Statistica Sinica **32** (2022), 2265-2293
doi:https://doi.org/10.5705/ss.202020.0358

ROBUST INFERENCE FOR PARTIALLY OBSERVED FUNCTIONAL RESPONSE DATA

Yeonjoo Park[1], Xiaohui Chen[2] and Douglas G. Simpson[2]

[1] *University of Texas at San Antonio and*
[2] *University of Illinois at Urbana-Champaign*

Abstract: Irregular functional data, in which densely sampled curves are observed over different ranges, pose a challenge for modeling and inference, and sensitivity to outlier curves is a concern in applications. Motivated by applications in quantitative ultrasound signal analysis, this study investigates a class of robust M-estimators for partially observed functional data, including functional location and quantile estimators. The consistency of the estimators is established under general conditions on the partial observation process. Under smoothness conditions on the class of M-estimators, asymptotic Gaussian process approximations are established and used for large-sample inference. The large-sample approximations justify using a bootstrap approximation for robust inferences about the functional response process. The performance of the proposed estimators is demonstrated by means of simulations and an analysis of irregular functional data from quantitative ultrasound analysis.

Key words and phrases: Bootstrap, functional central limit theorem, functional quantile, L^2-norm test, trend analysis.

1. Introduction

With advances in instrumentation and the ability to acquire data densely over a continuum, function-valued data acquisition is becoming increasingly common in many fields; see, for example, Ramsay and Silverman (2005) and Horváth and Kokoszka (2012). Earlier works on functional data focused in large part on regular functional data, in which the functional samples are densely collected over a common domain, or on sparse functional data, in which the functional response for each subject is sparsely sampled, with a small number of irregularly spaced measurements over the domain. In recent years, applications have emerged that produce partially observed functional data. Here, each individual trajectory is collected only over individual-specific subinterval(s), densely or sparsely, within the whole domain of interest. Several recent works have begun addressing the es-

Corresponding author: Douglas G. Simpson, University of Illinois at Urbana-Champaign, Champaign, IL 61820, USA. E-mail: dgs@illinois.edu.

timation of covariance functions for short functional segments observed at sparse and irregular grid points, called "functional snippets" (Descary and Panaretos (2019); Lin and Wang (2020); Lin, Wang and Zhong (2020); Zhang and Chen (2022)), or for fragmented functional data observed on small subintervals (Delaigle et al. (2021)). For densely observed partial data, existing studies have focused on estimating the unobserved part of curves (Kraus (2015); Delaigle and Hall (2016); Kneip and Liebl (2019)), prediction (Liebl (2013); Goldberg, Ritov and Mandelbaum (2014)), classification (Delaigle and Hall (2013); Stefanucci, Sangalli and Brutti (2018); Mojirsheibani and Shaw (2018); Kraus and Stefanucci (2018); Park and Simpson (2019)), functional regression (Gellar et al. (2014)), and inferences (Gromenko, Kokoszka and Sojka (2017); Kraus (2019)).

Robustness to atypical curves or deviations from Gaussian variation is a practical concern in modeling and inference, especially for partially observed functional data. For example, Park and Simpson (2019) demonstrated that t-type heavy-tailed models for functional data outperform Gaussian methods for the probabilistic classification of quantitative ultrasound (QUS) measurements, which extract diagnostic information on biological tissue, such as a tumor, from ultrasound radio frequency backscattering signals. In QUS analysis, the backscattered spectrum is captured by a transducer by scanning the region of interest. The attenuation adjusted backscatter coefficient (BSC) comprises a functional curve spanning the frequency range of the transducer.

Wirtzfeld et al. (2015) presented QUS data from an inter-laboratory diagnostic ultrasound study in which two types of induced mammary tumors were scanned using multiple transducers of varying center frequencies: 4T1 tumors in mice, and MAT tumors in rats. Figure 1 shows a subset of the data. The resulting BSC curves are observed over varying frequency ranges, depending on the transducers used in the scanning. At the same time, several curves show atypical behaviors, especially at the lower frequency ranges in the 4T1 group.

The above example illustrates issues that motivate the two main goals of this study: (i) to develop a robust functional data analysis approach that is general enough to handle partially observed functional data; and (ii) to establish asymptotic properties to provide the foundation for associated robust inferences.

Several authors have studied robust estimation for balanced functional data. Works by Fraiman and Muniz (2001), Cuevas, Febrero and Fraiman (2007), and López-Pintado and Romo (2009, 2011) have extended the data-depth notion in robust multivariate data analysis to functional data, and defined depth-weighted robust estimators. Locantore et al. (1999), Gervini (2008), and Sinova, González-Rodríguez and Van Aelst (2018) developed robust estimators based on a fully

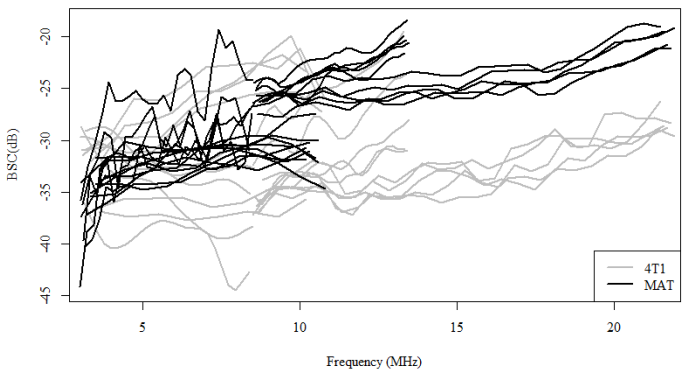

Figure 1. BSC data example by scanning two mammary tumors, 4T1 and MAT.

functional approach, investigating the robustness and asymptotic properties of the estimators. However, none of these methods can be applied directly to partially observed functional data.

We propose a new class of functional M-estimators by extending the class of M-estimators of Huber (2005) to include functional data. The approach considered differs from the recent functional location M-estimators developed by Sinova, González-Rodríguez and Van Aelst (2018), which impose a bounded M-estimator score function on the norm of the entire functional deviation from the location parameter function. Our approach builds a robust estimator in a cross-sectional manner to take advantage of all available curve data at each spatial location, while adapting to uneven patches in the response samples due to the partial observation of individual curves. The difference is discussed in detail in Section 2.2. Even with fully observed functional data, the cross-sectional approach considered here can adjust outlying patches in local spatial locations better than when applying a robust pseudo-norm to the entire function.

We employ a missing-data approach to deal with the partially observed functional data, extending the framework considered by Kraus (2015, 2019) and Park (2017) for functional mean and covariance estimation. We establish the asymptotic properties of the M-estimator, including the consistency and Gaussian process approximations. Furthermore, we adapt the results to develop a robust functional ANOVA test using L^2-norm statistics and a functional trend test for shape inferences, in each case implementing the inferences using a bootstrap approach. We use an influence analysis to investigate the robustness of the M-estimators to establish the bounded influence of outlying curves. Simulation studies and a

real-data analysis from a QUS study demonstrate the properties of the methods.

Section 2 defines the new class of functional M-estimators. The approach taken here has the advantage of being directly applicable to partially observed functional data, while bounding the influence of extreme curves. Section 3 establishes the theoretical properties of the M-estimators, including the consistency and the Gaussian process approximations of the estimates. These results are then used to develop the L^2-norm test and the functional trend test, and to support the bootstrap inferences for the practical implementation. The remaining sections include the simulations and real-data examples. Technical proofs and additional simulation results are included in the online Supplementary Material.

2. M-Estimator for Partially Observed Functional Data

2.1. Modeling assumptions

Let $X_1(t), \ldots, X_n(t)$ be functional samples observed over varying subsets, $\mathrm{S}_1, \ldots, \mathrm{S}_n$, of a compact set C. Similarly to Kraus (2015, 2019) and Park (2017), we consider the observed curves to be the result of filtering latent full information curves $Y_1(t), \ldots, Y_n(t)$ on C by independent indicator processes $\delta_1(t), \ldots, \delta_n(t)$, where

$$\delta_i(t) = \begin{cases} 1, & \text{if } Y_i(t) \text{ is observed;} \\ 0, & \text{if } Y_i(t) \text{ is unobserved;} \end{cases}$$

for $t \in \mathrm{C}$ and $i = 1, \ldots, n$. We make general assumptions about the nature of the filtering functions δ_i, and the modeling assumptions include the following:

M1: The stochastic processes, $(Y_i, \delta_i) := \{(Y_i(t), \delta_i(t)),\ t \in \mathrm{C}\}$, for $i = 1, \ldots, n$, are independent and identically distributed (i.i.d.) on $(\Omega, \mathscr{F}, \mathbb{P})$ and jointly $\mathscr{F}$-measurable.

M2: There are missing sampling variables $V_i = (V_{i1}, \ldots, V_{iK}) \in \mathcal{V}$, and there is a measurable missing scheme $h : \mathrm{C} \times \mathcal{V} \to \{0, 1\}$ such that (i) $V_1, \ldots, V_n$ are i.i.d. random variables with common distribution f, and (ii) $\delta_i(t) = h(t, V_i)$.

M3: $E(\delta_i(t)) = b(t)$, $t \in \mathrm{C}$, where $b(\cdot)$ is uniformly continuous and bounded away from zero, $\inf_{t \in \mathrm{C}} b(t) > 0$.

M4: $Y_i(t)$ and $\delta_i(t)$ are independent, for $i = 1, \ldots, n$.

An advantage of using robust estimators here is that it avoids restrictive moment assumptions on the process Y_i, thus enabling the analysis of partially observed processes from heavy-tailed or outlier-prone sampling distributions. Condition M2 is used to prove the uniform convergence of the average of the sample

indicator processes $\delta_i(t)$, for $i = 1, \ldots, n$, to $b(t)$. Kraus (2019) specified this sup-norm convergence of the averaged sample indicator processes as one of his conditions. Here, we impose only mild explicit conditions to derive the large-sample properties of the robust estimator; see Section 3. This condition is satisfied by a wide range of partial sampling structures, including the examples below.

