Statistica Sinica Preprint No: SS-2025-0307	
Title	Wild Bootstrap for Mean Response Inference in Functional Linear Regression Models
Manuscript ID	SS-2025-0307
URL	http://www.stat.sinica.edu.tw/statistica/
DOI	10.5705/ss.202025.0307
Complete List of Authors	Hyemin Yeon, Xiongtao Dai and Daniel Nordman
Corresponding Authors	Hyemin Yeon
E-mails	hyeon2@gmu.edu
Notice: Accepted author version.	

Statistica Sinica

Wild Bootstrap for Mean Response Inference in Functional Linear Regression Models

Hyemin Yeon[1], Xiongtao Dai[2], and Daniel Nordman[3]

[1] *George Mason University,* [2] *University of California, Berkeley,* [3] *Iowa State University*

Abstract: Functional regressors complicate inference in linear regression problems so that the bootstrap can play a useful role in quantifying uncertainty and calibrating intervals. The best bootstrap in practice, though, can depend on factors in the data as well as computational considerations and existing bootstraps can have limitations: residual bootstrap is computationally fast and simple but may fail when the errors are heterogeneous, while paired bootstrap applies more generally in functional linear regression at a cost of much higher computation. To bridge this gap, we develop a wild bootstrap method for functional linear regression, which is akin to a modified version of residual bootstrap but designed to have a wide scope of application like paired bootstrap, including to heteroscedastic errors. Its theoretical consistency is established and numerical studies suggest that wild bootstrap can provide accurate and computationally fast inference. Importantly, we also suggest a practical and effective approach of selecting truncation levels, specifically designed for mean response inference problems. The proposed bootstrap in functional linear regression is further illustrated through a weather data example, and an accompanying R package `BTSinFLRM` provides numerical implementations.

Key words and phrases: functional principal components analysis, heteroscedasticity, infinite dimensionality, resampling, scalar-on-function regression, stabilized volatility method

1. Introduction

1.1 Motivation

Bootstrap methods have been widely used and studied for statistical inference in classical linear regression models (LRMs) given as

$$Y = \alpha + \beta^\top X + \varepsilon. \tag{1.1}$$

where Y and X respectively represent a scalar response and a finite-dimensional vector regressor involving an intercept α and the slope vector β, while ε denotes an error. There are three main bootstrap methods for use in LRMs (i.e., residual, paired, and wild bootstraps), which all aim to re-create bootstrap versions $\{(Y_i^*, X_i^*)\}_{i=1}^n$ of observed data $\{(Y_i, X_i)\}_{i=1}^n$ for approximating sampling distributions. *Residual bootstrap* (RB) uses the same regressors $X_i^* = X_i$ as the original data and generates a response as $Y_i^* = \widehat{\mu}(X_i) + \varepsilon_i^*$ by using an estimator $\widehat{\mu}(X_i)$ in place of $\mu(X_i) \equiv \alpha + \beta^\top X_i$ and drawing a bootstrap error ε_i^* from centered residuals $\hat{\varepsilon}_i \equiv Y_i - \widehat{\mu}(X_i)$, while *paired bootstrap (PB)* creates each bootstrap observation (Y_i^*, X_i^*) by drawing independently from the original data pairs $\{(Y_j, X_j)\}_{j=1}^n$ (cf. Efron, 1979, Freedman, 1981). RB is computationally simple but strictly intended for for homoscedastic regression models, while PB is valid even under heteroscedastic models due to its nonparametric resampling, though the latter often can entail higher computational burdens, bias in bootstrap statistics, and potentially wider intervals compared to RB. *Wild bootstrap (WB)* was proposed to circumvent the respective limitations of RB and PB (cf. Liu, 1988, Wu, 1986). Like RB, WB uses the original data regressors $X_i^* = X_i$ and generates a response as $Y_i^* = \widehat{\mu}(X_i) + \varepsilon_i^*$; the key difference from RB is that the bootstrap error ε_i^* in WB is defined in a way to depend on X_i so that WB can handle heteroscedasticity in regression. Unlike PB, though, WB has also easy implementation; for example, WB does not require computing an inverse covariance matrix of regressors in each bootstrap resample $\{(Y_i^*, X_i^*)\}_{i=1}^n$ as in PB.

Recent interest has focused on extending such bootstrap methods to inference about a functional linear regression model (FLRM) given by

$$Y = \alpha + \langle \beta, X \rangle + \varepsilon, \tag{1.2}$$

where Y still denotes a scalar response (involving a model error) but X is instead a functional regressor. That is, X is a random element taking values in an infinite dimensional Hilbert space $\mathbb{H}$ with inner product $\langle \cdot, \cdot \rangle$, while the parameter $\beta \in \mathbb{H}$ is now called the slope function for determining the mean response $\mu(X) \equiv \alpha + \langle \beta, X \rangle$. Unlike with a classical LRM, the regressors X in a FLRM can assume an infinite dimension of values, which creates challenges in estimation, related to bias and scaling (cf. Cardot et al., 2007, Hall and Horowitz, 2007), and also complicates bootstrap inference. Due to the latter, fewer formal bootstrap developments exist for FLRMs compared to standard LRMs, though some progress has been made. In FLRMs under homoscedasticity, RB has been considered for interval estimation of the conditional mean, or similarly the projection $\langle \beta, X_0 \rangle$, of a new regressor X_0 by González-Manteiga and Martínez-Calvo (2011) and Yeon et al. (2023). Taking a different perspective from mean response inference, Khademnoe and Hosseini-Nasab (2024) proposed an RB approach for testing the significance of the slope β under homoscedastiticy. Under heteroscedastic errors, Yeon et al. (2024a) proposed a PB in FLRMs with bias correction, albeit with high computational costs. However, in contrast to LRMs, formal inference with WB for even mean inference has not established for FLRMs under possible heteroscedastic errors, which represents a methodological and theoretical gap for bootstrap inference under FLRMs.

1.2 Our contributions

This work aims to formally treat a wild bootstrap (WB) for inference for the mean response $\mu(X_0) \equiv \alpha + \langle \beta, X_0 \rangle$ at a new regressor X_0 in FLRMs. Similar to LRMs, the motivation for WB is that this approach has the potential to integrate the benefits of previous bootstrap methods for FLRMs. That is, WB can be valid for FLRMs under general assumptions (including heteroscedastic errors) much like PB, but WB also has potential to be computationally fast, like RB. Particularly with functional regressors, WB requires less computational effort and is more computationally scalable than PB.

To have a large scope of application, we prove the consistency of WB for mean response inference in FLRMs under mild assumptions. Numerical studies between our proposed WB and previous bootstraps (e.g., RB and PB) indicate that WB performs comparatively well. WB similarly exhibits good coverage accuracy under homoscedastic errors like RB but, while RB completely fails to maintain coverage under heteroskedastic errors, WB also has a coverage performance closer to PB under heteroskedasticity at a lower computational cost. Further, WB can show much better coverage than PB in cases where errors are both heterogeneous and highly skewed. While all existing bootstrap methods for FLRMs require some tuning parameter selection (i.e., truncation levels), WB additionally allows an intuitive and practical method for selecting these in practice. The main idea is to select such truncation levels by examining where WB intervals stabilize in both center and length, which is not a feasible strategy for other bootstraps, like PB (Yeon et al., 2024a), due to greater computational burdens and sensitives as truncation levels vary. This approach also allows tuning parameters to change with the regressor X_0 at which an interval estimator of a target mean $\mu(X_0)$ is desired, which is not applicable in other classical choices for tuning parameters, such as cross-validation (Hastie et al., 2009) or fraction of variance explained (Kokoszka and Reimherr, 2017).

To highlight some distinctions from previous bootstrap works with FLRMs, the development of WB here under possible heteroscedasticity involves a different theoretical treatment than with PB (Theorem 1) due to the unique resampling scheme of WB. For example, while the bootstrap observations from PB are exchangeable, such exchangeability does not hold for the WB data (e.g., $(Y_1^*, X_1) \overset{\mathsf{d}}{\neq} (Y_2^*, X_2)$) and the WB responses Y_i^* are defined in a distinct manner per functional regressor X_i. Consequently, theoretical validity must be approached differently (cf. Section S1.3 of the supplement for details). Additionally, most previous contributions on resampling for FLRMs have again focused mainly on RB and homoscedastic errors, such as González-Manteiga and Martínez-Calvo (2011) and Yeon et al. (2023) for mean response estimation and Khademnoe and Hosseini-Nasab (2024) and Lin and Lin (2025) on slope testing (i.e., $\beta = 0$). We likewise target mean responses which enables mean comparisons in a classically useful fashion under potential heteroscedasticity, as illustrated in Section 5.2, though the WB could potentially be applied for slope tests under heteroscedastic errors in future research. Also, while González-Manteiga and Martínez-Calvo (2011) considered a version of WB in addition to RB, our WB approach is quite different in that (i) our general framework accommodates heteroscedasticity and the associated scaling/studentization required in FLRMs and (ii) we consider WB for inference about mean responses directly rather than incomplete or biased versions of means. In contrast to previous bootstraps, we also propose an intuitive data-driven method for selecting tuning parameters.

1.3 Outline of the paper

Section 2 reviews regression in FLRMs, including statistical quantities involving mean parameters and projections. Section 3 then describes the WB method for inference and formally establishes its validity. Empirical coverage performances are investigated through numerical studies in Section 4, wherein previous bootstrap methods, RB and PB, are also compared. Through a data example in Section 5, we illustrate the implementation of WB, including practical selection of tuning parameters involved in the WB procedure. Section 6 offers concluding remarks. The WB method for FLRMs is implemented in the R package `BTSinFLRM` available online,[†] while proofs and further technical details are deferred to the supplement.

