Annual Review of Statistics and Its Application

Empirical Likelihood in Functional Data Analysis

Hsin-wen Chang[1] and Ian W. McKeague[2]

[1]Institute of Statistical Science, Academia Sinica, Taipei, Taiwan; email: hwchang@stat.sinica.edu.tw

[2]Department of Biostatistics, Columbia University, New York, NY, USA

Annu. Rev. Stat. Appl. 2025. 12:425–48

First published as a Review in Advance on November 12, 2024

The *Annual Review of Statistics and Its Application* is online at statistics.annualreviews.org

https://doi.org/10.1146/annurev-statistics-112723-034225

Keywords

dynamic correlation, functional linear models, functional ANOVA, pointwise and simultaneous inference

Abstract

Functional data analysis (FDA) studies data that include infinite-dimensional functions or objects, generalizing traditional univariate or multivariate observations from each study unit. Among inferential approaches without parametric assumptions, empirical likelihood (EL) offers a principled method in that it extends the framework of parametric likelihood ratio–based inference via the nonparametric likelihood. There has been increasing use of EL in FDA due to its many favorable properties, including self-normalization and the data-driven shape of confidence regions. This article presents a review of EL approaches in FDA, starting with finite-dimensional features, then covering infinite-dimensional features. We contrast smooth and nonsmooth frameworks in FDA and show how EL has been incorporated into both of them. The article concludes with a discussion of some future research directions, including the possibility of applying EL to conformal inference.

1. INTRODUCTION

Functional data analysis (FDA) studies data that include infinite-dimensional functions or objects, generalizing traditional univariate or multivariate observations from each datum/study unit. That is, in contrast to vector-valued data, FDA handles function-valued data or observations of a function $X(t)$, where t is allowed to vary over a time interval I (or a more general index set). For example, in animal research, if the interest is in analyzing video recordings of the movement of animals over some time period, FDA can treat each recording, $\{X(t), t \in I\}$, recognizing dependence among video frames and among pixels in each frame. FDA has the advantage of analyzing the collection of videos (say $\{X_1(t), \ldots, X_n(t), t \in I\}$), in contrast to the simplistic approach of examining summary statistics obtained from each video. Furthermore, FDA allows some intuitive operations, such as the sample mean $\sum_{i=1}^{n} X_i(t)/n$, to have a useful data-analytic role provided the recordings are prealigned. Due to a surge in the need to analyze such complex data, including but not limited to images, sounds, videos, and texts, FDA has become a vibrant research area, although its roots date back to the 1950s (see, e.g., Wang et al. 2016, and references therein). In addition to the infinite-dimensional aspect, FDA can also handle a mix of functional data and traditional univariate/multivariate data (Ramsay & Silverman 2005). These various FDA settings yield diverse types of finite-dimensional and infinite-dimensional parameters/features of interest.

Nonparametric and semiparametric approaches to examining these features are central topics in FDA, since parametric assumptions can be unrealistic given the complexity of the data. The need to avoid such assumptions is also a motivation behind the recent interest in conformal inference (see, e.g., Fontana et al. 2023). Among these approaches, empirical likelihood (EL) offers a principled method in that it extends the framework of parametric likelihood ratio–based inference via the nonparametric likelihood (see, e.g., Owen 1988, 2001). For example, to test whether a target feature $\theta(F)$ (as a function of the distribution F) equals a hypothesized value θ_0, EL constructs a likelihood ratio (termed the EL ratio),

$$R = \frac{\sup\{L(F) \mid \theta(F) = \theta_0, F \in \mathcal{F}\}}{\sup\{L(F) \mid F \in \mathcal{F}\}},$$

to quantify the distance between the maximal likelihood with and without the constraint $\theta(F) = \theta_0$, where $L(\cdot)$ is the nonparametric likelihood, and $\mathcal{F}$ is the set of distributions supported on the observed data. By relating the nonparametric likelihood to the target features via constrained optimization (see Sections 2.2 and 3.2 for more details), this construction avoids the model-specific constraints of parametric likelihood (see, e.g., Lazar 2021). Furthermore, it enables EL to have several major similarities to usual parametric likelihood ratio–based inference, but without the need to specify a family of distributions for the data. First, the likelihood ratio statistic $-2 \log R$ typically has the same asymptotic chi-squared distribution as in Wilks's theorem for parametric likelihood ratio statistics (Wilks 1938). Second, EL utilizes the same formulation of likelihood ratio tests and confidence regions (by inverting the test) as in parametric settings. Third, EL inherits many optimality properties of the ordinary parametric likelihood ratio, such as a large deviation principle (Kitamura 2007, Kitamura et al. 2012), a second-order local maximinity property (Mukerjee 1994, Bravo 2003), and Bahadur efficiency (Bahadur 1967, Otsu 2010).