Example 1 (Functional segments over random subintervals in C)**.** Define a random interval $S_i = [l_i, u_i] \subset \mathrm{C}$, where $l_i = \min(v_{i1}, v_{i2})$, $u_i = \max(v_{i1}, v_{i2})$, and v_{ij}, for $j = 1, 2$, are i.i.d. replicates of a random variable V supported on C. Then, Condition M2 is satisfied with $h(t, v_i) = \mathbf{1}(l_i \leqslant t \leqslant u_i)$. This framework can be extended to multiple random intervals per curve, $S_i = \cup_{k=1}^{K}[l_{ki}, u_{ki}]$, where l_{ki} and u_{ki}, for $k = 1, \ldots, K$, are i.i.d. from V. The latter corresponds to the example of fragmented functional data considered in Delaigle et al. (2021).

Example 2 (Functional segments over fixed subintervals in C)**.** Given a fixed set of intervals, $I_1, \ldots, I_m$, such that $\cup_{j=1}^{m} I_j = \mathrm{C}$, we can define $h(t, v_i) = \mathbf{1}(t \in I_{v_i})$, where v_i are i.i.d. from a uniform discrete random variable V on $\{1, \ldots, m\}$. The resulting scheme comprises a set of functional fragments observed over predetermined subintervals, as in the motivating example.

Example 3 (Dense functional snippets (Lin and Wang (2020)))**.** Define an interval $S_i = [l_i, l_i + d] \subset \mathrm{C} = [0, 1]$, where $0 < d < 1$ denotes the length of a subinterval, and l_i are i.i.d. copies of a random variable V_1 with support $[0, 1-d]$. Then, Condition M2 holds with $h(t, v_i) = \mathbf{1}(l_i \leqslant t \leqslant l_i + d)$. A further extension is to allow d_i, the subinterval length, to be drawn from a distribution supported on [0,1], and then to let $h(t, \boldsymbol{v}_i) = \mathbf{1}(l_i \leqslant t \leqslant l_i + d_i)$.

Condition M3 implies that the full range is covered by a sufficient portion of the data for sufficiently large sample sizes. For example, in the case of the random interval $S_i = [l_i, u_i]$, the support of V should have positive probabilities at both boundaries of C to ensure positive $b(t)$ bounded away from zero. Lastly, letting P denote the joint probability measure for (Y, δ), Condition M4 implies that $P = P_Y \cdot P_\delta$, where P_Y and P_δ denote the marginal probability measures for Y and δ on C, respectively. Along with Condition M3, this enables us to estimate the functional parameter of Y based on the partially observed functions X.

2.2. Marginal M-estimator

For partially observed samples $X_1(t), \ldots, X_n(t)$, we define the functional M-estimator $\hat{\theta}_n(t)$ under the cross-sectional approach by minimizing the criterion

marginally for all values of t in parallel,

$$\hat{\theta}_n(t) = \underset{h\in\mathbb{R}}{\operatorname{argmin}} \sum_{i=1}^{n} \delta_i(t)\rho\left(X_i(t) - h\right), \tag{2.1}$$

for $t \in \mathrm{C}$ satisfying $\sum_{i=1}^{n} \delta_i(t) > 0$, where $\rho(\cdot)$ is a real-valued loss function. Otherwise, the estimator is undefined. In other words, for fixed t, $\hat{\theta}_n(t)$ represents a pointwise M-estimator calculated based on the information observed at spatial location t. If we observe an undefined $\hat{\theta}_n(t)$ at a certain range in C under a finite sample size, it can be estimated using interpolation or smoothing methods when the smoothness and continuity of $\hat{\theta}_n(t)$ is assumed. In practice, discretized partial curves are observed on fine grids, and interpolation can be applied for the estimation.

Example 4 (Observation Weighted Mean Functions)**.** In the special case in which $\rho(x) = x^2$, the estimator $\hat{\theta}_n(t)$ reduces to the weighted sample mean function

$$\bar{X}_\delta(t) = \frac{\sum_{i=1}^{n} \delta_i(t)X_i(t)}{\sum_{i=1}^{n} \delta_i(t)}, \quad t \in \mathrm{C}.$$

Kraus (2019) showed the consistency and asymptotic Gaussianity of $\bar{X}_\delta$ for estimating the mean function of Y if Y and δ are independent, and assuming moment conditions on the underlying distribution of $X_i(\cdot)$.

Example 5 (Response function quantiles)**.** The spatially dependent median and quantile functions for the response function nonparametrically characterize the functional response distribution. The M-estimator framework enables consistent estimation of the quantile functions, $Q_\tau(\cdot)$, using sample estimates of the form

$$\hat{Q}_\tau(t) = \underset{h\in\mathbb{R}}{\operatorname{argmin}} \sum_{i=1}^{n} \delta_i(t)\rho_\tau\left(X_i(t) - h\right),$$

where for the quantile function of order $\tau \in (0, 1)$, we define $\rho_\tau(x) = x(\tau - I(x < 0))$, extending the univariate quantile estimator described, for example, in Koenker (2005).

For the general class of marginal M-estimators, the following conditions are employed for the loss function ρ to ensure the robustness of the estimator and to allow weaker distributional assumptions:

A1 $\rho : \mathbb{R} \to \mathbb{R}$ is a continuous even function and strictly increasing on $\mathbb{R}^+$, with $\rho(0) = 0$.

A2 ρ is at most linearly increasing in the tails; therefore, $|\rho(x_1) - \rho(x_2)| \leqslant L|x_1 - x_2|$, for some constant $L > 0$.

A3 ρ is differentiable and its derivative $\psi = \dot{\rho}$ is continuous.

A4 The second-order derivative $\dot{\psi} = \ddot{\rho}$ is almost everywhere differentiable and Lipschitz continuous; that is, $|\dot{\psi}(x_1) - \dot{\psi}(x_2)| \leqslant K|x_1 - x_2|$, for some constant $K > 0$.

To prove consistency, we require only conditions A1 and A2. Conditions A2 and A3 imply that ψ is bounded, which enables efficient estimation for heavy-tailed functional data, without assuming moment conditions. Weak convergence is established by assuming the additional conditions A3 and A4. More general conditions than A3 and A4 are needed, for example, to establish the weak convergence of the functional quantile estimators, although a smoothed approximate quantile M-estimator is covered by these conditions.

When the loss function ρ is differentiable, the functional M-estimator marginally solves the estimating equation

$$\frac{1}{n}\sum_{i=1}^{n} \delta_i(t)\psi(X_i(t) - \hat{\theta}_n(t)) = 0, \quad t \in \mathrm{C}. \tag{2.2}$$

The advantage of defining the M-estimator marginally is that it can use all available data to estimate the underlying functional location curve when the data are missing in a piecewise fashion. This is in contrast to the functional M-estimator of Sinova, González-Rodríguez and Van Aelst (2018), defined for complete data as

$$\hat{\theta}_n^{\mathbb{H}}(\cdot) = \underset{h \in \mathbb{H}}{\operatorname{argmin}} \frac{1}{n}\sum_{i=1}^{n} \rho\left(\|Y_i(\cdot) - h(\cdot)\|_{\mathbb{H}}\right),$$

where $\|\cdot\|_{\mathbb{H}}$ is the norm for the Hilbert space $\mathbb{H}$. The estimator $\hat{\theta}_n^{\mathbb{H}}$ is not directly applicable to partially observed functional data, because $Y_i(\cdot)$ is not fully observed. As an alternative approach, $\hat{\theta}_n^{\mathbb{H}}$ can be calculated from reconstructed complete curves by adopting a curve reconstruction method (Kneip and Liebl (2019)). We examine its estimation performance in our simulation studies (Section 5.1), and compare its results with those of the marginal M-estimator.

The marginal M-estimator developed here is applicable to partially observed functional data without intermediate steps, and provides consistent estimates of the functional location parameter under regularity conditions; see Section 3.

Furthermore, under the complete data framework with $\delta_i(t) = 1$, for $i = 1, \ldots, n$, $t \in \mathrm{C}$, the marginal approach offers an alternative wherein the robustness or outlier resistance of the estimator is controlled locally with the function, rather than on the overall norm of the function. This is further demonstrated in the simulation studies in Section 5.

2.3. Fisher consistency and invariance properties

In this section, we define the functional location parameter, a theoretical version of $\hat{\theta}_n(t)$, and investigate its properties. Given the joint probability measure P for (Y, δ), the functional location parameter $\theta(t)$ is defined as

$$\theta(t) = \operatorname*{argmin}_{h \in \mathbb{R}} E_P\big[\delta(t)\rho\big(Y(t) - h\big)\big], \quad t \in \mathrm{C}. \tag{2.3}$$

Under Conditions A1–A3, $\theta(t)$ also marginally satisfies

$$E_P[\delta(t)\psi\big(Y(t) - \theta(t)\big)] = 0, \quad t \in \mathrm{C}. \tag{2.4}$$

We show below that, under general conditions, $\hat{\theta}_n(\cdot)$ converges uniformly to $\theta(\cdot)$ as n increases and, furthermore, that $n^{-1/2}(\hat{\theta}_n(\cdot) - \theta(\cdot))$ is asymptotically a Gaussian process.

In the special case where Y has symmetric marginal distributions, the M-estimator estimates the same location as the mean and median functionals, assuming those exist. This generalizes the familiar Gaussian framework. We assume Θ represents a functional parameter set in a Riemann integrable $L^2(\mathrm{C})$ space, which includes piecewise continuous functions with a finite number of bounded jumps. In particular, we have the following result.

Proposition 1 (Symmetric marginal distributions)**.** *Under conditions of* M1 *and* A1–A3*, if the marginal density of $Y(t)$, for each $t \in \mathrm{C}$, is symmetric about a deterministic function $\alpha(t) \in \Theta$, that is, $Y(t) - \alpha(t)$ and $\alpha(t) - Y(t)$ have the same distribution, then $\theta(t) = \alpha(t)$.*

Proposition 1 implies that $\theta(t)$ represents the functional center when the marginal density of $Y(t)$ is symmetric for each $t \in \mathrm{C}$. Next, consider the special case where there is a shift location function such that subtracting the shift function from the function responses removes the spatial dependence of the marginal distribution of the response.

B1 [Shifted marginal stationarity] There exists a deterministic function $\alpha(t) \in \Theta$, such that the shifted process $Z(t) = Y(t) - \alpha(t)$ has constant marginal distributions $Z(t) \sim F_Z$, for $t \in C$.

