: The Lower Bound

This subsection specifies Study 3 and reports its complete results. The study illustrates the training term of the lower bound (16) of Proposition 5: for a fixed training block, the length error remains at least of the order of the mean squared error of the center, however long the calibration block.

The model is that of Study 2 with $\beta = .9$ and the sample mean as center. Proposition 5 gives the lower bound when $m \geq Kn^{\max(1,4\beta-2)} \log(n+1)$, which is $m \geq Kn^{1.6} \log(n+1)$ at $\beta = .9$, for a constant K that the proposition does not specify. The cells are therefore $n \in \{64, 128, 256, 512, 1024\}$ and $m = 2^{\lceil \log_2(Kn^{1.6}) \rceil}$ for $K \in \{1, 4, 16, 64\}$, from 1024 to 4,194,304. The grid omits the factor $\log(n+1)$, which varies only by a factor 1.66 over the tested n, less than the factor 4 between consecutive values of K.

The quantity of interest is the length error $\mathbb{E}|\operatorname{len} \widehat{C}_\alpha^M - 2z_\alpha| = 2\mathbb{E}|\widehat{S}_{(k)}^M - z_\alpha|$. To separate the contribution of the center from that of the calibration, let $g = \bar{G}_n$ be the error of the training mean and $a(g)$ the radius of Appendix C, the $(1-\alpha)$-quantile of $|Z - g|$ for $Z \sim N(0,1)$. Taking expectations in (32) gives

$$\left| 2\mathbb{E}|\widehat{S}_{(k)}^M - z_\alpha| - 2\mathbb{E}\{a(g) - z_\alpha\} \right| \leq 2\mathbb{E}|\widehat{S}_{(k)}^M - a(g)|.$$

The term $2\mathbb{E}\{a(g) - z_\alpha\}$, which we call the reference, is the length error of the interval calibrated on an infinitely long block, and it depends only on the center. The right-hand side, which we call the calibration error, is the error of calibrating on a block of length m. The display thus states that the length error differs from the reference by at most the calibration error. Since $g \sim N(0, v_n)$ exactly, the reference equals $4\int_0^\infty \{a(v_n^{1/2}u) - z_\alpha\}\phi(u)\,du$ and is computed by quadrature; by Lemma 15(iv) it is asymptotically $z_\alpha v_n$, and its values .054, .031, .018, .010 and .006 for $n = 64, \ldots, 1024$ are within 2.2% of $z_\alpha v_n$. The calibration error is estimated on each trajectory, with $a(g)$ solved for the realized g.

Figure 3 reports all 20 cells. As K grows, the calibration error halves from each value of K to the next, as the calibration rate $m^{-1/2}$ predicts, for example from .078 at $K = 1$ to .040, .020 and .010 at $n = 64$, and the length error approaches the reference at every training size. With $K = 64$, the calibration error is at most .20 times the reference, so that the length error lies within about 20% of the reference; it is in fact between .997 and 1.080 times the reference, and from $K = 16$ to $K = 64$ it changes by less than 1.96 standard errors of the change at every size. With $K = 64$, the length error decreases from .056 at $n = 64$ to .0065 at $n = 1024$, a factor of 8.57, against the factor 9.19 of $n^{-.8}$ over the same range. Once the calibration is accurate, the length error therefore stays at the level $z_\alpha v_n$ set by the center, which decreases only as the training block grows, at the rate $n^{-(2\beta-1)}$ of the mean squared error of the center.

E Application: Day-Ahead Load Forecast Errors

Every morning, the New York Independent System Operator (NYISO) publishes a forecast of the hourly electricity load of the New York control area (NYCA) for the following day (New York Independent System Operator, 2026b), and it later publishes the load that was actually served (New York Independent System Operator, 2026a). A prediction interval around the forecast states how far the actual load may fall from it, and such probabilistic load forecasts have become important to the planning and operation of energy systems (Hong & Fan, 2016). We construct the intervals by split conformal calibration of the percentage forecast error $Y = 100(A - F)/F$, where A is the actual load and F the forecast, with one daily series for each hour of the day. The errors are persistent and, as we show below, have long memory, so exchangeability fails, whereas the theory of Section 3 in principle allows for long memory. We use data from 2005 on, because the median error shifts by about 2.5 percentage points in January 2005. The years 2005–2013 served to choose and fix the design, which was then applied once to the 111,010 hourly targets from 1 January 2014 to 31 August 2026, to measure how often the 90% intervals contain the actual load and how wide they are.