2. Estimation and inference in FLRMs

Section 2.1 describes basic estimation in FLRMs, including the functional principal component regression (FPCR) estimator of the slope, while Section 2.2 briefly reviews scaled statistical quantities for mean responses under possible heteroscedasticity.

2.1 Functional principal component regression estimator

To describe a theoretical framework for estimation in FLRMs, we assume that the slope function β in (1.2) lies in an infinite-dimensional separable Hilbert space $\mathbb{H}$ equipped

[†] https://github.com/luckyhm1928

with inner product $\langle\cdot,\cdot\rangle$ and the induced norm $\|\cdot\|$ defined as $\|x\| \equiv \sqrt{\langle x, x\rangle}$ for $x \in \mathbb{H}$. The functional regressor X is a random element in $\mathbb{H}$ with a finite second moment $\mathsf{E}[\|X\|^2] < \infty$, while the response Y and error ε are both scalar-valued random variables with $\mathsf{E}[\varepsilon|X] = 0$. Here we do not assume zero means for both X and Y, as occurs in some theoretical developments (cf. Cardot et al., 2007, Hall and Horowitz, 2007, Yeon et al., 2023); see Cai and Hall (2006), Yeon et al. (2024a) for related discussions.

Similar to how a least squares estimator in a classical LRM (1.1) is defined by normal equations, a functional version of normal equations is also important to defining estimation in a FLRM (1.2). Write $\Gamma \equiv \mathsf{E}[(X - \mathsf{E}[X]) \otimes (X - \mathsf{E}[X])]$ for the covariance operator of X, where $x \otimes y : \mathbb{H} \to \mathbb{H}$ denotes the tensor product between two elements $x, y \in \mathbb{H}$ defined as $(x \otimes y)(z) = \langle z, x\rangle y \in \mathbb{H}$ for $z \in \mathbb{H}$. Denoting $\Delta \equiv \mathsf{E}[(Y - \mathsf{E}[Y])(X - \mathsf{E}[X])]$ as the cross-covariance function between X and Y, we obtain normal equations from (1.2) as $\Delta = \Gamma\beta$. The slope function is then identifiable, and hence given as $\beta = \Gamma^{-1}\Delta$, under a condition $\{x \in \mathbb{H} : \Gamma x = 0\} = \{0\}$ (cf. Cardot et al., 2003, 2007).

With a finite second moment assumption $\mathsf{E}[\|X\|^2] < \infty$, because the covariance operator $\Gamma \equiv \mathsf{E}[(X - \mathsf{E}[X]) \otimes (X - \mathsf{E}[X])]$ is self-adjoint, non-negative definite, and compact (Hsing and Eubank, 2015, Chapter 4), this admits a spectral decomposition as $\Gamma = \sum_{j=1}^{\infty} \gamma_j(\phi_j \otimes \phi_j)$, where γ_j and ϕ_j respectively denote the j-th eigenvalue and eigenfunction of Γ for an integer $j \geq 1$ (Hsing and Eubank, 2015, Theorem 7.2.6). The set $\{\phi_j\}_{j=1}^{\infty}$ of eigenfunctions forms an orthonormal system of $\mathbb{H}$, and the eigenvalues

$\{\gamma_j\}_{j=1}^{\infty}$ are a non-negative, non-increasing sequence such that $\gamma_j \to 0$ as $j \to \infty$. Writing $\Gamma^{-1} \equiv \sum_{j=1}^{\infty} \gamma_j^{-1}(\phi_j \otimes \phi_j)$, the slope function in (1.2) can be expressed as

$$\beta = \Gamma^{-1}\Delta = \sum_{j=1}^{\infty} \gamma_j^{-1}\langle \Delta, \phi_j\rangle \phi_j. \tag{2.3}$$

Based on a random sample $\{(Y_i, X_i)\}_{i=1}^n$ with $Y_i = \alpha + \langle \beta, X_i\rangle + \varepsilon_i$ from the model (1.2), the so-called FPCR estimator of β can then be stated as a truncated sample version of the slope function (2.3). To explain, define the sample counterparts of Γ and Δ as $\hat{\Gamma}_n \equiv n^{-1}\sum_{i=1}^n (X_i - \bar{X})^{\otimes 2}$ and $\hat{\Delta}_n \equiv n^{-1}\sum_{i=1}^n (Y_i - \bar{Y})(X_i - \bar{X})$, where $\bar{X} \equiv n^{-1}\sum_{i=1}^n X_i$ and $\bar{Y} \equiv n^{-1}\sum_{i=1}^n Y_i$ denote the sample means of $\{X_i\}_{i=1}^n$ and $\{Y_i\}_{i=1}^n$, and $x^{\otimes 2} \equiv x \otimes x$ for $x \in \mathbb{H}$. The sample covariance operator $\hat{\Gamma}_n$ also admits spectral decomposition as $\hat{\Gamma}_n \equiv \sum_{j=1}^n \hat{\gamma}_j \hat{\phi}_j^{\otimes 2}$, where $\hat{\gamma}_j$ and $\hat{\phi}_j$ are the j-th sample eigenvalue and sample eigenfunction. However, in contrast to inversion of Γ as in (2.3), inversion of $\hat{\Gamma}_n$ is ill-posed because $\hat{\Gamma}_n$ has finite rank in the sense that the image of $\hat{\Gamma}_n$ has finite dimension. This technical complication can be overcome by regularizing the inversion of $\hat{\Gamma}_n$ and obtaining the FPCR estimator $\hat{\beta}_{h_n}$ of β as

$$\hat{\beta}_{h_n} \equiv \hat{\Gamma}_{h_n}^{-1}\hat{\Delta}_n = \sum_{j=1}^{h_n} \hat{\gamma}_j^{-1}\langle \hat{\Delta}_n, \hat{\phi}_j\rangle \hat{\phi}_j, \tag{2.4}$$

where $\hat{\Gamma}_{h_n}^{-1} \equiv \sum_{j=1}^{h_n} \hat{\gamma}_j^{-1}\hat{\phi}_j^{\otimes 2}$ is a finite approximation of $\Gamma^{-1} \equiv \sum_{j=1}^{\infty} \gamma_j^{-1}\phi_j^{\otimes 2}$, and the truncation level h_n represents the number of eigenpairs used in estimation (Cai and Hall, 2006, Cardot et al., 1999, 2007, Hall and Horowitz, 2007). The intercept parameter α is then estimated by $\hat{\alpha}_{h_n} \equiv \bar{Y} - \langle \hat{\beta}_{h_n}, \bar{X}\rangle$.

2.2 Target statistical quantities for mean response

While the FPCR estimator $\hat{\beta}_{h_n}$ is consistent (cf. Yeon et al., 2024b, Theorem S3), seminal work by Cardot et al. (2007) has shown that it is impossible for $a_n(\hat{\beta}_{h_n} - \beta)$ to converge in distribution to a non-degenerate random element taking values in $\mathbb{H}$ for any diverging scaling sequence $\{a_n\}$. This feature complicates inference about the slope function β directly. Importantly, however, meaningful inference about the conditional mean $\mu(X) \equiv \alpha + \langle \beta, X \rangle$ of a response in the FLRM (1.2) is possible by combining an estimator with an appropriate scaling term to define a statistical quantity with a well-defined distribution in large samples. We next briefly review such target quantities for inference (Cardot et al., 2007, Yeon et al., 2023, 2024a). Let X_0 denote a new regressor that is independent of the sample $\{(Y_i, X_i)\}_{i=1}^n$ and identically distributed as X. Then, a main problem of interest is the inference about the true mean

$$\mu(X_0) \equiv \alpha + \langle \beta, X_0 \rangle \tag{2.5}$$

of a response given X_0, which is estimated by its sample counterpart

$$\hat{\mu}_{h_n}(X_0) \equiv \hat{\alpha}_{h_n} + \langle \hat{\beta}_{h_n}, X_0 \rangle. \tag{2.6}$$

We can also consider related inference about the projection $\langle \beta, X_0 - \mathsf{E}[X] \rangle$ of the slope function β onto a new predictor $X_0 - \mathsf{E}[X]$ centered by the regressor mean $\mathsf{E}[X]$. This projection is useful for assessing the effect $\langle \beta, X_0 \rangle$ of the slope at a new predictor X_0 relative to an overall slope effect $\langle \beta, \mathsf{E}[X] \rangle$, in which estimation of the model intercept in (2.5) (or the unconditional mean response $\mathsf{E}[Y]$ in α) also becomes unnecessary. We then consider statistical quantities based on differences between estimated and true mean responses or projections given by

$$T_{\mathrm{mean},n}(X_0) \equiv \sqrt{\frac{n}{\hat{s}_{h_n}(X_0)}}\{\hat{\mu}_{h_n}(X_0) - \mu(X_0)\}, \tag{2.7}$$
$$T_{\mathrm{proj},n}(X_0) \equiv \sqrt{\frac{n}{\hat{s}_{h_n}(X_0)}}(\langle\hat{\beta}_{h_n}, X_0 - \bar{X}\rangle - \langle\beta, X_0 - \mathsf{E}[X]\rangle)$$

where $\mu(X_0)$ and $\hat{\mu}_{h_n}(X_0)$ are from (2.5)-(2.6). Here $\sqrt{n/\hat{s}_{h_n}(X_0)}$ denotes an appropriate scaling factor with general and possibly heteroscedastic FLRMs, as defined by