We then obtain the following proposition under the generalized distribution of Y.

Proposition 2 (Shifted marginal stationarity)**.** *Under conditions* M1*,* M4*,* A1*,* A2*, and* B1 *with the translation function $\alpha(t)$, there exists a constant c depending only on F_Z, such that $\theta(t) = \alpha(t) + c$.*

Consequently, in this setting, $\theta(t)$ is a well-defined location parameter that inherits any smoothness or bounded jumps up to an additive constant depending on $\alpha(t)$.

Remark 1 (τ-quantile function)**.** In general, without assuming either symmetry or shifted marginal stationarity, if ρ is defined as in Example 5, then $\theta(t)$ represents a τ-quantile functional.

2.4. Robust functionals and influence functions

We define the weighted M-functional,

$$M(t, h, P) = E_P\big[\delta(t)\big\{\rho\big(Y(t) - h(t)\big) - \rho\big(Y(t)\big)\big\}\big], \tag{2.5}$$

and $\theta(t)$ equivalently marginally minimizes $M(t, h, P)$, for $t \in \mathrm{C}$; see Section 3.2 of Huber (2005) for the univariate case. Under Conditions A1–A2, the expectation in (2.5) exists for every probability measure P, and we assume the following general conditions:

D1 $\sup_{\theta\in\Theta} \sup_{t\in\mathrm{C}} |M(t, \theta, P_n) - M(t, \theta, P)| \xrightarrow{p} 0$, where P_n denotes a sequence of measures, converges weakly to a measure P.

D2 For every $\epsilon > 0$, $\inf_{\theta^*\in\Theta} \inf_{t\in\mathrm{C}}\{M(t, \theta^*, P) - M(t, \theta, P) : |\theta^*(t) - \theta(t)| > \epsilon\} > 0$.

Condition D1 requires the uniform convergence of the weighted M-functional over the parameter space $\theta \in \Theta$ and $t \in \mathrm{C}$. For example, if P_n denotes the empirical measure of $\{Y_i(t), \delta_i(t), t \in \mathrm{C}\}_{i=1}^n$, then for given $\theta \in \Theta$, uniform convergence over $t \in \mathrm{C}$ holds under Condition A2. In the univariate case, Chapter 5 of van der Vaart (2007) provides other possible assumptions to replace uniform convergence over the parameter space. Condition D2 implies that, for $t \in \mathrm{C}$, only $\theta(t)$ yields a minimum value of $M(t, h, P)$. Thus, it is a well-separated point of minimum at t. It holds under Condition A1, and the derivations of the influence function and the large-sample properties are based on the above conditions on the functional M.

The outlier sensitivity of an estimator is often measured using the influence function (Hampel (1974)). Doing so here, we consider contaminated curve distributions that may show atypical behavior in two ways: extreme or outlier fluctuations in the process Y, or outlying behavior in the missing process, such as dependence between Y and δ. For convenience, let $T(P)(t)$ denote the distribution-dependent functional corresponding to the parameter $\theta(t)$. We then consider the behavior of T for contaminated distributions of the form

$$P_\varepsilon = (1-\varepsilon)P + \varepsilon\Delta_{(Y^*,\delta^*)}, \tag{2.6}$$

where $\Delta_{(Y^*,\delta^*)}$ is a point mass distribution concentrated on (Y^*,δ^*).

We first establish the continuity of T uniformly over the contaminating distribution, a robustness property that holds when the score function ψ is bounded. Note that, by definition, P_ε converges weakly to P as $\varepsilon \to 0$.

Theorem 1 (Contamination Robustness)**.** *Conditions* M1*,* M4*,* A1–A2*, and* D1–D2 *imply*

$$\lim_{\varepsilon\downarrow 0} \sup_{t\in C,(Y^*,\delta^*)} |T(P_\varepsilon)(t) - T(P)| = 0.$$

Next, we extend the notion of the functional influence function, adapting the definition of Gervini (2008) as $IF_T(Y^*,\delta^*)(t) = \lim_{\varepsilon\downarrow 0}\varepsilon^{-1}\{T(P_\varepsilon)(t) - T(P)(t)\}$ if the limit exists, where P_ε is given in (2.6). The corresponding gross-error sensitivity (see Gervini (2008)) with the sup-norm metric is given by $\gamma_T^\infty = \sup\{\sup_{t\in \mathrm{C}} |IF_T(Y^*,\delta^*)(t)| : \text{any } (Y^*,\delta^*)\}$.

Theorem 2 (Influence Robustness)**.** *Under* M1*,* M4*, and* A1–A4*, if we assume uniform continuity of the functional* $T(P)(t)$ *and* $\inf_{t\in \mathrm{C}} |E_P\big[\delta(t)\dot{\psi}(X(t),\theta(t))\big]| > 0$*, then*

$$IF_T(Y^*,\delta^*)(t) = \frac{\delta^*(t)\psi(Y^*(t),\theta(t))}{-E_P\big[\delta(t)\dot{\psi}(X(t),\theta(t))\big]}, \quad t\in \mathrm{C}, \tag{2.7}$$

and the boundedness of ψ *implies* $\gamma_T^\infty < \infty$.

Hence, the boundedness of the marginal score function ψ implies a bounded effect of heavy-tailed variation or outliers in the process Y, or of the dependent missing process on the functional location parameter.

3. Large-Sample Approximations

3.1. Uniform consistency

In establishing the consistency and asymptotic Gaussian approximations for the class of functional M-estimators, a key step is to develop an entropy bound,

which is used to establish the sup-norm convergence for the averaged indicator processes $\delta(t)$. In particular, we establish the convergence of

$$W_n = \sup_{t\in \mathrm{C}} \left| n^{-1} \sum_{i=1}^{n} [\delta_i(t) - b(t)] \right|, \tag{3.1}$$

where, marginally for each $t \in \mathrm{C}$, $\delta_i(t) \sim \mathsf{Ber}(b(t))$ are i.i.d., and the functions $t \mapsto \delta_i(t)$ are sampled from a general class on C satisfying Condition M2.

To bound the size of W_n, we need to control the size of the function class, $\mathcal{G} := \{h(t,\cdot) : t \in \mathrm{C}\}$. Under the missing-data sampling scheme in Condition M2, given a missing scheme $h : \mathrm{C} \times \mathcal{V} \to \{0,1\}$, for any $g \in \mathcal{G}$, there is a $t \in \mathrm{C}$ such that $g(v) = h(t,v)$, for $v \in \mathcal{V}$. Let $H : \mathcal{V} \to \{0,1\}$ be a measurable envelope for $\mathcal{G}$; that is, $H(v) \geqslant \sup_{g\in\mathcal{G}} g(v) = \sup_{t\in \mathrm{C}} h(t,v)$, for all $v \in \mathcal{V}$. Define the uniform entropy integral as in van der Vaart and Wellner (1996),

$$J(\delta, \mathcal{G}, H) = \int_0^\delta \sup_Q \sqrt{1 + \log(N(\mathcal{G}, L^2(Q), \varepsilon \|H\|_{Q,2}))}\, \mathrm{d}\varepsilon, \tag{3.2}$$

where the supremum runs over all finitely discrete probability measures on $(\mathcal{V}, \mathcal{B}(\mathcal{V}))$, and $N(\mathcal{G}, L^2(Q), \varepsilon)$ is the ε-covering number of $\mathcal{G}$ under the metric induced by $L^2(Q)$.

Lemma 1 (Expectation bound on W_n)**.** *If Condition* M2 *holds and* $J(1, \mathcal{G}, H) < \infty$*, then there is a universal constant* $C > 0$ *such that*

$$E[W_n] \leqslant C \frac{J(1,\mathcal{G},H)}{\sqrt{n}} \max\left\{1, \frac{J(1,\mathcal{G},H)}{\sqrt{n}}\right\}.$$

Corollary 1. *(i) If* $\mathcal{G}$ *is a finite class of functions (i.e.,* $|\mathcal{G}| < \infty$*), then the Hoeffding inequality and union bound imply that* $J(1,\mathcal{G},H) \lesssim \sqrt{\log|\mathcal{G}|}$*, and Lemma* 1 *yields*

$$E[W_n] \lesssim \sqrt{\frac{\log|\mathcal{G}|}{n}} \max\left\{1, \sqrt{\frac{\log|\mathcal{G}|}{n}}\right\} \lesssim \frac{1}{\sqrt{n}}.$$

(ii) If $\mathcal{G}$ *is a VC-type class; that is, there are constants* $A, v > 0$ *such that*

$$\sup_Q N(\mathcal{G}, L^2(Q), \varepsilon\|H\|_{Q,2}) \leqslant \left(\frac{A}{\varepsilon}\right)^v \quad \textit{for all } \varepsilon \in (0,1],$$

then

$$J(\delta,\mathcal{G},H) \lesssim \delta\sqrt{v \log\left(\frac{A}{\delta}\right)} \quad \textit{for all } \delta \in (0,1].$$

Then, Lemma 1 *implies that there is a constant* $C(v, A) > 0$ *depending only on* v *and* A *such that*

$$E[W_n] \leqslant \frac{C(v, A)}{\sqrt{n}}.$$

In either case (i) or (ii), we get the uniform rate of convergence $n^{-1/2}$ *for estimating* $b(t)$ *using* $n^{-1}\sum_{i=1}^{n}\delta_i(t)$*, that is,*

$$E[W_n] = O(n^{-1/2}).$$

Remark 2. Example 2 corresponds to case (i), a finite class of functions $\mathcal{G}$, and Example 1 and Example 3 correspond to case (ii), a VC-type class of $\mathcal{G}$. Thus, $E[W_n] = O(n^{-1/2})$ holds for all of the examples presented above.

Based on Lemma 1, the following result establishes the uniform consistency of the M-estimator.

Theorem 3 (Uniform Consistency)**.** *Under conditions* M1–M4*,* A1*,* A2*, and* D1–D2*,* $\hat{\theta}_n(t)$ *converges to* $\theta(t)$ *uniformly over* $t \in \mathrm{C}$*.*

Remark 3. As a special case, we obtain the uniform consistency of the functional quantile estimators of Example 5 for partially observed functional data.

