mathematics

MDPI

Article

Conformal Prediction Intervals for Semi-Functional Partial Linear Regression Under β-Mixing Dependence

Jeza Allohibi [1,2]

[1] Department of Mathematics, College of Science, Taibah University, Madinah 42353, Saudi Arabia; jlohibi@taibahu.edu.sa
[2] Energy, Industry, and Advanced Technologies Research Center, Taibah University, Madinah 42353, Saudi Arabia

Abstract

We study prediction intervals for the semi-functional partial linear model (SFPLM) under stationary, geometrically β-mixing dependence. We analyze a split conformal procedure based on a three way data partition with buffer gaps, a functional principal component projection semi-metric on the functional covariate, profiled least squares estimation of the parametric component, kernel estimation of the nonparametric component and of the conditional standard deviation, and a studentized absolute residual score. Marginal validity of split conformal prediction with a trained score under β-mixing is available from generic results of Oliveira et al. and of Barber and Pananjady, without any buffer and without accuracy requirements on the fitted estimators. Our main result is complementary to those guarantees: a finite sample marginal lower coverage bound whose theoretical finite-sample coverage penalty decomposes additively into seven interpretable components expressed in the structural primitives of the SFPLM, quantifying the price of replacing the ideal SFPLM score by the estimated score inside the proof. The penalty plays no role in the computation of the interval, involves unknown structural constants, and is not an operational correction. The bound requires no parametric error model, but it is not assumption free; it holds under explicit structural conditions, including geometric β-mixing, conditionally centered sub-Gaussian errors, a fractal small ball regime for the projected functional covariate, and local regularity of the score distribution. Simulations under a protocol frozen before outcome computation, spanning mild and strong dependence, a misspecification stress test, and a dependent-score design with exactly quantified score autocorrelation, show near nominal coverage for all methods, with the gapped and contiguous variants statistically indistinguishable in coverage. Studentization showed no systematic coverage advantage, while interval-length differences were systematic. The value of the analysis lies in the explicit model-specific estimation layer of the coverage decomposition, not in a numerical gain over naive split conformal.

Keywords: conformal prediction; semi-functional partial linear regression; β-mixing dependence; functional principal component projection; small ball probability; prediction intervals; functional data analysis; studentized residual

MSC: 62G08; 62G15; 62R10; 62M10; 62G20

Check for updates

Academic Editor: Brani Vidakovic

Received: 13 June 2026
Revised: 2 August 2026
Accepted: 13 August 2026
Published: 4 September 2026

Copyright: © 2026 by the author. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

1. Introduction

Functional partial linear regression has emerged over the past two decades as a flexible model class for problems in which a scalar response depends jointly on a finite dimensional

covariate vector and a square integrable random function. The semi-functional partial linear model (SFPLM) of [1] writes

$$Y_i = X_i^\top \beta_0 + m(W_i) + \varepsilon_i, \qquad i = 1, \ldots, n, \tag{1}$$

where $X_i \in \mathbb{R}^p$, W_i takes values in a separable Hilbert space $\mathcal{H}$, $\beta_0 \in \mathbb{R}^p$ is a parameter of interest, $m : \mathcal{H} \to \mathbb{R}$ is a smooth nonparametric regression operator, and ε_i is an unobserved error. The SFPLM combines the interpretability of a parametric scalar component with the flexibility of a nonparametric functional component, and has been studied in chemometrics (near infrared spectroscopy [2]), in biomedical signal analysis (electroencephalography), and in functional time series prediction. Estimation theory for the SFPLM is by now well developed. $\sqrt{n}$-asymptotic normality of $\widehat{\beta}$ and almost complete convergence of $\widehat{m}(\cdot)$ are available under independent observations [1], under α-mixing dependence [3], and under various missing data and functional time series structures [4]. Distribution free inference (in the sense of requiring no parametric error model) for the response Y_{n+1} at a new (X_{n+1}, W_{n+1}), however, has received much less attention.

In applications where $\{(X_i, W_i, Y_i)\}_{i=1}^n$ arises as a stationary β-mixing time series, the practitioner typically requires a prediction interval for Y_{n+1} rather than a point estimate, together with a finite sample coverage guarantee that does not rely on a parametric error distribution or on a normal approximation. The standard plug in interval $\widehat{Y}^{\mathrm{pred}} \pm z_{1-\alpha/2}\,\check{\sigma}(W_{n+1})$ requires correct specification of the conditional error distribution; bootstrap intervals require correct calibration of the resampling scheme [5]. Conformal prediction [6] provides an attractive distribution free alternative, but its standard validity guarantee assumes that the calibration sample and the future point are jointly exchangeable, an assumption that is violated by the temporal dependence intrinsic to the applications above.

A growing body of work addresses conformal prediction beyond exchangeability. For finite-dimensional time series, ref. [7] establish approximate validity of block-permutation conformal inference under stationarity and weak dependence (see also [8] for the related counterfactual and synthetic control setting); refs. [9,10] propose ensemble batch prediction intervals with asymptotic coverage; refs. [11,12] study online and adaptive variants. For functional data, ref. [13] construct distribution free prediction bands under exchangeability of functional observations, and ref. [14] develop conformal bands for functional time series. Most recently, ref. [15], abbreviated BCRT 2023 below, give a unified non exchangeability bound. The conformal coverage gap is at most an average of total variation distances between the joint law and its position swapped versions. This generic bound applies to any non exchangeable setting, including stationary β-mixing data, but the resulting penalty is not a function of the structural primitives of the SFPLM problem; in particular, it does not involve the functional small ball probability $\psi(h)$, the kernel bandwidth h_n, the FPC projection dimension q, or the studentization rate. To be operational in a SFPLM context, the BCRT framework must be specialized to surface those primitives explicitly. Most directly, Oliveira, Orenstein, Ramos, and Romano [16] establish generic split conformal coverage statements under non exchangeability, including stationary β-mixing settings, for an arbitrary score function fitted on the training block (Theorems 1 and 4 of the published version). Very recently, Barber and Pananjady [17] analyzed unmodified split conformal under temporal dependence through switch coefficients. For a score trained on an initial block of the same series, their Theorem 4 with Corollary 2 gives a finite sample coverage bound under stationary β-mixing, and their Theorem 2 proves a matching lower bound result in the pretrained memoryless setting. These generic results settle marginal validity of split conformal for dependent data at the level of a trained score, with no buffer requirement and no requirement on the accuracy of the fitted regression or

scale estimators. What they do not provide is a decomposition expressed in the estimation primitives of a specific semiparametric model. Our contribution is a SFPLM specific analysis in which the coverage penalty is decomposed into terms involving the FPC semi-metric, the small ball probability, the bandwidth, the buffer gap, and the studentization rate. The dependence-only components of this decomposition overlap conceptually with the generic dependence-only bounds of [16,17], which may be sharper under their own conditions; they are not the source of the model-specific novelty here.

To the best of our knowledge, a finite sample conformal coverage penalty decomposition of this form has not been developed for the semi-functional partial linear model under β-mixing dependence. The present paper undertakes such a specialization. We propose a dependence aware split conformal prediction procedure for the SFPLM under stationary geometric β-mixing, based on a three way data partition (training, gap, calibration, gap, test) with a buffer length g_n chosen to control conditional dependence between the training estimators and the calibration scores. Estimation uses a profiled least squares estimator for β_0, a functional Nadaraya Watson estimator for m on the FPC projection semi-metric $d_q(w, w') = \|P_q(w - w')\|_{L^2}$, and a kernel estimator for the conditional standard deviation $\sigma(\cdot)$. The conformal score is the studentized absolute residual.

Our principal result is a finite sample coverage statement with a theoretical finite-sample coverage penalty. Under the assumptions stated in Section 2, with $\psi(h) \asymp h^q$ in a fractal small ball regime and the gap g_n chosen logarithmically in the calibration sample size, the proposed procedure satisfies

$$\mathbb{P}\big(Y_{n+1} \in \widehat{C}_n(X_{n+1}, W_{n+1})\big) \;\geq\; 1 - \alpha - \Delta_n, \tag{2}$$

where Δ_n decomposes additively into seven interpretable components, namely a β-mixing Yu-coupling cost, a calibration empirical process cost under the block independent surrogate, a gap cost induced by the buffer between training, calibration and test, a regression perturbation cost reflecting the uniform error of the plug in estimator $\widehat{m}$, a studentization cost reflecting the uniform error of $\check{\sigma}$, the standard split conformal rounding $1/(n_2 + 1)$, and a polynomially small good event failure probability. Each summand is a function of quantities chosen by the user (h_n, g_n, n_1, n_2), of the proof-side block lengths $(b_n, b_{2,n})$ of (A8′) and (C3), and of structural quantities $(\psi(\cdot), \beta(\cdot), s, M_S, \sigma_{\min})$ that arise from the SFPLM model. Δ_n quantifies, inside the proof, the gap between the nominal level and the guaranteed marginal coverage; it plays no role in the computation of the interval, and because its summands involve unknown structural constants it cannot be evaluated numerically or used as a correction in practice. Detailed expressions are given in Sections 2 and 4; complete proofs appear in the Theoretical Supplement.

We summarize the contributions.

(i) A dependence aware split conformal procedure for the SFPLM under β-mixing, with a three way partition, buffer length g_n, and studentized residual score on the FPC projection semi-metric.

(ii) A finite sample coverage bound of the form $\mathbb{P}(Y_{n+1} \in \widehat{C}_n) \geq 1 - \alpha - \Delta_n$, with the theoretical finite-sample coverage penalty Δ_n decomposing into dependence, gap, regression perturbation, studentization, rounding, and good event failure terms, each expressed in a structural primitive of the SFPLM. The model-specific content of the bound is its estimation-aware layer, which prices the transfer from the ideal SFPLM score to the implemented estimated score; the relation of its dependence-only components to the generic bounds of [16,17] is discussed in Section 8.3.

(iii) A uniform deviation theory for the SFPLM kernel estimator on the FPC projection semi-metric under β-mixing, including oracle, plug in, and variance estimation results,

formulated as high probability statements with explicit failure event probabilities suitable for direct insertion into the conformal coverage analysis; it appears to be new in the finite sample β-mixing form stated here and is of independent interest.

(iv) A simulation study, pre-specified and frozen before outcome computation, covering mild and strong dependence regimes, a boundary stress test, and a dependent-score design whose ideal score has exactly quantified serial dependence, together with gap and projection sensitivity studies and dependence diagnostics. All methods, gapped and contiguous, studentized and plain, attain near nominal coverage in these regimes; the evidence and its honest interpretation are reported in Sections 6 and 8.

The contribution is not a generic conformal framework, but a problem specific specialization for the SFPLM in which the coverage penalty is expressed through the functional small ball probability, the bandwidth, the FPC projection dimension, the mixing rate, and the studentization error. Three limitations should be flagged at the outset. First, the guarantee is distribution free only in a restricted sense. No parametric error model and no normal approximation are imposed, but the bound is not assumption free, and the structural conditions of Section 2 (geometric β-mixing, conditionally centered sub-Gaussian errors, the fractal small ball regime, a fixed FPC dimension q, a variance floor $\sigma_{\min}$, and local regularity of the score density near its $(1-\alpha)$-quantile) are essential to the proof. Second, in the simulation regimes of Section 6 the proposed procedure does not show a numerical advantage over naive split conformal. Both attain the nominal level and produce intervals of comparable length, and we report this openly. Third, Δ_n is not claimed to be sharp, and its dependence-only components are not claimed to improve on generic bounds; the contribution of Δ_n is its interpretability, in that each component names the model primitive it depends on. The theory addresses fixed p and fixed FPC dimension q, and the autoregressive residual diagnostic in Section 6 violates the conditional noise structure of our assumptions and is reported as a boundary stress test; high dimensional and growing-q extensions are noted as open questions in Section 8.

The remainder of the paper is organized as follows. Section 2 fixes the model, the FPC semi-metric, the small ball regime, and the assumption block; Section 3 specifies the procedure and tuning rules; Section 4 states the coverage theorem, the supporting uniform deviation theorems, and the auxiliary lemmas; Section 5 outlines the proof strategy, whose technical centerpiece is a tailored β-mixing deviation bound for SFPLM scores in the FPC fractal regime, with complete proofs in the Theoretical Supplement; Sections 6 and 7 report the simulation study and the Tecator illustration; and Sections 8 and 9 discuss related work and conclude.

2. Model, Notation and Assumptions

2.1. Model and Data Structure

Let $\mathcal{H}$ denote a separable real Hilbert space, typically $L^2([0,1])$ with the standard inner product. The observations are indexed by $i = 1, 2, \ldots$ and take values in $\mathbb{R}^p \times \mathcal{H} \times \mathbb{R}$ for a fixed integer $p \geq 1$. We write $Z_i = (X_i, W_i, Y_i)$ for the i-th observation. The semi-functional partial linear model assumes

$$Y_i = X_i^\top \beta_0 + m(W_i) + \varepsilon_i, \qquad i = 1, 2, \ldots, \tag{3}$$

where $\beta_0 \in \mathbb{R}^p$ is an unknown parameter vector, $m : \mathcal{H} \to \mathbb{R}$ is an unknown nonparametric regression operator, and ε_i is an unobserved scalar error with $\mathbb{E}[\varepsilon_i \mid X_i, W_i] = 0$. The conditional variance function is denoted by

$$\sigma^2(w) := \mathbb{E}\Big[\varepsilon_i^2 \,\Big|\, W_i = w\Big], \qquad w \in \mathcal{H}, \tag{4}$$

and the conditional standard deviation by $\sigma(w) := \sqrt{\sigma^2(w)}$. The sequence $\{Z_i\}_{i\geq 1}$ is assumed strictly stationary throughout.

Given an observed sample $\{Z_i\}_{i=1}^n$ and a future point Z_{n+1} drawn from the same stationary law, the prediction problem we address is the construction of a measurable interval $\widehat{C}_n(X_{n+1}, W_{n+1}) \subseteq \mathbb{R}$, depending only on $\{Z_i\}_{i=1}^n$ and on (X_{n+1}, W_{n+1}), satisfying

$$\mathbb{P}\big(Y_{n+1} \in \widehat{C}_n(X_{n+1}, W_{n+1})\big) \;\geq\; 1 - \alpha - \Delta_n \tag{5}$$

for a user specified miscoverage level $\alpha \in (0,1)$ and a theoretical coverage penalty Δ_n that vanishes under the asymptotic regime of Section 2.6.

2.2. β-Mixing Dependence

Temporal dependence is modeled through the β-mixing coefficient. For a strictly stationary sequence $\{Z_i\}$, let $\sigma(Z_a, \ldots, Z_b)$ denote the σ-algebra generated by $Z_a, \ldots, Z_b$. The β-mixing coefficient at lag $k \geq 1$ is

$$\beta(k) := \sup_{j\geq 1} \mathbb{E}\left[\sup_{A \in \sigma(Z_{j+k}, Z_{j+k+1}, \ldots)} \big| \mathbb{P}\big(A \,\big|\, \sigma(Z_1, \ldots, Z_j)\big) - \mathbb{P}(A) \big| \right]. \tag{6}$$

We work throughout under the geometric mixing rate

$$\beta(k) \;\leq\; B_0\, \rho^k, \qquad k \geq 1, \tag{7}$$

for constants $B_0 > 0$ and $\rho \in (0,1)$. This regime is standard in functional time-series analysis; we refer to [2,3] for examples in the SFPLM setting.

2.3. Functional Semi-Metric and FPC Projection

The semi-metric used for kernel estimation in $\mathcal{H}$ is the FPC projection semi-metric. Let Σ_W denote the covariance operator of W on $\mathcal{H}$, and let $\{\phi_k\}_{k\geq 1}$ be an L^2-orthonormal sequence of eigenfunctions of Σ_W ordered by decreasing eigenvalues. For a fixed integer $q \geq 1$ we write P_q for the orthogonal projection onto $\operatorname{span}\{\phi_1, \ldots, \phi_q\}$ and define

$$d_q(w, w') \;:=\; \big\| P_q(w - w') \big\|_{L^2}, \qquad w, w' \in \mathcal{H}. \tag{8}$$

The map d_q is a semi-metric on $\mathcal{H}$. It is symmetric and satisfies the triangle inequality, but $d_q(w, w') = 0$ does not imply $w = w'$. The FPC dimension q is held *fixed* throughout this paper; growing-q regimes are outside the scope of the present work and are noted in Section 8 as an open direction.

For $w \in \mathcal{H}$ and $h > 0$, write $\mathbb{P}_w(h) := \mathbb{P}(d_q(W, w) \leq h)$. Let $\mathcal{S}_F \subset \mathbb{R}^q$ denote the projected covariate support set defined in (A5*) below, identified with the support of $P_q W$ in the q-dimensional FPC coordinate system; uniform deviation theory in this paper is developed over $\mathcal{S}_F$. Let $\mathcal{S}_F^+ := \{w \in \mathcal{H} : P_q w \in \mathcal{S}_F\}$ as in (A5*+); $\mathcal{S}_F^+$ is identified through P_q with $\mathcal{S}_F$, so any quantity uniform over $\mathcal{S}_F^+$ depends on w only through $P_q w$. Set

$$\psi(h) \;:=\; \inf_{z \in \mathcal{S}_F} \mathbb{P}\big(d_q(W, P_q^{-1} z) \leq h\big), \qquad \Psi(h) \;:=\; \sup_{z \in \mathcal{S}_F} \mathbb{P}\big(d_q(W, P_q^{-1} z) \leq h\big), \tag{9}$$

where $P_q^{-1} z$ denotes any element of $\mathcal{H}$ projecting to z (both quantities depend on z only through the FPC coordinates). Equivalently, with the identification $\mathcal{S}_F^+ \leftrightarrow \mathcal{S}_F$, $\psi(h) =$

$\inf_{w\in\mathcal{S}_F^+}\mathbb{P}_w(h)$ and $\Psi(h)=\sup_{w\in\mathcal{S}_F^+}\mathbb{P}_w(h)$. The fractal small ball regime assumes that ψ and Ψ are of the same polynomial order:

$$c_\psi\, h^q \;\le\; \psi(h) \;\le\; \Psi(h) \;\le\; C_\psi\, h^q, \qquad h\in(0,h_0], \tag{10}$$

for constants $0<c_\psi\le C_\psi<\infty$ and $h_0>0$. Equation (10) holds, for instance, when W has a bounded joint density with respect to Lebesgue measure on $\mathbb{R}^q$ in the FPC coordinate system.