$$\hat{s}_{h_n}(x) \equiv \langle\hat{\Lambda}_n\hat{\Gamma}_{h_n}^{-1}(x - \bar{X}), \hat{\Gamma}_{h_n}^{-1}(x - \bar{X})\rangle = \|\hat{\Lambda}^{1/2}\hat{\Gamma}_{h_n}^{-1}(x - \bar{X})\|^2, \quad x \in \mathbb{H}, \tag{2.8}$$

in terms of a sample covariance operator

$$\hat{\Lambda}_n \equiv n^{-1}\sum_{i=1}^{n}\left(X_i\hat{\varepsilon}_{i,k_n} - n^{-1}\sum_{i=1}^{n}X_i\hat{\varepsilon}_{i,k_n}\right)^{\otimes 2}$$

based on residuals

$$\hat{\varepsilon}_{i,k_n} \equiv Y_i - \hat{\mu}_{k_n}(X_i), \quad i = 1, \ldots, n \tag{2.9}$$

with $\hat{\mu}_{k_n}(X_i) \equiv \hat{\alpha}_{k_n} + \langle\hat{\beta}_{k_n}, X_i\rangle$ formed by a FPCR estimator $\hat{\beta}_{k_n}$ in (2.4) using a truncation level k_n; for greatest generality here, the truncation level k_n used for defining residuals need not be the same as the truncation h_n for the target FPCR estimator $\hat{\beta}_{h_n}$ and the sample covariance operator $\hat{\Lambda}_n$ will be consistent for Λ whenever $\hat{\beta}_{k_n}$ is consistent for β (cf. Yeon et al., 2024b, Lemma S23). The scaling $\hat{s}_{h_n}(x)$ in (2.8) estimates its population counterpart

$$s_{h_n}(x) \equiv \langle \Lambda \Gamma_{h_n}^{-1}(x - \mathsf{E}[X]) \Gamma_{h_n}^{-1}(x - \mathsf{E}[X]) \rangle = \|\Lambda^{1/2}\Gamma_{h_n}^{-1}(x - \mathsf{E}[X])\|^2, \quad x \in \mathbb{H}, \quad (2.10)$$

with $\Lambda \equiv \mathsf{E}[(X\varepsilon)^{\otimes 2}]$ (cf. Yeon et al., 2024b, Proposition S4), where this scaling for FLRMs can be motivated by recalling that a standard normal limit follows in a classical LRM (1.1) for a scaled LRM projection $\sqrt{n/s(X_0)}\{(\hat{\beta}^{LSE})^\top X_0 - \beta^\top X_0\}$, involving the least squares estimator $\hat{\beta}^{LSE}$ of the slope $\beta \in \mathbb{R}^p$, a new predictor $X_0 \in \mathbb{R}^p$, and analog scaling as $s(x) \equiv x^\top \Gamma^{-1} \Lambda \Gamma^{-1} x = \|\Lambda^{1/2}\Gamma^{-1}x\|^2_{\mathbb{R}^p}$, $x \in \mathbb{R}^p$, with $\Gamma \equiv \mathsf{E}[XX^\top]$ and $\Lambda \equiv \mathsf{E}[(X\varepsilon)(X\varepsilon)^\top]$ (Freedman, 1981). For the FLRMs, the inverse covariance operator Γ^{-1} is first truncated due to its unboundedness, which leads to analog scaling $s_{h_n}(x)$ for the heteroscedastic FLRMs as in (2.10).

Central limit theorems (CLTs) in homoscedastic FLRMs have been studied by Cardot et al. (2007), Yeon et al. (2023), and Yeon (2026), who work with the projection statistic $T_{\mathrm{proj},n}(X_0)$ in (2.7) upon replacing $\hat{s}_{h_n}(X_0)$ from (2.8) by a homoscedastic version. However, sampling distributions and the scaling form of projection statistics can change under under heteroscedastic FLRMs in general. For this reason, we follow the framework of Yeon et al. (2024a) for defining the scaled quantities in (2.7), which admit large-sample standard normal limits for either homoscedastic or hetereroscedastic FLRMs. For completeness, we provide this CLT in Theorem S1 of the supplement.

3. Wild bootstrap (WB) in heteroscedastic FLRMs

From the Introduction, recall the motivation for WB with FLRMs is that this approach can find a compromise of strengths in existing bootstrap approaches, i.e., RB is computationally fast but valid only for homoscedastic FLRMs, while PB is more generally valid but can be conservative (wider intervals) and computationally expensive for large data due to its fully non-parametric resampling. In contrast to RB and PB, the WB aims re-creates a bootstrap sample in which regressors match those $\{X_i\}_{i=1}^n$ of the original data (i.e., like RB which is beneficial for computation) *and* bootstrap responses $\{Y_i^*\}_{i=1}^n$ are formed by resampling residuals in a way that allows for heteroscedastic behavior (i.e., which is advantageous for wide validity). Namely, the WB method creates a data sample $\mathcal{D}_n^* \equiv \{(Y_i^*, X_i)\}_{i=1}^n$ of regressor-response pairs, in which bootstrap responses are generated as

$$Y_i^* = \hat{\mu}_{g_n}(X_i) + \varepsilon_i^*, \quad i = 1, \ldots, n$$

with $\hat{\mu}_{g_n}(X_i) \equiv \hat{\alpha}_{g_n} + \langle \hat{\beta}_{g_n}, X_i \rangle$ formed by a FPCR estimator $\hat{\beta}_{g_n}$ in (2.4) using a truncation level g_n; for greatest generality, the truncation level g_n for defining the WB version $\hat{\mu}_{g_n}(X_i)$ of the response mean $\mu(X_i)$ need not be the same as the truncation h_n defining the target FPCR estimator $\hat{\beta}_{h_n}$ from (2.4) or the truncation k_n defining residuals $\{\hat{\varepsilon}_{i,k_n}\}_{i=1}^n$ from (2.9). WB errors $\{\varepsilon_i^*\}_{i=1}^n$ are generated independently, though not identically as with RB, in a way that these have mean zero $\mathsf{E}^*[\varepsilon_i^*] = 0$ and a variance $\mathsf{E}^*[(\varepsilon_i^*)^2] = \hat{\varepsilon}_{i,k_n}^2$ in the bootstrap world, where E^* denotes expectation under the bootstrap generation and $\hat{\varepsilon}_{i,k_n}$ denotes the i-th (original data) residual defined in (2.9), $i = 1, \ldots, n$. Some specifications for WB errors $\{\varepsilon_i^*\}_{i=1}^n$ are given below.

From these bootstrap data, the WB analog of the FPCR estimator $\hat{\beta}_{h_n} \equiv \hat{\Gamma}_{h_n}^{-1}\hat{\Delta}_n$ from (2.4) is given as $\hat{\beta}_{h_n}^* \equiv \hat{\Gamma}_{h_n}^{-1}\hat{\Delta}_n^*$, where $\hat{\Delta}_n^* \equiv n^{-1}\sum_{i=1}^n (Y_i^* - \bar{Y}^*)(X_i - \bar{X})$ denotes the WB cross-covariance function between $\{Y_i^*\}_{i=1}^n$ and $\{X_i\}_{i=1}^n$ with $\bar{X} \equiv n^{-1}\sum_{i=1}^n X_i$ and $\bar{Y}^* \equiv n^{-1}\sum_{i=1}^n Y_i^*$, while the WB intercept is similarly computed as $\hat{\alpha}^* \equiv \bar{Y}^* - \langle \hat{\beta}_{h_n}^*, \bar{X}\rangle$; we then construct the WB estimator of the conditional mean as

$$\hat{\mu}_{h_n}^*(X_0) \equiv \hat{\alpha}^* + \langle \hat{\beta}_{h_n}^*, X_0 \rangle. \tag{3.11}$$

The WB then approximates the sampling distribution of a target statistical quantity $T_{\mathrm{mean},n}(X_0)$ or $T_{\mathrm{proj},n}(X_0)$ from (2.7) at a new predictor X_0 with a bootstrap counterpart

$$\begin{aligned} T_{\mathrm{mean},n}^*(X_0) &\equiv \sqrt{\frac{n}{\hat{s}_{h_n}^*(X_0)}}\{\hat{\mu}_{h_n}^*(X_0) - \hat{\mu}_{g_n}(X_0)\}, \\ T_{\mathrm{proj},n}^*(X_0) &\equiv \sqrt{\frac{n}{\hat{s}_{h_n}^*(X_0)}}(\langle \hat{\beta}_{h_n}^*, X_0 - \bar{X}\rangle - \langle \hat{\beta}_{g_n}, X_0 - \bar{X}\rangle), \end{aligned} \tag{3.12}$$

which involve scaling $\hat{s}_{h_n}^*(X_0)$ computed from the bootstrap sample; regarding the latter, a collection of WB residuals is computed as

$$\hat{\varepsilon}_{i,k_n}^* \equiv Y_i^* - \hat{\mu}_{k_n}^*(X_i), \quad i = 1, \ldots, n,$$

using the WB regression estimator $\hat{\beta}_{k_n}^* = \hat{\Gamma}_{h_n}^{-1}\hat{\Delta}_n^*$ with truncation level k_n, and then

$$\hat{s}_{h_n}^*(x) \equiv \langle \hat{\Lambda}_n^* \hat{\Gamma}_{h_n}^{-1}(x - \bar{X}), \hat{\Gamma}_{h_n}^{-1}(x - \bar{X})\rangle = \|(\hat{\Lambda}_n^*)^{1/2}\hat{\Gamma}_{h_n}^{-1}(x - \bar{X})\|^2, \quad x \in \mathbb{H}, \tag{3.13}$$

defines the bootstrap scaling with

$$\hat{\Lambda}_n^* \equiv n^{-1}\sum_{i=1}^{n}\left(X_i\hat{\varepsilon}_{i,k_n}^* - n^{-1}\sum_{i=1}^{n}X_i\hat{\varepsilon}_{i,k_n}^*\right)^{\otimes 2}$$

as an empirical covariance from the WB sample products $\{X_i\hat{\varepsilon}_{i,k_n}^*\}_{i=1}^n$.