3.2. Functional central limit theorem

We first derive a general functional central limit theorem for the functional sample mean under the missing-data framework, previously studied by Park (2017) and Kraus (2019). Then we adapt the result to obtain an asymptotic Gaussian process approximation for the proposed M-estimators. Let C be a compact subset in a general metric space equipped with a metric d, and let $V(t)$, for $t \in \mathrm{C}$, be a mean-square continuous process defined on a probability space $(\Omega, \mathscr{F}, P)$. We suppose that $V(t, \cdot)$ is measurable for each $t \in \mathrm{C}$, and $V(\cdot, \omega)$ is continuous for each $\omega \in \Omega$. We consider the second-order stationary process V with mean zero and covariance function γ (i.e., $\gamma(s, t) = \mathrm{Cov}(V(s), V(t)), s, t \in \mathrm{C}$), denoted by $V \sim \mathrm{SP}(0, \gamma)$. We define the process $Z_n(t)$ as

$$Z_n(t) = \frac{\sqrt{n}\sum_{i=1}^{n}\delta_i(t)V_i(t)}{\sum_{j=1}^{n}\delta_j(t)}, \quad t \in \mathrm{C}.$$

The following result is adapted from the functional central limit theorem of Kraus (2019), which specifies a key step, that is, the sup-norm convergence of the averaged sample indicator processes in (3.1), as one of technical conditions. However, we establish the asymptotic Gaussianity using Lemma 1 under the more explicit

and practical Condition M2.

Theorem 4 (Functional Central Limit Theorem for Partially Observed Data)**.** *Let $V_1, \ldots, V_n$ be i.i.d. samples of the second-order stationary process V. Under M2–M4 with replacement of Y by V, we have*

$$Z_n \rightsquigarrow GP(0, \vartheta),$$

where $\vartheta(s,t) = \gamma(s,t)v(s,t)b(s)^{-1}b(t)^{-1}$, for $s,t \in \mathrm{C}$, and $v(s,t) = E_{P_\delta}[\delta(s)\delta(t)]$.

3.3. Gaussian process approximation of M-estimator

Building on the uniform consistency of the marginal M-estimators and the functional central limit theorem, the results of this section establish that robust M-estimators have Gaussian process large-sample approximations under weaker distributional conditions than the functional sample mean. For notational simplicity, we denote $\psi(x-\theta)$ by $\psi(x,\theta)$.

Theorem 5 (Asymptotic Normality of M-estimator)**.** *Under conditions* M1–M4, A1–A4*, and* D1–D2*, and if $E_{P_Y}[\dot{\psi}(Y(t),\theta(t))]$ exists and is nonsingular almost everywhere on* C*,*

$$\sqrt{n}\big(\hat{\theta}_n(t) - \theta(t)\big) \rightsquigarrow GP(0, \xi),$$

$\xi(s,t) = \varphi(s,t)E_{P_Y}[\dot{\psi}(Y(s),\theta(s))]^{-1}E_{P_Y}[\dot{\psi}(Y(t),\theta(t))]^{-1}$, *where* $\varphi(s,t) = \mathrm{Cov}\big\{\psi(Y(t),\theta(t)),\, \psi(Y(s),\theta(s))\big\}v(s,t)b(s)^{-1}b(t)^{-1}$ *with* $v(s,t) = E_{P_\delta}[\delta(s)\delta(t)]$.

3.4. Robust inferences

The uniform Gaussian approximation provides a tool for developing (i) a robust functional ANOVA (fANOVA) test for the equality of the location parameters in several populations, and (ii) a trend test to detect functional trends in observed curves. Specifically, we follow up the fANOVA-type test by performing a trend test to determine whether there is a specific systematic trend over t in the group differences, for example, a constant or a linear trend in the intergroup differences across the functional domain.

3.4.1. L_2-test on location functions

Statistical tests on robust location functions can be developed using the preceding asymptotic results. An important example is testing the equality of location functions in several populations using the null hypothesis $H_0 : \theta_1(t) = \cdots = \theta_k(t)$, with $\theta_g(t)$, for $g = 1, \ldots, k$, representing the functional location parameter (2.3) of population g.

Under fully observed functional data structures, Shen and Faraway (2004),

Cuevas, Febrero and Fraiman (2004), and Zhang and Liang (2014) developed robust fANOVA tests of this type. Under partially observed data structures, Kraus (2019) developed a mean-based fANOVA test on the functional population mean. Our applications motivate robust testing for partially observed data, combining the two issues investigated by the previous authors.

Let $X_{g1}(t), \ldots, X_{gn_g(t)}$, for $g = 1, \ldots, k$, denote partially observed functional curves for group g, and assume $\mathrm{C} = [0, 1]$, without loss of generality. Extending the fANOVA under balanced data by Shen and Faraway (2004), we derive an L^2-norm-based test for testing the equality of robust location functions with a test statistic, $T_n = \int_{t\in\mathrm{C}} SSR_n(t)dt/trace(\hat{\xi})$, where $n = \sum_{g=1}^{k} n_g$, $SSR_n(s) = \sum_{g=1}^{k} n_g[\hat{\theta}_g(t) - \bar{\theta}_{\cdot}(t)]^2$ with the functional M-estimator $\hat{\theta}_g$ for group g, and the grand mean $\bar{\theta}_{\cdot}(t) = \sum_{g=1}^{k} n_g\hat{\theta}_g/n$. Here, $\hat{\xi}$ represents a consistent estimator of the asymptotic covariance of the functional M-estimator in Theorem 5.

Corollary 2. *Assume that $n_g \to \infty$, $n_g/n = a_g > 0$, for $g = 1, \ldots, k$, $trace(\xi) < \infty$, and $\xi(t, t) > 0$, for any $t \in \mathrm{C}$, where $\xi(s, t)$ denotes the asymptotic covariance function of the functional M-estimator derived in Theorem 5. Then, under the null hypothesis of equal location functions and under the same conditions of Theorem 5, we have $T_n \xrightarrow{d} T_0$, where*

$$T_0 \overset{d}{=} \sum_{r=1}^{\infty} \lambda_r A_r, \quad A_r \overset{i.i.d.}{\sim} \chi^2_{k-1},$$

where $\lambda_r = \kappa_r/trace(\xi)$, for $r = 1, \ldots, \infty$, are the scaled eigenvalues with κ_r, for $r = 1, \ldots, \infty$, the decreasing-ordered eigenvalues of the covariance function $\xi(s, t)$.

In practice, we calculate the test statistic by plugging in the estimated covariance function from a bootstrap procedure, avoiding the complications associated with estimating the asymptotic covariance function and its eigenvalues; see Section 4.

3.4.2. Functional trend test

The Gaussian approximation for the M-estimator functionals enables a corresponding approximation for inference on trends or probes, even if the data are only partially observed, as described above.

Corollary 3. *Under the same conditions of Theorem 5, let $c = \langle\theta(\cdot), \phi(\cdot)\rangle$, where $\phi(\cdot)$ is a fixed Riemann integrable L^2-function on C, and $\langle\cdot, \cdot\rangle$ represents the inner product of the fixed functions over C, $\langle f, g\rangle = \int_{\mathrm{C}} f(t)g(t)dt$. Define $\zeta_n =$*

$\sqrt{n}(\langle\hat{\theta}_n(\cdot), \phi(\cdot)\rangle - c)$. *Then, under* $tr(\xi) < \infty$,

$$\zeta_n \rightsquigarrow N(0, \tau^2),$$

where $\tau^2 = \int\int_{\mathrm{C}} \phi(s)\xi(s,t)\phi(t)dsdt$.

In the context of balanced, complete functional data, Ramsay and Silverman (2005) called parameters such as $\langle\theta(\cdot), \phi(\cdot)\rangle$ probes, generalizing the concept of a contrast, and discussed asymptotic confidence intervals. Here, we derive such intervals for the partially observed data framework. These intervals provide information on whether or not a trend of interest is present. We can also perform it as a follow-up to the robust fANOVA test to detect systematically distinct behaviors among the functional location parameters. Application examples are presented in Sections 5 and 6.

4. Bootstrap Inference

This section provides a bootstrap approach to perform the trend test. The key is to jointly resample the Y and δ processes simultaneously. Under the assumption of missing completely at random in Condition M4, it might seem ideal to generate partially observed pseudo samples by resampling Y and δ over C independently. However, this is impossible in practice, owing to the missing information on the unobserved segments of $X_i(t)$, for $i = 1, \ldots, n$. Instead, we generate pseudo observations by resampling partially observed curves from the data. The resulting bootstrap recovers the missing-at-random dependence structure asymptotically.

Suppose that $\mathbf{Y}(t) = [Y_1(t), \ldots, Y_n(t)]^T$ and $\boldsymbol{\delta}(t) = [\delta_1(t), \ldots, \delta_n(t)]^T$ are the observed Y and δ information, respectively. Let $\mathbf{U} = [\mathbf{U}_1, \ldots, \mathbf{U}_n]^T$ denote an $(n \times n)$ matrix, where $\mathbf{U}_i \sim$ multinomial $(1, \mathrm{rep}(1/n, n))$. Here, $\mathbf{U}_i$ tells which functional curve is chosen for the ith bootstrap sample. Then, the bootstrapped functional vector is generated by $\mathbf{Y}^*(t) = [Y_1^*(t), \ldots, Y_n^*(t)]^T = \mathbf{U}\mathbf{Y}(t)$ and $\boldsymbol{\delta}^*(t) = [\delta_1^*(t), \ldots, \delta_n^*(t)]^T = \mathbf{U}\boldsymbol{\delta}(t)$. The joint resampling uses the same $\mathbf{U}_i$ to generate Y_i^* and δ_i^*, which corresponds to X_i^*, eventually, where $\mathbf{X}^*(t) = [X_1^*(t), \ldots, X_n^*(t)]^T = \mathbf{U}\mathbf{X}(t)$. The bootstrap algorithm for the robust L^2-test is as follows.