Remark 1 (Notation and factorization through P_q)**.** *All deviation theory in this paper is developed in the q-dimensional FPC coordinate system. Equivalently, d_q factorizes through the projection P_q, and uniform deviation, small ball, and entropy quantities depend on W only through P_qW. Under (A5*) below, the support set $\mathcal{S}_F\subset\mathbb{R}^q$ is the support of P_qW in this coordinate system, and (A5**+) identifies $\mathcal{S}_F^+$ with $\mathcal{S}_F$. We use $\mathcal{S}_F$ throughout uniform estimates and write $\mathcal{S}_F^+$ only when emphasizing that an estimate is interpreted as a property of P_qW.*

2.4. Kernel and Estimators (Preview)

The kernel weighted estimators of m and σ^2 used throughout the paper rely on a kernel function $K:[0,\infty)\to[0,\infty)$ with the following standing properties. K is bounded, Lipschitz, supports $[0,1]$, and *type-II*, in the sense that there exists $k_0>0$ such that $K(u)\ge k_0$ for $u\in[0,1/2]$. The role of the type-II condition is to ensure that the population denominator $\mathbb{E}[K(d_q(w,W)/h_n)]$ is bounded below by a constant multiple of $\psi(h_n/2)$ on $\mathcal{S}_F$. Standard examples include the Epanechnikov, biweight, and triweight kernels.

The estimators themselves, a profiled least squares estimator $\widehat{\beta}$ for β_0, a functional Nadaraya Watson estimator $\widehat{m}$ for m on d_q, and a kernel estimator $\widehat{\sigma}^2$ for the conditional variance σ^2 defined in (4), together with the truncated variant $\check{\sigma}(w):=\max\{\widehat{\sigma}(w),\sigma_{\min}/2\}$ used to studentize the conformal score, are formally defined in Section 3. Their uniform convergence rates and the corresponding high probability events are stated in Section 4 and proved in the Theoretical Supplement.

2.5. Data Split and Block Length Notation

The dependence aware split conformal procedure uses a three way split of the observed sequence into a training set, a calibration set, and a test point, with two buffer gaps:

$$\underbrace{Z_1,\dots,Z_{n_1}}_{\mathcal{D}_{\text{train}}}\ \underbrace{Z_{n_1+1},\dots,Z_{n_1+g_n}}_{\text{gap}}\ \underbrace{Z_{n_1+g_n+1},\dots,Z_{n_1+g_n+n_2}}_{\mathcal{D}_{\text{cal}}}\ \underbrace{Z_{n_1+g_n+n_2+1},\dots}_{\text{gap}}\ \underbrace{Z_{n+1}}_{\text{test}}. \tag{11}$$

The training sample $\mathcal{D}_{\text{train}}$ has size $n_1\to\infty$ and is used to compute $(\widehat{\beta},\widehat{m},\check{\sigma})$. The calibration sample $\mathcal{D}_{\text{cal}}$ has size $n_2\to\infty$ and is used to compute the conformal quantile. The buffer length g_n controls the conditional dependence between the training estimators and the calibration scores, and between the calibration scores and the future score.

For empirical process arguments we partition each set into blocks. The training side block length is b_n with $\mu_1:=\lfloor n_1/b_n\rfloor$ blocks; the calibration side block length is $b_{2,n}$ with $\mu_2:=\lfloor n_2/b_{2,n}\rfloor$ blocks. Under geometric β-mixing the canonical choices b_n, $b_{2,n}$, and g_n all grow logarithmically in the corresponding sample size; explicit forms for $b_{2,n}$ and g_n are given in (C3)–(C4) below, while the training block length b_n may be any sequence satisfying (A8′), canonically $b_n=\lceil 3\log n_1/\log(1/\rho)\rceil$, for which $\mu_1\beta(b_n)\le B_0 n_1^{-2}\le n_1^{-1}$ for n_1 large.

2.6. Assumptions

The following block of assumptions is used throughout the paper. Each theorem in Section 4 states the precise sub-block under which it holds; here we list the full block, separated into structural conditions (A) on the data generating process and the estimators, and tuning conditions (C) on the conformal procedure.

(A1) *Stationarity and geometric β-mixing.* The sequence $\{Z_i\}_{i\geq 1}$ is strictly stationary and satisfies (7) for some $B_0 > 0$ and $\rho \in (0,1)$.

(A2′) *Conditionally centered sub-Gaussian errors (martingale-difference form).* Let $\mathcal{F}_{i-1} := \sigma(Z_1, \ldots, Z_{i-1})$ and $\mathcal{G}_i := \mathcal{F}_{i-1} \vee \sigma(X_i, W_i)$. There exists $\nu_\varepsilon > 0$ such that, for every i,

$$\mathbb{E}[\varepsilon_i \mid \mathcal{G}_i] = 0, \qquad \mathbb{E}[\varepsilon_i^2 \mid \mathcal{G}_i] = \sigma^2(W_i), \qquad \mathbb{E}\big[\exp(\lambda \varepsilon_i) \mid \mathcal{G}_i\big] \leq \exp(\lambda^2 \nu_\varepsilon^2/2) \text{ for all } \lambda \in \mathbb{R}.$$

In particular $\mathbb{E}[\varepsilon_i \mid X_i, W_i] = 0$, and the conditional variance of ε_i given the past and the contemporaneous covariates depends only on W_i. This martingale-difference form is strictly stronger than the contemporaneous moment condition $\mathbb{E}[\varepsilon_i \mid X_i, W_i] = 0$; it is needed because the proofs of Lemma 2 and Theorem 4 require the within-block cross-covariances of kernel-weighted error summands to vanish at the small ball variance scale $\psi(h_n)$, which contemporaneous centering alone does not deliver. It holds whenever the errors are generated conditionally independently given the covariate path, as in Experiments 1–2 of Section 6, and it fails for autoregressive error recursions such as the boundary stress test of Experiment 3.

(A3) *Fixed-dimensional covariate.* The dimension $p \geq 1$ is fixed and $\mathbb{E}[\|X_i\|_2^2] < \infty$.

(A4) *Smoothness of m.* There exist $L > 0$ and $s \in (0,1]$ such that $|m(w) - m(w')| \leq L\, d_q(w,w')^{s \wedge 1}$ for all $w, w' \in \mathcal{S}_F^+$.

(A4$_\sigma$) *Smoothness of σ^2.* There exist $L_\sigma > 0$ and $s_\sigma \in (0,1]$ such that $|\sigma^2(w) - \sigma^2(w')| \leq L_\sigma\, d_q(w,w')^{s_\sigma \wedge 1}$ for $w, w' \in \mathcal{S}_F^+$.

(A5*) *Compact projected support with uniform interior volume.* The projected covariate $P_q W$ has compact support $\mathcal{S}_F := \operatorname{supp}(P_q W) \subset \mathbb{R}^q$ in the q-dimensional FPC coordinate system, with covering numbers $N(\eta, \mathcal{S}_F, d_q) \leq A\, \eta^{-q}$ for some $A < \infty$ and $\eta \in (0, \eta_0]$. Furthermore, $\mathcal{S}_F$ satisfies a uniform interior-volume condition: there exist $r_0 > 0$ and $c_{\text{cone}} \in (0,1]$ such that, for every $z \in \mathcal{S}_F$ and every $h \in (0, r_0]$,

$$\operatorname{Leb}_q\big(B(z,h) \cap \mathcal{S}_F\big) \;\geq\; c_{\text{cone}}\, \omega_q\, h^q, \tag{12}$$

where $B(z,h)$ is the Euclidean ball in $\mathbb{R}^q$ and ω_q is the volume of the unit ball. All uniform bounds in this paper are taken over $\mathcal{S}_F$.

(A5*$^+$) *Identification of $\mathcal{S}_F^+$.* We adopt the convention $\mathcal{S}_F^+ := \{w \in \mathcal{H} : P_q w \in \mathcal{S}_F\}$. By (A5*), $\mathcal{S}_F^+$ is identified through P_q with $\mathcal{S}_F$, and any uniform bound over $\mathcal{S}_F^+$ depends on w only through $P_q w$.

(A6*) *Density-bounded fractal regime.* The law of $P_q W$ admits a Lebesgue density f_Z on $\mathcal{S}_F$ with $0 < c_f \leq f_Z(z) \leq C_f < \infty$ for $z \in \mathcal{S}_F$. Equivalently, ψ and Ψ defined in (9) satisfy the fractal small ball regime (10) with constants $c_\psi := c_f\, \omega_q\, c_{\text{cone}}$ and $C_\psi := C_f\, \omega_q$, where ω_q is the volume of the unit Euclidean ball in $\mathbb{R}^q$ and $c_{\text{cone}} \in (0,1]$ accounts for the boundary geometry of $\mathcal{S}_F$.

(A7) *Kernel.* K is bounded, Lipschitz, supported on $[0,1]$, and type-II with constant $k_0 > 0$.

(A8′) *Bandwidth–block-length condition.* $h_n \to 0$, $n_1 \psi(h_n) / \log n_1 \to \infty$, $n_1 \psi(h_n) / \big[b_n^2\, (\log n_1)^3\big] \to \infty$, and $\mu_1\, \beta(b_n) \leq n_1^{-1}$ for all sufficiently large n_1. The third power of $\log n_1$ (rather than the second) is needed because the variance estimation proof of Theorem 4 truncates the sub-exponential summands $\varepsilon_i^2 - \sigma^2(W_i)$

at level of order $\log n_1$; the condition is harmless in practice, being satisfied by every polynomial bandwidth under (A6*), in particular by the rate-optimal choice of Corollary 1.

(A9) *Fixed FPC dimension.* $q \geq 1$ is fixed and does not depend on n.

(A10) *Finite sample tail bound for* $\widehat{\beta}$. There exist a constant $C_\beta < \infty$ and an integer N_β, depending only on the structural quantities of (A), such that for all $n_1 \geq N_\beta$,

$$\mathbb{P}\Big(\|\widehat{\beta} - \beta_0\|_2 > C_\beta \sqrt{\log n_1 / n_1} \Big) \;\leq\; n_1^{-1}.$$

In particular $\|\widehat{\beta} - \beta_0\|_2 = O_P(\sqrt{\log n_1 / n_1})$. We assume the tail form, rather than the customary in-probability statement $\|\widehat{\beta} - \beta_0\|_2 = O_P(n_1^{-1/2})$ (which is faster by a $\sqrt{\log n_1}$ factor but supplies no explicit event), because the explicit good event bookkeeping of Theorem 1 requires an event $\Omega_\beta := \{\|\widehat{\beta} - \beta_0\|_2 \leq C_\beta \sqrt{\log n_1 / n_1}\}$ with a fixed constant and a polynomially small failure probability; an O_P statement alone cannot supply such an event. We invoke (A10) as a high level condition supporting the finite sample bookkeeping of Theorem 1. It is reasonable in the present setting, where the parametric dimension p is fixed, the design is conditionally sub-exponential by (A11A), and the errors are conditionally centered sub-Gaussian and geometrically β-mixing by (A1) and (A2′); a tail of the displayed form then follows for the profiled least squares estimator from the same blocking and Bernstein type concentration used for the kernel scores in the Theoretical Supplement. We state it as an assumption rather than reproduce that derivation, which is standard but tangential to the conformal analysis.

(A11A) *Conditional sub-exponential covariates.* For each component $j = 1, \ldots, p$ and uniformly over $w \in \mathcal{S}_F^+$, the conditional law of $X^{(j)}$ given $W = w$ has ψ_1-Orlicz norm bounded by $C_X < \infty$.

(A12) *Bounded local mean.* $\sup_{w \in \mathcal{S}_F^+} \|\mathbb{E}[X \mid W = w]\|_2 \leq M_X < \infty$.

(A12$^+$) *Bounded conditional absolute moments of* ε. Implied by (A2′); we record explicitly that $\sup_{w \in \mathcal{S}_F^+} \mathbb{E}\big[\, |\varepsilon|^k \,\big|\, W = w \,\big] \leq M_k$ for $k = 1, 2, 3, 4$.

($\mathrm{A}_{\sigma_{\min}}$) *Variance bounded below.* $\sigma(w) \geq \sigma_{\min} > 0$ for $w \in \mathcal{S}_F^+$.

(C1) *Score density.* The ideal score $S = |\varepsilon| / \sigma(W)$ has a Lebesgue density f_S in a fixed neighborhood $[q_{1-\alpha} - \xi,\, q_{1-\alpha} + \xi]$ of $q_{1-\alpha}$ with $0 < m_S \leq f_S(t) \leq M_S < \infty$ on this neighborhood.

(C2) *Quantile finiteness.* $\alpha \in (0, 1)$ is fixed and $q_{1-\alpha} := \inf\{t : F_S(t) \geq 1 - \alpha\}$ is finite.

(C3) *Calibration block length.* $b_{2,n} = \lceil C_b \log n_2 / \log(1/\rho) \rceil$ for some $C_b \geq 3$, with $\mu_2 = \lfloor n_2 / b_{2,n} \rfloor$.

(C4) *Gap length.* $g_n = \lceil C_g \log n / \log(1/\rho) \rceil$ for some $C_g \geq 3$, with $n_1 + g_n + n_2 + g_n \leq n$.

(C5) *Sample-size compatibility.* $n_1, n_2 \to \infty$ jointly, with $(\log n)^3 / n_1 \to 0$, equivalently $\log n = o(n_1^{1/3})$, and the bandwidth–block-length condition (A8′) holds for h_n and b_n as functions of n_1. The growth condition $(\log n)^3 / n_1 \to 0$ ensures that the calibration inflation term $L_n = C_X^* C_\beta (\log n)^{3/2} n_1^{-1/2}$ of Table 1, and with it the perturbation rate η_n, vanishes as $n \to \infty$; it excludes regimes in which n_2 grows super-exponentially in n_1, where the eventual-threshold statement of Theorem 1 would fail. It is satisfied by every polynomial relationship between n_1 and n_2, including all designs used in Section 6. This explicit growth condition is necessary because joint growth alone does not imply the rate required by Theorem 1.

The labels (A·) and (C·) are referenced throughout Section 4 and the Theoretical Supplement; whenever a result requires only a sub-block of these conditions, the relevant indices are cited explicitly in the statement.

2.7. Rate Notation

The following deterministic rate symbols are used throughout the paper. They are functions of the structural quantities of the SFPLM model, of the quantities (h_n, g_n, n_1, n_2) chosen by the user, and of the proof-side block lengths $(b_n, b_{2,n})$ of (A8′) and (C3), which parametrize the coupling and blockwise concentration arguments but are not inputs of the implemented procedure (see the Remark after Algorithm 1).

Table 1. Symbol–definition–rate summary used in Section 4 and the Theoretical Supplement; rates evaluated under (A1)–(A8′) and the FPC fractal regime $\psi(h) \asymp h^q$.

Symbol	Definition	Rate
$r_{m,n}$	$\sup_{w\in\mathcal{S}_F} \lvert\widehat{m}(w) - m(w)\rvert$	$h_n^{s\wedge 1} + \sqrt{\log n_1/(n_1\,\psi(h_n))} + \sqrt{\log n_1/n_1}$
L_n	calibration X-control inflation	$C_X^*\,C_\beta\,(\log n)^{3/2}\,n_1^{-1/2}$
$r_{\mathrm{pred},n}$	$r_{m,n} + L_n$	$r_{m,n} + L_n$
$r_{\sigma,n}$	$\sup_{w\in\mathcal{S}_F} \lvert\breve{\sigma}(w) - \sigma(w)\rvert$	$\max\{h_n^{s_\sigma\wedge 1},\, h_n^{s\wedge 1}\} + \sqrt{\log n_1/(n_1\,\psi(h_n))} + \sqrt{\log n_1/n_1}$
ω_n	rate bounding $\sup_t \lvert\widehat{F}_{\mathrm{cal}}(t) - F_S(t)\rvert$ w.h.p.	$\sqrt{b_{2,n}\,\log n_2/n_2}$
A_n	affine offset $2\,r_{\mathrm{pred},n}/\sigma_{\min}$	$O\big(r_{\mathrm{pred},n}/\sigma_{\min}\big)$
B_n	affine slope $2\,r_{\sigma,n}/\sigma_{\min}$	$O\big(r_{\sigma,n}/\sigma_{\min}\big)$
η_n	$A_n + B_n(q_{1-\alpha} + \xi/2)$	$O\big(r_{\mathrm{pred},n}/\sigma_{\min} + q_{1-\alpha}\,r_{\sigma,n}/\sigma_{\min}\big)$
Δ_n	coverage penalty (Theorem 1)	see Equation (13)

The theoretical finite-sample coverage penalty Δ_n has the form

$$\begin{aligned}\Delta_n \;&= C_1\,\frac{n_2}{b_{2,n}}\,\beta(b_{2,n}) + C_3\,\sqrt{\frac{b_{2,n}\,\log n_2}{n_2}} + C_2\,\beta(g_n)\\ &\quad + c_m\,M_S\,\frac{r_{\mathrm{pred},n}}{\sigma_{\min}} + c_\sigma\,M_S\,q_{1-\alpha}\,\frac{r_{\sigma,n}}{\sigma_{\min}} + \frac{1}{n_2+1} + \delta_{\mathrm{good},n},\end{aligned} \tag{13}$$

where $\delta_{\mathrm{good},n}$ is the polynomially small failure probability of the good event (training and calibration/future) enumerated in Section 4.1, and the constants $C_1, C_2, C_3, c_m, c_\sigma, M_S, q_{1-\alpha}, \sigma_{\min}$ are structural quantities of (A) and (C). Because these constants are unknown, Δ_n is a proof-side quantity; it is not evaluated by the procedure and does not modify the interval. The interpretation of the seven summands, and the full statement of what Δ_n is and is not, are given after Theorem 1 in Section 4.3.