In terms of constructing the WB errors $\{\varepsilon_i^*\}_{i=1}^n$, one popular approach is to use random multipliers, which corresponds to the so-called *multiplier WB*. Independently of the original data, let $\mathcal{W}_n \equiv \{W_i\}_{i=1}^n$ denote a set of independent random variables, generated for use in bootstrap to have mean zero and variance one. Then, WB errors $\{\varepsilon_i^*\}_{i=1}^n$ are defined as $\varepsilon_i^* \equiv W_i\hat{\varepsilon}_{i,k_n}$, $i = 1,\ldots,n$, with $\hat{\varepsilon}_{i,k_n}$ denoting the i-th (original data) residual from (2.9). By this construction, the WB errors possess the required moment properties in bootstrap: $\mathsf{E}^*[\varepsilon_i^*] = 0$ and $\mathsf{E}^*[(\varepsilon_i^*)^2] = \hat{\varepsilon}_{i,k_n}^2$. Our numerical studies show that the choice of multipliers may affect the performance of WB, though the common choices of multipliers below tend to induce similar results, where differences diminish with increasing sample size.

Example 1. Basic choices of multipliers include a two-point distribution with $\mathsf{P}(W_i = -(\sqrt{5}-1)/2) = (\sqrt{5}+1)/(2\sqrt{5}) = 1 - \mathsf{P}(W_i = (\sqrt{5}+1)/2)$ (Cao-Abad, 1991, Härdle and Marron, 1991, Mammen, 2000); a standard normal distribution with $W_i \sim \mathsf{N}(0,1)$; and a mean zero distribution with second and third moments of one, e.g., $W_i = V_i/2 + (V_i^2-1)/2$ for a variable $V_i \sim \mathsf{N}(0,1)$ (Härdle and Mammen, 1991, Mammen, 1993).

We now present a consistency result to verify that the WB method can validly approximate the distribution of target statistics $T_{\text{mean},n}(X_0)$ and $T_{\text{proj},n}(X_0)$, respectively, for inference about the mean response $\mu(X_0)$ and projection $\langle \beta, X_0 - \mathsf{E}[X]\rangle$. Let P^* denote bootstrap probability, as induced by WB resampling given the data $\{(Y_i, X_i)\}_{i=1}^n$.

Theorem 1. *In addition to the assumptions of the CLT in Theorem S1 of the supplement, we suppose that Conditions (L), (R), and (W1)–(W2) (including Condition (A4) for g_n) hold, which are listed in Section S1.2 of the supplement. Then, letting $T_n(X_0)$ denote either $T_{\text{mean},n}(X_0)$ or $T_{\text{proj},n}(X_0)$ from* (2.7) *and $T_n^*(X_0)$ denote the WB counterpart from* (3.12)*, the WB approximation of $T_n(X_0)$ by $T_n^*(X_0)$ is valid: as $n \to \infty$,*

$$\sup_{y \in \mathbb{R}} |\mathsf{P}^*(T_n^*(X_0) \le y | X_0) - \mathsf{P}(T_n(X_0) \le y | X_0)| \xrightarrow{\mathsf{P}} 0.$$

Similar to RB results in Yeon et al. (2023), the consistency of WB here could also be extended to cases where when the new predictor X_0 depends on the observed regressors $\{X_i\}_{i=1}^n$ or when the new X_0 and observed $\{X_i\}_{i=1}^n$ regressors do not share the same distribution, though we do not pursue repeating technical details here. As an example, with assumptions/arguments as in Yeon et al. (2023, Propositions S3, S4, S6, and S7), Theorem 1 holds for X_0 given as an average of some observed regressors $X_1, \ldots, X_L$ for some $L \le n$; see also Yeon et al. (2024a, Remark 7). We have also relegated the technical conditions for Theorem 1 to the supplement, with full details provided in Section S1.2 of the supplement. Conditions (W1)–(W2) above are WB error assumptions that hold for the multipliers in Example 1. For illustration, we may provide a concrete example that satisfies the remaining process conditions. In essence, these involve the eigenvalue decay of regressors, the smoothness of the slope function, the heteroscedastic error structure, and the growth rate for the truncation parameter h_n.

Example 2. In the space $\mathbb{H} = L^2([0,1])$ of all square integrable functions from $[0,1]$ to $\mathbb{R}$, we consider the Fourier basis functions $\{\phi_j\}_{j=1}^{\infty}$ as eigenfunctions, while the corresponding eigenvalues $\{\gamma_j\}_{j=1}^{\infty}$ are determined by their gaps $\delta_j \equiv \gamma_j - \gamma_{j+1} \asymp j^{-a}$ with decay rate $a > 2$ and $\gamma_1 \equiv \sum_{j=1}^{\infty} \delta_j$. The distribution of the random function X is then defined through the (normalized) functional principal component (FPC) scores $\gamma_j^{-1/2}\langle X, \phi_j\rangle = \xi W_j$, where $\{W_j\}_{j=1}^{\infty}$ are iid standard normal variables independent of the latent random variable ξ with $\mathsf{E}[\xi^{10}] < \infty$. This construction leads to dependence in principal component scores and satisfies Conditions (A1)–(A4) along with Condition (S). We next consider heteroscedastic errors with conditional variance $\sigma^2(X) \equiv \mathsf{E}[\varepsilon^2|X] = \|X\|^2$, which fulfill Conditions (C), (D), and (L); for example, the log normal error as in the simulation studies (Section 4) satisfies this. A polynomial decay rate b for the slope function β, i.e., $|\langle\beta, \phi_j\rangle| \asymp j^{-b}$, together with truncation choice such that $h_n \asymp n^{1/v}$ with $\min\{7, 2a+1\} < v < a + 2b - 1$ yields Condition $B(u)$ with $u > 7$ and the growth rate $n^{-1/2}h_n^{7/2}(\log h_n)^4 = o(1)$ for h_n in Theorem 1.

4. Numerical studies

Here we examine the performance of confidence intervals constructed by WB as well as by other bootstrap methods (i.e., RB and PB); these are implemented in the R package `BTSinFLRM` accompanying the paper. In numerical studies, we focus on inferring the projection $\mu(X_0) \equiv \langle \beta, X_0 \rangle$ by assuming both response and regressor have zero mean, for simplicity. Section 4.1 describes the simulation design, while Section 4.2 presents coverage rates and computational timings. Take-aways are summarized in Section 4.3.

4.1 Simulation designs

We consider different scenarios to investigate the performance of bootstrap methods. The slope function $\beta = \sum_{j=1}^{J} \beta_j \phi_j$ under consideration is defined by its Fourier series truncated up to a large integer $J = 15$, where $\{\beta_j\}$ denote the Fourier coefficients of β and $\{\phi_j\}_{j=1}^{J}$ represents the first J Fourier basis functions $\{1, \sin(2\pi t), \cos(2\pi t), \ldots, \}$ on $[0, 1]$. Here, we define $\beta_j = 3W_{\beta,j} j^{-b}$ with $b = 3.5$, where $\{W_{\beta,j}\}$ are independent variables with the identical distribution $\mathsf{P}(W_{\beta,j} = -1) = 1/2 = \mathsf{P}(W_{\beta,j} = 1)$. The functional regressors $\{X_i\}_{i=1}^{n}$ are generated as independent copies of the random element X, which is constructed based on the following truncated Karhunen–Loève expansion

$$X \overset{\mathsf{d}}{=} \sum_{j=1}^{J} \sqrt{\gamma_j} \xi_j \phi_j. \tag{4.14}$$

We take the Fourier basis functions $\{\phi_j\}_{j=1}^J$ for the eigenfunctions, while the eigenvalues are constructed based on the eigengaps with polynomial decay rate as $\delta_j = \gamma_j - \gamma_{j+1} = 2j^{-a}$ where $\gamma_1 = 2\sum_{j=1}^{\infty} j^{-a}$ and $a = 2.5$. All functions above are evaluated at the same 50 equi-distant discretized points in $[0, 1]$. For brevity, we consider only one combination of a and b as above, as these values are less consequential than the distribution of FPC scores and errors in the model (Yeon et al., 2023, 2024a). As in Example 2, the FPC scores $\{\xi_j\}_{j=1}^J$ are defined as $\xi_j = \xi W_j$, where ξ is a common latent and $\{W_j\}$ are independent standard normal random variables, which are independent of ξ; this gives non-normal and dependent FPC scores $\{\xi_j\}_{j=1}^J$. As for the distribution of the latent variable ξ that determines the moment of the regressor X, we focus on a $\mathtt{t}(4)$-distribution for ξ; this represents a difficult case in the sense that coverage accuracy of bootstrap intervals is expected to increase as ξ has higher moments (cf. Yeon et al., 2023, 2024a). The functional regressors $\{X_i\}_{i=1}^n$ and a new regressor X_0 are then generated following the expansion in (4.14).