EL has other important advantages. Specifically, it is self-normalized, so that there is typically no need for variance estimation as in many traditional inferential methods. In addition, unlike many non-EL alternatives, EL-based confidence regions have data-driven shape and orientation—they preserve the range of the observations (Hall & La Scala 1990), and even monotonicity (if there is any) of the functional data (Chang & McKeague 2022). The data-driven nature of EL can be attributed to the fact that it searches for the estimated target feature within the convex hull of the data. When the sample size is small, however, this means EL can perform quite differently from its

large sample behavior (see, e.g., Chen & Huang 2013). Fortunately, in this case, the performance of EL can be boosted by resampling techniques such as bootstrap (see, e.g., Owen 1988), and statistical software packages have been developed to facilitate the computing, even in complex FDA settings (see, e.g., the R packages in Wang et al. 2018, Chang & McKeague 2022). Because of these favorable qualities of EL, there has been increasing interest in using EL for FDA. This article presents a timely review of this topic. Although there have been reviews of EL concerning its basic theory and practical use (Lazar 2021), comparison to related methods such as generalized method of moments (Parente & Smith 2014), and in regression (Chen & Van Keilegom 2009) and time series (Nordman & Lahiri 2014), and there have been reviews of FDA that focus on regression (Morris 2015), smoothing approaches (Wang et al. 2016), and functional data with dependence among the functional observations (Koner & Staicu 2023), none of these reviews have addressed the use of EL in FDA.

While EL in FDA is a relatively new topic, EL in longitudinal data analysis (LDA) has a longer history. Like FDA, each study unit in LDA has repeated measurements, but longitudinal data are typically viewed as having fewer measurements per study unit compared with functional data (see, e.g., Zhou & Lin 2014). Nevertheless, there are methods nowadays that can handle longitudinal data and functional data in a unified framework, for example by viewing longitudinal data as sparse functional data (see, e.g., Hu et al. 2021). Since FDA views the repeated observations per study unit as a sample path of a stochastic process or random field, we include in this review EL-based LDA methods that take the same approach.

In FDA, there can be finite- and/or infinite-dimensional features of interest. The finite-dimensional are less common than the infinite-dimensional, but there are important examples, such as functional correlation measures and regression coefficients for finite-dimensional covariates in partial functional linear models (PFLMs) (see Section 2.1 for details). While infinite-dimensional features are more common, a fully fledged treatment of such features involves sophisticated methods, such as empirical process theory (e.g., van der Vaart 2000). Therefore, many articles restrict attention to pointwise inference (see, e.g., Lian 2012). However, implementing such inference alone does not provide reliable confidence regions (for the whole target function) or accurate type I error rates. Simultaneous inference is still needed for accurate type I error rate control, but is less explored in the literature due to the challenges. To sum up, it is important to distinguish between pointwise and simultaneous inference for infinite-dimensional features, as we do in the following. We refer to **Figure 1** for a taxonomy of the FDA topics discussed in this article.

While traditional FDA has focused mostly on smooth target functions and smooth data, discontinuous functional features have been handled by nonsmooth approaches in FDA (see, e.g., Zhu et al. 2014) that are different from change point methods (the differences are explained in the last paragraph of Section 4.2). These features arise where there are edges among different regions in the functional data, or when information at each time point (or point in the domain of the functional data) is obtained based on a specific threshold. An example is the incidence of sedentary behavior as a function of time, based on accelerometer readings. Sedentary behavior is commonly defined as having an accelerometer reading below a preestablished cut-point, such as the 100 intensity count (per minute) used by Matthews et al. (2008). This binary categorization leads to a step function over time, as individuals shift between sedentary and nonsedentary states depending on whether the reading crosses the threshold or not. Another example is an occupation time curve, which shows the amount of time spent above each activity level across the full range of accelerometer readings (Chang & McKeague 2022); **Figure 2** shows an example based on data from the 2005–2006 US National Health and Nutrition Examination Survey (US Natl. Cent. Health Stat. 2006). Simultaneous inference beyond estimation based on this kind of nonsmooth functional