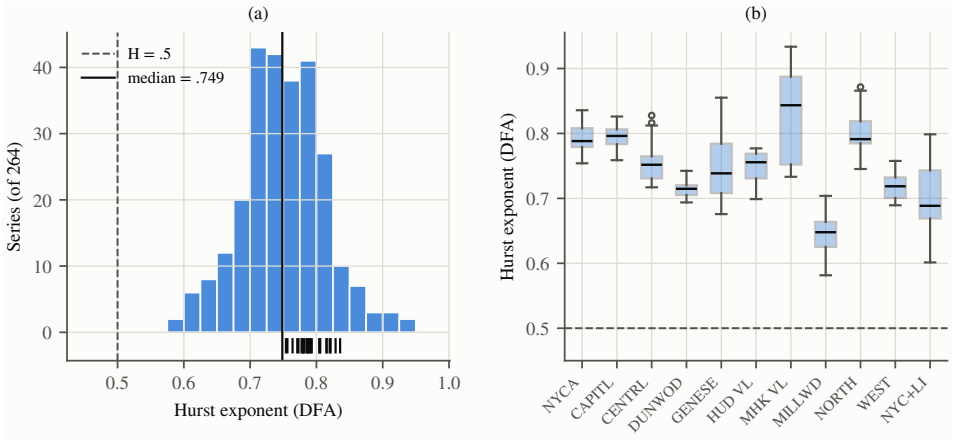

Figure 6: Application, Hurst exponents: (a) histogram; (b) box plots of the estimates by area.

For the target day D, the calibration block consists of the m days ending on $D-2$, all of whose hours have ended before the forecast for D is posted, and the training block of the n days immediately before it, with $n = m \in \{365, 730, 1461\}$. The target is thus two days after the calibration block, the fixed horizon $h = 2$ of Appendix A.4. The origin rolls daily, and every fit is recomputed at every origin. Missing days are skipped, and $k = \lceil (m' + 1)(1 - \alpha) \rceil$ with $\alpha = .1$ and m' the number of available calibration days.

- CMR uses a constant covariate. The center $\widehat{\mu}$ is the training median of Y, the radius $\widehat{S}^M_{(k)}$ is the k-th smallest value of $|Y_t - \widehat{\mu}|$ over the calibration block, and the interval for the load is $[F\{1 + (\widehat{\mu} - \widehat{S}^M_{(k)})/100\},\ F\{1 + (\widehat{\mu} + \widehat{S}^M_{(k)})/100\}]$.
- CQR fits linear quantile regressions of Y at the levels .05 and .95 on the forecast, standardized on the training block, and on two annual harmonics of the day of the year d, $\cos(2\pi jd/365.25)$ and $\sin(2\pi jd/365.25)$ for $j = 1, 2$, and calibrates them with the score of Section 2.1. Each fit is a linear program, accepted only if a feasible dual solution certifies its objective to within 10^{-6}; no fit failed.
- The width of the hindsight interval, defined after the evaluation, is a proxy for the oracle width. With a constant covariate, the oracle CMR interval of Section 2.2 is $\mu \pm q_\alpha$, with μ the median of Y under the stationary law and q_α the $(1 - \alpha)$-quantile of $|Y - \mu|$, and its width is $2q_\alpha$. CMR estimates μ from the n training errors and q_α from the m calibration errors. The hindsight interval $\tilde{\mu} \pm \tilde{q}$ of an hour estimates both from all N errors of that hour in the evaluation period ($N \geq 4613$): $\tilde{\mu}$ is their median and $\tilde{q}$ the $\lceil (N+1)(1-\alpha) \rceil$-th smallest value of $|Y - \tilde{\mu}|$. Under the model of Section 2, the errors are ergodic (Appendix B.1), so $\tilde{\mu}$ and $\tilde{q}$ converge to μ and q_α as N grows when μ is unique and Assumption 2 holds; for finite N, the hindsight width $2\tilde{q}$ is an estimate of $2q_\alpha$ with an error of its own. The hindsight interval uses the targets themselves, so it cannot be used for prediction, and it covers .900 of them by construction.