The data pairs $\{(Y_i, X_i)\}_{i=1}^n$ are generated with different sample sizes $n \in \{50, 200, 1000\}$, where the scalar responses $\{Y_i\}_{i=1}^n$ defined by error terms ε_i that follow a centered log normal distribution conditional on X_i. The latter conditional distribution, written terms of a generic error ε and regressor X, is given by $\log(\varepsilon + \exp[\mu(X) + \tau^2/2]) \sim \mathsf{N}(\mu(X), \tau^2)$, where the mean $\mu(X) \equiv \{\log[\sigma^2(X)/(e^{\tau^2} - 1)] - \tau^2\}/2$ is defined as either $\sigma^2(X) = \mathrm{tr}(\Gamma) = \sum_{j=1}^J \gamma_j$ in the homoscedastic scenario or as $\sigma^2(X) = \|X\|^2$

in the heteroscedastic scenario. This error distribution allows a non-trivial effect of higher moments because the (conditional) skewness of the error ε, given by $\mathsf{skew}[\varepsilon|X] = (e^{\tau^2}+2)(e^{\tau^2}-1)^{1/2}$, exponentially increases over different parameters τ^2, while the lower moments $\mathsf{E}[\varepsilon|X] = 0$ and $\mathsf{E}[\varepsilon^2|X] = \sigma^2(X)$ do not depend on τ^2. We report results with $\tau^2 = 0.1$ and $\tau^2 = 3$, which respectively give skewness 1 and 100, approximately.

Based on 1000 Monte Carlo iterations per sample size and model configuration, we compute the empirical coverage rates, as well as average widths, of 95% confidence intervals for a projection $\langle \beta, X_0 \rangle$ from the symmetrized intervals from each bootstrap methods (cf. Hall, 1988); for example, we construct WB intervals from approximating the absolute value of projection statistic in (2.7) through the absolute value of bootstrap projection statistic in (3.12). The three types of WB multipliers listed in Example 1 are considered and denoted as WB1 (two-point), WB2 (standard normal), and WB3 (Mammen, 1993), respectively. We compare these WB intervals against RB and PB intervals as well as intervals by a direct standard normal (e.g., CLT) approximation (cf. Section 2.2). Bootstrap methods are implemented based on 1000 resamples and truncation levels $k_n = g_n = [1.5 \cdot n^{1/6}] \in \{3, 4, 6\}$, $h_n \in \{1, \ldots, 15\}$, where $[\cdot]$ denotes the closest integer, where values of k_n and g_n are motivated by consistency conditions for $\hat{\beta}_{k_n}$.

4.2 Performance of WB bootstrap intervals

Figure 1 shows empirical coverage probabilities over both scenarios of homoscedastic and heteroscedastic errors, focusing on a log-variance parameter $\tau^2 = 0.1$ (or error skewness of 1). Under homoscedasticity, WB intervals exhibit accuracy that closely mirrors that of the RB intervals. Under heteroscedasticity, WB intervals maintain relatively good performance in coverage, while RB intervals are relatively poor in contrast (i.e., coverages for RB are so low as to often not plot in Figures 1-2) and CLT-based approximation intervals also perform poorly. PB intervals tend to have better accuracy than WB intervals under heteroscedasticity over smaller sample sizes $n \in \{50, 200\}$, though WB intervals achieve comparable coverage accuracy for increasing sample sizes.

Figure 2 summarizes empirical coverages, focusing on heteroscedastic errors with a large log-variance parameter $\tau^2 = 3$ (or error skewness of 100). In contrast to the lower skewness case in Figure 1, the coverage results in Figure 2 indicate that the WB can be preferable to PB in terms of coverage accuracy under heavy skewness. In this setting, PB intervals can become rather conservative, as indicated in Figure 3 which shows average lengths of intervals in connection to Figure 2. Here, the lengths of PB intervals are longer and less stable compared to WB intervals, which may owe to the fact that the PB method resamples regressor variables, unlike WB, which may create more difficulties under heavy skewness.

As raised by a reviewer, regarding the effect of the truncation level h_n, WB does often perform worse than PB in terms of coverage accuracy for small values of h_n in Figures 1-2. This aspect attributes to non-parametric resampling scheme of the PB, which can induce wider (more conservative) intervals for such h_n values (cf. Figure 3). However, these results also generally suggest that WB intervals need higher truncation levels h_n to become stable in length, after which these intervals also become more stable in coverage. This insight can be helpful toward selecting a truncation level h_n for the

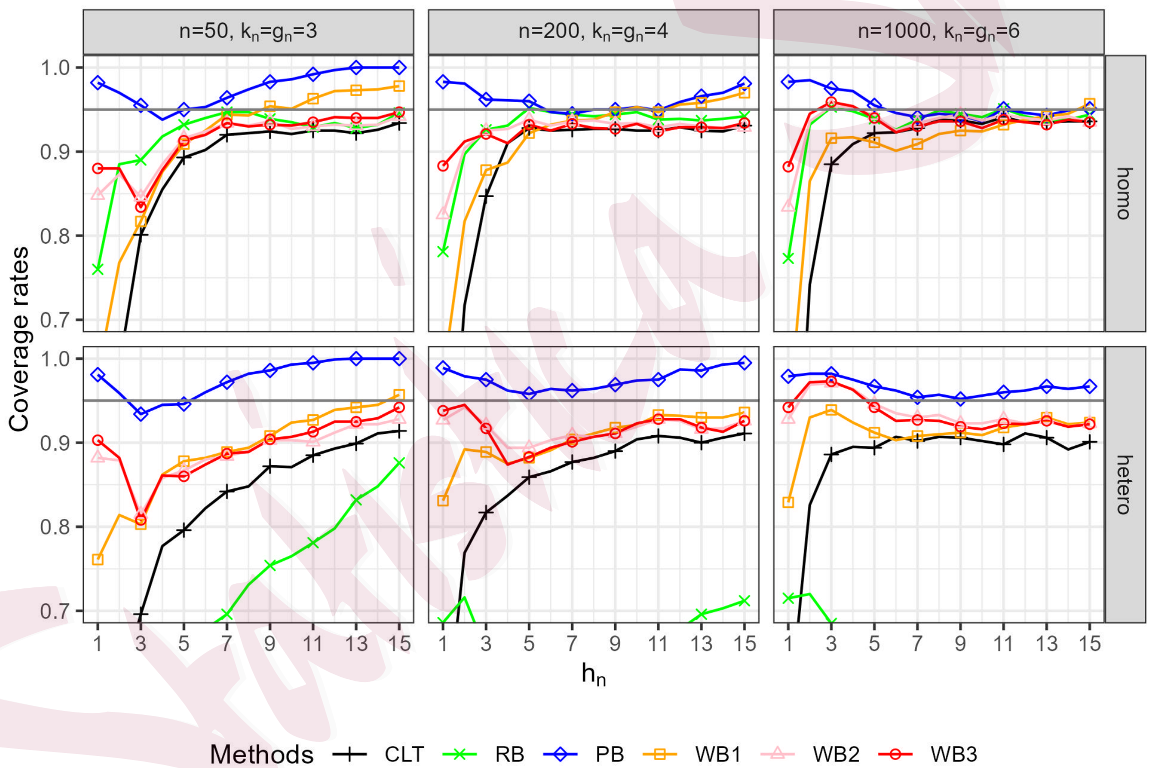

Figure 1: Under log-variance parameter $\tau^2 = 1$, empirical coverage rates of 95% intervals for the projection $\langle \beta, X_0 \rangle$ over truncation levels $h_n \in \{1, \ldots, 15\}$.

WB in application; we revisit this point in Section 5.1 where practical implementation is discussed. In contrast, the length behaviors in Figure 3 (as well as other length studies in the supplement, (cf. Figure S1) indicate that PB intervals tend to be more sensitive than WB to the choice of the truncation level h_n, where instability is apparent when sample sizes are small and the error skewness is strong. The tuning parameter h_n for PB remains challenging to choose, and intervals can be quite wide.

Remark 1. Following a reviewer's suggestion, we conducted additional numerical studies to examine the effects of k_n and g_n in greater detail. Specifically, simulations were conducted under the same setup as in Figure 2, but by fixing either k_n and g_n while varying the other as well as h_n. The results are represented in Figure S5 of the supplement.

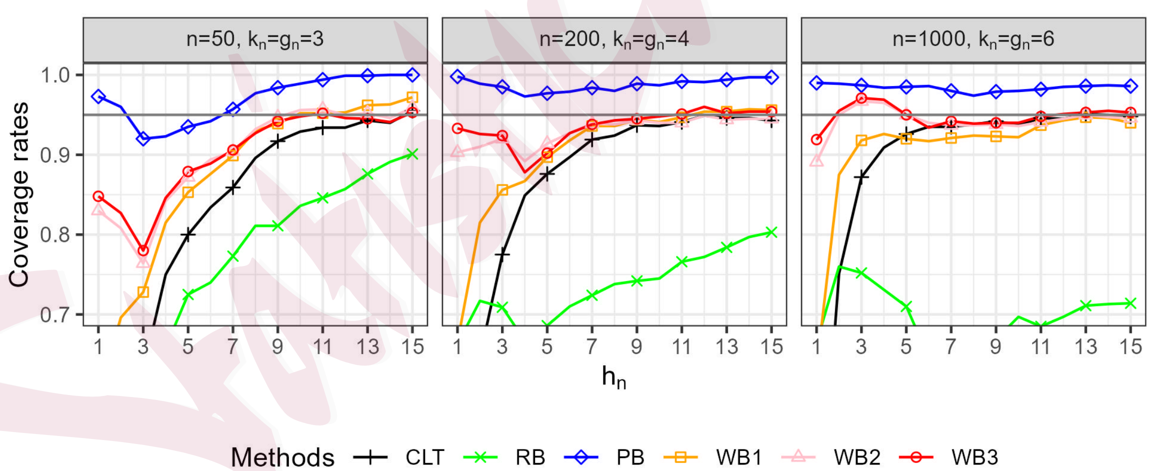

Figure 2: Under log-variance parameter $\tau^2 = 2$, empirical coverage rates of 95% intervals for the projection $\langle \beta, X_0 \rangle$ over truncation levels $h_n \in \{1, \ldots, 15\}$.