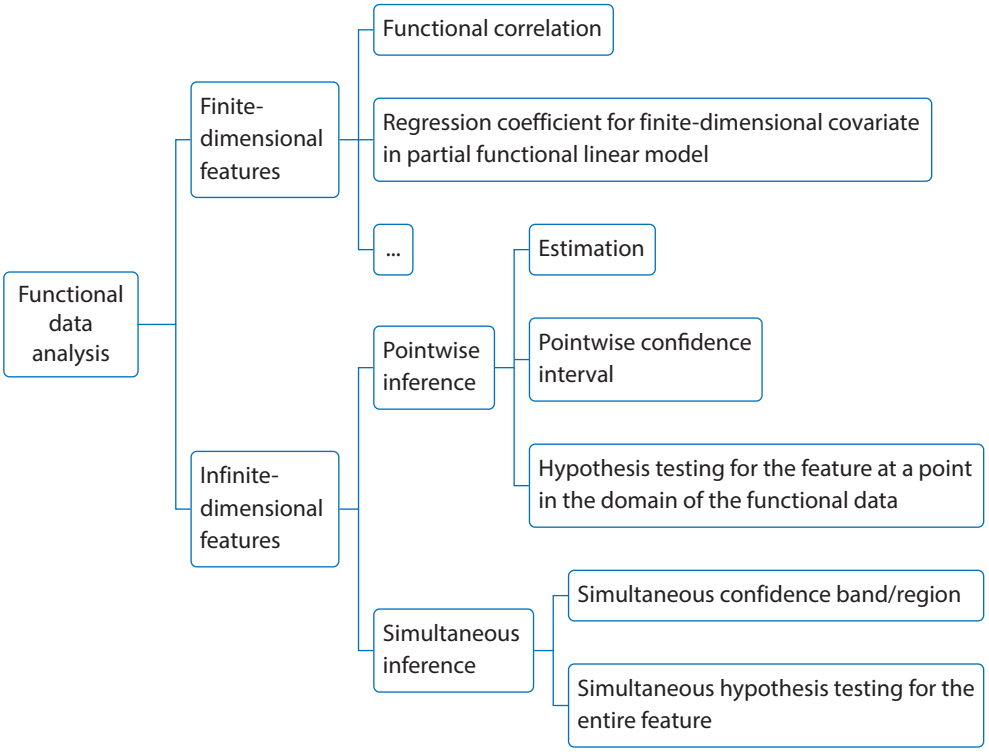

Figure 1

Taxonomy diagram for finite- and infinite-dimensional features in functional data analysis (FDA) discussed in the article. Note that for infinite-dimensional features, simultaneous inference is needed for accurate type I error rate control.

data is challenging but necessary. In particular, Chang & McKeague (2022) have shown that traditional smoothing approaches for constructing simultaneous confidence bands for the functional mean under a fixed dense design can lead to misleading results near points of discontinuity. In the following, we elaborate on the distinction between smooth and nonsmooth methods.

The study of EL in FDA inherits the aforementioned challenges in inference about infinite-dimensional features, and the need to distinguish between smooth and nonsmooth methods. A computational demand with EL lies in the constrained optimization involved in the likelihood ratio. To facilitate practical implementation, many computational techniques and software packages have been devised. Readers are directed to Lazar (2021), Chang (2020), and Kim et al. (2024) for examples of such packages, and the following for those based on functional data.

Functional data are usually assumed to live in some function space, much as a random variable takes its values on the real line. Among the many possible choices of function spaces for FDA, $L^2(I)$ is often used (Wang et al. 2016), where I was introduced in the first paragraph of Section 1, and $L^2(I)$ denotes the space of all square integrable functions on I endowed with the inner product $\langle x, y\rangle = \int_I x(t)y(t)\mathrm{d}t$. Other choices of the function space (and metric) need to be guided by known properties of the functional data. For instance, for analyzing continuous curves, we could use the space $C(I)$ of real-valued continuous functions on I endowed with the supremum norm $\|x\| = \max_{t\in I}|x(t)|$ (Dette et al. 2020). Another example is the space of functions of bounded variation equipped with the supremum norm, which is general enough for handling bounded, monotonic, but not necessarily continuous, curves, and is well matched with inferential methods with the