Remark 4. In the procedure, we plug in the bootstrap variance $\hat{\xi}_n^*(t,t)$ to calculate T_n^* and use this as a test statistic. Under the joint bootstrap, for $t \in \mathrm{C}$, the conditional second moments of the bootstrapped samples $\mathrm{Cov}(\rho(Y_i^*(t))|\mathbf{Y}(t))$ and $\mathrm{Cov}(\delta_i^*(t)|\boldsymbol{\delta}(t))$ converge to $\mathrm{Cov}(\rho(Y_i(t)))$ and $\mathrm{Cov}(\delta_i(t))$, respectively, as

Algorithm 1. Bootstrap approximation for testing the equality location parameters

Calculate $\hat{\theta}_g(t)$ from observed samples $\{X_{gi}(t)\}_{i=1}^{n_g}$, $g = 1, \ldots, k$ and calculate $\bar{\theta}_\cdot(t) = \sum_{g=1}^{k} n_g \hat{\theta}_g(t)/n$
for g=1,...,k **do**
  for b=1,...,B **do**
    Sample pseudo partially observed curves $\{X_{gi}^*(t)\}_{i=1}^{n_g}$ with replacement from $\{X_{gi}(t)\}_{i=1}^{n_g}$
    Calculate $\hat{\theta}_g^{*(b)}(t)$ from the pseudo observations $\{X_{gi}^*(t)\}_{i=1}^{n_g}$
  end for
end for
Based on $\hat{\theta}_g^{*(b)}(t)$, $g = 1, \ldots, k$, $b = 1, \ldots, B$, for $t \in \mathrm{C} = [0, 1]$, calculate $\hat{\xi}_n^*(t,t) = (n/kB)\sum_{g=1}^{k}\sum_{b=1}^{B}(\hat{\theta}_g^{*(b)}(t) - \bar{\theta}_g^{*(\cdot)}(t))^2$, where $\bar{\theta}_g^{*(\cdot)}(t) = (1/B)\sum_{b=1}^{B}\hat{\theta}_g^{*(b)}(t)$
Calculate the test statistic T_n^* by replacing $trace(\hat{\xi})$ with $\int_{\mathrm{C}} \hat{\xi}_n^*(t,t)dt$
For $t, s \in \mathrm{C}$, calculate $\hat{\xi}_n^*(t,s) = (n/kB)\sum_{g=1}^{k}\sum_{b=1}^{B}(\hat{\theta}_g^{*(b)}(t) - \bar{\theta}_g^{*(\cdot)}(t))(\hat{\theta}_g^{*(b)}(s) - \bar{\theta}_g^{*(\cdot)}(s))$ and calculate λ_r^*, $r = 1, \ldots, R$, in Corollary 2 based on the bootstrapped eigenvalues κ_r^*, $r = 1, \ldots, R$ of $\hat{\xi}_n^*(t,s)$ and the bootstrapped trace $\int_{\mathrm{C}} \hat{\xi}_n^*(t,t)dt$
Approximate the p-value of the location test based on T_n and $T_0^* = \sum_{r=1}^{R}\lambda_r^*\chi_{k-1}^2$.

$B, n \to \infty$. In addition, $Y_i^*(t)$ and $\delta_i^*(t)$ are asymptotically uncorrelated, with correlation approaching zero as $B, n \to \infty$. Thus, T_n^* converges to T_n, by Slutsky's theorem. In the same manner, T_0^* converges to T_0, which justifies using the bootstrapped procedure for the robust inferential test. The simulations of Section 5 provide empirical confirmation of the accuracy of the bootstrap inference.

Algorithm 2. Bootstrap confidence interval for functional trend

for b=1,...,B **do**
  Sample pseudo partially observed curves $\{X_i^*(t)\}_{i=1}^{n}$ with replacement from $\{X_i(t)\}_{i=1}^{n}$
  Calculate the M-estimator $\hat{\theta}_n^{*(b)}(t)$ from the pseudo observations $\{X_i^*(t)\}_{i=1}^{n}$
  Project $\hat{\theta}_n^{*(b)}(t)$ to the direction of interest $\phi(t)$, and calculate $\hat{c}^{*(b)}$.
end for
Based on $\hat{c}^{*(b)}$, $b = 1, \ldots, B$, calculate the $(1-\alpha)100\%$ bootstrap confidence interval using the $100\alpha/2$th and $100(1-\alpha/2)$th percentiles of the bootstrap distribution $\hat{\theta}_n^*(t)$.

If the bootstrapped confidence interval of the projection coefficient excludes zero, the test for the trend is statistically significant, otherwise it is not.

5. Simulation Study

We present two simulation studies. In Section 5.1, the finite-sample behavior of the M-estimator is examined by comparing the estimation accuracy of our marginal approach with that of existing functional approaches (i) under complete functional data without missing data and (ii) under partially observed functional data. In particular, for the incomplete case, the functional estimators are applied to reconstructed curves. In Section 5.2, we investigate the performance of the bootstrap inference for the trend test with different structures of partially observed functional data.

5.1. Simulation study: estimation accuracy

We first generate functional data under six scenarios to investigate the estimation accuracy and perform the comparative study. In total, 80 independent curves following $X(t) = \mu(t) + \sigma(t)\epsilon(t)$, for $t \in [0,1]$, are generated for each scenario by varying the assumptions on $\sigma(t)$ or $\epsilon(t)$, or by adding artificial contamination under the fixed smooth location function $\mu(t)$. Here, $\sigma(t)$ represents the magnitude of the noise, and $\epsilon(t)$ denotes the error process. The goal is to estimate $\mu(t)$ under various settings. Examples of the simulated data under the different scenarios are shown in the Supplementary Material.

For models 1–3, we generate data using the Gaussian, t_3, and Cauchy processes, respectively, assumed on $\epsilon(t)$ with the constant $\sigma(t) = 2$ over $[0,1]$. An exponential spatial correlation structure is assumed on the noise process, where $\mathrm{Cor}(\epsilon(t_1), \epsilon(t_2)) = \exp(-|t_1 - t_2|/d)$. Here, the range parameter d determines the spatial dependence within a curve. We use the value 0.3, but studies with other values show similar results. All curves are simulated at 100 equidistant points in $[0,1]$.

Model 4 considers data with t_3 independent errors with random scales, where $\sigma(t)$, for $t \in [0,1]$, is generated from $N(2, 10^2)$. Models 5 and 6 generate partially contaminated data, where $X(t) = \mu(t) + \sigma(t)\epsilon(t)$, for $t \in [0, 0.2) \cup (0.4, 1]$, under a Gaussian process with a constant scale, as in models 1–3, and $X(t) = \mu(t) + \zeta(t)$, for $t \in [0.2, 0.4]$. For model 5, we consider a Cauchy distributed independent noise process $\zeta(t)$ with a unit scale and, for model 6, we assume a Cauchy distributed contamination under an exponential spatial correlation with a unit scale.

To calculate the proposed M-estimator, we use the Huber loss function and estimate the location parameter using constant or scaled robust tuning parameters. The former uses a constant tuning parameter, say c, on $t \in [0,1]$, and the latter applies varying robust tuning parameters, $c(t) = r * \mathrm{MAD}(t)$, for $t \in [0,1]$,

Table 1. Relative median ISE of the mean with respect to that of the proposed M-estimator under an unscaled robust tuning parameter for models 1–6.

model1	model2	model3	model4	model5	model6
0.78	1.50	79.9	1.64	98.6	16.4

where r controls the overall robustness, and $\mathrm{MAD}(t)$ indicates the marginal median absolute deviation (MAD) of the response function values at t. In the simulation, we choose c as 0.8 to make the marginal estimates close to the marginal median values. For the scaled approach, we set $r = 0.2$ to make a fair comparison, considering that $\sigma = 2$.

For the comparative study, we consider two competitors, the functional M-estimator developed by Sinova, González-Rodríguez and Van Aelst (2018), and the functional median proposed by Gervini (2008). Under complete data, they can be applied directly to the generated data using the same Huber loss function, with a robust tuning parameter of 0.8 for the functional M-estimator. To evaluate the performance, we calculate the integrated square error (ISE), $\mathrm{ISE}(\hat{\mu}) = \sum_{t=1}^{100}[\hat{\mu}(s_t) - \mu(s_t)]^2/100$, over 500 repetitions.

5.1.1. Estimation accuracy for complete functional data

The estimation accuracies under complete data are displayed in Figure 2. Here, the two gray boxes represent the results from the marginal M-estimators, with "M" and "Sc.M" denoting an M-estimator with a constant tuning parameter and with a scaled tuning parameter, respectively. The "Func.M" and "Med." indicate the functional M-estimator of Sinova, González-Rodríguez and Van Aelst (2018) and the median of Gervini (2008), respectively. The results from the sample mean are excluded from the visualization owing to exceedingly large ISEs in all scenarios, except for the Gaussian case. Instead, we present the relative ratios of the median ISE of the mean with respect to that of the proposed M-estimator in Table 1. Similar relative ratios are found with respect to other robust estimators; thus, they are not included in the paper.