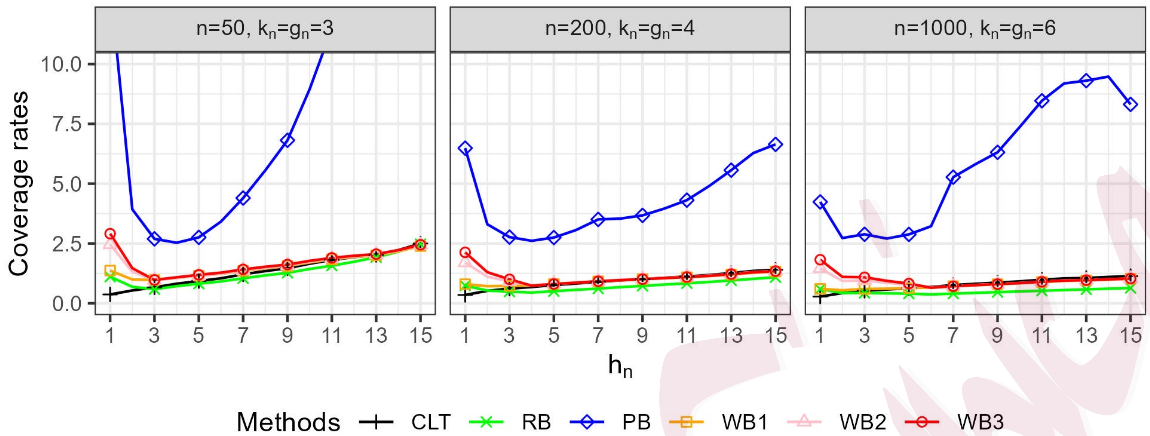

Figure 3: Under log-variance parameter $\tau^2 = 2$, average lengths of 95% intervals for the projection $\langle \beta, X_0 \rangle$ over truncation levels $h_n \in \{1, \ldots, 15\}$.

Overall, large differences between g_n and k_n tend to produce over- or under-coverage in WB intervals, while g_n close to k_n leads to more reasonable behavior in WB intervals. This finding supports our WB implementation with g_n being equal to k_n in Section 5.1. We refer to Section S3.3 of the supplement for more details.

Remark 2. One could construct WB statistics akin to (3.12) but without bootstrap-level studentization, i.e., applying original data scaling $\hat{s}_{h_n}(X_0)$ from (2.8) in place of bootstrap scaling $\hat{s}^*_{h_n}(X_0)$ from (3.13). Additional numerical studies have indicated that the coverage accuracy of this WB version is not robust to the choice of the multipliers, potentially because this WB does not mimic the actual studentization in the target statistic (2.7). We refer to Section S3.2 of the supplement for more discussion.

To examine computational speeds among bootstraps for functional regression, we consider the effects of different sample sizes $n \in \{50, 200, 1000\}$ and ten different discretization sizes $M \in \{25, 50, \ldots, 225, 250\}$ for regressor curves. Bootstrap confidence intervals were computed using three functions `RBinFLRM`, `PBinFLRM`, `WBinFLRM` found in the R package `BTSinFLRM`, as companion software to this paper, in one representative simulation scenario ($\sigma^2(X) = \mathrm{tr}(\Gamma)$ and $\tau^2 = 1$ with truncation levels $k_n = g_n = [2n^{1/6}]$ and $h_n = g_n + 1$). Computing speeds for bootstrap intervals were timed on a machine equipped with Intel Xeon Gold 6144 Processor (with base frequency 3.5GHz) using the R package `microbenchmark` (Mersmann, 2025). Figure 4 displays the average timing speeds from 1000 simulations. As expected, WB and RB require similar amounts of computing time, due to their resampling constructions with fixed regressors. However, PB can be more time demanding than other bootstrap methods, particularly when either sample or discretization sizes is large, owing to its more involved resampling scheme with regressors. In this sense, WB can have computational advantages over PB in heteroscedastic cases with large samples or fine discretizations.

4.3 Summary

We emphasize that the proposed WB is designed for mean response inference without the intent to outperform the existing RB and PB. Rather, the WB serves as an useful alternative to both RB and PB by offering key benefits: (i) WB applies to both homoscedastic and heteroscedastic models, a clear benefit relative to the RB; (ii) for large datasets, it shows both methodological and computational advantages over the PB; (iii) WB performs better than the PB for some complex cases, e.g., for largely skewed errors, where PB intervals are excessively conservative; and (iv) WB shows stability in interval length as a function of tuning parameters, much more so than PB, which helps toward selection of these parameters in practice (cf. Section 5.1).

5. Practical implementation

Section 5.1 overviews choices of the truncation levels k_n, g_n, h_n in practice with WB inference for mean responses. This approach is then illustrated in Section 5.2 through a United States weather dataset.

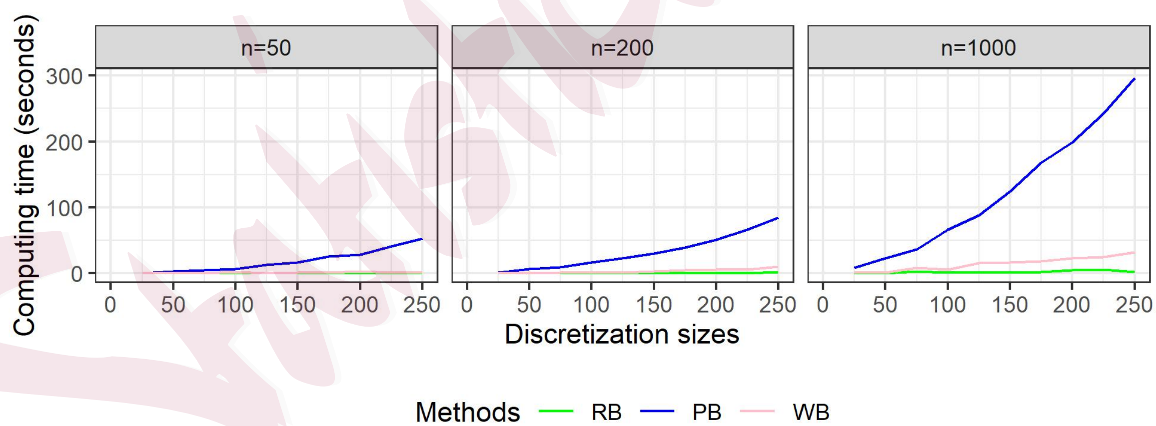

Figure 4: Computing timings of each bootstrap method in seconds over discretization sizes $M \in \{25, 50, \ldots, 225, 250\}$.

5.1 Selection of truncation levels

Tuning parameter selection is an important problem in regularization problems, including the methods based on the FPCR estimator. There have been some attempts to define an optimal truncation level based on estimation error (Cai and Hall, 2006, Cai et al., 2018, Hall and Horowitz, 2007), though such choices have no optimality guarantees for bootstrap inference. Additionally, in our context with bootstrap, appropriate truncation levels may also depend on the mean response (or projection) target of interest at a regressor X_0 value. An optimal choice for bootstrap coverage accuracy requires more investigation, outside the scope of the current study.

Instead, we suggest some practical guidance for choosing truncation levels k_n, h_n, g_n with the proposed WB procedure based on theoretical and numerical findings. Firstly, the truncation k_n for calculating residuals may be chosen by cross-validation (e.g., based on prediction error), as the requirement for the corresponding estimator $\hat{\beta}_{k_n}$ is simply consistency (cf. Theorem 1). Secondly, we set $g_n = k_n$, because $\hat{\beta}_{g_n}$ also requires only consistency and the WB intervals perform well for g_n equal to k_n in our numerical studies (cf. Remark 1). Lastly, for h_n used to infer the mean response $\mu(X_0)$, we recommend selecting a value larger than g_n for the following reasons: the theoretical result in Theorem 1 advises against choosing h_n smaller than g_n (i.e., by Condition (R) given in the supplement), while numerical findings in Section 4 indicate that the WB intervals perform better and also exhibit stability for sufficiently large h_n exceeding

g_n. The latter point suggests that a value of h_n may be selected by inspecting where WB intervals appear stable. With this in mind, we suggest a concrete rule below for selecting h_n. The strategy follows the spirit of the so-called "minimum volatility method" of Politis et al. (1999, Section 9.3.2), which aims to compute intervals over a series of bandwidths and then select the one where the resulting interval appears to exhibit the lowest volatility. Our heuristic also closely resembles the notion of choosing a truncation parameter where the fraction of variance explained in classical FPCA appears stable (e.g., Kokoszka and Reimherr, 2017, Section 12.2).

To elaborate further, with $k_n = g_n$ chosen as described above, let H denote a finite set of consecutive candidate integers for selecting h_n (e.g., $H = g_n + \{0, 1, \ldots, 20\}$). We write w_h and c_h for the width and center of the WB intervals corresponding to $h \in H$. To identify truncation levels where these two interval characteristics stabilize (i.e., widths and centers cease to change dramatically), we consider the following sets:

$$H_w \equiv \{h \in H : |w_{h+1} - w_h| \leq \rho_w\}, \qquad H_c \equiv \{h \in H : |c_{h+1} - c_h| \leq \rho_c\},$$

where $\rho_w, \rho_c > 0$ are some small thresholds for defining stability. Denoting $H_{\text{stable}} \equiv H_w \cap H_c$ and picking a small integer $r \geq 0$, we determine a truncation level h_n as

$$\hat{h}_n = \hat{h}_n(r) \equiv \min\{h \in H_{\text{stable}} : h, h+1, \ldots, h+r \in H_{\text{stable}}\}. \tag{5.15}$$

Our selection strategy is called the stabilized volatility method (SVM) hereafter.