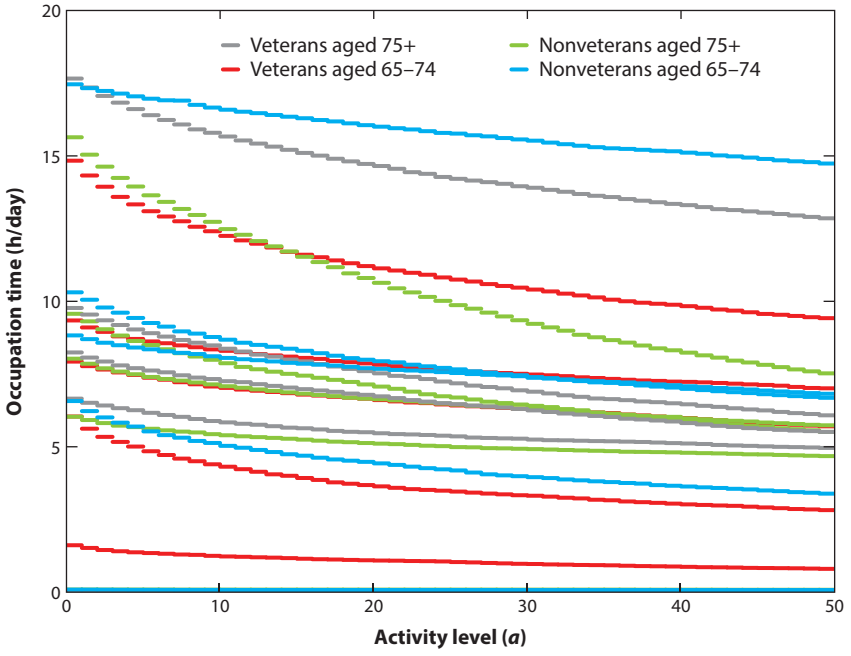

Figure 2

Example of nonsmooth functional data: occupation time from systematically selected (the smallest and largest observations, and every quartile at $a = 0$) veterans aged 75 and older, nonveterans aged 75 and older, veterans aged 65–74, and nonveterans aged 65–74, based on the NHANES physical activity data. Note the curves have jump discontinuities by the definition of occupation time (Chang & McKeague 2022) and the use of gridded values of accelerometer readings. Abbreviation: NHANES, National Health and Nutrition Examination Survey.

supremum norm formulation (Chang & McKeague 2022). Due to its common usage, the $L^2(I)$ framework is assumed in the following unless otherwise specified.

This article is organized as follows. Section 2 reviews EL-based inference for finite-dimensional features in FDA. We start by introducing FDA frameworks that involve finite-dimensional features in Section 2.1, and then present EL approaches and their advantages in Section 2.2. Section 3 reviews EL-based inference for infinite-dimensional features in FDA. Again, we start by introducing relevant FDA frameworks in Section 3.1, and then present EL approaches and their advantages in Section 3.2. In doing so, we carefully distinguish between pointwise and simultaneous inference, as previously discussed. Section 4 carefully differentiates between smooth (Section 4.1) and nonsmooth (Section 4.2) frameworks in FDA, illustrates how EL can be incorporated, and reviews the use of these frameworks in the settings discussed in Sections 2 and 3. In Section 5, various directions for future research are highlighted.

2. EMPIRICAL LIKELIHOOD–BASED INFERENCE FOR FINITE-DIMENSIONAL FUNCTIONAL DATA ANALYSIS FEATURES

2.1. Examples of Finite-Dimensional Features in Functional Data Analysis

There are numerous studies of finite-dimensional features in FDA, such as functional correlation and partial functional linear regression. Here we use the following examples to illustrate the basic ideas. These examples are chosen because EL-based methods are available, as reviewed in

Section 2.2. The examples should be read sequentially, as notation in one example may also be used in the next.

Example 1 (Functional correlation). There are many existing functional correlation measures that aim at capturing the correlation structure between pairs of functional datasets. We restrict our attention to dynamical correlation, due to its explicit representation and stable results over various choices of bandwidth and number of repeated measurements (Dubin & Müller 2005). Recalling notation from Section 1, for a pair of random functions $\{X(t), Y(t), t \in I\}$ in $L^2(I)$, the dynamical correlation is defined as the scalar

$$\rho = \mathrm{E}(\langle X^*, Y^* \rangle), \tag{1.}$$

where X^* and Y^* denote standardized versions of X and Y, respectively, and the notation $\langle x, y \rangle = \int_I x(t)y(t)\mathrm{d}t$ has been defined in Section 1. Estimation, hypothesis testing, and confidence intervals for ρ have been proposed in the literature since Dubin & Müller (2005). However, there are challenges in non-EL-based inference for ρ, including intractability of the asymptotic variance of the estimator of ρ, and poor confidence interval coverage when the data are irregularly spaced.

Example 2 (Functional regression). There are many functional regression methods for different combinations of response/covariate data types, such as scalar-on-function regression (meaning the response is scalar and the covariate is functional; see Equation 2), function-on-scalar regression (Reiss et al. 2010), and function-on-function regression (see, e.g., Kim et al. 2018, and Equation 10 in Section 3.1). These methods also differ according to the relationships between the response and covariates, such as linear, nonlinear (Yao & Müller 2010), and nonparametric (see Equation 12 in Section 3.1). The well-established linear scalar-on-function regression, also known simply as the functional linear model, is given by

$$Y = \beta_0 + \int_0^1 \gamma(t)X(t)\mathrm{d}t + \varepsilon \tag{2.}$$

for a scalar response Y and functional covariate $X(t)$, and $\gamma(\cdot)$ is the unknown coefficient function. Here ε is an independent random error with zero mean and finite variance, and the integral in Equation 2 can be viewed as summing over a potentially unlimited number of predictors in a conventional multiple regression if we think of each t as indexing a separate scalar covariate. Finite-dimensional features often occur in variants of the PFLM, which extends Equation 2 to handle mixed-type predictors. More specifically, the PFLM can be specified as

$$Y = \boldsymbol{\beta}^\top \boldsymbol{Z} + \int_0^1 \gamma(t)X(t)\mathrm{d}t + \varepsilon, \tag{3.}$$

where $\boldsymbol{Z}$ is a p-dimensional covariate vector, and $\boldsymbol{\beta}$ is a $p \times 1$ vector of unknown regression coefficients. In the PFLM, $\boldsymbol{\beta}$ is a finite-dimensional feature of interest, and its inference has been thoroughly explored (see, e.g., Shin 2009). However, non-EL-based methods need to estimate the complex limiting covariance matrix of the estimator of $\boldsymbol{\beta}$.