Under the Gaussian model (model 1), robust estimators achieve similar estimation accuracy to that of the functional mean. However, under the heavy-tailed or contaminated scenario, the sample mean with large ratios fails. Now, for the comparison between robust estimators in Figure 2, we see that under (a) Gaussian and (b) t_3 errors, all four estimators achieve similar estimation accuracy. Under (c) Cauchy noise, existing functional approaches perform slightly better. However, this is not surprising because the discretized curves are generated under a multivariate Cauchy distribution, where the discretized functional approach is

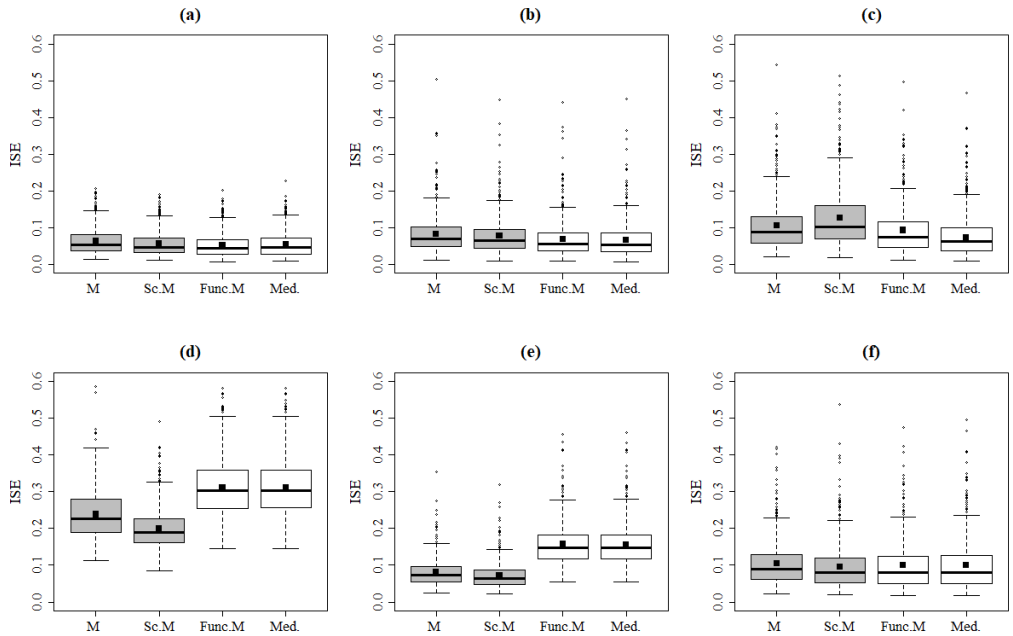

Figure 2. Box plots of the ISE over 500 replications from the marginal M-estimator (M), marginal scaled M-estimator (Sc.M), functional M-estimator (Func.M), and functional median (Med.) under completely observed data from (a) Gaussian, (b) t_3, (c) Cauchy, (d) independent t_3 with random scales, Gaussian partially contaminated by (e) Cauchy independent noise, and by (f) Cauchy processes. Square dots represent mean values.

meant to be optimal. Nevertheless, the marginal approach still achieves comparable performance. Under the constant σ in models 1–3, the M-estimator with a constant tuning parameter seems slightly more stable than the estimator with a scaled parameter, but it does not seem to be a significant difference. The plot (d) displays the estimation error from the marginally independent noise, and we see that the two competitors fall behind the marginal approach in terms of estimation accuracy. We observe the same pattern under the model of partial contamination by marginally independent Cauchy noise in (e). Under the spatially correlated contamination in (f), we again see similar performance among the four methods. In contrast to the comparable estimation errors among the estimators under models 1–3, the distinction in performance is apparent under models 4 and 5. Furthermore, for the model of random noise scale, the M-estimator with a scaled tuning parameter slightly outperforms that with an unscaled parameter. In summary, our proposed marginal approach provides comparable or superior performance in terms of estimation accuracy under various scenarios, as

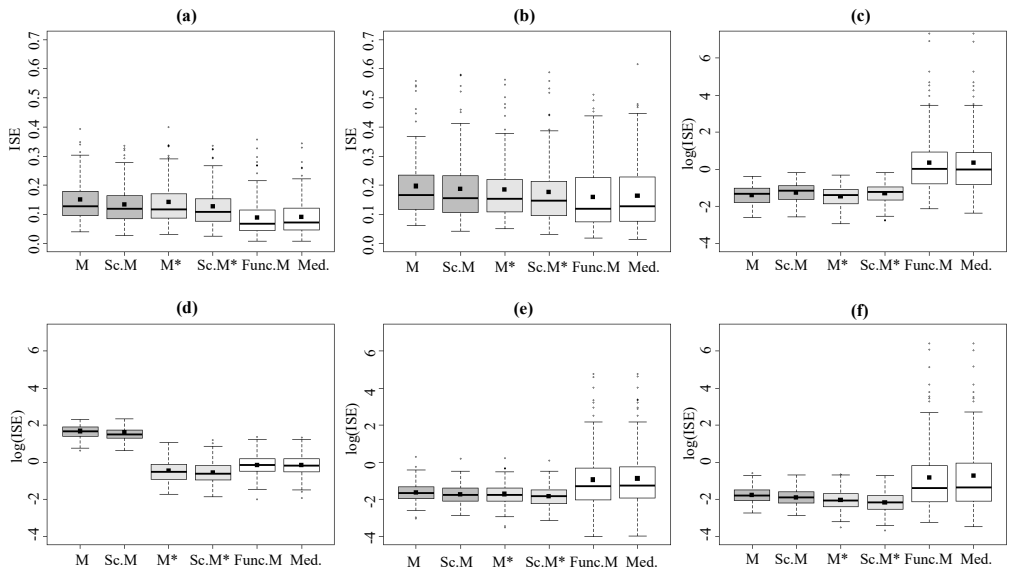

Figure 3. Box plots of the ISE or log transformed ISE over 500 replications from the marginal M-estimator (M), marginal scaled M-estimator (Sc.M), marginal M-estimator under pre-smoothed curves (M*), marginal scaled M-estimator under pre-smoothed curves (Sc.M*), functional M-estimator (Func.M), and functional median (Med.) under partially observed data over random intervals from (a) Gaussian, (b) t_3, (c) Cauchy, (d) independent noise t_3 with random scales, and Gaussian partially contaminated by (e) Cauchy independent noise and by (f) Cauchy processes. Square dots represent mean values.

compared with existing methods.

5.1.2. Estimation accuracy for partially observed functional data

To investigate the estimation performance under incomplete data with missing segments, we apply two sampling frameworks to each generated set of curves: (i) a partial sampling process under random interval sampling (Example 1), where, v_{1i} and v_{2i} are generated from Beta(0.3,0,3) and $\delta_i(t) = \mathbf{1}(\min(v_{i1}, v_{i2}) \leqslant t \leqslant \max(v_{i1}, v_{i2}))$, for $i = 1, \ldots, 80$; and (ii) a random missing process under a fixed number of intervals (Example 2), I_j, for $j = 1, 2, 3$, $t \in [0, 1]$ satisfying $\cup_{j=1}^{3} I_j = [0, 1]$. We randomly assign one of the three to the ith curve, for $i = 1, \ldots, 80$. Note that we perform the analysis for $t \in [\varepsilon, 1 - \varepsilon]$ with $\varepsilon = 0.01$ and a scaled beta distribution to ensure the Condition M3, as discussed in Section 2.1.

Because the two functional competitors are not directly applicable to incomplete curves, we apply them to reconstructed curves by adopting the consistent

reconstruction technique of Kneip and Liebl (2019), as described in Section 2.2. Note that Kneip and Liebl (2019) employ FPCA to reconstruct unobserved segments, and consistent estimation is fulfilled by nonparametric smoothing, for example, a local linear smoother, on observed fragments. Thus, for a fair comparison, we apply the proposed M-estimators to pre-smoothed curves, especially under robust nonparametric smoothing (Fried, Einbeck and Gather (2007)). For each simulated set, the same bandwidth is used in both local linear smoothing and robust smoothing.

Figure 3 presents the estimation accuracy from the partially sampled data under random intervals with two more estimators, "M*" and "Sc.M*" calculated based on pre-smoothed curves via robust smoothing. The results from the partial sampling process under a fixed number of intervals are provided in the Supplementary Material with very similar findings to those of the random interval cases.

The box plots from (a) Gaussian and (b) $t(3)$ show similar results to those observed from the regular structured data, with slightly lower errors from the existing functional approach under the Gaussian model. We also observe similar errors of the marginal M-estimator under raw and robust pre-smoothed data. Next, we find interesting results under other distributional settings, and box plots are generated in log transformed ISE, owing to some severely large ISEs calculated from the competitors. Under the (c) Cauchy distribution or Gaussian data with Cauchy contaminations, either (e) Cauchy fragmented or (f) Cauchy marginal noise, we examine the failure of Func.M and Med. based on reconstructed curves, with very large ISEs ranging from 0.03 to 600. This is due to the unstable reconstruction from heavy-tailed partial observations, demonstrating the failure of the regular smoothing technique and FPCA for data with potential outliers. Under these settings, the proposed M-estimator performs significantly better, with similar performance using smoothed or raw data. In contrast, under (d) marginal $t(3)$ heterogeneous scales of noise, applying the functional approach for reconstructed curves works well through smoothing, because it alleviates marginal peaks by using information from nearby neighbors. Nonetheless, the marginal M-estimator based on robust pre-smoothed data outperforms the functional approach based on the ISE.

5.2. Simulation study: robust inference

In the second simulation study, we investigate the validity of the bootstrap-based inference in a functional trend test. The coverage probability and the length of the bootstrapped confidence intervals are investigated under five models and five sampling structures, including a partially observed framework. For

the first three models, we borrow models 1–3 from the first simulation by assuming a Gaussian, t_3, and Cauchy process, respectively, in the error process, but with $\mu(t) = \phi_0(t) + 2\phi_1(t) + 0.5\phi_2(t)$, for $t \in [0, 1]$, where $\phi_0(t)$, $\phi_1(t)$, and $\phi_2(t)$ representing orthonormal constant, linear, and quadratic basis functions, respectively. The other two models follow the contamination scenario in the first simulation, namely, Gaussian curves contaminated by a Cauchy process on [0,0.3] and [0.7,1], respectively, with a constant noise scale. For each scenario, 80 curves are generated at 100 equidistant points over $[0, 1]$.

We consider two further partial sampling frameworks, in addition to the two missing procedures considered in the previous simulation: (iii) a dense functional snippet (Example 3), with length of subintervals set as $d = 0.2$ and $\delta_i(t) = \mathbf{1}(l_i \leqslant t \leqslant l_i + d)$, where l_i is drawn from Unif(0, 0.8) following Lin and Wang (2020); and (iv) fragmented functional data on sparse and irregular grid points. Although condition M2 requires careful extension, not covered here, to include a sparse irregular sampling scheme, we examine the performance of the robust inference for potential study of the extension. The data-generation process and results under the sparse design are provided in the Supplementary Material.

We calculate bootstrapped 95% confidence intervals of the projection coefficients to constant, linear, and quadratic functions under the M-estimator, scaled M-estimator, and the sample mean function with 800 bootstrapped samples. Then, the coverage probabilities are estimated from 500 repetitions based on the number of true coefficients in the confidence intervals. We also calculate the median length of the intervals.