Hurst exponents are estimated by detrended fluctuation analysis (Peng et al., 1994) for the 264 hourly error series of NYCA and its ten zone areas (the eleven NYISO load zones, with New York City and Long Island combined) over the evaluation period. Figure 6 shows their histogram, with the 24 NYCA series marked below the axis, and their box plots by area, 24 hourly series each: all 264 estimates exceed .5, with median .749, so the errors have long memory.

Figure 7 shows the results. At $n = m = 730$, CMR and CQR both cover .895 of the 111,010 targets, .005 below the nominal level, and with $n = m = 365$ and 1461 their coverage lies between .886 and .899. Coverage varies over the years, from .833 in 2016 to .934 in 2019 for CMR. The mean width of CMR is 5.86%, 5.81% and 5.90% of the forecast for $n = m = 365$, 730 and 1461, or 1017 MW at 730, and that of CQR 6.01%, 5.86% and 5.96%. The mean width of the hindsight interval is 5.91%, or 1036 MW (Figure 7(c)), and 5.82% and 5.79% when it is computed from 2014–2019

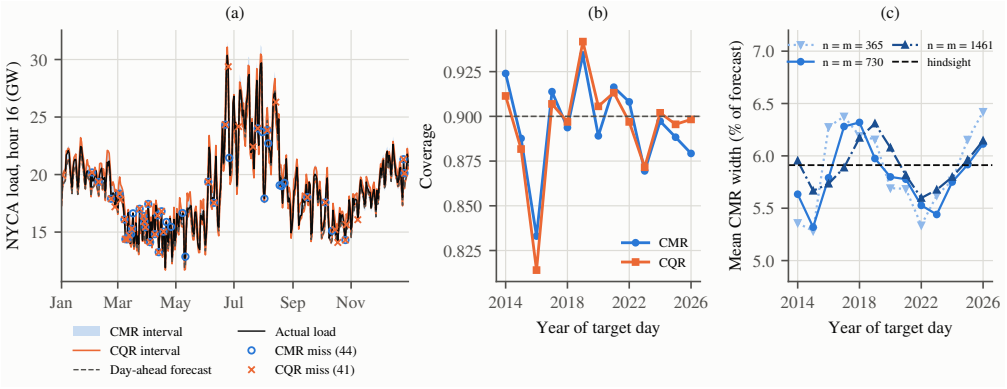

Figure 7: Application, NYCA, evaluation period 2014–2026. (a) Hour beginning 16:00, target days in 2025, $n = m = 730$: actual load, forecast, CMR interval (band) and CQR interval (lines), in GW, with the misses of each method marked. (b) Coverage by year of CMR and CQR, $n = m = 730$, over all 24 hours; the dashed line is the nominal level; 2026 covers January to August. (c) Mean width of CMR by year, in percent of the forecast, for $n = m = 365$, 730 and 1461; the dashed line is the mean width of the hindsight interval.

and from 2020–2026 only. On average, and relative to the mean hindsight width, the CMR width at a single origin with $n = m = 730$ differs from the hindsight width of its hour by 9%, and the hindsight widths of an hour computed from 2014–2019 and from 2020–2026 differ by 13%.

The results make two points. First, although the errors have long memory, both procedures cover close to the nominal level of 90% over more than twelve years of targets that were not used to design them. Second, and more important, the mean width of CMR is within 2% of the hindsight width at every block size: to the accuracy of this proxy, the calibrated interval is on average as wide as the oracle interval. Since the hindsight width is itself an estimate, the comparison concerns average widths, not the width at a single origin.