Example 3 (Dependence across study units). While studies of functional regression typically focus on independently and identically distributed (i.i.d.) random errors $\varepsilon_1, \ldots, \varepsilon_n$ as copies of ε, some authors consider dependent errors (here n is the number of study units). While such dependence often arises in social and earth science applications, where observations over time are usually serially correlated, it is also possible that some seemingly independent data have dependence structure, for instance when nearby (in terms of sample

index i) study units are collected in the same batch. One popular approach to handling these situations is to assume that the random errors ε_i follow a pth order autoregressive model:

$$\varepsilon_i = a_1\varepsilon_{i-1} + a_2\varepsilon_{i-2} + \cdots + a_p\varepsilon_{i-p} + e_i, \qquad 4.$$

where the e_i are i.i.d. zero-mean white noise with finite variance, and $\boldsymbol{a} = (a_1, a_2, \ldots, a_p)^{\mathrm{T}}$ is the finite-dimensional feature of interest (see, e.g., Dabo-Niang & Guillas 2010). A common first step to conduct the relevant functional regression is to test if there exists error correlation at all—that is, testing

$$H_0 : \boldsymbol{a} = \boldsymbol{0} \text{ versus } H_1 : \boldsymbol{a} \neq \boldsymbol{0}. \qquad 5.$$

Here EL is called for to circumvent the estimation of the complicated asymptotic variance of the Durbin–Watson test statistic (Durbin & Watson 1950), one of the standard tests for serial correlation in linear regression.

In all the above examples, to construct the confidence region or conduct the hypothesis test of interest, a typical approach is by showing asymptotic normality of an estimator for the target feature and using the estimated variance of that estimator to gauge the accuracy of the estimation; this is called normal approximation (NA) or Wald-type approach in the EL literature (see, e.g., Zhou & Lin 2016, Chang & McKeague 2022). However, the variance estimation is rather complicated in many advanced settings, such as those of FDA (see, e.g., Wang et al. 2018) and nonparametric or semiparametric regression (see, e.g., You et al. 2006). In addition, NA-based confidence regions are elliptical and do not adapt to skewness in the data. In contrast, EL is self-normalized, thereby circumventing the need for variance estimation, and the shape of EL-based confidence regions is determined by the data instead of a priori. We discuss this in the next subsection.

2.2. Empirical Likelihood Approach and Its Advantages

Let $\boldsymbol{\theta} \in \Theta$ be a d-dimensional feature of interest, for some $d < \infty$. Following the tradition originated by Qin & Lawless (1994) for nonfunctional data, EL handles the case when the truth $\boldsymbol{\theta}_0$ satisfies an estimating equation $\mathrm{E}[\boldsymbol{M}(\boldsymbol{X}, \boldsymbol{\theta}_0)] = 0$, where $\boldsymbol{M}(\boldsymbol{X}, \boldsymbol{\theta})$ is a smooth function of the functional data $\boldsymbol{X} = \boldsymbol{X}(t)$ and the feature. EL-based hypothesis testing typically handles $H_0 : \boldsymbol{\theta} = \tilde{\boldsymbol{\theta}}$ versus $H_1 : \boldsymbol{\theta} \neq \tilde{\boldsymbol{\theta}}$, where $\tilde{\boldsymbol{\theta}}$ is a candidate value for $\boldsymbol{\theta}$. The EL test relies on the EL ratio statistic $-2\log R(\tilde{\boldsymbol{\theta}}, \boldsymbol{X}_{1:n})$ (see next paragraph for more details), a nonparametric version of the likelihood ratio statistic, where $\boldsymbol{X}_{1:n} = \{\boldsymbol{X}_1(t), \ldots, \boldsymbol{X}_n(t), t \in I\}$ is the set of functional data at hand. Note that $\boldsymbol{X}_{1:n}$ is typically assumed to consist of independent copies of $\boldsymbol{X}(t)$, but generalizations to dependent data have also been considered, such as for handling Example 3. The statistic $-2\log R(\tilde{\boldsymbol{\theta}}, \boldsymbol{X}_{1:n})$ is always nonnegative (because a likelihood ratio is bounded between 0 and 1), and it evaluates how far the candidate value $\tilde{\boldsymbol{\theta}}$ of the feature deviates from the truth. The test statistic is compared with some cutoff value q as the upper α-quantile of some calibration/reference distribution, where $\alpha \in [0, 1]$ is a user-chosen type I error rate. This reference can be obtained from the large-sample limit of $-2\log R(\tilde{\boldsymbol{\theta}}, \boldsymbol{X}_{1:n})$.