2.8. Three Remarks

Remark 2. *Assumption (A9) fixes the FPC dimension q. The fractal small ball condition (10) depends on q through the exponent, and the constants c_ψ, C_ψ, the entropy constant A in (A5*), and the kernel discretization cost in the proofs all depend on q. A growing-q regime would require a separate analysis in which $q = q_n$ is selected from data and the rate constants acquire additional q_n-factors. Such an extension is outside the scope of this paper and is noted in Section 8 as an open direction.*

Remark 3. *Assumption (A2′) imposes a sub-Gaussian conditional error rather than the milder sub-exponential alternative. The reason is that the variance estimator $\widehat{\sigma}^2$ is the kernel weighted mean of squared residuals, and concentration of $\widehat{\sigma}^2$ requires concentration of ε^2. Under sub-Gaussian ε, the random variable ε^2 is sub-exponential, and the proof of Theorem 4 uses the same Bernstein architecture as the proof of Theorem 2. Under merely sub-exponential ε, ε^2 is only sub-Weibull of order $1/2$, and the corresponding concentration carries an additional logarithmic factor. We adopt (A2′) for clarity; a sub-Weibull extension is noted as a remark in Section 8.*

Remark 4. *Assumption (C1) imposes density bounds on the score $S = |\varepsilon|/\sigma(W)$ only on a fixed neighborhood of the population quantile $q_{1-\alpha}$. It is not a global distributional assumption on ε or on W. Its role is to translate empirical-quantile fluctuations into coverage perturbations through the local Lipschitz bound $|F_S(\widehat{q}) - F_S(q)| \leq M_S\,|\widehat{q} - q|$, and to invert the empirical CDF deviation into a quantile rate via the lower bound $f_S \geq m_S$. The condition is mild and verifiable in concrete SFPLM applications under standard tail conditions on ε.*

3. Dependence Aware Conformal Prediction Procedure

The estimators $(\widehat{\beta}, \widehat{m}, \check{\sigma})$ are computed on the training subsample, the conformal quantile on the calibration subsample, and the final interval is evaluated at a future point, using the notation $(\mathcal{H}, d_q, P_q, \psi, h_n, b_n, b_{2,n}, g_n)$ fixed in Section 2.

3.1. Index Sets and Data Split

Following Section 2.5, we partition the index set $\{1, \ldots, n\}$ into three contiguous blocks separated by two buffer gaps. Set

$$\mathcal{I}_1 := \{1, \ldots, n_1\}, \qquad \mathcal{I}_2 := \{n_1 + g_n + 1, \ldots, n_1 + g_n + n_2\}, \tag{14}$$

with the future / test index $n + 1 := n_1 + g_n + n_2 + g_n + 1$. The training subsample is $\mathcal{D}_{\text{train}} := \{Z_i : i \in \mathcal{I}_1\}$ and the calibration subsample is $\mathcal{D}_{\text{cal}} := \{Z_i : i \in \mathcal{I}_2\}$. The gap indices in $\{n_1 + 1, \ldots, n_1 + g_n\} \cup \{n_1 + g_n + n_2 + 1, \ldots, n\}$ are not used by the procedure; their role is to control conditional dependence between $\mathcal{D}_{\text{train}}$ and $\mathcal{D}_{\text{cal}}$, and between $\mathcal{D}_{\text{cal}}$ and the future point, in the analysis of the Theoretical Supplement.

3.2. FPC Projection in Practice

The semi-metric d_q of (8) is defined through the population projection P_q onto the leading q eigenfunctions of the covariance operator Σ_W. In implementation, P_q is replaced by an empirical projection $\widehat{P}_q$ obtained from $\mathcal{D}_{\text{train}}$ as follows. Let $\bar{W}_{n_1}(\cdot) := n_1^{-1} \sum_{i \in \mathcal{I}_1} W_i(\cdot)$ be the empirical mean curve, and let

$$\widehat{\Sigma}_W \;:=\; \frac{1}{n_1 - 1} \sum_{i \in \mathcal{I}_1} (W_i - \bar{W}_{n_1}) \otimes (W_i - \bar{W}_{n_1})$$

be the empirical covariance operator. The first q empirical eigenfunctions $\widehat{\phi}_1, \ldots, \widehat{\phi}_q$ are computed from the spectral decomposition of $\widehat{\Sigma}_W$, and $\widehat{P}_q$ is the orthogonal projection onto $\operatorname{span}\{\widehat{\phi}_1, \ldots, \widehat{\phi}_q\}$. The corresponding empirical scores at any point $w \in \mathcal{H}$ are $\widehat{\xi}_k(w) := \langle w - \bar{W}_{n_1}, \widehat{\phi}_k \rangle_{L^2}$, $k = 1, \ldots, q$.

Theorem 1 is stated for the population FPC projection P_q and the population semi-metric d_q it induces. The implemented procedure estimates the projection from the training sample through the empirical FPCA above; this introduces an eigenfunction-estimation error that the theorem does not cover. A rigorous extension would require, at minimum, eigen-gap conditions on the covariance operator Σ_W and operator-perturbation control of Davis–Kahan type relating $\widehat{P}_q$ to P_q, propagated through the small-ball and uniform-deviation arguments; we do not develop such an extension, and no empirical-FPCA coverage theorem is claimed. The sensitivity study of Section 6.8 compares the implemented empirical-projection pipeline with the population-projection pipeline on identical data with paired seeds and finds practically negligible differences in these settings (largest paired coverage difference at most 0.0005; largest paired length difference -0.0131 ± 0.0050, about 0.3% of the interval length, at $\rho = 0.85$, $n_1 = 250$). This is sensitivity evidence only; it neither extends nor substitutes for the theorem.

3.3. Profiled Estimator of β_0

The estimator $\widehat{\beta}$ is computed by a Robinson style profiled least squares construction on $\mathcal{D}_{\text{train}}$. Define the kernel weighted local means

$$\widehat{r}_Y(w) := \frac{\sum_{i\in\mathcal{I}_1} K\Big(\frac{d_q(w,W_i)}{h_n}\Big) Y_i}{\sum_{i\in\mathcal{I}_1} K\Big(\frac{d_q(w,W_i)}{h_n}\Big)}, \qquad \widehat{r}_X(w) := \frac{\sum_{i\in\mathcal{I}_1} K\Big(\frac{d_q(w,W_i)}{h_n}\Big) X_i}{\sum_{i\in\mathcal{I}_1} K\Big(\frac{d_q(w,W_i)}{h_n}\Big)}, \tag{15}$$

where $\widehat{r}_X(\cdot)$ is taken componentwise. Set the residualized variables

$$\widetilde{Y}_i := Y_i - \widehat{r}_Y(W_i), \qquad \widetilde{X}_i := X_i - \widehat{r}_X(W_i), \qquad i \in \mathcal{I}_1, \tag{16}$$

and define

$$\widehat{\beta} := \Big(\sum_{i\in\mathcal{I}_1} \widetilde{X}_i \widetilde{X}_i^\top\Big)^{-1} \sum_{i\in\mathcal{I}_1} \widetilde{X}_i \widetilde{Y}_i. \tag{17}$$

Under (A1)–(A4) and (A11A), the matrix $n_1^{-1}\sum_{i\in\mathcal{I}_1} \widetilde{X}_i\widetilde{X}_i^\top$ converges in probability to a positive definite limit, so the inverse in (17) exists with probability tending to one. In implementation, when the empirical Gram matrix is poorly conditioned at finite n_1, a small ridge $\lambda_n I_p$ may be added before inversion; the theory below treats $\lambda_n = 0$ and assumes nonsingularity.

3.4. Functional Estimator of m

Given $\widehat{\beta}$, the regression operator m is estimated by a functional Nadaraya Watson smoother on the partial residuals $Y_i - X_i^\top\widehat{\beta}$:

$$\widehat{m}(w) := \frac{\sum_{i\in\mathcal{I}_1} K\Big(\frac{d_q(w,W_i)}{h_n}\Big)(Y_i - X_i^\top\widehat{\beta})}{\sum_{i\in\mathcal{I}_1} K\Big(\frac{d_q(w,W_i)}{h_n}\Big)}, \qquad w \in \mathcal{S}_F^+. \tag{18}$$

For analysis we also introduce the oracle version

$$\widetilde{m}(w) := \frac{\sum_{i\in\mathcal{I}_1} K\Big(\frac{d_q(w,W_i)}{h_n}\Big)(Y_i - X_i^\top\beta_0)}{\sum_{i\in\mathcal{I}_1} K\Big(\frac{d_q(w,W_i)}{h_n}\Big)}, \tag{19}$$

which uses the unknown β_0 in place of $\widehat{\beta}$. The oracle estimator $\widetilde{m}$ is not used in computation; it appears only in the analysis of the Theoretical Supplement, where the rate $r_{m,n}$ is decomposed into an oracle uniform deviation (Theorem 2) and a plug-in transfer term (Theorem 3).

3.5. Conditional Variance Estimator and Truncation

Define the training residuals

$$\widehat{\varepsilon}_i := Y_i - X_i^\top\widehat{\beta} - \widehat{m}(W_i), \qquad i \in \mathcal{I}_1, \tag{20}$$

and the kernel-NW variance estimator

$$\widehat{\sigma}^2(w) \;:=\; \frac{\displaystyle\sum_{i\in\mathcal{I}_1} K\Big(\frac{d_q(w,W_i)}{h_n}\Big)\,\widehat{\varepsilon}_i^{\,2}}{\displaystyle\sum_{i\in\mathcal{I}_1} K\Big(\frac{d_q(w,W_i)}{h_n}\Big)}, \qquad w\in\mathcal{S}_F^{+}. \tag{21}$$

For studentization in the conformal score below we use the truncated estimator

$$\check{\sigma}(w) \;:=\; \max\big\{\widehat{\sigma}(w),\, \sigma_{\min}/2\big\}. \tag{22}$$

The truncation in (22) prevents division by an arbitrarily small variance estimate. Under ($\mathrm{A}_{\sigma_{\min}}$) and Theorem 4, the truncation is inactive on a high probability event, on which $\check{\sigma}\equiv\widehat{\sigma}$; off that event the truncation guarantees that the prediction interval below remains well defined.

3.6. Calibration Scores and Conformal Quantile

The studentized absolute residuals on the calibration subsample are

$$\widehat{S}_i \;:=\; \frac{\big|\,Y_i - X_i^{\top}\widehat{\beta} - \widehat{m}(W_i)\,\big|}{\check{\sigma}(W_i)}, \qquad i\in\mathcal{I}_2, \tag{23}$$

and the corresponding ideal scores, used only in analysis, are $S_i := |\varepsilon_i|/\sigma(W_i)$. Given the miscoverage level $\alpha\in(0,1)$, set

$$k_\alpha \;:=\; \big\lceil (1-\alpha)(n_2+1)\big\rceil, \tag{24}$$

and define the empirical conformal quantile

$$\widehat{q}_{1-\alpha} \;:=\; \text{the } k_\alpha\text{-th order statistic of } \big\{\widehat{S}_i : i\in\mathcal{I}_2\big\}. \tag{25}$$

The choice $k_\alpha = \lceil (1-\alpha)(n_2+1)\rceil$ is the standard split conformal convention and yields the rounding term $1/(n_2+1)$ in (13). When $k_\alpha > n_2$, which can occur only for $n_2 < (1-\alpha)/\alpha$, we adopt the standard convention $\widehat{q}_{1-\alpha} := +\infty$, in which case the interval below is the whole real line and covers trivially.

3.7. Prediction Interval

For a new pair $(x,w)\in\mathbb{R}^p\times\mathcal{S}_F^{+}$, the point predictor is

$$\widehat{y}(x,w) \;:=\; x^{\top}\widehat{\beta} + \widehat{m}(w), \tag{26}$$

and the prediction interval is

$$\widehat{C}_n(x,w) \;:=\; \Big[\,\widehat{y}(x,w) - \widehat{q}_{1-\alpha}\,\check{\sigma}(w),\; \widehat{y}(x,w) + \widehat{q}_{1-\alpha}\,\check{\sigma}(w)\,\Big]. \tag{27}$$

Theorem 1 provides a finite sample lower bound on $\mathbb{P}\big(Y_{n+1}\in\widehat{C}_n(X_{n+1},W_{n+1})\big)$ of the form $1-\alpha-\Delta_n$, with Δ_n as in (13).

3.8. Algorithm

The Algorithm 1 summarizes the procedure; its inputs are the observed sample and the tuning parameters fixed in Section 2, and its output is the prediction interval $\widehat{C}_n(\cdot,\cdot)$ above.

Algorithm 1 Dependence aware split conformal prediction interval for the SFPLM under β-mixing.

Input: Sample $\{(X_i, W_i, Y_i)\}_{i=1}^{n}$; FPC dimension $q \geq 1$; kernel K as in (A7); miscoverage $\alpha \in (0,1)$; bandwidth h_n; gap length g_n; lower variance bound $\sigma_{\min} > 0$.
Output: Prediction interval $\widehat{C}_n(x, w)$ at any test point $(x, w) \in \mathbb{R}^p \times \mathcal{S}_F^+$.
Step 1. Partition $\{1, \ldots, n\}$ into $\mathcal{I}_1$, gap_1, $\mathcal{I}_2$, gap_2 and the test index $n+1$ as in (14).
Step 2. Compute the empirical FPC projection $\widehat{P}_q$ from $\{W_i\}_{i \in \mathcal{I}_1}$ as in Section 3.2.
Step 3. Compute $\widehat{\beta}$ from (17) using the profiled-LS construction (15), (16).
Step 4. Compute $\widehat{m}(\cdot)$ from (18) on $\mathcal{S}_F^+$.
Step 5. Compute the training residuals (20), $\widehat{\sigma}^2(\cdot)$ from (21), and the truncated $\check{\sigma}(\cdot)$ from (22).
Step 6. For each $i \in \mathcal{I}_2$ compute the studentized score $\widehat{S}_i$ from (23).
Step 7. Compute the empirical conformal quantile $\widehat{q}_{1-\alpha}$ from (24), (25).
Step 8. Output the interval $\widehat{C}_n(x, w)$ in (27).

Remark 5 (Operational vs. proof-side quantities)**.** *The gap length g_n is the only blocking quantity the procedure uses: it determines the index sets of Step 1. The block lengths b_n and $b_{2,n}$ of (A8′) and (C3) are proof devices. They parametrize the Yu-coupling and blockwise concentration arguments behind Theorem 1 and appear in the first two summands of Δ_n, but they are not required to compute the interval; no step of Algorithm 1 evaluates them. This mirrors the analysis-side parameters of the generic dependent-conformal literature, namely the proof-side block parameters inside the Theorem 4 penalty of [16] and the deletion quantities τ, τ^* of Theorem 4 of [17], which are likewise optimized inside the bounds rather than supplied to the procedure; neither source imposes an operational block or buffer.*

3.9. Operational Tuning

The operational conditions on (h_n, g_n, q) are stated in (A8′), (C4), and (A9); condition (C3) constrains the proof-side block length $b_{2,n}$ (Remark 5). For implementation, the following operational rules are used as defaults in Section 6.

(T1) Bandwidth. The preliminary rate $h_n^* = (\log n_1 / n_1)^{1/(q+2(s\wedge 1))}$ (with $s \wedge 1 = 1$ as the working value) is multiplied by the median pairwise FPC score distance computed on $\mathcal{D}_{\text{train}}$, as a scale invariant adjustment. Sensitivity to a $[h_n^*/2, 2h_n^*]$ band is reported in Section 6.

(T2) Condition (C3) in practice (analysis constant). The block length $b_{2,n}$ is a proof-side quantity (Remark 5); no implementation step evaluates it. When a value is needed for the analysis, for instance in the rate proxies of Section 6.7, we use $b_{2,n} = \lceil 3 \log n_2 / \log(1/\rho) \rceil$ as in (C3) under a known geometric mixing rate ρ. Without a specified ρ, the block length rule of [18] gives a data driven analogue; the two choices give comparable values in our settings.

(T3) Gap length. $g_n = \lceil 3 \log n / \log(1/\rho) \rceil$ under known ρ, doubled when in doubt. The cost of a slightly larger gap is at most a small reduction in the effective n_2.

(T4) FPC dimension. q is fixed before computing $\widehat{\beta}$ and $\widehat{m}$. We use $q = 3$ in the simulation and real data studies; sensitivity to alternative fixed q values is left to future work. In practice, q may be selected from data using a cumulative eigenvalue rule on $\widehat{\Sigma}_W$ targeting at least 90% of the spectral mass.

The rules (T1)–(T4) are stated as operational defaults; they are not adaptive choices supported by formal theory in the present paper. Adaptive selection of h_n or q from data is left to future work (Section 8).