Using the same simulation setup as that for Figure 2, we conduct numerical studies

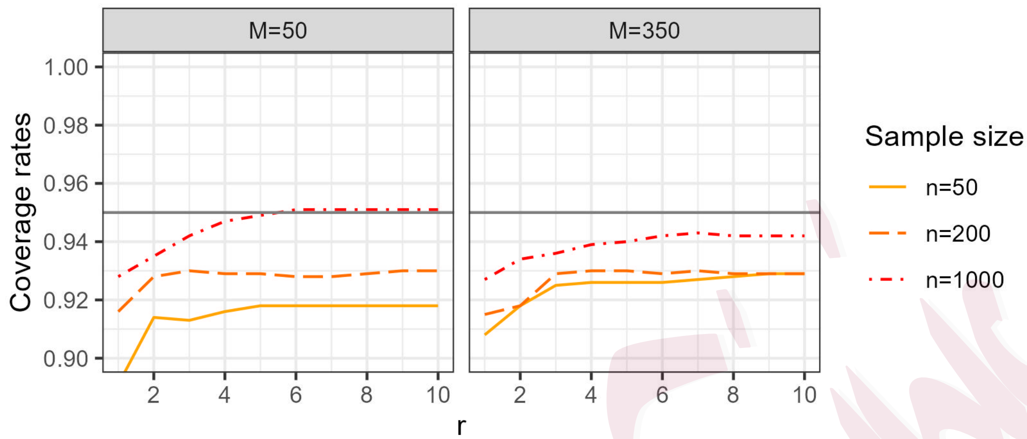

Figure 5: Empirical coverages of 95% WB intervals based on standard normal multipliers and truncation $\hat{h}_n(r)$ selected by SVM (5.15). Data generation matches that of Figure 2, where scenarios with a denser regressor grid are also considered (right).

to demonstrate the coverage performance of WB intervals based on SVM-selected truncation levels. With $\rho_w = \rho_c = 0.01$, Figure 5 displays empirical coverage rates of WB intervals over different $r \in \{1, 2, \ldots, 10\}$ for defining the criterion (5.15); here we consider three different sample sizes n along with two discretization levels $M \in \{50, 350\}$ for regressor curves. Note that the coverage results appear reasonable, showing a stable pattern of behavior over various r for all sample sizes n or levels M. These findings suggest that the proposed SVM for selecting the truncation level h_n with WB can be useful. In the next section, we further demonstrate this selection approach, which can be combined with a graphical visualization.

5.2 Real data analysis

To illustrate the proposed WB method, we consider a US weather dataset, called nClimGrid-Daily, provided by the National Centers for Environmental Information. The data contain daily temperatures (°C) and precipitation (mm) on gridded fields in the US.[‡] In particular, we extract data for 2022, where each functional regressor X_i represents daily average temperatures of a state i in the US, while the associated response Y_i is the total precipitation for $i = 1, \ldots, 47$; three states (Alaska, Alabama, and Hawaii) do not appear in the data so that the total sample size is $n = 47$. As daily observations are available, the discretization size is large as $M = 365$. We examine four new regressor curves $\{X_{0,l}\}_{l=1}^{4}$ as the average curves of a region in the US: Northeast, South, North Central, and West as defined by the United States Census Bureau. The supplement shows the temperature curves of each state and their averages $\{X_{0,l}\}_{l=1}^{4}$.

We first examine some distributional features of the data, in terms of heteroscedasticity and skewness, that may suggest the choice of bootstrap method. A pilot truncation k_n is first obtained as $k_n = 6$ by cross-validation based on the prediction error as suggested in Section 5.1. To examine whether the errors have constant variance over different regions, we compute the residuals $\hat{\varepsilon}_{i,k_n} \equiv Y_i - \bar{Y} - \langle \hat{\beta}_{k_n}, X_i - \bar{X} \rangle$, and Figure 6 provides a scatterplot of residuals versus the predicted values $\hat{Y}_{i,k_n} \equiv \bar{Y} + \langle \hat{\beta}_{k_n}, X_i - \bar{X} \rangle$. This figure suggests some heterosceadsticity in the data, especially due to West region

[‡]https://www.ncei.noaa.gov/products/land-based-station/nclimgrid-daily

(as does a study of estimated standard deviations shown in the supplment). In the presence of potential heteroscedasticity, we might expect that RB intervals should not be used (or may work poorly) and that PB or WB intervals would be more reasonable. We next consider kernel-estimated densities from residuals in each region to examine skewness; these (as shown in the supplement) exhibit some slight skewness, again particularly in West region. As discussed in Section 4, potential skewness in errors can suggest that the WB method can become preferable to PB for coverage accuracy. Further, as the discretization size $M = 365$ is relatively large for these data, WB is also computationally faster to implement than PB, as shown in Figure 4.

We will apply the results in Theorem 1 to construct (symmetrized) WB intervals of

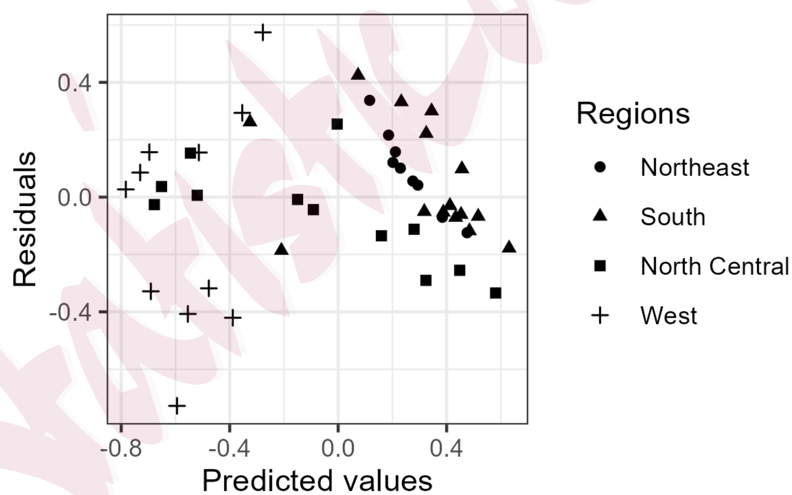

Figure 6: A scatter plot of residuals versus predicted values with $k_n = 6$ from the US weather dataset.

the centered projections $\{\langle \beta, X_{0,l} - \mathsf{E}[X]\rangle\}_{l=1}^4$ for all four regions with normal multipliers. This allows us to compare the original projections $\{\langle \beta, X_{0,l}\rangle\}_{l=1}^4$ with the global average projection $\langle \beta, \mathsf{E}[X]\rangle$, and hence, the mean responses $\{\mu(X_{0,l})\}_{l=1}^4$ with the global mean $\mu(X)$. That is, the resulting intervals have the interpretation of indicating how the mean response (total precipitation) per region may differ from a common global average. For bootstrap inference, 1000 bootstrap resamples are used along with $g_n = k_n = 6$ as recommended in Section 5.1.

To illustrate application of the selection rule for h_n from Section 5.1 for WB intervals, we focus on the West region (i.e., the value of $X_{0,l}$ for $l = 4$) to guide a choice of h_n, where heteroscedasticity and skewness are more pronounced. The bootstrap intervals for the West region, computed for different values of h_n, are presented in Figure 7. RB intervals are narrow in general, though less accuracy for this method would be expected under heteroscedasticity. The widths of the PB intervals tend to be wide and exhibit the least stable pattern in widths; this also matches behavior observed in numerical studies (Section 4). Note that the WB intervals stably behave like the RB counterparts, though with larger widths to incorporate heteroscedasticity. Figure 7 visually suggests that the WB intervals (pink) begin to stabilize in both center and width, when the truncation level h_n reaches 9 or higher.

To quantitatively select h_n by the SVM of Section 5.1, we compute $\hat{h}_n = \hat{h}_n(r)$ in (5.15) with $H \equiv \{6, \ldots, 20\}$ and $\rho_w = \rho_c = 0.01$ for the candidate sets. It is worthwhile recalling from the right panel of Figure 5 that the proposed SVM was demonstrated as reasonable, even for densely observed regressors as in these weather data. Our findings indicate that, for any $r \in \{1, \ldots, 10\}$, the selected truncation level for the West region is $\hat{h}_n = 9$, as also visually identified from Figure 7. Figure S9 of the supplement shows the intervals for the other regions based on each bootstrap, where the best empirical choice of h_n for the inference of $\mu(X_{0,l})$ may change with or depend on the new regressor $X_{0,l}$ (i.e., region) and the selected values for WB were 9 for Northeast, 12 for South, and 10 for North Central, e.g., for all $r \in \{2, 3, 4, 5, 6\}$.