This limit is typically chi-squared, but exceptions can occur in the following two cases: The first is when $\tilde{\boldsymbol{\theta}}$ is on the boundary of Θ, a classical case when Wilks's theorem can also fail (see, e.g., Anastasiou & Reinert 2020, and references therein); the second is when there are (possibly infinite-dimensional) nuisance parameters in $\boldsymbol{M}(\boldsymbol{X}, \cdot)$ (see, e.g., Hjort et al. 2009), which is often estimated using plug-in of estimates of the nuisance parameters. When the limiting distribution of $-2\log R(\tilde{\boldsymbol{\theta}}, \boldsymbol{X}_{1:n})$ is not chi-squared, it can be approximated by bootstrap, a type of calibration based on repeated sampling. More specifically, we can obtain quantities that are asymptotically equivalent to $-2\log R(\tilde{\boldsymbol{\theta}}, \boldsymbol{X}_{1:n})$ and are easy/fast to compute. Then, we calculate these quantities

using a bootstrap sample resampled from the original data, and do this B times (e.g., $B = 1{,}000$). Each bootstrap sample can be generated based on sampling n curves with replacement from the original data $\boldsymbol{X}_{1:n}$, or from more complex functions of the original data, such as the residuals in regression settings. This results in B values of the aforementioned quantities, which leads to a bootstrap estimate of the limit of $-2\log R(\tilde{\boldsymbol{\theta}}, \boldsymbol{X}_{1:n})$.

The EL-based confidence set (a set of possible $\boldsymbol{\theta}$) $C(\boldsymbol{X}_{1:n}) \subset \Theta$ is constructed in a similar way as any other confidence set/conformal set

$$C(\boldsymbol{X}_{1:n}) = \{\tilde{\boldsymbol{\theta}} : s(\tilde{\boldsymbol{\theta}}, \boldsymbol{X}_{1:n}) \leq q'\} \tag{6.}$$

that contrasts a score $s(\tilde{\boldsymbol{\theta}}, \boldsymbol{X}_{1:n})$ (like a conformal score in conformal inference) with a threshold q'. For EL, we set the score $s(\tilde{\boldsymbol{\theta}}, \boldsymbol{X}_{1:n})$ in Equation 6 to be the EL ratio statistic $-2\log R(\tilde{\boldsymbol{\theta}}, \boldsymbol{X}_{1:n})$. Setting q' as q typically leads to an asymptotic level $1-\alpha$ guarantee as $\mathrm{P}(\boldsymbol{\theta}_0 \in C(\boldsymbol{X}_{1:n})) \to 1-\alpha$; that is, the probability that the confidence set contains the true value of the feature is (asymptotically) $1-\alpha$. Note that the EL-based confidence set can be used to test H_0 versus H_1, too, by using the rejection region $\{\boldsymbol{\theta}_0 \notin C(\boldsymbol{X}_{1:n})\}$. To distinguish this from the EL test mentioned in the beginning of this paragraph, the term "EL test" in this article specifically refers to the testing procedure that relies on comparing the EL ratio statistic with the critical value q'—that is, rejecting H_0 when $-2\log R(\tilde{\boldsymbol{\theta}}, \boldsymbol{X}_{1:n}) > q'$. Despite this distinction, many EL-based methods focus on constructing/computing confidence sets, since the extension to hypothesis testing procedures is trivial. Furthermore, the construction/computation of a confidence set is more challenging than that of the corresponding EL test, because the former typically involves solving for the roots $\boldsymbol{\theta}_B$ of $-2\log R(\boldsymbol{\theta}_B, \boldsymbol{X}_{1:n}) = q$ for defining the boundary of the confidence region. This disparity in the challenge of construction/computation between EL confidence set and EL test is less seen in NA-based methods, where there is typically a similar level of difficulty in confidence set and test construction. This does not necessarily mean NA-based methods are easier, as NA-based test statistics can be hard to derive due to the intractability of the relevant variance estimate, as mentioned in Section 2.1; we give examples later in this section.