4. Main Theoretical Results

Throughout this section, the constants $C, C_0, C_1, C_2, C_3, c_m, c_\sigma, c_X^*, C_X^*, C_\beta, D_{\min}$ depend on the structural quantities of (A) and (C) but not on $n_1, n_2, h_n, b_n, b_{2,n}, g_n$; their numerical values may change from line to line.

4.1. Coverage Penalty and Good Event Notation

The theoretical finite-sample coverage penalty Δ_n in (13) is built from deterministic rate symbols of Table 1 and a probabilistic good event failure probability. Recall

$$\begin{aligned}
r_{m,n} &= h_n^{s\wedge 1} + \sqrt{\frac{\log n_1}{n_1\,\psi(h_n)}} + \sqrt{\frac{\log n_1}{n_1}},\\
r_{\sigma,n} &= \max\{h_n^{s_\sigma\wedge 1},\, h_n^{s\wedge 1}\} + \sqrt{\frac{\log n_1}{n_1\,\psi(h_n)}} + \sqrt{\frac{\log n_1}{n_1}},\\
L_n &= C_X^*\, C_\beta\, (\log n)^{3/2}\, n_1^{-1/2}, \qquad r_{\mathrm{pred},n} = r_{m,n} + L_n,\\
\omega_n &= \sqrt{\frac{b_{2,n}\,\log n_2}{n_2}},\\
A_n &= \frac{2\, r_{\mathrm{pred},n}}{\sigma_{\min}}, \qquad B_n = \frac{2\, r_{\sigma,n}}{\sigma_{\min}},\\
\eta_n &= A_n + B_n\,(q_{1-\alpha} + \xi/2) = O\!\left(\frac{r_{\mathrm{pred},n}}{\sigma_{\min}} + \frac{q_{1-\alpha}\, r_{\sigma,n}}{\sigma_{\min}}\right).
\end{aligned}$$

The good event failure probability $\delta_{\mathrm{good},n}$ collects the failure probabilities of all the high probability events used in Theorem 1: the denominator and stochastic numerator events of Lemmas 1 and 2, the training side estimation events behind Theorems 3 and 4 (in particular the $\widehat{\beta}$ event of (A10), the training X-control event Ω_X^{tr}, and the $\widehat{\sigma}$ event), the calibration / future X-control event Ω_X^{cal} of Lemma 4, and the calibration empirical CDF event of Lemma 3 Step 1. Under (A1), (A8′), (C3), and (C5), $\mu_1\,\beta(b_n) \le n_1^{-1}$ (part of (A8′)) and $\mu_2\,\beta(b_{2,n}) \le B_0\, n_2^{-2}$ (using the geometric mixing bound of (A1)), so the aggregate good event failure probability satisfies

$$\delta_{\mathrm{good},n} = O(n_1^{-1}) + O(n_2^{-2}) + O(n^{-A_{\mathrm{tail}}}),$$

with the individual components being $\mu_1\,\beta(b_n) + O(n_1^{-2})$ from the denominator and stochastic numerator events of Lemmas 1 and 2 (which together comprise the training side event failure $\delta_{T,n}$ of Theorem 2, included as a denominator/numerator component of $\delta_{\mathrm{good},n}$), $O(n_1^{-1})$ from the $\widehat{\beta}$ event of (A10), the training X-control event, and the variance event behind Theorem 4, $O(n_2^{-2}) + \mu_2\,\beta(b_{2,n}) = O(n_2^{-2})$ from the calibration empirical CDF event of Lemma 3 Step 1, and $O(n^{-A_{\mathrm{tail}}})$ from the calibration/future X-control event Ω_X^{cal} of Lemma 4 for any chosen polynomial tail exponent $A_{\mathrm{tail}} \ge 2$. The $O(n_2^{-2})$ component is not in general dominated by the training side terms, since (C5) allows n_2 to grow much more slowly than n_1; it is therefore displayed separately. As is standard, $\delta_{\mathrm{good},n}$ denotes a generic good event failure probability that is enlarged by fixed numerical constants whenever union bounds are applied; in particular, repeated uses of the same good event failure inside the proof of Theorem 1—e.g., a localization failure absorption in Step 2 followed by an inclusion failure absorption in Step 3—are absorbed into the displayed $\delta_{\mathrm{good},n}$ summand of Δ_n and do not require a separate term.

4.2. Generic Dependent Split-Conformal Validity

Before stating the main theorem, we record what generic theory already provides, so that the contribution of the SFPLM-specific analysis is delimited precisely.

Remark 6 (Generic dependent split-conformal validity)**.** *Under the respective conditions of Oliveira et al. and Barber and Pananjady, generic dependent split-conformal results provide marginal coverage guarantees for a score trained on the training block and then applied to calibration and test observations. These results require no operational buffer and do not require consistency of the fitted regression or scale estimators. In particular, they cover the contiguous M2 procedure of Section 6 under their respective conditions. See Oliveira, Orenstein, Ramos and Romano [16], Theorems 1 and 4, for marginal coverage, and Barber and Pananjady [17], Theorem 4 and Corollary 2, for the split-trained setting with memory $L = 0$. Theorem 2 of [17] gives a matching lower-bound result for the pretrained memoryless setting. The contribution of Theorem 1 below is not to re-establish marginal validity but to expose, for the SFPLM score, an estimation-aware decomposition in the model's own primitives (Section 4.3).*

4.3. SFPLM-Specific Estimation-Aware Coverage Decomposition

Theorem 1 (SFPLM-specific estimation-aware coverage decomposition)**.** *Assume the SFPLM (3) together with (A1)–(A12$^+$), ($A_{\sigma_{\min}}$), and (C1)–(C5). Let $\widehat{C}_n(\cdot,\cdot)$ be the prediction interval defined in (27) from the procedure of Section 3, built from the estimated studentized score, with the convention $\widehat{q}_{1-\alpha} := +\infty$ when $k_\alpha > n_2$, and let $s^\star(x,w,y) = |y - x^\top\beta_0 - m(w)|/\sigma(w)$ denote the ideal SFPLM score. Fix $\alpha \in (0,1)$. There exists an integer N_0, depending only on the structural quantities of (A) and (C) (in particular through the requirements $B_n \le 1/2$, $2\eta_n \le \xi/2$, $(C_0/m_S)\,\omega_n + 1/(m_S n_2) \le \xi/2$, and $k_\alpha \le n_2$, together with the finitely many sample-size thresholds internal to the cited lemmas, e.g., N_β of (A10)), such that for all $n \ge N_0$,*

$$\mathbb{P}\big(Y_{n+1} \in \widehat{C}_n(X_{n+1}, W_{n+1})\big) \;\ge\; 1 - \alpha - \Delta_n, \tag{28}$$

where Δ_n is given explicitly by (13), namely

$$\begin{aligned}\Delta_n \;&= C_1\,\frac{n_2}{b_{2,n}}\,\beta(b_{2,n}) + C_3\sqrt{\frac{b_{2,n}\,\log n_2}{n_2}} + C_2\,\beta(g_n)\\ &\quad + c_m\,M_S\,\frac{r_{\mathrm{pred},n}}{\sigma_{\min}} + c_\sigma\,M_S\,q_{1-\alpha}\,\frac{r_{\sigma,n}}{\sigma_{\min}} + \frac{1}{n_2+1} + \delta_{\mathrm{good},n}.\end{aligned}$$

The seven summands of Δ_n arise from a transfer argument anchored at the ideal score $s^\star$. The first two terms $C_1(n_2/b_{2,n})\beta(b_{2,n})$ and $C_3\sqrt{b_{2,n}\log n_2/n_2}$ are dependence costs internal to the calibration sample. The first is the Yu-coupling contribution for replacing the dependent calibration block sequence by an independent block surrogate, and the second is the calibration empirical process cost under that surrogate, reflecting the effective block size $b_{2,n}$ in the empirical CDF of the ideal scores. The third term $C_2\,\beta(g_n)$ aggregates the costs of two distinct Berbee–Yu couplings across buffers of length g_n: one decouples the calibration and future observations from the training σ-field across the first gap, and the other decouples the future test point from the calibration block across the second gap; the two couplings are separate proof steps with distinct roles, and their costs, each at most $\beta(g_n)$, are absorbed into the single constant C_2. The next two summands, $c_m M_S\, r_{\mathrm{pred},n}/\sigma_{\min}$ and $c_\sigma M_S\, q_{1-\alpha} r_{\sigma,n}/\sigma_{\min}$, are the price of replacing (β_0, m, σ) by the estimators $(\widehat{\beta}, \widehat{m}, \check{\sigma})$ of Section 3, that is, of comparing the implemented score with $s^\star$. The sixth summand $1/(n_2+1)$ is the standard split conformal rounding cost from the order statistic quantile in (25). The seventh summand $\delta_{\mathrm{good},n}$ records the cost of conditioning on the high probability good event. The estimation terms appear because the theorem tracks this oracle-to-estimated

transfer, not because marginal conformal validity of a trained score requires estimation accuracy. The generic results of [16,17] (Remark 6) give dependence-only marginal validity for the trained score, gap-free, with no estimation layer. What the transfer buys is a decomposition expressed in SFPLM primitives, the bandwidth h_n, the small-ball function ψ, the FPC dimension q, the variance floor $\sigma_{\min}$, and the smoothness indices (s, s_σ), separating the model-estimation layer from the dependence layer.

We state precisely what Δ_n is and is not. Δ_n is a theoretical finite-sample coverage penalty. It quantifies, in the proof, the gap between the nominal level and the guaranteed marginal coverage; it plays no role in the computation of the interval. In particular: (i) Δ_n does not modify the conformal quantile; the implemented interval uses the ordinary split-conformal order statistic $\widehat{q}_{1-\alpha} = \widehat{S}_{(k_\alpha)}$ with $k_\alpha = \lceil (1-\alpha)(n_2+1) \rceil$; (ii) Δ_n does not widen the implemented interval; (iii) its summands involve unknown structural constants $(C_1, C_2, C_3, c_m, c_\sigma, M_S, \sigma_{\min}$, the mixing prefactors, and the threshold $N_0)$, so Δ_n is not operational: it cannot be numerically evaluated or used as a correction in practice; (iv) the proxy table of Section 6.7 reports observable rate proxies with those constants set to one; the proxies illustrate rate behaviour only, they are not values of Δ_n and not a certified finite-sample bound, and they are never summed; (v) the additive decomposition is not claimed to be sharp: it stacks union bounds and couplings, and its dependence-only components overlap conceptually with generic dependence-only bounds ([16], Theorems 1 and 4; [17], Theorem 4 and Corollary 2), which may be sharper under their own conditions (Theorem 2 of [17] proves sharpness in the pretrained memoryless setting). A practically usable quantile adjustment under dependence is future work and is not developed here.

4.4. Oracle Uniform Deviation for $\widetilde{m}$

Theorem 2 (Oracle uniform deviation for $\widetilde{m}$)**.** *Assume (A1), (A2′), (A4)–(A7), (A8′), and (A9). There exist constants $L > 0$, $C_0, C, D_{\min} > 0$ depending only on the structural quantities of (A) and on q, and an event $\Omega_T^{\#}$ with*

$$\mathbb{P}\big((\Omega_T^{\#})^c\big) \;\le\; \delta_{T,n} \;:=\; C\, n_1^{-2} + 4\,\mu_1\,\beta(b_n),$$

such that on $\Omega_T^{\#}$,

$$\sup_{w \in \mathcal{S}_F} \big|\widetilde{m}(w) - m(w)\big| \;\le\; L\, h_n^{s \wedge 1} \;+\; \frac{C_0}{D_{\min}} \sqrt{\frac{\log n_1}{n_1\, \psi(h_n)}}. \tag{29}$$

Equivalently, in O_P form,

$$\sup_{w \in \mathcal{S}_F} \big|\widetilde{m}(w) - m(w)\big| \;=\; O_P\!\left(h_n^{s \wedge 1} + \sqrt{\frac{\log n_1}{n_1\, \psi(h_n)}} \right).$$

The bound (29) is path by path on $\Omega_T^{\#}$. The two terms reflect the bias of the functional Nadaraya Watson smoother under (A4) and the stochastic deviation of its kernel weighted numerator under β-mixing in the FPC fractal regime (A6*). The proof, given in the Theoretical Supplement, combines the denominator lower bound of Lemma 1, the stochastic numerator deviation of Lemma 2, and the deterministic bias bound of (A4).

4.5. Plug-in Uniform Deviation for $\widehat{m}$

Theorem 3 (Plug-in uniform deviation for $\widehat{m}$). *Assume the conditions of Theorem 2, together with (A10), (A11A), and (A12). There exists $C^* > 0$ such that, in O_P form,*

$$\sup_{w\in\mathcal{S}_F} \left|\widehat{m}(w) - m(w)\right| \;=\; O_P\!\left(h_n^{s\wedge 1} + \sqrt{\frac{\log n_1}{n_1\,\psi(h_n)}} + \sqrt{\frac{\log n_1}{n_1}} \right) \;=\; O_P(r_{m,n}). \tag{30}$$

A high probability formulation: on a measurable event of probability at least $1 - \delta_{T,n} - O(n_1^{-1})$,

$$\sup_{w\in\mathcal{S}_F} \left|\widehat{m}(w) - m(w)\right| \;\le\; L\,h_n^{s\wedge 1} + \frac{C_0}{D_{\min}}\sqrt{\frac{\log n_1}{n_1\,\psi(h_n)}} + C^*\sqrt{\frac{\log n_1}{n_1}}.$$

Since $\psi(h_n) \le 1$, the term $\sqrt{\log n_1/n_1}$ is always dominated by $\sqrt{\log n_1/(n_1\,\psi(h_n))}$, so the plug in rate matches the oracle rate; we display the term separately because it tracks the distinct $\widehat{\beta}$-transfer source.

4.6. Uniform Consistency of $\check{\sigma}$

Theorem 4 (Uniform consistency of $\check{\sigma}$). *Assume the conditions of Theorem 3, together with ($A4_\sigma$) and ($A_{\sigma_{\min}}$). Then*

$$\sup_{w\in\mathcal{S}_F} \left|\check{\sigma}(w) - \sigma(w)\right| \;=\; O_P(r_{\sigma,n}), \qquad r_{\sigma,n} \;=\; \max\!\left\{h_n^{s_\sigma\wedge 1},\, h_n^{s\wedge 1}\right\} + \sqrt{\frac{\log n_1}{n_1\,\psi(h_n)}} + \sqrt{\frac{\log n_1}{n_1}}. \tag{31}$$

A high probability counterpart with explicit constants and failure probability $O\big(\delta_{T,n} + n_1^{-1}\big)$ is given in the Theoretical Supplement. Under $s_\sigma \ge s \wedge 1$ and (A8′), the rate of $\check{\sigma}$ matches the rate of $\widehat{m}$ in Theorem 3; otherwise the slower of the two bias rates governs the maximum.

4.7. Auxiliary Lemmas

The following five lemmas are the building blocks of Theorems 1–4.

Lemma 1 (Denominator lower bound). *Assume (A1), (A5*), (A6*), (A7), and (A8′). Define the kernel denominator*

$$\widehat{D}(w) \;:=\; \frac{1}{n_1}\sum_{i\in\mathcal{I}_1} K\Big(\frac{d_q(w,W_i)}{h_n}\Big). \tag{32}$$

There exists a constant $D_{\min} > 0$ depending on (k_0, c_ψ, C_ψ, q) and an event Ω_1^* with

$$\mathbb{P}\big((\Omega_1^*)^c\big) \;\le\; C\,n_1^{-2} + 2\,\mu_1\,\beta(b_n)$$

such that on Ω_1^*,

$$\inf_{w\in\mathcal{S}_F} \widehat{D}(w) \;\ge\; D_{\min}\,\psi(h_n). \tag{33}$$

Lemma 2 (Stochastic numerator deviation). *Assume (A1), (A2′), (A5*), (A6*), (A7), (A8′), and (A9). Define*

$$\widehat{N}_1(w) \;:=\; \frac{1}{n_1}\sum_{i\in\mathcal{I}_1} K\Big(\frac{d_q(w,W_i)}{h_n}\Big)\,\varepsilon_i. \tag{34}$$

There exists a constant $C_0 > 0$ depending on the structural quantities of (A), on q, and on the chosen κ, and an event $\Omega_n^\#$ with

$$\mathbb{P}\big((\Omega_n^\#)^c\big) \;\le\; C\,n_1^{-2} + 2\,\mu_1\,\beta(b_n) + 2\,n_1^{1-\kappa}$$

for any chosen $\kappa \geq q+3$, *such that on* $\Omega_n^{\#}$,

$$\sup_{w \in \mathcal{S}_F} \left|\widehat{N}_1(w)\right| \;\leq\; C_0 \sqrt{\frac{\psi(h_n)\,\log n_1}{n_1}}. \tag{35}$$

Lemma 2 is the technical centerpiece of the paper. The proof in the Theoretical Supplement integrates alternating block Yu coupling on the training sample, Bernstein concentration for β-mixing summands with sub-Gaussian errors, chaining over an (h_n/n_1)-net of $\mathcal{S}_F$ in the FPC projection semi-metric, and the small ball variance bound implied by (A6*). On combining the denominator lower bound (33) with (35), the stochastic numerator divided by the denominator yields the $\sqrt{\log n_1/(n_1\,\psi(h_n))}$ regression error rate of Theorem 2.