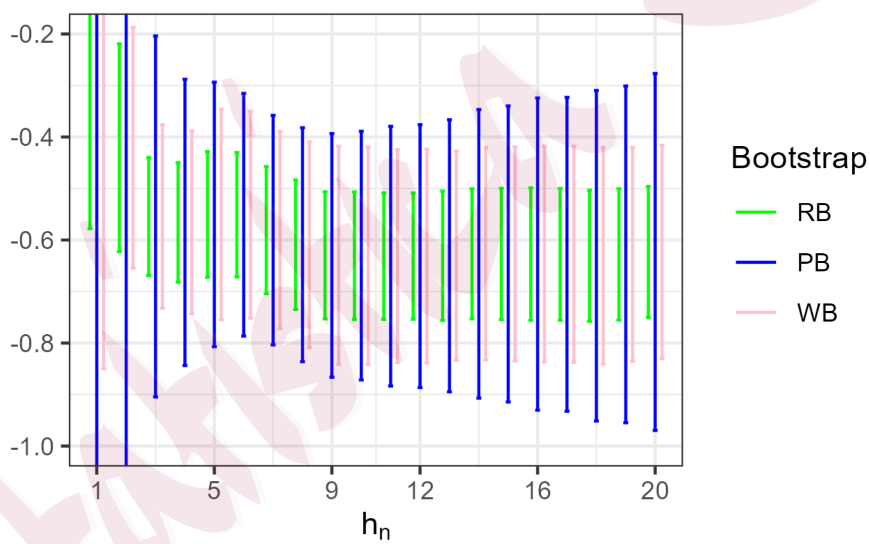

Figure 7: Bootstrap intervals for centered projection $\langle \beta, X_{0,4} - \mathsf{E}[X] \rangle$ for West from the US weather dataset with $k_n = 6 = g_n$ and different $h_n \in H \equiv \{1, \ldots, 20\}$. The plot is cropped within $[-1.0, -0.2]$ due to too large widths of PB intervals.

In Table 1, we provide the WB intervals based on the chosen truncation levels h_n. The results support a conclusion that the projections $\langle \beta, X_{0,l} \rangle$ from each region differ substantially from the global mean projection $\langle \beta, \mathsf{E}[X] \rangle$, as none of the WB intervals contain zero. More specifically, the mean total precipitation in the Northeast and South appears higher than the global mean, whereas the North Central and West regions experience less rainfall on average compared to the the total precipitation across all 47 observed states. See the supplement for further numerical summaries. As these data seemingly involve complicated error structures (i.e., with heteroscedastic and skewed distributions), the WB intervals are helpful toward drawing valid inferences.

6. Concluding remarks and outlook

Uncertainty quantification is an important, but challenging, problem in functional linear regression models (FLRMs). We proposed and formally established a new wild bootstrap (WB) method for mean responses in FLRMs. As suggested by both the-

Table 1: 95% WB intervals for centered projections $\{\langle \beta, X_{0,l} - \mathsf{E}[X] \rangle\}_{l=1}^{4}$ from weather data using selected truncations $k_n = 6 = g_n$ and $h_n = 9$ for Northeast ($l = 1$), $h_n = 12$ for South ($l = 2$), $h_n = 10$ for North Central ($l = 3$), and $h_n = 9$ for West ($l = 4$).

Northeast	South	North Central	West
$[0.239, 0.466]$	$[0.267, 0.445]$	$[-0.220, -0.029]$	$[-0.842, -0.418]$

oretical findings and numerical results, WB combines the strengths of both residual bootstrap (RB) and paired bootstrap (PB) with functional regressors: WB retains a low computational burden by keeping regressors fixed in the bootstrap scheme, similar to RB, while retaining validity under both homoscedasticity and heteroscedasticity, akin to PB. In particular, WB serves as a useful alternative to RB in heteroscedastic cases and to PB when the data structure is large, while also outperforming both methods in coverage accuracy under heavy skewness in heteroscedastic errors and also allowing for a selection method for tuning parameters due to stability in WB interval behavior. Additionally, we have prepared an accompanying R package, which provides implementations of WB as well as RB and PB for setting confidence intervals for mean responses in FLRMs.

We conclude by outlining several potential extensions of interest. As Khademnoe and Hosseini-Nasab (2024) and Lin and Lin (2025) apply bootstrap for testing the slope $H_0 : \beta = 0$ in homoscedastic FLRMs, it may be possible to consider WB for extending such slope tests under heteroscedastic errors. Further, bootstrap developments could also be meaningfully considered for inference about more complicated FLRMs such as with functional response (Dette and Tang, 2024) or as with generalized (Müller and Stadtmüller, 2005), partial (Shin, 2009), or high-dimensional (Xue and Yao, 2021) FLRMs. Finally, further investigation is warranted in developing bootstrap inference in FLRMs with sparsely observed functional regression due to unique challenges from sparseness (Zhou et al., 2023). Such research could enrich inference in FLRMs.

Supplementary Materials

This supplement contains all technical details, additional simulation results, and extra figures regarding data example.

Acknowledgements

We are thankful to associate editor and two anonymous reviewers for their constructive and helpful feedback. A large part of the work by Hyemin Yeon was carried out while he was at Kent State University, and research is partially supported by the Department of Mathematical Sciences at Kent State University (Yeon) and NSF DMS-2515719 (Nordman).

References

Cai, T. T. and P. Hall (2006). Prediction in functional linear regression. *Annals of Statistics 34*(5), 2159–2179.

Cai, T. T., L. Zhang, and H. H. Zhou (2018). Adaptive functional linear regression via functional principal component analysis and block thresholding. *Statistica Sinica 28*(4), 2455–2468.

Cao-Abad, R. (1991). Rate of convergence for the wild bootstrap in nonparametric regression. *Annals of Statistics 19*(4), 2226–2231.

Cardot, H., F. Ferraty, and P. Sarda (1999). Functional linear model. *Statistics & Probability Letters 45*(1), 11–22.

Cardot, H., F. Ferraty, and P. Sarda (2003). Spline estimators for the functional linear model. *Statistica Sinica*, 571–591.

Cardot, H., A. Mas, and P. Sarda (2007). CLT in functional linear regression models. *Probability Theory and Related Fields 138*(3-4), 325–361.

Dette, H. and J. Tang (2024). Statistical inference for function-on-function linear regression. *Bernoulli 30*(1), 304–331.

Efron, B. (1979). Bootstrap methods: Another look at the jackknife. *Annals of Statistics 7*(1), 1–26.

Freedman, D. A. (1981). Bootstrapping regression models. *Annals of Statistics 9*(6), 1218–1228.

González-Manteiga, W. and A. Martínez-Calvo (2011). Bootstrap in functional linear regression. *Journal of Statistical Planning and Inference 141*(1), 453–461.

Hall, P. (1988). On symmetric bootstrap confidence intervals. *Journal of the Royal Statistical Society Series B: Statistical Methodology 50*(1), 35–45.

Hall, P. and J. L. Horowitz (2007). Methodology and convergence rates for functional linear regression. *Annals of Statistics 35*(1), 70–91.

Härdle, W. and E. Mammen (1991). Bootstrap methods in nonparametric regression. In *Nonparametric Functional Estimation and Related Topics*, pp. 111–123. Springer.

Härdle, W. and J. S. Marron (1991). Bootstrap simultaneous error bars for nonparametric regression. *Annals of Statistics*, 778–796.

Hastie, T., R. Tibshirani, and J. Friedman (2009). *The Elements of Statistical Learning*. Springer New York.

Hsing, T. and R. Eubank (2015). *Theoretical Foundations of Functional Data Analysis, with an Introduction to Linear Operators*, Volume 997. John Wiley & Sons.

Khademnoe, O. and S. M. E. Hosseini-Nasab (2024). On the validity of the bootstrap hypothesis testing in functional linear regression. *Statistical Papers 65*(4), 2361–2396.

Kokoszka, P. and M. Reimherr (2017). *Introduction to Functional Data Analysis*. Chapman and Hall/CRC.

Lin, Y. and Z. Lin (2025). Hypothesis testing for functional linear models via bootstrapping. *Bernoulli 31*(4), 3309–3330.

Liu, R. Y. (1988). Bootstrap procedures under some non-iid models. *Annals of Statistics 16*(4), 1696–1708.

Mammen, E. (1993). Bootstrap and wild bootstrap for high dimensional linear models. *Annals of Statistics 21*(1), 255–285.

Mammen, E. (2000). Resampling methods for nonparametric regression. *Smoothing and Regression: Approaches, Computation, and Application*, 425–450.

Mersmann, O. (2025). *microbenchmark: Accurate Timing Functions*. R package version 1.5.0.

Müller, H.-G. and U. Stadtmüller (2005). Generalized functional linear models. *Annals of Statistics 33*(2), 774–805.

Politis, D., J. Romano, and M. Wolf (1999). *Subsampling*. Springer Series in Statistics. Springer New York.

Shin, H. (2009). Partial functional linear regression. *Journal of Statistical Planning and Inference 139*(10), 3405–3418.

Wu, C.-F. J. (1986). Jackknife, bootstrap and other resampling methods in regression analysis. *Annals of Statistics 14*(4), 1261–1295.

Xue, K. and F. Yao (2021). Hypothesis testing in large-scale functional linear regression. *Statistica Sinica 31*(2), 1101–1123.

Yeon, H. (2026). Gaussian and bootstrap approximation for functional principal component regression. *To appear in Electronical Journal of Statistics*.

Yeon, H., X. Dai, and D. J. Nordman (2023). Bootstrap inference in functional linear regression models with scalar response. *Bernoulli 29*(4), 2599 – 2626.

Yeon, H., X. Dai, and D. J. Nordman (2024a). Bootstrap inference in functional linear regression models with scalar response under heteroscedasticity. *Electronic Journal of Statistics 18*(2), 3590 – 3627.

Yeon, H., X. Dai, and D. J. Nordman (2024b). Supplement to "bootstrap inference in functional linear regression models with scalar response under heteroscedasticity". *Iowa State University Digital Repository*, <https://dr.lib.iastate.edu/handle/20.500.12876/jrl8QQqr>.

Zhou, H., F. Yao, and H. Zhang (2023). Functional linear regression for discretely observed data: from ideal to reality. *Biometrika 110*(2), 381–393.

Department of Mathematical Sciences, Kent State University

E-mail: hyeon1@kent.edu

ORCiD: 0009-0008-3585-2103

Division of Biostatistics, University of California, Berkeley

E-mail: xiongtao.dai@hotmail.com

Department of Statistics, Iowa State University

E-mail: dnordman@iastate.edu