The construction of $-2\log R(\tilde{\boldsymbol{\theta}}, \boldsymbol{X}_{1:n})$ proceeds as follows. The EL ratio for the test is just like a parametric likelihood ratio, with the numerator maximizing a nonparametric likelihood subject to constraint under H_0, and the denominator maximizing the nonparametric likelihood subject to constraint defined by the union of the spaces under H_0 and H_1. That is, the EL ratio for $\boldsymbol{\theta}$ is

$$\mathcal{R}(\tilde{\boldsymbol{\theta}}, \boldsymbol{X}_{1:n}) = \frac{\sup\{L(F) : E_F[\boldsymbol{M}(\boldsymbol{X}, \tilde{\boldsymbol{\theta}})] = 0, F \in \mathcal{F}\}}{\sup\{L(F) : F \in \mathcal{F}\}}, \tag{7.}$$

where $F(\cdot)$ is a candidate for the cumulative distribution function of $\boldsymbol{X}$, $E_F(\cdot)$ denotes taking expectation with respect to $F(\cdot)$, $\mathcal{F}$ is the set of distributions supported on the data $\boldsymbol{X}_{1:n}$,

$$L(F) = \prod_{i=1}^{n} p_i \tag{8.}$$

is the nonparametric likelihood, p_i is the point mass of $F(\cdot)$ at $\boldsymbol{X}_i$, and we follow the convention $\sup \emptyset = 0$. Note that Equation 7 can be used as a general form of the EL ratio for any type of data, including the case that $\boldsymbol{X}$ has a functional form. As mentioned in Section 1, Equation 7 indeed avoids assuming specific forms in relating the likelihood to the target features by optimizing the nonparametric likelihood $L(F)$ subject to constraints imposed on the target features. This flexibility may look too good to be true at first, and one may doubt its implementability. Note that the denominator of Equation 7 is obviously $\prod_{i=1}^{n}(1/n)$. The numerator can be shown by

Lagrange's method to be $\prod_{i=1}^{n} \tilde{p}_i$, where $\tilde{p}_i = \{n[1 + \tilde{\boldsymbol{\lambda}}^\top \tilde{\boldsymbol{M}}_i]\}^{-1}$, $\tilde{\boldsymbol{M}}_i = \boldsymbol{M}(\boldsymbol{X}_i, \tilde{\boldsymbol{\theta}})$, and $\tilde{\lambda}$ satisfies the estimating equation $\sum_{i=1}^{n} \tilde{p}_i \tilde{\boldsymbol{M}}_i = 0$. It follows that

$$-2 \log \mathcal{R}(\tilde{\boldsymbol{\theta}}, \boldsymbol{X}_{1:n}) = 2 \sum_{i=1}^{n} \log\left(1 + \tilde{\boldsymbol{\lambda}}^\top \tilde{\boldsymbol{M}}_i\right).$$

By the same reasoning as in EL for a multivariate mean (Owen 2001, p. 70), any feasible solution for $\tilde{\boldsymbol{\theta}}$ in computing $\mathcal{R}(\tilde{\boldsymbol{\theta}}, \boldsymbol{X}_{1:n})$ lies in the convex hull of $\{\tilde{\boldsymbol{M}}_1, \ldots, \tilde{\boldsymbol{M}}_n\}$; this is what we mean by EL respecting the range restrictions inherent in the data, as mentioned in Section 1. The limiting distribution of $-2 \log R(\tilde{\boldsymbol{\theta}}, \boldsymbol{X}_{1:n})$ can typically be shown to be $\chi^2(d)$, a chi-squared distribution with d degree of freedom. Remarkably, this nonparametric analogue of Wilks's theorem holds across various kinds of problems (some examples are provided in the next paragraph), rendering the calibration of the confidence set in Equation 6 quite standard—just take q to be the upper α-quantile of $\chi^2(d)$. Exceptions are discussed in the first paragraph of Section 2.2.

An example of what $-2 \log R(\tilde{\boldsymbol{\theta}}, \boldsymbol{X}_{1:n})$ typically looks like across different values of $\tilde{\boldsymbol{\theta}}$ is shown in **Figure 3**. This example is taken from Sang et al. (2019, section 5.2), who study the dynamical correlation ρ (see Example 1) between two electrodes, C5 and P8, placed on central and temporal scalp regions, respectively, each containing 256 electroencephalogram (EEG) measurements from 76 alcoholic subjects. **Figure 3** shows that $-2 \log R(\tilde{\rho}, \boldsymbol{X}_{1:n})$ attains its minimum at $\hat{\rho}$, the point estimator for dynamical correlation, and increases as $\tilde{\rho}$ moves away from $\hat{\rho}$. The two bounds of the confidence interval for ρ are then found as the two $\tilde{\rho}$s that equate $-2 \log R(\tilde{\rho}, \boldsymbol{X}_{1:n})$ with q, which is the upper 0.05-quantile of $\chi^2(1)$ in this case.