Lemma 3 (Calibration quantile consistency)**.** *Assume (A1), (C1), and (C3). Let* $S_i := |\varepsilon_i|/\sigma(W_i)$ *denote the ideal score,* $\widehat{q}_{1-\alpha}^{\,\mathrm{ideal}}$ *the* $\lceil (1-\alpha)(n_2+1)\rceil$*-th order statistic of* $\{S_i\}_{i\in\mathcal{I}_2}$, *and* $\widehat{q}_{1-\alpha}$ *the empirical conformal quantile of* (25). *Then*

$$\left|\,\widehat{q}_{1-\alpha}^{\,\mathrm{ideal}} - q_{1-\alpha}\,\right| \;=\; O_P\!\left(\frac{1}{m_S}\sqrt{\frac{b_{2,n}\,\log n_2}{n_2}}\right). \tag{36}$$

If, in addition, the assumptions of Lemma 4 hold, then on the good event of Lemma 4,

$$\left|\,\widehat{q}_{1-\alpha} - q_{1-\alpha}\,\right| \;=\; O_P\!\left(\sqrt{\frac{b_{2,n}\,\log n_2}{n_2}} + \frac{r_{\mathrm{pred},n}}{\sigma_{\min}} + q_{1-\alpha}\,\frac{r_{\sigma,n}}{\sigma_{\min}}\right). \tag{37}$$

Lemma 4 (Affine score perturbation)**.** *Assume the conditions of Theorem 3 and Theorem 4, together with (A11A). Define the calibration/future X-control event*

$$\Omega_X^{\mathrm{cal}} \;:=\; \Big\{\max_{i\in\mathcal{I}_2\cup\{n+1\}} \|X_i\|_2 \;\leq\; C_X^*\,\log n\Big\}, \tag{38}$$

where C_X^* *depends on (A11A) and on a chosen polynomial tail exponent* $A_{\mathrm{tail}} \geq 2$, *with* $\mathbb{P}\big((\Omega_X^{\mathrm{cal}})^c\big) \leq 2\,p\,n^{-A_{\mathrm{tail}}}$. *Set*

$$A_n \;:=\; \frac{2\,r_{\mathrm{pred},n}}{\sigma_{\min}}, \qquad B_n \;:=\; \frac{2\,r_{\sigma,n}}{\sigma_{\min}}. \tag{39}$$

On the intersection Ω_{train} *of the good events of Lemmas 1 and 2, Theorems 3 and 4, and* Ω_X^{cal}, *the studentized scores satisfy the global affine bound*

$$(1-B_n)\,S_i \;-\; A_n \;\leq\; \widehat{S}_i \;\leq\; A_n \;+\; (1+B_n)\,S_i, \qquad i\in\mathcal{I}_2\cup\{n+1\}, \tag{40}$$

with no restriction on the value of S_i. *Equivalently,* $|\widehat{S}_i - S_i| \leq A_n + B_n\,S_i$ *globally on* Ω_{train}. *The good event failure probability* $\delta_{\mathrm{good},n}$, *which appears in Theorem 1, is the union bound aggregate of all high probability events listed in Section 4.1, and, under (A8′) and (C3), satisfies* $\delta_{\mathrm{good},n} = O(n_1^{-1}) + O(n_2^{-2}) + O(n^{-A_{\mathrm{tail}}})$, *with the individual components* $\mu_1\,\beta(b_n) + O(n_1^{-2})$, $O(n_1^{-1})$, $O(n_2^{-2})$, *and* $O(n^{-A_{\mathrm{tail}}})$ *corresponding to the denominator/numerator, training estimation, calibration empirical CDF, and calibration / future X-control events, respectively.*

The bound (40) is global in S_i. It does not require S_i to lie in any neighborhood of $q_{1-\alpha}$, and in particular no density assumption on S is invoked in its statement. The multiplicative slope B_n controls the scale distortion contribution $|\varepsilon_i|\cdot|1/\check{\sigma} - 1/\sigma|$ in the tails through the cancellation $|\varepsilon_i| = S_i\,\sigma(W_i)$, and the additive offset A_n controls the prediction error

contribution $|a_i|/\check{\sigma}$, both uniformly in $i \in \mathcal{I}_2 \cup \{n+1\}$ on Ω_{train}. Since $B_n = 2\, r_{\sigma,n}/\sigma_{\min} \to 0$ under (A8′) and the rate definitions of Section 2.7, we work throughout Section 4 and the Theoretical Supplement for n large enough that $B_n \le 1/2$; the bound $1 - B_n \ge 1/2$ then guarantees non degeneracy of the lower affine slope used by the order statistic monotonicity argument of Lemma 3 Step 3.

Lemma 5 (Dependence handling). *Assume (A1) and (C1)–(C4). The ideal score* $S_i := |\varepsilon_i|/\sigma(W_i)$ *is a deterministic function of* Z_i *alone, and the ideal conformal quantile* $\widehat{q}^{\text{ideal}}_{1-\alpha}$ *is a deterministic functional of* $\{S_i : i \in \mathcal{I}_2\}$*; both depend on the data through* $(Z_i)_{i \in \mathcal{I}_2 \cup \{n+1\}}$ *only. The marginal ideal score conformal coverage satisfies*

$$\mathbb{P}\Big(S_{n+1} \le \widehat{q}^{\text{ideal}}_{1-\alpha} \Big) \ge 1 - \alpha - \frac{1}{n_2+1} - C_1 \frac{n_2}{b_{2,n}}\, \beta(b_{2,n}) - C_3 \sqrt{\frac{b_{2,n} \log n_2}{n_2}} - C_2\, \beta(g_n), \tag{41}$$

where C_1, C_2, C_3 *are finite positive constants that may depend on the fixed structural quantities* $(B_0, \rho, m_S, M_S, \alpha)$ *and on the fixed tuning-rule constants* (C_b, C_g) *from (C3)–(C4), but* not *on* $n_1, n_2, h_n, b_n, b_{2,n}$*, or* g_n*.*

Lemma 5 is a specialization to the SFPLM setting of the non exchangeability framework of [15]. The lemma does not claim exact rank exchangeability of the calibration scores under β-mixing; the term $C_1(n_2/b_{2,n})\beta(b_{2,n})$ reflects the total variation cost of replacing the calibration joint law by a block independent surrogate via Yu coupling, and the term $C_3\sqrt{b_{2,n}\log n_2/n_2}$ is the calibration empirical process cost under the block independent surrogate the empirical CDF deviation rate of the cal scores from F_S at effective block size $b_{2,n}$. Both costs are made fully explicit in the proof of the Theoretical Supplement.

4.8. Rate Under the FPC Fractal Small Ball Regime

Corollary 1. *Under the assumptions of Theorem 1 and the FPC fractal small ball regime (10), the bandwidth optimal choice* $h_n^* \asymp (\log n_1/n_1)^{1/(q+2(s\wedge 1))}$ *satisfies*

$$r_{m,n} \asymp \Big(\frac{\log n_1}{n_1}\Big)^{(s\wedge 1)/(q+2(s\wedge 1))}, \tag{42}$$

up to the dominated $\sqrt{\log n_1/n_1}$ *term. With* $b_{2,n} = \lceil C_b \log n_2/\log(1/\rho)\rceil$ *and* $g_n = \lceil C_g \log n/\log(1/\rho)\rceil$ *for* $C_b, C_g \ge 3$ *as in (C3)–(C4), under* $s_\sigma \ge s \wedge 1$*, and assuming additionally that* n_1 *and* n_2 *grow polynomially in* n *(for instance proportionally to* n*, as in the canonical splits of Section 3.9; this is used to dominate* L_n *and the logarithmic factors), the theoretical finite-sample coverage penalty takes the simplified form*

$$\Delta_n = O\Bigg(\Big(\frac{\log n_1}{n_1}\Big)^{(s\wedge 1)/(q+2(s\wedge 1))} + \frac{\log n_2}{\sqrt{n_2}} + \beta(g_n) + \frac{1}{n_2+1} + \delta_{\text{good},n} \Bigg). \tag{43}$$

The leading two terms in (43) are the kernel regression perturbation rate and the calibration empirical process cost $\log n_2/\sqrt{n_2}$, which arises from the simplification $\sqrt{b_{2,n}\log n_2/n_2} = \log n_2/\sqrt{n_2}$ under $b_{2,n} \asymp \log n_2$ as in (C3). For small q the kernel rate dominates; for larger q, the calibration empirical process term may dominate. The $\beta(g_n)$ and $1/(n_2+1)$ terms are polynomially small under the geometric mixing rate of (A1). Under (A8′), $\mu_1\,\beta(b_n) \le n_1^{-1}$, so $\delta_{\text{good},n} = O(n_1^{-1}) + O(n_2^{-2}) + O(n^{-A_{\text{tail}}})$, while its displayed components remain the denominator/numerator event failures, the training estimation event failures, the calibration empirical CDF event failure, and the calibration/future X-control event failures.

4.9. Relation to BCRT 2023

Remark 7. *Theorem 1 can be viewed as a problem specific specialization, to the SFPLM under β-mixing, of the generic non exchangeability framework of [15], whose coverage gap is bounded by the average of total variation distances between the joint law of the augmented calibration sample and its position swapped versions. The proof of Lemma 5 is BCRT compatible in spirit but proceeds through a direct total variation coupling to a block independent surrogate and a β-mixing empirical CDF inversion, rather than through an exact exchangeable rank argument. The relation of Δ_n to [15] and to the dependence-only analyses of [16,17], including all tightness and novelty statements, is consolidated in Section 8.3.*

5. Outline of the Proofs

Complete proofs of Lemmas 1–5 and Theorems 2–4 and 1 are given in the Theoretical Supplement (Section S1), in statement order. The proof has two halves: a training half that establishes uniform deviation bounds for the kernel estimators on high probability events, and a calibration half that analyzes the conformal quantile under dependence and assembles the coverage bound by a transfer argument anchored at the ideal score $s^\star$.

Training half. Lemma 1 lower-bounds the kernel denominator uniformly over $\mathcal{S}_F$ through the type-II kernel property and the fractal small ball regime, controlling the stochastic deviation by alternating-block Yu coupling [19], Bernstein concentration for the independent block sums [20,21], and chaining over an (h_n/n_1)-net whose cardinality is fixed by (A5*). Lemma 2 applies the same architecture to the kernel weighted error numerator with a $\sqrt{C \log n_1}$ truncation of the sub-Gaussian errors. Theorem 2 adds the Hölder bias bound under (A4); Theorem 3 transfers it to $\widehat{m}$ through the profiled structure and the $\widehat{\beta}$ event of (A10), the inflation L_n tracking evaluation at calibration and future covariates; Theorem 4 repeats the argument for the variance estimator and propagates the truncation $\check{\sigma} \geq \sigma_{\min}/2$.

Calibration half and assembly. Lemma 3 controls the empirical CDF of the ideal calibration scores around $q_{1-\alpha}$ by alternating $b_{2,n}$-blocks, Yu coupling at cost $C_1(n_2/b_{2,n})\beta(b_{2,n})$, blockwise concentration at rate ω_n, and CDF-to-quantile inversion through (C1). Lemma 4 compares $\widehat{S}$ with $s^\star$ on the good event by a global affine perturbation (offset A_n, slope B_n), shifting the conformal quantile by at most η_n. Lemma 5 prices the dependence between calibration block and future point through two separate Berbee–Yu couplings across the two buffers, each at cost $\beta(g_n)$. The proof of Theorem 1 conditions on the good events (cost $\delta_{\text{good},n}$), localizes the estimated quantile near $q_{1-\alpha}$ through a calibration-measurable event, transfers coverage from the implemented to the ideal-score interval through Lemma 4 (the strict positivity $q_{1-\alpha} > 0$ used in c_σ follows from (C1) and score nonnegativity), and controls the ideal-score conformal step by Lemmas 3 and 5 with the order-statistic rounding $1/(n_2+1)$; summing the costs yields Δ_n as displayed in (13).

6. Simulation Study

The simulation study was conducted under a protocol specified and frozen before any outcome comparison: every design choice (data generating mechanisms, methods, sample sizes, seeds, replication count, and evaluation rules) was fixed and hash-stamped before any publication run, and no design parameter was changed after outcomes were observed. Each cell uses 500 Monte Carlo replications, chosen before execution from a Monte Carlo precision target; all 36,500 attempted method replications across the study completed with zero failures, and a scripted post-run gate checked completeness, seed integrity and precision flags. The complete numerical tables are collected in a separate numerical supplement (Tables S1–S6). The empirical claim is conservative: in every regime studied, all methods, gapped and contiguous, studentized and plain, attain coverage

close to the nominal level, and the value of the proposed procedure lies in the theoretical decomposition of Theorem 1, not in empirical dominance.

6.1. Design and Methods Compared

We compare five split conformal variants at miscoverage level $\alpha = 0.10$, all built on the estimators of Section 3:

(M1) Gapped, studentized (the proposed procedure): the three way split (training, gap, calibration, gap, test) with the studentized absolute residual score (23).
(M2) Contiguous, studentized: identical estimators and score, but $g_n = 0$, so the calibration and test segments follow the training segment without a buffer.
(M3) Gapped, non-studentized: as M1 with the plain absolute residual score $|Y_i - X_i^\top \widehat{\beta} - \widehat{m}(W_i)|$.
(M4) Contiguous, non-studentized: as M2 with the plain absolute residual score.
(M5) Gapped block bootstrap: a non-conformal competitor on M1's split; calibration studentized residuals are block bootstrapped (block length $\lceil \log n_2 \rceil$, 200 bootstrap replicates) and the median of the bootstrap $(1-\alpha)$-quantiles is used as the half width multiplier.

The M1–M2 contrast isolates the buffer gap; the M1–M3 and M2–M4 contrasts isolate studentization; M5 benchmarks a resampling alternative. Table 2 summarizes the design. Four experiments are run: Experiment 1 (baseline), Experiment 2 (strong dependence with heteroskedasticity), Experiment 3 (a boundary stress test whose autoregressive residual noise violates (A2′) by design), and Experiment 4, a new dependent-score experiment whose ideal score sequence has exactly quantified serial dependence.

Table 2. Simulation design. All experiments use 500 Monte Carlo replications per cell, level $\alpha = 0.10$, test size 200, $K_{\text{true}} = 5$ generating components, $q = 3$ estimated components, the Epanechnikov kernel, and the dense backend. Gap lengths are the operational values applied by the split. Base seeds: 42 (Exp. 1), 4242 (Exp. 2), 9999 (Exp. 3), 7777 (Exp. 4). Experiments 1–2 initialize all latent chains from their exact stationary laws (burn-in 0); Experiment 3 uses a verified 200-step burn-in; Experiment 4 draws the magnitude chain from its exact stationary distribution. Variance floors: $\sigma_{\min} = 0.5$ (Exps. 1–3, submitted configuration) and $\sigma_{\min} = 2/5$ with floor $1/5$ (Exp. 4).

Experiment	ρ	n_1	n_2	Test	Operational Gaps
Exp. 1	0.3, 0.7	1000	1000	200	20, 66
Exp. 2	0.85, 0.9	1000	200, 500	200	138, 141, 214, 219
Exp. 3	0.85 ($\psi_\varepsilon = 0.85$)	1000	500	200	141
Exp. 4	0.3, 0.85	1000	500	200	19, 141

6.2. Data Generating Mechanism

The simulation design separates the truth generating mechanism from the theoretical FPC dimension used by the estimators. The truth has $K_{\text{true}} = 5$ active basis directions, while the projected covariate used by the regression and the semi-metric d_q has *theoretical FPC dimension* $q = 3$, in agreement with the conditions of Theorem 1 (cf. Section 2.6). The construction is chosen so that the projected covariate $P_q W$ has bounded support on $\mathbb{R}^3$, in agreement with (A5*) and (A6*).

Curves $W_i : [0,1] \to \mathbb{R}$ are sampled on a 100-point grid through the orthonormal sine basis $\phi_k(t) = \sqrt{2}\,\sin(\pi k t)$, $k = 1, \ldots, K_{\text{true}}$.

Independent componentwise AR-to-uniform scores. For each component $k = 1, \ldots, K_{\text{true}}$, draw a stationary scalar Gaussian AR(1) process $G_{i,k}$ marginally $\mathcal{N}(0,1)$,

$$G_{i,k} \;=\; \rho\, G_{i-1,k} \;+\; \sqrt{1-\rho^2}\, e_{i,k}, \qquad e_{i,k} \overset{\text{iid}}{\sim} \mathcal{N}(0,1) \text{ across } i,k,$$

with the AR(1) sequences mutually independent across k. Apply the componentwise probability integral transform $U_{i,k} := \Phi(G_{i,k})$, so that $U_{i,k}$ is uniform on $[0,1]$ marginally, and define the FPC scores by the affine map

$$\xi_{i,k} := \sqrt{12\,\lambda_k}\,(U_{i,k} - 1/2), \qquad \lambda_k = k^{-2},$$

so that $\xi_{i,k}$ is uniform on $[-\sqrt{3\lambda_k}, \sqrt{3\lambda_k}]$ with mean zero and variance λ_k. The componentwise PIT is a fixed measurable transformation and therefore preserves geometric β-mixing of each scalar AR(1); independence across k preserves geometric β-mixing of the joint sequence $\xi_i = (\xi_{i,1}, \ldots, \xi_{i,K_{\text{true}}})$. Curves are reconstructed by $W_i(t) = \sum_{k=1}^{K_{\text{true}}} \xi_{i,k}\,\phi_k(t)$.