Figure 4 displays the results from the projected coefficients to the quadratic trend. The three plots in the first row illustrate the empirical coverage probabilities under regular, random interval, and dense functional snippet cases, and the results from the other partial structures are provided in the Supplementary Material. In (a) and (b), we observe that the coverage probabilities from the robust estimators are always around 95%, but the overall probabilities from the mean tend to be less than the desired 95%. In particular, under the Cauchy or contaminated model, an inference from the functional mean may fail to detect the true quadratic trend. In (c), we see coverage probabilities lower than the desired 95%, but still around 90% from our proposed inference, owing to the relatively small effective sample size at each t with the length of subintervals set as $d = 0.2$ in $[0, 1]$. Interestingly, a coverage probability close to 95% is observed through the mean inference on the Cauchy data, but this is not surprising, considering the wide length of the bootstrapped confidence intervals in (f). The plots of (d), (e), and (f) visualize the median length of the confidence interval from each estima-

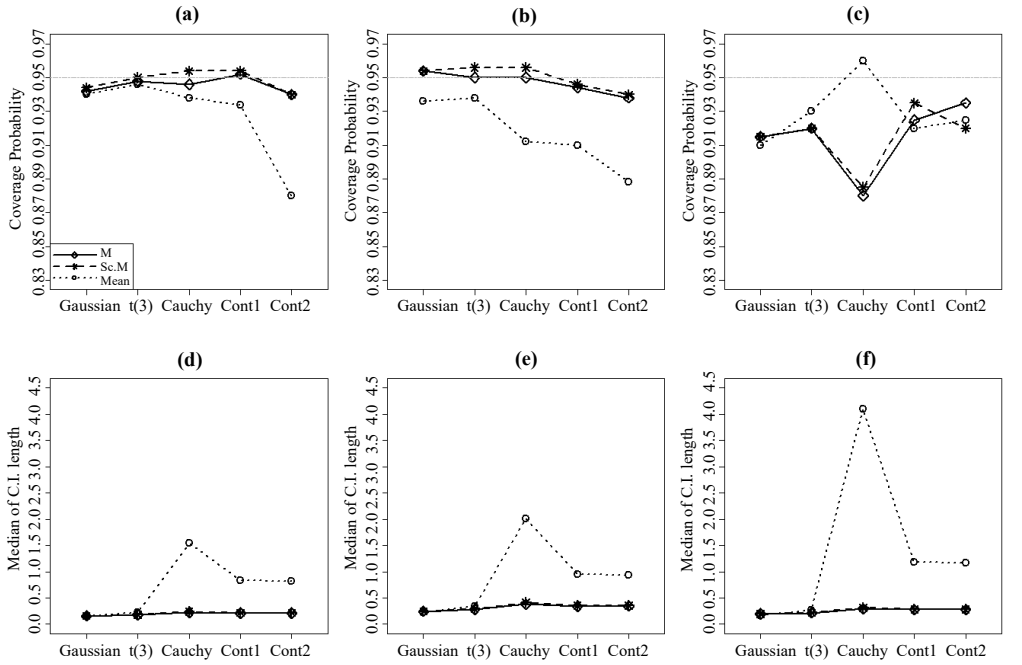

Figure 4. Coverage probabilities of bootstrapped confidence intervals of projection coefficients to quadratic function under Gaussian, t_3, Cauchy, and two contaminated data from M-estimator (M), scaled M-estimator (Sc.M), and Mean functions over 500 repetitions; (a) regular structure, (b) partially observed structure under random intervals, and (c) dense functional snippets. Median length of bootstrapped confidence intervals of projection coefficient under (d) regular structure, partially observed structure under (e) random intervals, and under (f) dense functional snippet.

tor. The inference from the functional mean seems less informative and unstable, with the wide length of the interval. However, the results from the proposed M-estimator seem stable, regardless of the distribution assumptions and missing structures.

6. Example: Quantitative Ultrasound Analysis

We illustrate the estimation and inference of the M-estimator using QUS data. Wirtzfeld et al. (2015) presented data and results from diagnostic ultrasound studies using multiple transducers to scan mammary tumors (4T1) and benign fibrous masses (MAT) in rats and mice. In this experiment, five transducers are used for a noninvasive scan of each animal. Specifically, two transducers, 9L4 and 18L6, from Siemens, cover 3–10.8 MHz, L14-5 from Ultrasonix uses

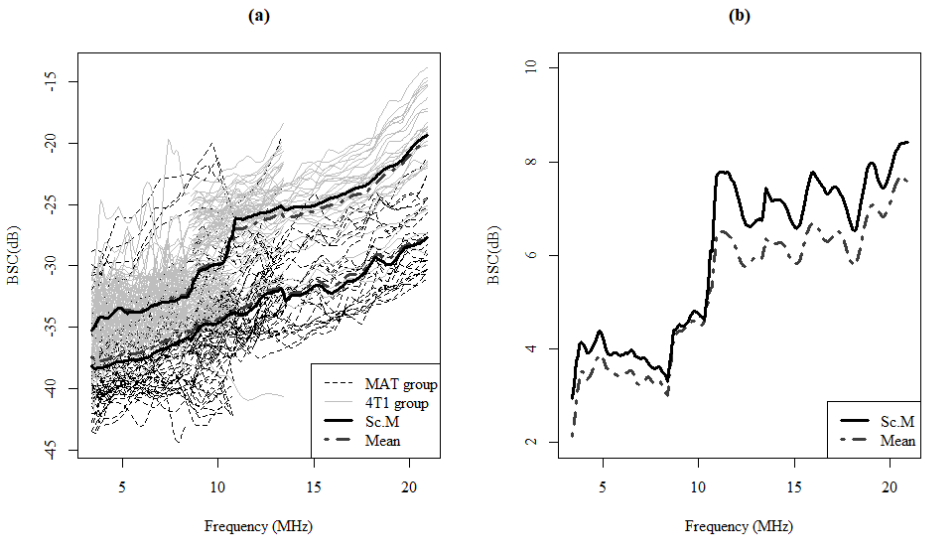

Figure 5. QUS data. (a) BSC curves from MAT and 4T1 tumors with proposed functional M-estimator and functional mean for each group. (b) Marginal group differences of M-estimator and mean.

frequencies 3–8.5 MHz, and MS200 and MS400 from VisualSonics cover higher frequencies, 8.5–13.5 MHz and 8.5–21.9 MHz. The experiment aims to detect significant differences in the behavior of the BSC curves between two distinct tumors, and to investigate the consistency among frequency ranges or transducers in such detection. To address this, we calculate the functional M-estimator and perform robust inferential tests.

Figure 5 (a) shows the estimated group location parameters from the marginal M-estimator under a Huber loss with a scaled robust tuning parameter, and from the functional mean for the two tumor types. We observe remarkable jumps at 8.5 MHz and 10.8 MHz in the group of the 4T1 tumor, corresponding to frequencies where a change in the variety of transducers is observed. However, the jump in the functional mean at 10.8 MHz is weaker than the jump in the M-estimator, owing to multiple outlying curves in the 4T1 group, which have suspiciously small values or abnormal behaviors compared with the majority.

To demonstrate the significant distinction in the behaviors of the BSC from the two different tumors, Wirtzfeld et al. (2015) applied separate fANOVA to subsets of the data, consisting of BSC curves collected from the same transducer, spanning the same frequency ranges, to avoid a partial sampling issue. They also test the equality of the functional mean parameters, which might not be valid with

Table 2. Estimated projection coefficients to basis functions. 95% and 99% bootstrapped confidence intervals in round brackets and square brackets, respectively. Bracket with * indicates an interval not including zero.

	Quadratic	Linear	Constant	Step1	Step2	Step3
M-estimator	−0.15	1.55	6.00	0.22	0.66	−0.11
	$(-0.52, 0.17)$	$(0.54, 2.38)^*$	$(4.90, 6.97)^*$	$(-0.05, 0.52)$	$(0.29, 1.00)^*$	$(-0.46, 0.31)$
	$[-0.65, 0.29]$	$[0.18, 2.63]^*$	$[4.51, 7.20]^*$	$[-0.14, 0.64]$	$[0.18, 1.09]^*$	$[-0.65, 0.44]$
Scaled M-estimator	−0.22	1.48	5.98	0.24	0.58	−0.06
	$(-0.50, 0.11)$	$(0.56, 2.31)^*$	$(4.88, 6.91)^*$	$(-0.01, 0.49)$	$(0.27, 0.86)^*$	$(-0.42, 0.26)$
	$[-0.61, 0.23]$	$[0.20, 2.55]^*$	$[4.50, 7.15]^*$	$[-0.10, 0.56]$	$[0.17, 0.97]^*$	$[-0.55, 0.40]$
Mean	-0.16	1.32	5.28	0.24	0.34	-0.07
	$(-0.44, 0.15)$	$(0.58, 2.05)^*$	$(4.33, 6.13)^*$	$(0.02, 0.46)^*$	$(0.05, 0.62)^*$	$(-0.38, 0.25)$
	$[-0.53, 0.24]$	$[0.30, 2.29]^*$	$[4.07, 6.35]^*$	$[-0.05, 0.52]$	$[-0.02, 0.71]$	$[-0.47, 0.32]$

outlying curves. Thus, we now perform an L^2-type robust fANOVA test following the bootstrap procedure in Section 4, and detect a significant group difference with a p-value < 0.0001 ($T_n^* = 31.32$). This enables a full-scale analysis with higher power by using all curves in one test. Then, as a follow-up, we examine the systematic trend in the functional difference between the two tumors; Figure 5 (b) presents a seemingly increasing trend. However, an inferential justification is needed to make a conclusion. At the same time, we also want to investigate the effect of the transducers in the BSC measures. To this end, we calculate bootstrapped confidence intervals of the projection coefficients corresponding to the selected basis and step functions. We specifically consider constant, linear, and quadratic basis functions, as well as three step functions, Step1, Step2, and Step3, where Step1 has a jump at 8.5 MHz, Step2 has a jump at 10.8 MHz, and Step3 has a jump at 13.5 MHz. The step functions are defined based on known transducer information. The inferences based on the coefficients of the first three basis functions enable us to identify a general trend, where higher frequencies separate two groups more efficiently than lower frequencies do. At the same time, the coefficients of the three step functions provide information to discover the transducer effect. We adopt the Huber function in the M-estimator with constant and scaled robust tuning parameters, as discussed in Section 5. For the unscaled one, we choose $c = 0.8$, and for the scaled one, $r = 0.4$, considering the overall estimated MAD over the whole frequency range.