Projected support bounded. For the theoretical FPC dimension $q = 3$, the projected covariate is $P_q W_i = (\xi_{i,1}, \xi_{i,2}, \xi_{i,3}) \in \mathbb{R}^3$, supported on the bounded rectangle $\mathcal{S}_F = \prod_{k=1}^{3}[-\sqrt{3\lambda_k}, \sqrt{3\lambda_k}] \subset \mathbb{R}^3$, with constant joint density on $\mathcal{S}_F$ (independent uniforms). Hence (A5*) and (A6*) hold by construction with $c_f = C_f$ equal to the inverse Lebesgue volume $\prod_{k=1}^{3}(2\sqrt{3\lambda_k})^{-1}$. The remaining components $(\xi_{i,4}, \xi_{i,5})$ contribute to the truth W_i via ϕ_4, ϕ_5 but are ignored by the FPC projected estimators, which use $\widehat{P}_q$ at $q = 3$ as in Section 3.2; this is the operational distinction between the truth dimension $K_{\text{true}} = 5$ and the theoretical FPC dimension $q = 3$.

Scalar covariate. The covariate $X_i \in \mathbb{R}^3$ follows an analogous componentwise AR-to-uniform construction: $H_{i,j} = \rho\,H_{i-1,j} + \sqrt{1-\rho^2}\,e'_{i,j}$ with $H_{i,j} \sim \mathcal{N}(0,1)$ marginally and the scalar processes mutually independent across j, and $X_{i,j} := \sqrt{12}\,(\Phi(H_{i,j}) - 1/2)$ is uniform on $[-\sqrt{3}, \sqrt{3}]$ with mean zero and unit variance. We fix $\beta_0 = (1,\, 0.5,\, -0.3)^\top$ and $m(W_i) = \langle W_i, \phi_1\rangle + 0.5\,\sin(\langle W_i, \phi_2\rangle)$; both the response and the smooth component depend on W_i only through the first $q = 3$ projected scores, in agreement with the d_q-Lipschitz condition (A4).

Conditional variance. The conditional variance function takes one of two forms:

$$\sigma^2(W_i) = \begin{cases} 0.25 + 0.5\,\langle W_i, \phi_1\rangle^2 & \text{(Experiment 1, baseline)}, \\ 0.16 + 1.5\,|\langle W_i, \phi_1\rangle| & \text{(Experiment 2, strong dep. + heteroskedasticity)}, \end{cases} \tag{44}$$

both bounded above on $\mathcal{S}_F$ since $\langle W, \phi_1\rangle = \xi_1$ has bounded support. In Experiments 1–2, errors are drawn conditionally independently given the covariate path, $\varepsilon_i \mid W_i \sim \mathcal{N}(0, \sigma^2(W_i))$; in particular $\mathbb{E}[\varepsilon_i \mid \mathcal{F}_{i-1} \vee \sigma(X_i, W_i)] = 0$ and $\mathbb{E}[\varepsilon_i^2 \mid \mathcal{F}_{i-1} \vee \sigma(X_i, W_i)] = \sigma^2(W_i)$, so the martingale-difference form of (A2′) holds. For the boundary stress test (Experiment 3), the conditional independence structure is replaced by an AR(1) residual recursion

$$\varepsilon_i = \psi_\varepsilon\,\varepsilon_{i-1} + \sqrt{1-\psi_\varepsilon^2}\,\nu_i, \qquad \nu_i \mid W_i \sim \mathcal{N}(0,\, \sigma^2(W_i)), \qquad \psi_\varepsilon = 0.85, \tag{45}$$

with σ^2 the strong heteroskedastic form of (44). Experiment 3 violates the conditional noise structure of (A2′); we report it as a boundary diagnostic, not as support for the theorems.

The response is $Y_i = X_i^\top \beta_0 + m(W_i) + \varepsilon_i$.

Experiment 4: a dependent-score construction with exact properties. Experiments 1–2 have serially dependent covariates but conditionally independent errors, so their ideal score sequences are independent. Experiment 4 closes this gap. It makes the ideal score sequence itself serially dependent, with the dependence level known in closed form, while satisfying every directly verifiable assumption of the theorem block. Covariates and the strong heteroskedastic variance form $\sigma^2(W_i) = 4/25 + (3/2)\,|\langle W_i, \phi_1\rangle|$ are as in Experiment 2. The error is

$$\varepsilon_i = \sigma(W_i)\,s_i\,R_i, \qquad R_i = \frac{r_{H_i}\,U_i}{c_\delta}, \qquad c_\delta = \sqrt{301/300},$$

where s_i are iid Rademacher signs, U_i are iid uniform on $[1 - 1/10, 1 + 1/10]$, and H_i is a stationary three state Markov chain, independent of the covariates, with magnitude states $r = (1/5, 1, 9/5)$, transition matrix

$$P = \begin{pmatrix} 13/20 & 1/14 & 39/140 \\ 7/20 & 1/2 & 3/20 \\ 1/20 & 13/14 & 3/140 \end{pmatrix}, \qquad \pi = \big(49/116,\ 23/58,\ 21/116\big),$$

started exactly from its stationary law π. Every row j of P satisfies $\sum_k P_{jk} r_k^2 = 1$ and $\mathbb{E}[U_i^2] = 301/300 = c_\delta^2$, so $\mathbb{E}[R_i^2 \mid \text{past}] = 1$ exactly and the conditional moment structure of (A2′) is preserved: ε_i is a conditionally centered martingale difference with conditional variance $\sigma^2(W_i)$. The ideal score is $S_i = |\varepsilon_i|/\sigma(W_i) = R_i$, a strictly stationary sequence with exact lag-one autocorrelation

$$\operatorname{corr}(S_i, S_{i+1}) = \frac{75456}{444365} \approx 0.1698,$$

computed in closed form from (r, P, π) and verified numerically to that value in the frozen protocol. Because U_i has a Lebesgue density, S_i has a piecewise constant density that is strictly positive on the neighborhood $[q_{0.90} - \xi,\ q_{0.90} + \xi]$ with $\xi = 1/(10\, c_\delta)$ of the population quantile $q_{0.90} = 3117/(1750\, c_\delta)$, with density value $175\, c_\delta/348$ there, so (C1) holds with exact constants. The chain H_i is a finite state aperiodic irreducible Markov chain and is geometrically β-mixing, and the variance floor holds with $\sigma_{\min} = 2/5$ exactly ($\inf_w \sigma^2(w) = 4/25$). The asymptotic tuning conditions (A8′)/(C5) and the high level $\widehat{\beta}$-deviation assumption (A10) are not finite sample verifiable claims and remain assumptions here, as in the theory; we therefore do not describe Experiment 4 as unconditionally satisfying all theorem assumptions. The condition $\mathbb{E}[R_i^2 \mid \text{past}] = 1$ restricts the admissible dependence constructions; the frozen design supplies a nonzero and exactly quantified absolute-score dependence level, and no claim is made that its autocorrelation is maximal over all (A2′)-compatible processes. Sampling uses four mutually independent random number streams per replication (covariates, chain, magnitudes, signs).

The response in every experiment is $Y_i = X_i^\top \beta_0 + m(W_i) + \varepsilon_i$.

6.3. Estimation, Tuning and Variance Floors

The FPC dimension is fixed at $q = 3$. The empirical FPC projection $\widehat{P}_q$ is computed from the training curves as in Section 3.2, and FPC scores are obtained by Euclidean inner product on the discretized grid (a finite dimensional embedding for d_q). The bandwidth is $h_n^* = (\log n_1/n_1)^{1/(q+2)}$ multiplied by the median pairwise d_q-distance on $\mathcal{D}_{\text{train}}$, the kernel is Epanechnikov, and the gap length for the gapped methods is $g_n = \lceil 3 \log n / \log(1/\rho) \rceil$ computed by the two pass rule of the frozen protocol (operational values in Table 2).

The variance floors are reported per experiment. In Experiment 1, $\inf_w \sigma(w) = 0.5$, and the implementation floor $\check{\sigma} \geq 0.25$ equals $\sigma_{\min}/2$ for $\sigma_{\min} = 0.5$, matching the theoretical truncation rule. In Experiments 2–3, $\inf_w \sigma(w) = 0.4$, while the submitted configuration $\sigma_{\min} = 0.5$ with floor 0.25 is retained for continuity with the original study; the floor 0.25 lies below the realised $\sigma(W_i)$ in every replication of the main grid (zero floor activations were recorded), so the discrepancy is numerically inactive. Experiment 4 uses the theorem-compatible values exactly: $\sigma_{\min} = 2/5$ and floor $1/5 = \sigma_{\min}/2$.

In Experiments 1 and 2, the studentized ideal score $S = |\varepsilon|/\sigma(W)$ has a smooth Lebesgue density in a neighbourhood of the population 0.90-quantile because the noise is conditionally Gaussian, $\sigma(W)$ is bounded away from zero on the compact projected support $\mathcal{S}_F$, and the joint law of (ε, W) has a Lebesgue density by construction, so (C1) holds. In Experiment 4, (C1) holds with the exact constants displayed in Section 6.2. Experiment 3

is retained only as a boundary stress test outside the conditional noise structure of (A2′), so (C1) need not be verified there.

We report per cell the empirical test coverage, its Monte Carlo standard error, and the mean and median interval length, over 500 Monte Carlo replications. The replication count was fixed in advance from a planning target of Monte Carlo half width 0.005 for coverage; the achieved standard errors (at most 0.0024 across all cells) meet the target in every cell.

6.4. Main Coverage and Length Results

Table 3 summarizes the main grid, nine cells (Table 2) by five methods, through the quantities that support the principal claims; the complete five-method coverage and length tables are Tables S1 and S2 of the Numerical Supplement. Figure 1 displays the coverage results with 95% Monte Carlo error bars.

Table 3. Main grid summary (500 replications per cell; Monte Carlo standard errors in parentheses). Paired Δ is the mean per-replication paired difference coverage (M1) − coverage (M2) with its paired Monte Carlo standard error ($\times 10^3$); no paired difference is statistically resolved at the 5% level. Length Δ% is the percentage excess of the plain-score M3 mean length over the studentized M1 mean length. Complete five-method coverage, length, paired and diagnostics tables are Tables S1–S3 and S6 of the Numerical Supplement.

Cell	M1 Coverage	M2 Coverage	Paired $\Delta \times 10^3$	M3 vs. M1 Length Δ%	M5 Coverage
Exp. 1, $\rho = 0.3$, $n_2 = 1000$	0.9001 (0.0011)	0.9003 (0.0011)	−0.17 (0.64)	+2.7	0.8990 (0.0011)
Exp. 1, $\rho = 0.7$, $n_2 = 1000$	0.9004 (0.0011)	0.9010 (0.0010)	−0.68 (1.21)	+2.9	0.8995 (0.0011)
Exp. 2, $\rho = 0.85$, $n_2 = 200$	0.9005 (0.0013)	0.8993 (0.0014)	+1.24 (2.10)	+3.5	0.8954 (0.0013)
Exp. 2, $\rho = 0.85$, $n_2 = 500$	0.9012 (0.0012)	0.8989 (0.0012)	+2.34 (1.64)	+3.8	0.8992 (0.0012)
Exp. 2, $\rho = 0.9$, $n_2 = 200$	0.8984 (0.0013)	0.9025 (0.0015)	−4.17 (2.40)	+3.2	0.8934 (0.0013)
Exp. 2, $\rho = 0.9$, $n_2 = 500$	0.9009 (0.0012)	0.8996 (0.0012)	+1.34 (1.85)	+3.6	0.8989 (0.0012)
Exp. 3, $\rho = 0.85$, $n_2 = 500$	0.9001 (0.0023)	0.8987 (0.0022)	+1.43 (3.07)	−2.3	0.8973 (0.0024)
Exp. 4, $\rho = 0.3$, $n_2 = 500$	0.9007 (0.0010)	0.9018 (0.0010)	−1.14 (0.59)	+3.1	0.8992 (0.0010)
Exp. 4, $\rho = 0.85$, $n_2 = 500$	0.9003 (0.0011)	0.8992 (0.0010)	+1.17 (1.51)	+3.3	0.8986 (0.0011)

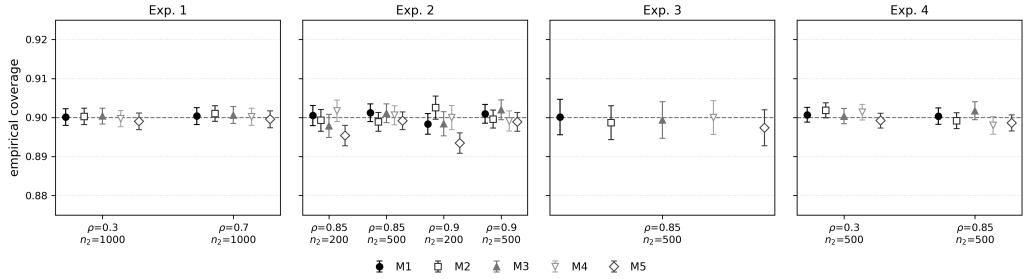


Figure 1. Empirical coverage by method and cell at $\alpha = 0.10$ (nominal 0.90, dashed line), 500 replications per cell; error bars are ±1.96 Monte Carlo standard errors. All 45 method cells lie within 0.007 of the nominal level.

Three regularities summarize the grid. First, coverage is close to nominal everywhere: across all 45 method cells the empirical coverage ranges from 0.8934 to 0.9025. The gapped and contiguous studentized methods are statistically indistinguishable in coverage in every cell, including both cells of the dependent-score Experiment 4 (M1 0.9007 and 0.9003 versus M2 0.9018 and 0.8992, with Monte Carlo standard errors near 0.0010). A per-replication paired comparison on identical data confirms this: no paired M1−M2 coverage difference is statistically resolved at the 5% level in any of the nine cells, and the largest absolute z-distance from zero is 1.94 (Table 3; the complete paired analysis is Table S6 of the Numerical Supplement). Second, studentization showed no systematic coverage advantage; all methods remained close to nominal coverage, while interval-length differences were systematic: in the heteroskedastic Experiment 2 the studentized M1

intervals are about 3.5% shorter on average than the plain M3 intervals (Table 3; Table S2 of the Numerical Supplement), and the same ordering holds for M2 versus M4. Third, the block bootstrap M5 is the only method with cells below 0.895 (minimum 0.8934 at $\rho = 0.90$, $n_2 = 200$), a mild undercoverage consistent with bootstrap quantile bias at small calibration sizes.

6.5. Score Dependence Diagnostics

Nominal marginal coverage does not establish exchangeability or independence of the conformity scores, so the frozen protocol specified in advance direct diagnostics of the score series that the methods actually use. Figure 2 summarizes the score autocorrelations; the complete diagnostics table, with autocorrelations, the Geyer initial positive sequence integrated autocorrelation time (IAT), effective sample size (ESS), and the Ljung–Box statistic, for the ideal standardized score and for the estimated scores of M1 and M3 on the calibration block, is Table S3 of the Numerical Supplement.

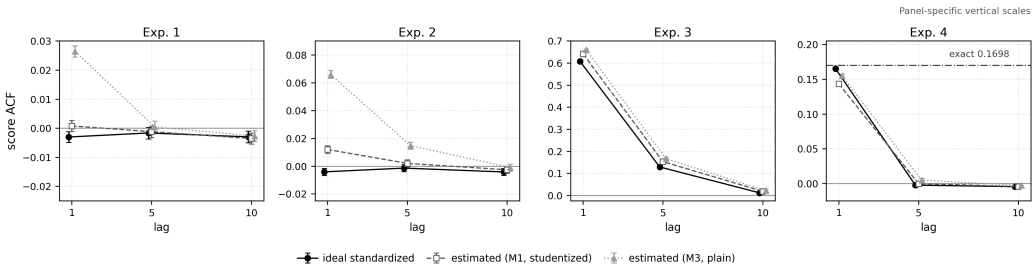


Figure 2. Mean score autocorrelations at lags 1, 5 and 10 on the calibration block (gap geometry), by experiment and score series, with 95% Monte Carlo intervals. In Experiment 4 the horizontal line marks the exact ideal-score lag-one autocorrelation $75456/444365 \approx 0.1698$.

The diagnostics separate the four regimes cleanly. In Experiments 1–2 the ideal scores are serially independent by construction and the measured autocorrelations are near zero; the estimated plain scores (M3) inherit mild positive dependence from the covariates through the fitted conditional mean (lag-one autocorrelation up to about 0.07 in Experiment 2), which studentization largely removes (M1 about 0.01). In Experiment 4 the estimated scores carry the designed dependence (measured mean lag-one autocorrelation 0.14–0.16 against the exact ideal value 0.1698), and ESS is reduced accordingly; coverage nevertheless remains nominal for all methods, consistent with Remark 6. In the boundary Experiment 3 the score dependence is strong (lag-one autocorrelation about 0.6, median IAT about 5, median ESS about 100 of $n_2 = 500$), and coverage still remains near nominal; since Experiment 3 violates (A2′), this stability is an empirical observation, not a theorem consequence.

6.6. Gap Sensitivity

The buffer gap is priced in Δ_n through $C_2\,\beta(g_n)$, but its operational cost is data discarded. Two designs, specified before the results were computed, separate these effects for M1 on the Experiment 2–4 DGPs at $\rho = 0.85$ ($\psi_\varepsilon = 0.85$ for the Experiment 3 DGP): Design A holds n_1, n_2 fixed and grows the generated series, so the gap costs generation length but no estimation data; Design B holds the total budget fixed at $N = 1987$, so larger gaps shrink n_1 and n_2. Gaps $g \in \{0, 70, 141, 282\}$ span zero to twice the theoretical rule. Table 4 reports the fixed-budget endpoints, Figure 3 the full coverage profiles, and the complete 24-row two-design table is Table S4 of the Numerical Supplement; the two designs are never pooled.

Table 4. Fixed-budget gap cost (Design B endpoints; total budget $N = 1987$, 500 replications per cell): no gap versus twice the theoretical gap. Discarded denotes observations inside the two buffer gaps. The full 24-row two-design sensitivity table is Table S4 of the Numerical Supplement.

DGP	g	n_1/n_2	Discarded	Coverage (MC SE)	Length (MC SE)
Exp. 2 DGP	0	1191/596	0	0.8989 (0.0012)	3.943 (0.010)
Exp. 2 DGP	282	815/408	564	0.9039 (0.0012)	4.008 (0.011)
Exp. 3 DGP	0	1191/596	0	0.8993 (0.0023)	4.233 (0.017)
Exp. 3 DGP	282	815/408	564	0.9011 (0.0025)	4.350 (0.021)
Exp. 4 DGP	0	1191/596	0	0.9033 (0.0010)	4.079 (0.009)
Exp. 4 DGP	282	815/408	564	0.9041 (0.0011)	4.095 (0.009)

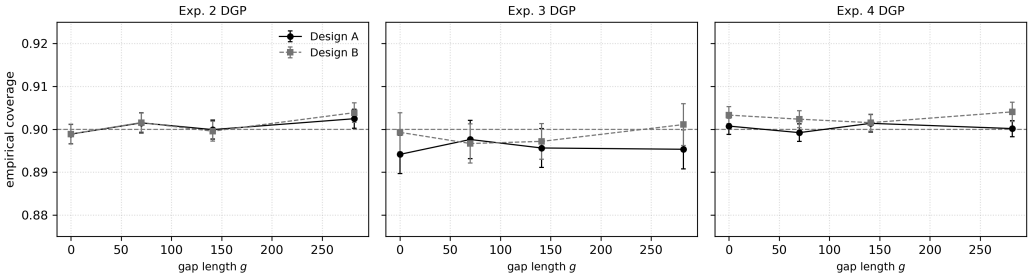


Figure 3. Gap sensitivity of coverage for the studentized gapped method under Design A (fixed n_1, n_2) and Design B (fixed total budget), 500 replications per point; error bars are ± 1.96 Monte Carlo standard errors.

Coverage is flat in the gap in both designs and all three DGPs, including the dependent-score DGP. Under the fixed budget of Design B, doubling the theoretical gap ($g = 282$) discards 28.4% of the observations; coverage remained close to nominal in every cell, and the entire cost appears as interval length, which grows by between 0.4% (Experiment 4 DGP) and 2.8% (Experiment 3 DGP) relative to $g = 0$. The buffer gap was therefore empirically close to costless under Design A, and no detectable coverage benefit from the buffer gap was found at the tested sample sizes and under the frozen data-generating processes; its role remains the support of the theoretical decomposition, and Section 8.3 draws the resulting practical guidance.

6.7. Observable Rate Proxies for the Coverage Penalty

Δ_n involves unknown structural constants and cannot be evaluated (Section 4.3). What can be measured is the rate behaviour of its components with the constants set to one. Table 5 reports these observable proxies for the gapped studentized method in each main grid cell: the three design constants $P_{\text{block}}, P_{\text{calEP}}, P_{\text{gap}}$ and the rounding term, together with measured estimation-error proxies P_{mean} (total conditional-mean error against the simulation truth, supremum over the test block, scaled by $\sigma_{\min}$) and P_σ (scaled conditional-scale error times the cell's median conformal quantile).

Two patterns are visible. The coupling and gap proxies are negligible at these geometric mixing rates (10^{-6} and below), while the calibration empirical process proxy and the estimation proxies are the non-trivial components, consistent with the discussion after Corollary 1. The estimation proxies shrink from $n_2 = 200$ to $n_2 = 500$ cells at fixed n_1 only weakly, as they are driven by the training size; across experiments they grow with the heteroskedasticity strength. These are diagnostics of rate behaviour only; they are not values of Δ_n and are never summed. Variance-floor activations and zero-neighbour kernel denominators, which are observable warning signs related to the good event, were zero in all replications of all cells.

Table 5. Observable rate proxies for the components of the theoretical finite-sample coverage penalty Δ_n (gapped method only), with unknown structural constants set to one. $P_{\text{block}} = (n_2/b_2)\rho_\star^{b_2}$, $P_{\text{calEP}} = \sqrt{b_2 \log n_2 / n_2}$, $P_{\text{gap}} = \rho_\star^g$, and $P_{\text{round}} = 1/(n_2+1)$ are design constants with $\rho_\star$ the frozen per-experiment mixing-rate proxy; P_{mean} and P_σ are measured against the simulation truth (medians over replications). These are diagnostics of rate behaviour only; they are not values of Δ_n, are not a certified finite-sample bound, and are never summed.

Cell	P_{block}	P_{calEP}	P_{gap}	P_{round}	P_{mean} (Med)	P_σ (Med)
Exp. 1, $\rho = 0.3$, $n_2 = 1000$	2.2×10^{-8}	0.353	3.5×10^{-11}	1.0×10^{-3}	1.224	0.861
Exp. 1, $\rho = 0.7$, $n_2 = 1000$	1.2×10^{-8}	0.638	6.0×10^{-11}	1.0×10^{-3}	1.227	0.867
Exp. 2, $\rho = 0.85$, $n_2 = 200$	2.5×10^{-7}	1.611	1.8×10^{-10}	0.005	1.581	1.100
Exp. 2, $\rho = 0.85$, $n_2 = 500$	3.3×10^{-8}	1.196	1.1×10^{-10}	0.002	1.552	1.101
Exp. 2, $\rho = 0.9$, $n_2 = 200$	1.6×10^{-7}	2.000	1.6×10^{-10}	0.005	1.602	1.099
Exp. 2, $\rho = 0.9$, $n_2 = 500$	2.2×10^{-8}	1.483	9.5×10^{-11}	0.002	1.597	1.105
Exp. 3, $\rho = 0.85$, $n_2 = 500$	3.3×10^{-8}	1.196	1.1×10^{-10}	0.002	2.207	1.565
Exp. 4, $\rho = 0.3$, $n_2 = 500$	1.3×10^{-7}	0.446	1.2×10^{-10}	0.002	1.915	1.198
Exp. 4, $\rho = 0.85$, $n_2 = 500$	3.3×10^{-8}	1.196	1.1×10^{-10}	0.002	1.987	1.224

6.8. Projection Sensitivity

Theorem 1 concerns the population projection P_q; the implemented pipeline uses the empirical $\widehat{P}_q$ (Section 3.2). A paired comparison, specified in the frozen protocol, runs both pipelines on identical data with identical seeds on the Experiment 2 DGP at $\rho \in \{0.3, 0.85\}$ and $n_1 \in \{250, 1000\}$. Table S5 of the Numerical Supplement reports the paired differences for all cells.

The differences are practically negligible: all paired coverage differences are within ± 0.0005, and the largest paired length difference is -0.0131 ± 0.0050 (about 0.3% of the interval length) at the smallest training size with strong dependence. This is sensitivity evidence only; it does not extend Theorem 1 to the empirical projection.

6.9. Computational Complexity

The estimator evaluates kernel weights between test points and training points; a dense implementation stores the full weight matrix. The compactly supported Epanechnikov kernel makes an exact sparse implementation possible in which only in-bandwidth pairs are stored. The sparse backend does not change the worst-case $O(n^2)$ time and memory complexity, since the number of in-bandwidth pairs can grow quadratically; its practical benefit depends on the bandwidth, the neighbourhood density, the projection dimension q, and the data geometry. Both backends were verified to produce identical results to within floating point tolerance on common inputs before benchmarking. Table 6 reports single pass wall times and peak memory.

At the simulation sizes the dense backend is faster (at $n = 1000$: 0.061 s dense versus 0.087 s sparse), so the dense backend is used throughout the simulation study. The observed benefit of the sparse backend appears at larger sizes: at $n = 10{,}000$ it is about 5.6 times faster (1.00 s versus 5.60 s) and uses about 20 times less peak memory (0.16 GB versus 3.32 GB), and it extends to sizes where the dense weight matrix would exceed the memory budget (up to $n = 40{,}000$ in 6.0 s here). These observed runtimes do not alter the worst-case complexity statement.

Table 6. Complexity benchmarks: one full estimator pass (kernel construction, regression and variance smoothing, 500 query evaluations) at $q = 3$; means and standard deviations over five repetitions, subprocess-isolated. Dense and sparse backends are compared at common sizes; sparse-only sizes are reported separately and are not compared against dense. The exact sparse backend does not change the worst-case $O(n^2)$ complexity; its practical benefit depends on the bandwidth, the neighbourhood density (mean in-bandwidth neighbours shown), the projection dimension, and the data geometry.

n	Backend	Wall Time (s)	Peak Memory (GB)	h	Mean Neighbours
1000	dense	0.061 ± 0.010	0.11 ± 0.00	0.816	46
1000	sparse	0.087 ± 0.011	0.06 ± 0.00	0.816	46
2000	dense	0.240 ± 0.035	0.23 ± 0.00	0.701	61
2000	sparse	0.131 ± 0.009	0.07 ± 0.00	0.701	61
5000	dense	1.300 ± 0.012	0.89 ± 0.00	0.613	102
5000	sparse	0.420 ± 0.023	0.10 ± 0.00	0.613	102
10,000	dense	5.602 ± 1.452	3.32 ± 0.08	0.536	140
10,000	sparse	0.996 ± 0.142	0.16 ± 0.01	0.536	140
20,000 (sparse-only size)	sparse	2.710 ± 0.224	0.35 ± 0.01	0.475	197
40,000 (sparse-only size)	sparse	6.045 ± 0.455	0.80 ± 0.05	0.422	276

6.10. Interpretation

The study, conducted under a protocol frozen before any outcome comparison, supports four statements, reported without embellishment. First, all methods attain near nominal coverage in every regime studied, including Experiments 3–4. The near-nominal coverage of both the gapped and contiguous procedures is consistent with the generic results of Remark 6; that theory does not imply equality of their finite-sample coverage estimates or interval lengths, and the observed null is not evidence against the gap's role inside the proof architecture of Theorem 1. Second, studentization showed no systematic coverage advantage, while interval-length differences were systematic; studentization is a length and adaptivity device under heteroskedasticity, and it removes most of the estimation induced score dependence visible in the diagnostics. Third, the buffer gap was empirically close to costless under Design A, and no detectable coverage benefit from the buffer gap was found at the tested sample sizes and under the frozen data-generating processes; at fixed budget its cost is a modest length increase, and we make no claim of practical superiority for the gapped procedure in these regimes. Fourth, the strong violation of (A2′) in Experiment 3 did not destabilize coverage in these designs; this is an empirical robustness observation about these DGPs, not a claim of validity beyond the assumption block.

6.11. Limitations of the Simulation Evidence

(i) The DGPs are controlled, with bounded support uniform margin FPC scores and conditionally Gaussian noise (Experiments 1–2), AR(1) residuals (Experiment 3), or the exactly specified hidden Markov magnitude construction (Experiment 4). Real data may exhibit non stationarity, time varying mixing, seasonality, or unbounded covariates, all outside the assumption block of Section 2.

(ii) The FPC dimension is fixed at $q = 3$ throughout; data driven q and growing q_n are outside the present theory and not assessed.

(iii) The empirical FPC projection $\widehat{P}_q$ is used in simulations, while the theory of Section 4 is stated for the population projection P_q; the paired sensitivity evidence of Section 6.8 quantifies but does not close this gap (Section 3.2).

(iv) Nominal coverage in these regimes does not establish exchangeability of the scores, and the diagnostics of Section 6.5 document nonzero score dependence in Experiments 3–4; no claim of empirical dominance of any method over any other is made.

(v) The real data evidence in the next section is illustrative only and is not a substantive chemometric finding.

7. Real Data Illustration

We illustrate the procedure on the Tecator near infrared (NIR) spectroscopy dataset [22], a functional covariate benchmark widely used in the SFPLM literature [2]. The illustration is operational. It shows that the procedure runs end to end on real functional data with auxiliary scalar covariates under an ordered split. It is not evidence of theorem validity, and no β-mixing claim is made for this dataset. The comparison below uses a genuinely contiguous baseline. Reusing the gapped method's training window would make the two arms nearly identical by construction, so the shared-window arrangement is included only as a sensitivity arm.

7.1. Dataset and Preprocessing

The Tecator near infrared spectroscopy dataset [22] is publicly available through CMU StatLib (with attribution to Tecator AB). It contains $n_{\text{total}} = 240$ meat samples analysed on a Tecator Infratec Food and Feed Analyzer in the near infrared transmission range 850–1050 nm; for each sample the release records a 100-channel absorbance spectrum together with reference water, fat, and protein content (the 22 derivative PCA auxiliary columns in the release are not used here). We set

$$\begin{aligned} Y_i &= \text{fat content } (\%), \\ X_i &= \big(\text{water content, protein content}\big) \in \mathbb{R}^2, \\ W_i &= \text{absorbance spectrum on a 100-point grid.} \end{aligned}$$

Fat content ranges from 0.9% to 58.5% with mean 18.5%.

Preprocessing is the standard SFPLM pipeline. Each spectrum W_i is centered around the training sample mean curve. The empirical FPC eigenfunctions $\widehat{\phi}_1, \ldots, \widehat{\phi}_q$ are computed by SVD of the centered training spectra, and the empirical projection $\widehat{P}_q$ is fixed at $q = 3$ before calibration, matching the simulation choice of Section 6.3. The scalar covariates X_i are centered and scaled to unit variance using the training subsample only. The Tecator absorbance spectra and scalar chemistry covariates are physically constrained to bounded ranges, in agreement with the compact projected support hypothesis of (A5*); the procedure of Section 3 is therefore applied as an operational illustration on a bounded support benchmark, and is not intended to demonstrate validity beyond the support regime considered in Section 4 and the Theoretical Supplement.

7.2. Real-Data Design

All arms evaluate on the common final test block, observations 201–240 in dataset order ($N_{\text{test}} = 40$), at $\alpha = 0.10$ with $q = 3$. The two primary arms are:

Primary M1 (gapped) : training 1–120, buffer gap 121–130, calibration 131–190, buffer gap 191–200, test 201–240 ($n_1 = 120$, $n_2 = 60$, $g = 10$).

Primary M2 (contiguous): training 21–140, calibration 141–200, test 201–240 ($n_1 = 120$, $n_2 = 60$, no gap): the calibration block immediately follows the training block and immediately precedes the test block.

Because the primary arms train on different windows of the ordered dataset, differences between them reflect the combined effect of the training window and the split geometry; they are not attributed solely to the buffer gap. Two labelled sensitivity arms complete the design: S1 applies a gap inside M2's window (training 21–140, gap

141–150, calibration 151–200, $n_2 = 50$), and S2 is the shared-window arrangement (training 1–120, calibration 121–180, observations 181–200 unused).

Each arm is run in two specifications: the full SFPLM (spectra as the functional covariate; water and protein content as scalar covariates) and a spectra-only specification with $p = 0$ scalar covariates, which reports the procedure's behaviour when the near-collinear scalar chemistry covariates are omitted. All estimators, the bandwidth rule, and the studentized score are as in Section 6.3, fitted on each arm's training block only. The splits are deterministic, so there is no Monte Carlo uncertainty for these numbers and none is reported.

7.3. Results

Table 7 reports all arms and specifications; Figure 4 displays the primary arms.

Table 7. Tecator analysis (single ordered benchmark; $\alpha = 0.10$, common test block 201–240; no Monte Carlo uncertainty exists for these deterministic splits and none is invented). Primary arms: M1 with train 1–120, buffer gaps 121–130 and 191–200, calibration 131–190; M2 genuinely contiguous with train 21–140 and calibration 141–200. S1 and S2 are labelled sensitivity arms (S1: M2's training window with a gap and $n_2 = 50$; S2: the shared-window split). Because the primary arms train on different windows, differences between them are not attributed solely to the buffer gap.

Arm	Specification	Coverage	Mean Length	Median Length	$\widehat{q}_{1-\alpha}$
primary_M1	full SFPLM	0.975 (39/40)	6.515	6.651	2.4228
primary_M1	spectra only	0.800 (32/40)	42.552	37.228	2.3453
primary_M2	full SFPLM	0.975 (39/40)	4.919	4.821	1.8725
primary_M2	spectra only	0.775 (31/40)	39.040	37.325	2.0193
S1	full SFPLM	1.000 (40/40)	5.660	5.548	2.1549
S1	spectra only	0.775 (31/40)	39.040	37.325	2.0193
S2	full SFPLM	0.975 (39/40)	6.515	6.651	2.4228
S2	spectra only	0.800 (32/40)	42.224	36.941	2.3272

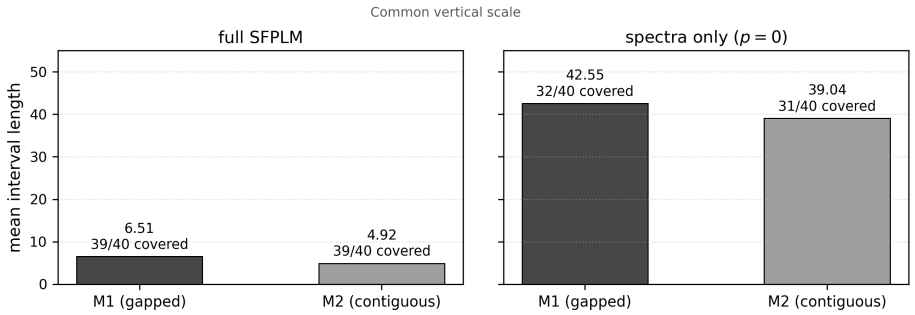


Figure 4. Tecator analysis, primary arms: mean interval length and test coverage (covered out of 40) for the gapped (M1) and contiguous (M2) primary arms, in the full SFPLM specification (**left**) and the spectra-only specification (**right**). Lengths in % fat units.