Table 2 shows the estimated coefficients of the functional group difference projected to six basis functions and the corresponding 95% and 99% bootstrapped confidence intervals based on 3,000 replications. The discretized curves in the data are densely collected, but do not share common grids, so interpolation is applied to each curve at an equally spaced grid of 176 points over 3–21.6 MHz. The computation time on a 3.60 GHz Intel(R) Core(TM) i7-7700 CPU is 234

seconds for the derivation of the bootstrapped confidence intervals from the M-estimator with $n_{4T1} = 115$, $n_{MAT} = 149$.

First, we observe that the results of the M-estimator from the scaled and constant tuning parameters look almost the same, except for the discrepancy in the estimated coefficients of the quadratic term. However, the quadratic trend is statistically insignificant from both bootstrapped inferences; thus, the two estimators derive fundamentally the same conclusion. Then, a significant linear trend is detected in the group differences with positive coefficients from the M-estimator, which implies that higher frequencies are more efficient for detecting group differences than lower frequencies are. The finding is the same for the mean approach, but a shrunk estimate is observed owing to the effect of outliers.

To examine the transducer effect, we examine the results from three step functions. The change point at 8.5 MHz (Step1) turns out to be statistically insignificant from the robust estimators, but the 95% confidence interval from the mean does not include zero, implying significant distinct behavior at this jump. For the second change point, robust M-estimators detect a significant positive jump at 10.8 MHz with confidence, with the lower bound far from zero, but the inference from the mean function fails to detect such a change in the 99% confidence interval. Although the inference from the 95% confidence interval detects a significant jump, it lacks confidence, with a lower bound very close to zero. Again, this different conclusion is due to multiple outliers in the 4T1 group, and the mean function underestimates the jump at this change point. The last change point between two VisualSonics transducers turns out to be statistically insignificant from both estimators. In conclusion, the BSC curves significantly distinguish between different tumors, along with all frequency ranges, and higher frequencies separate them more efficiently than lower frequencies do. Furthermore, we see a significant positive jump at 10.8 MHz, which implies the high efficiency of the VisualSonics transducers in terms of tissue distinction compared with the others.

7. Discussion

We propose a class of robust M-estimators applicable to partially observed functional data. We show that our estimator is consistent and asymptotically follows the Gaussian process with root-n rates of convergence under a key condition for the sup-norm convergence of the indicator process. In addition, robust inferential tools are developed under asymptotic normality that can be performed in practice with bootstrap procedures. The validity of the bootstrap inference is supported by the convergence of the conditional second moments of the boot-

strapped samples, as well as by the simulation studies, where the true trend is detected with the desired coverage probability under heavy-tailed or contaminated distributions with various structures of missingness. In terms of estimation accuracy, numerical simulation experiments demonstrate that the proposed estimator outperforms existing functional robust estimators, even in the case of complete data.

The proposed partial sampling framework is particularly appealing, because various types of recent incomplete data structures satisfy our assumptions, including dense functional snippets (Lin and Wang (2020)) and fragmented functional data (Delaigle et al. (2021)). These connections demonstrate the wider applicability of the methods developed here. In addition, based on our simulation studies, further extension to segmented data recorded at sparse and irregular design points is a promising direction for further development.

Supplementary Material

The online Supplementary Material includes (i) technical proofs of the propositions, lemmas, and theorems, and (ii) figures and detailed results from the simulation studies.

Acknowledgments

This work was supported by National Science Foundation CAREER Award DMS-1752614 (X. Chen), University of Illinois Research Board Award RB18099 (X. Chen), and National Institutes of Health Grant R01CA226528-01A1 (D. G. Simpson).

References

Cuevas, A., Febrero, M. and Fraiman, R. (2004). An ANOVA test for functional data. *Computational Statistics & Data Analysis* **47**, 111–122.

Cuevas, A., Febrero, M. and Fraiman, R. (2007). Robust estimation and classification for functional data via projection-based depth notions. *Computational Statistics* **22**, 481–496.

Delaigle, A. and Hall, P. (2013). Classification using censored functional data. *Journal of the American Statistical Association* **108**, 1269–1283.

Delaigle, A. and Hall, P. (2016). Approximating fragmented functional data by segments of markov chains. *Biometrika* **103**, 779—799.

Delaigle, A., Hall, P., Huang, W. and Kneip, A. (2021). Estimating the covariance of fragmented and other related types of functional data. *Journal of the American Statistical Association* **116**, 1383–1401.

Descary, M.-H. and Panaretos, V. M. (2019). Recovering covariance from functional fragments. *Biometrika* **106**, 145–160.

Fraiman, R. and Muniz, G. (2001). Trimmed means for functional data. *Test* **10**, 419–440.

Fried, R., Einbeck, J. and Gather, U. (2007). Weighted repeated median smoothing and filtering. *Journal of the American Statistical Association* **102**, 1300–1308.

Gellar, J. E., Colantuoni, E., Needham, D. M. and Crainiceanu, C. M. (2014). Variable-domain functional regression for modeling icu data. *Journal of the American Statistical Association* **109**, 1425–1439.

Gervini, D. (2008). Robust functional estimation using the median and spherical principal components. *Biometrika* **95**, 587–600.

Goldberg, Y., Ritov, Y. and Mandelbaum, A. (2014). Predicting the continuation of a function with applications to call center data. *Journal of Statistical Planning and Inference* **147**, 53–65.

Gromenko, O., Kokoszka, P. and Sojka, J. (2017). Evaluation of the cooling trend in the ionosphere using functional regression with incomplete curves. *The Annals of Applied Statistics* **11**, 898–918.

Hampel, F. R. (1974). The influence curve and its role in robust estimation. *Journal of the American Statistical Association* **69**, 383–393.

Horváth, L. and Kokoszka, P. (2012). *Inference for Functional Data with Applications*. Springer.

Huber, P. J. (2005). *Robust Statistics*. John Wiley & Sons, Hoboken.

Kneip, A. and Liebl, D. (2019). On the optimal reconstruction of partially observed functional data. *The Annals of Statistics* **48**, 1692–1717.

Koenker, R. (2005). *Quantile Regression*. Cambridge, New York.

Kraus, D. (2015). Components and completion of partially observed functional data. *Journal of the Royal Statistical Society, Series B (Statistical Methodology)* **77**, 777–801.

Kraus, D. (2019). Inferential procedures for partially observed functional data. *Journal of Multivariate Analysis* **173**, 583–603.

Kraus, D. and Stefanucci, M. (2018). Classification of functional fragments by regularized linear classifiers with domain selection. *Biometrika* **106**, 161–180.

Liebl, D. (2013). Modeling and forecasting electricity spot prices: A functional data perspective. *The Annals of Applied Statistics* **7**, 1562–1592.

Lin, Z. and Wang, J.-L. (2020). Mean and covariance estimation for functional snippets. *Journal of the American Statistical Association*, to appear.

Lin, Z., Wang, J.-L. and Zhong, Q. (2020). Basis expansions for functional snippets. *Biometrika* **108**, 709–726..

Locantore, N., Marron, J. S., Simpson, D. G., Tripoli, N., Zhang, J.-T. and Cohen, K. L. (1999). Robust principal component analysis for functional data (with discussion). *Test* **8**, 1–73.

López-Pintado, S. and Romo, J. (2009). On the concept of depth for functional data. *Journal of the American Statistical Association* **104**, 718–734.

López-Pintado, S. and Romo, J. (2011). A half-region depth for functional data. *Computational Statistics and Data Analysis* **55**, 1679–1695.

Mojirsheibani, M. and Shaw, C. (2018). Classification with incomplete functional covariates. *Statistics & Probability Letters* **139**, 40–46.

Park, Y. (2017). *Effect Size Estimation and Robust Classification for Irregularly Sampled Functional Data*. Ph.D. thesis, University of Illinois at Urbana-Champaign. `http://hdl.handle.net/2142/98126`.

Park, Y. and Simpson, D. G. (2019). Robust probabilistic classification applicable to irregularly sampled functional data. *Computational Statistics and Data Analysis* **131**, 37–49.

Ramsay, J. and Silverman, B. W. (2005). *Functional Data Analysis*. Springer.

Shen, Q. and Faraway, J. (2004). An F type test for linear models with functional responses. *Statistica Sinica* **14**, 1239–1257.

Sinova, B., González-Rodríguez, G. and Van Aelst, S. (2018). M-estimators of location for functional data. *Bernoulli* **24**, 2328–2357.

Stefanucci, M., Sangalli, L. M. and Brutti, P. (2018). PCA-based discrimination of partially observed functional data, with an application to AneuRisk65 data set. *Statistica Neerlandica* **72**, 246–264.

van der Vaart, A. W. and Wellner, J. A. (1996). *Weak Convergence and Empirical Processes: With Applications to Statistics*. Springer-Verlag, New York.

van der Vaart, A. W. (2007). *Asymptotic Statistics*. Cambridge University Press, Cambridge.

Wirtzfeld, L. A., Ghoshal, G., Rosado-Mendez, I. M., Nam, K., Park, Y., Pawlicki, A. D. et al. (2015). Quantitative ultrasound comparison of MAT and 4T1 mammary tumors in mice and rats across multiple imaging systems. *Journal of Ultrasound in Medicine* **34**, 1373–1383.

Zhang, A. and Chen, K. (2022). Nonparametric covariance estimation for mixed longitudinal studies, with applications in midlife women's health. *Statistica Sinica* **32**, 345–365.

Zhang, J.-T. and Liang, X. (2014). One-way ANOVA for functional data via globalizing the pointwise F-test. *Scandinavian Journal of Statistics* **41**, 51–71.

Yeonjoo Park

University of Texas at San Antonio, San Antonio, TX 78249, USA.

E-mail: yeonjoo.park@utsa.edu

Xiaohui Chen

University of Illinois at Urbana-Champaign, Champaign, IL 61820, USA.

E-mail: xhchen@illinois.edu

Douglas G. Simpson

University of Illinois at Urbana-Champaign, Champaign, IL 61820, USA.

E-mail: dgs@illinois.edu

(Received September 2020; accepted April 2021)