In the full SFPLM specification both primary arms cover 39 of 40 test observations (0.975); the contiguous arm produces shorter intervals on this dataset (mean length 4.92 versus 6.51 % fat units). Given the compositional caveat above, we read this as the combined effect of the more recent training window and the absence of a buffer, not as a cost of the gap alone; S1, which shares M2's training window but inserts a gap ($n_2 = 50$), covers 40 of 40 with mean length 5.66, between the two primary arms. S2 uses the shared-window arrangement and gives 0.975 coverage with mean length 6.51, identical to primary M1; this arrangement is uninformative about the gap. In the spectra-only specification, coverage drops to 31–32 of 40 and mean lengths grow by a factor of 6–8. In this ordered split, including water and protein yielded substantially shorter intervals

and higher empirical coverage than the spectra-only specification. Because water, protein and fat are compositionally related, this comparison does not identify an independent incremental contribution of the scalar predictors.

7.4. Interpretation and Limitations

The illustration supports three statements. The procedure runs end to end on real spectroscopy data and produces sensible intervals in both specifications. The gap versus no-gap comparison on this small ordered dataset is confounded by the training window and is reported with that caveat; no practical superiority claim in either direction is made. In this ordered split, including the scalar chemistry covariates yielded substantially shorter intervals and higher empirical coverage than the spectra-only specification; the compositional caveat of Section 7.3 applies, and no independent incremental contribution of the scalar predictors is identified.

The limitations of Section 6.11 apply with added force: the dataset is sample collection ordered but not a β-mixing time series; $N_{\text{test}} = 40$ is small and the splits are deterministic, so coverage values are coarse (1/40 resolution); the empirical FPC projection is used while the theory concerns the population projection (Section 3.2); the water and protein covariates are nearly collinear with fat content in mass percent space, which shrinks the residual scale of the full SFPLM specification; and the illustration is not a substantive chemometric contribution.

8. Discussion

8.1. Summary of the Contribution

The paper develops a finite sample conformal prediction interval for the semi-functional partial linear model (3) under geometric β-mixing: a split conformal procedure (Section 3) with an ordered three way partition train | gap | calibration | gap | test, an FPC projection semi-metric d_q, and a studentized residual score. The guarantee imposes no parametric error model and no normal approximation, but it is not assumption free; it holds under the structural conditions of Section 2.6. Theorem 1 states the marginal lower coverage bound $\Pr\{Y_{n+1} \in \widehat{C}_n(X_{n+1}, W_{n+1})\} \geq 1 - \alpha - \Delta_n$, where the theoretical finite-sample coverage penalty Δ_n decomposes into the seven interpretable summands of (13), each naming the model primitive it depends on, with (A8′) tying all seven to a single rate calculus; the penalty is a proof-side quantity that plays no role in computing the interval (Section 4.3). The deviation lemmas proved in the Theoretical Supplement, especially Lemmas 1–5, are the SFPLM specific technical content, assembled into the coverage bound by an oracle-transfer argument anchored at the ideal score.

8.2. Why All Methods Cover Empirically

The study of Section 6, pre-specified and frozen before outcome computation, shows that the gapped and contiguous procedures, studentized and plain, all attain near nominal empirical coverage across the regimes considered. We do not interpret this as evidence against the theoretical contribution. The near-nominal coverage of both procedures is consistent with generic dependent split-conformal theory (Remark 6), which provides marginal validity for a trained score under β-mixing without any buffer and without estimation accuracy requirements; that theory does not imply equality of their finite-sample coverage estimates or interval lengths. Two structural observations add texture. First, the studentized score is a ratio in which the fitted conditional mean and scale absorb much of the covariate induced variation, and the diagnostics of Section 6.5 show that studentization removes most of the estimation induced score dependence visible in the plain scores; studentization showed no systematic coverage advantage, while interval-length differences

were systematic. Second, conformal coverage is a marginal probability statement that aggregates over the test covariates and the calibration draw, so local over and under coverage can cancel in the aggregate; nominal marginal coverage therefore neither requires nor certifies exchangeability of the score sequence, which is why the protocol specified the score dependence diagnostics in advance.

8.3. Relation to Generic Dependent Conformal Theory, Tightness, and What Is Model-Specific

Marginal validity of split conformal prediction with a score trained on the training block and then applied to calibration and test observations under dependence is provided by generic theory under the cited papers' exact conditions: Oliveira, Orenstein, Ramos and Romano [16] (Theorems 1 and 4 in the published JMLR numbering) prove it through decoupling, concentration and a β-mixing instantiation, and Barber and Pananjady [17] (arXiv:2510.02471v2; Theorem 4 with Corollary 2 for the split-trained setting, memory $L = 0$) give switch-coefficient penalties; in the pretrained memoryless setting, Theorem 2 of [17] proves the bound sharp up to the stated constant. In both papers the analysis quantities (block parameters; deletion parameters τ, τ^*) are optimized inside the bounds: no operational buffer or block is required of the algorithm, and no requirement is placed on the accuracy of the fitted regression or scale estimators. The contiguous baseline is not covered by the manuscript's SFPLM-specific gapped theorem as currently formulated, but it is covered by generic dependent split-conformal results under their respective conditions. An earlier draft of this manuscript phrased this as an absence of guarantees for M2; that phrasing was incorrect and has been removed.

Against that backdrop, this paper's contribution is not marginal validity and not a sharper dependence penalty. The dependence-only components of Δ_n (coupling, calibration empirical process, buffer gap) overlap conceptually with generic dependence-only bounds and are not the source of the manuscript's model-specific novelty; generic bounds may be sharper under their own conditions. The additive decomposition, built from union bounds and couplings, is not claimed to be sharp. The same assessment applies to the generic non exchangeability framework of [15], whose total variation coupling bound covers stationary β-mixing data among many other settings; we do not claim that Δ_n is uniformly tighter than a mechanical BCRT-plus-coupling bound. What is genuinely SFPLM-specific is the oracle-transfer layer: an explicit finite sample account of the price of replacing the ideal score $|Y - X^\top \beta_0 - m(W)|/\sigma(W)$ by the implemented estimated score, expressed in the model's own primitives (bandwidth, small-ball function, FPC dimension, variance floor, and smoothness indices), together with the explicit good-event bookkeeping this requires.

The simulation evidence is consistent with this division of labour and is reported without embellishment: across all four designs, including the new dependent-score experiment whose ideal score has exactly quantified serial dependence (lag-one autocorrelation $75,456/444,365 \approx 0.17$) and which satisfies the verifiable assumption block including the score-density condition, the gapped and contiguous studentized methods were statistically indistinguishable in coverage, and studentization showed no systematic coverage advantage (its effects were on interval length). The buffer gap's cost and benefit were both small at the sample sizes studied: under a fixed total budget, doubling the theoretical gap discarded 28% of the observations while coverage was unchanged and the length grew by at most about 3%. We therefore make no claim of practical superiority for the gapped procedure in these regimes; its role is to support the explicit theoretical decomposition, whose value is interpretability of the model-specific estimation layer, transparent pricing of the proof architecture, and the diagnostics it motivates. Tightening the model-specific bound, and any practically usable quantile adjustment under dependence, are future work.

For completeness within the wider literature, the exchangeability based statements of [6,23–25], the time series proposals of [7,9–12], and the functional conformal work of [13,14] do not address the dependent SFPLM setting with an FPC projection semi-metric; the contribution here is that joint specialization, with an estimation-aware decomposition.

8.4. Limitations

The theory and methodology have five explicit limitations. First, the sample dimension p of the parametric covariate X and the FPC dimension q are fixed and not data adaptive; the deviation lemmas of the Theoretical Supplement use p and q as fixed constants in their constants and do not track high dimensional or growing-q regimes. Second, the procedure is calibrated under geometric β-mixing (A1); polynomial β-mixing rates lead to a strictly weaker Δ_n and would require revisiting the block length condition (C3) and the bandwidth block length condition (A8′). Third, the noise process is assumed to be conditionally sub-Gaussian as in (A2′); the autoregressive residual diagnostic of Section 6 violates this conditional noise structure and is reported as a boundary stress test, not as evidence for the theorem. Fourth, Theorem 1 concerns the population projection P_q; see Section 3.2 for the empirical-FPCA limitation and the paired sensitivity evidence. Fifth, the support assumption (A5*) requires the projected covariate P_qW to have compact support on $\mathbb{R}^q$ and the density bounded fractal regime (A6*) requires f_Z to be bounded above and below on $\mathcal{S}_F$; an extension to unbounded projected support, or to densities that decay or blow up at the boundary, would require a triangular array uniform deviation theory in which the support set is allowed to grow with n, and is outside the scope of the present paper.

8.5. Future Work

Four directions are natural extensions of the present results. First, a polynomial-β-mixing analogue of Theorem 1 would replace the geometric Yu-coupling cost by its polynomial counterpart and would require a corresponding change to (C3); the form of Δ_n would be preserved but with slower rates on the coupling and calibration empirical process terms. Second, relaxing the conditional sub-Gaussian assumption (A2′) to a sub-Weibull or polynomial tail assumption would broaden applicability to heavier tailed financial and chemometric noise, at the cost of a larger studentization perturbation summand in Δ_n. Third, allowing the FPC dimension q to grow with n would require a joint rate calculus that ties the bandwidth h_n to q_n and to the spectral decay of Σ_W; the bandwidth block length condition (A8′) is the right entry point for this extension. Fourth, a practically usable quantile adjustment under dependence, and adaptive selection of h_n, q, and g_n from data, are operationally desirable but are not supported by the present theory and should be addressed in separate work.

9. Conclusions

This paper addressed the construction of finite sample prediction intervals for the semi functional partial linear model under geometric β-mixing dependence, without imposing a parametric error model or a normal approximation; the coverage guarantee relies on the explicit structural assumptions of Section 2. The methodology is a split conformal procedure with an ordered three way partition with buffer gaps, an FPC projection semi metric, profiled and kernel estimators, and a studentized residual score, implementable from a single ordered sample using the operational rules of Section 3.

The main theoretical contribution is Theorem 1, a finite sample marginal coverage bound, valid for $n \geq N_0$, whose theoretical finite-sample coverage penalty Δ_n decomposes into seven interpretable components expressed in the model's primitives. The penalty

quantifies the oracle-to-estimated score transfer inside the proof; it involves unknown structural constants, plays no role in computing the interval, and is not an operational correction. Generic dependent split-conformal theory establishes marginal validity for a trained score without any buffer (Remark 6); the model-specific content of the present analysis is the estimation-aware layer of the decomposition and the supporting uniform deviation theory for SFPLM kernel estimators under β-mixing.

The simulation study, conducted under a protocol frozen before any outcome comparison (500 replications per cell, all methods, four designs including a dependent-score construction with exactly quantified score autocorrelation) found near nominal coverage for every method in every cell; the gapped and contiguous variants were statistically indistinguishable in coverage, studentization showed no systematic coverage advantage while interval-length differences were systematic, and doubling the buffer gap under a fixed budget discarded 28% of the data with no coverage change and a length increase of at most about 3%. The Tecator analysis shows that the procedure runs end to end on real spectroscopy data; its gap versus no-gap comparison is confounded by the training window and is reported with that caveat.

Several technical extensions remain open, including a theoretical accounting of the empirical FPCA estimation error, a joint rate calculus for growing FPC dimension q_n, a high dimensional treatment of the parametric covariate X, adaptive selection of h_n, q, and g_n, sub Weibull or polynomial tail noise relaxations, a practically usable quantile adjustment under dependence, and rolling or online conformal prediction for functional time series.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/math14173201/s1.

Funding: This scientific paper is derived from a research grant funded by Taibah University, Madinah, Kingdom of Saudi Arabia, with grant number (1225-15-448).

Data Availability Statement: The Tecator near infrared spectroscopy dataset used in the real data illustration is publicly available through the CMU StatLib archive.

Acknowledgments: The author used ChatGPT (OpenAI; GPT-5.5) only for language polishing. The author gratefully acknowledges the support of the Energy, Industry, and Advanced Technologies Research Center, Taibah University, Madinah, Saudi Arabia, for funding this research project.

Conflicts of Interest: The author declares no conflicts of interest.

References

1. Aneiros-Pérez, G.; Vieu, P. Semi-functional partial linear regression. *Stat. Probab. Lett.* **2006**, *76*, 1102–1110. [CrossRef]
2. Ferraty, F.; Vieu, P. *Nonparametric Functional Data Analysis: Theory and Practice*; Springer Series in Statistics; Springer: New York, NY, USA, 2006.
3. Aneiros-Pérez, G.; Vieu, P. Nonparametric time series prediction: A semi-functional partial linear modeling. *J. Multivar. Anal.* **2008**, *99*, 834–857. [CrossRef]
4. Ling, N.; Vieu, P. Nonparametric modelling for functional data: Selected survey and tracks for future. *Statistics* **2018**, *52*, 934–949. [CrossRef]
5. Aneiros, G.; Raña, P.; Vieu, P.; Vilar, J. Bootstrap in semi-functional partial linear regression under dependence. *Test* **2018**, *27*, 659–679. [CrossRef]
6. Vovk, V.; Gammerman, A.; Shafer, G. *Algorithmic Learning in a Random World*; Springer: New York, NY, USA, 2005.
7. Chernozhukov, V.; Wüthrich, K.; Zhu, Y. Exact and robust conformal inference methods for predictive machine learning with dependent data. In Proceedings of the 31st Conference on Learning Theory (COLT 2018), Stockholm, Sweden, 6–9 July 2018; PMLR 75; pp. 732–749.
8. Chernozhukov, V.; Wüthrich, K.; Zhu, Y. An exact and robust conformal inference method for counterfactual and synthetic controls. *J. Am. Stat. Assoc.* **2021**, *116*, 1849–1864. [CrossRef]
9. Xu, C.; Xie, Y. Conformal prediction interval for dynamic time-series. In Proceedings of the 38th International Conference on Machine Learning (ICML 2021), Virtual, 18–24 July 2021; PMLR 139; pp. 11559–11569.

10. Xu, C.; Xie, Y. Sequential predictive conformal inference for time series. In Proceedings of the 40th International Conference on Machine Learning (ICML 2023), Honolulu, HI, USA, 23–29 July 2023; PMLR 202; pp. 38707–38727.
11. Stankevičiūtė, K.; Alaa, A.M.; van der Schaar, M. Conformal time-series forecasting. In Proceedings of the Advances in Neural Information Processing Systems 34 (NeurIPS 2021), Virtual, 6–14 December 2021; pp. 6216–6228.
12. Zaffran, M.; Féron, O.; Goude, Y.; Josse, J.; Dieuleveut, A. Adaptive conformal predictions for time series. In Proceedings of the 39th International Conference on Machine Learning (ICML 2022), Baltimore, MD, USA, 17–23 July 2022; PMLR 162; pp. 25834–25866.
13. Diquigiovanni, J.; Fontana, M.; Vantini, S. Conformal prediction bands for multivariate functional data. *J. Multivar. Anal.* **2022**, *189*, 104879. [CrossRef]
14. Ajroldi, N.; Diquigiovanni, J.; Fontana, M.; Vantini, S. Conformal prediction bands for two-dimensional functional time series. *Comput. Stat. Data Anal.* **2023**, *187*, 107821. [CrossRef]
15. Barber, R.F.; Candès, E.J.; Ramdas, A.; Tibshirani, R.J. Conformal prediction beyond exchangeability. *Ann. Stat.* **2023**, *51*, 816–845. [CrossRef]
16. Oliveira, R.I.; Orenstein, P.; Ramos, T.; Romano, J.V. Split conformal prediction and non-exchangeable data. *J. Mach. Learn. Res.* **2024**, *25*, 1–38.
17. Barber, R.F.; Pananjady, A. Predictive inference for time series: Why is split conformal effective despite temporal dependence? *arXiv* **2025**, arXiv:2510.02471v2.
18. Politis, D.N.; White, H. Automatic block-length selection for the dependent bootstrap. *Econom. Rev.* **2004**, *23*, 53–70. Correction in *Econom. Rev.* **2009**, *28*, 372–375.
19. Yu, B. Rates of convergence for empirical processes of stationary mixing sequences. *Ann. Probab.* **1994**, *22*, 94–116. [CrossRef]
20. Merlevède, F.; Peligrad, M.; Rio, E. Bernstein inequality and moderate deviations under strong mixing conditions. In *High Dimensional Probability V: The Luminy Volume*; IMS Collections; Institute of Mathematical Statistics: Beachwood, OH, USA, 2009; Volume 5, pp. 273–292.
21. Rio, E. *Asymptotic Theory of Weakly Dependent Random Processes*; Probability Theory and Stochastic Modelling; Springer: Berlin, Germany, 2017; Volume 80.
22. Borggaard, C.; Thodberg, H.H. Optimal minimal neural interpretation of spectra. *Anal. Chem.* **1992**, *64*, 545–551. [CrossRef]
23. Lei, J.; G'Sell, M.; Rinaldo, A.; Tibshirani, R.J.; Wasserman, L. Distribution-free predictive inference for regression. *J. Am. Stat. Assoc.* **2018**, *113*, 1094–1111. [CrossRef]
24. Romano, Y.; Patterson, E.; Candès, E.J. Conformalized quantile regression. In Proceedings of the Advances in Neural Information Processing Systems 32 (NeurIPS 2019), Vancouver, BC, Canada, 8–14 December 2019; pp. 3538–3548.
25. Barber, R.F.; Candès, E.J.; Ramdas, A.; Tibshirani, R.J. Predictive inference with the jackknife+. *Ann. Stat.* **2021**, *49*, 486–507. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.