Note that the above EL-based inference does not involve variance estimation, as mentioned in Section 2.1. This major advantage enables Sang et al. (2019) to conduct inference for ρ in Equation 1 based on the $\chi^2(1)$ limiting distribution of the EL ratio statistic under irregular and regular designs, where regular design in FDA means the random functions are observed over a

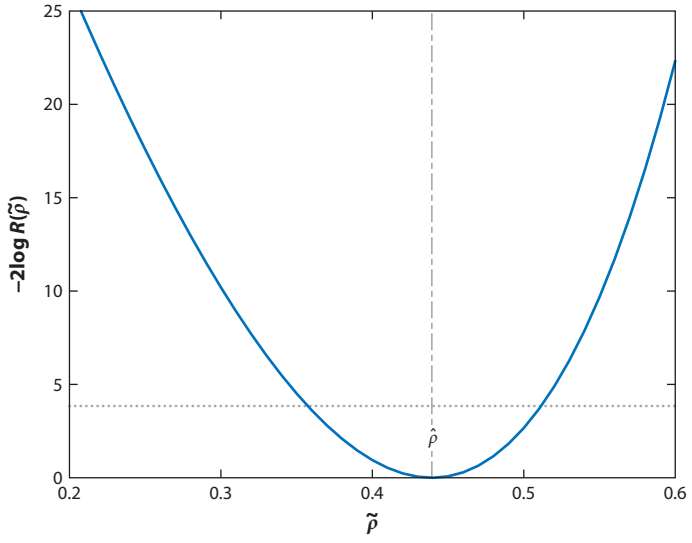

Figure 3

$-2 \log R(\tilde{\rho}, \boldsymbol{X}_{1:n})$ (*solid curve*) across different values of $\tilde{\rho}$ and a vertical line at $\hat{\rho}$ (*dashed*) based on a study of the dynamical correlation ρ (see Example 1) between two electrodes, each containing 256 electroencephalogram measurements from 76 alcoholic subjects. The horizontal dotted line is placed at the upper 0.05-quantile of $\chi^2(1)$.

common grid of time points across all subjects [see Cao (2018) for the R code for Sang et al. (2019)]. This advantage is also leveraged by Hu et al. (2018) in building confidence regions for the regression coefficients $\boldsymbol{\beta}$ in the PFLM (Equation 3) based on the $\chi^2(p)$ limiting distribution of the EL ratio statistic, under the cases of complete or missing responses. Hu & Liang (2022) further extend the work of Hu et al. (2018) to the case of the single-index PFLM:

$$Y = g(\boldsymbol{\beta}^\top \boldsymbol{Z}) + \int_0^1 \gamma(t)X(t)\mathrm{d}t + \varepsilon, \qquad 9.$$

with missing responses or/and parts of the covariates, where $g(\cdot)$ is an unknown link function. Li et al. (2021) circumvent the estimation of the complicated asymptotic variance in an extended Durbin–Watson statistic for Equation 3 with the error structure given in Equation 4 by developing an EL-based test for Equation 5 to determine the existence of serial correlation. Yang et al. (2023) further extend the work of Li et al. (2021) by relaxing the linearity assumption of the $\boldsymbol{\beta}^\top \boldsymbol{Z}$ term in Equation 3. Note that all the references in this paragraph face the challenge of infinite-dimensional nuisance parameters in $\boldsymbol{M}(\boldsymbol{X}, \cdot)$, which are estimated using plug-in, as mentioned in the first paragraph of Section 2.2.

The optimality of EL is typically reflected in smaller confidence regions or more powerful hypothesis tests over competing approaches, and confirmed by simulation studies. Depending on the specific problem at hand and the simulation settings, EL can have similar or better (i.e., closer to nominal level) empirical accuracy compared with Wald-type approaches, given moderate to large sample sizes. Among the references given in the previous paragraph, Sang et al. (2019) obtain better empirical coverages for the EL over the percentile bootstrap approach proposed by Dubin & Müller (2005), Hu et al. (2018) find that EL-based confidence intervals for scalar-valued β have shorter interval lengths than NA-based confidence intervals, and Hu & Liang (2022) show improved empirical coverage compared with NA approaches. As for serial correlation testing in Equation 5, Li et al. (2021) demonstrate similar empirical accuracy and power properties of EL and a Durbin–Watson-type test (that involves estimation of the complicated asymptotic variance) under Equation 3 with Equation 4's error structure. In contrast, Yang et al.'s (2023) EL extensions of the work of Li et al. (2021) have slightly better empirical power than the Ljung–Box test statistic (Ljung & Box 1978) under nonnormal white noise.

3. EMPIRICAL LIKELIHOOD–BASED INFERENCE FOR INFINITE-DIMENSIONAL FUNCTIONAL DATA ANALYSIS FEATURES

3.1. Examples of Infinite-Dimensional Features in Functional Data Analysis

Infinite-dimensional features in FDA arise in studies of functional means, functional ANOVA (analysis of variance), and functional regression. The following examples are chosen because EL-based methods are available (reviewed in Section 3.2). Note that similar issues explained in Section 2.1 for NA-based inference apply here, too.

Example 4 (Functional mean). For a random function $\{Y(t), t \in I\}$ in some chosen function space, a basic question is to infer about its mean function $\mu(t) = \mathrm{E}[Y(t)]$. This feature is infinite-dimensional because it is a function over I. Estimation, hypothesis testing, and confidence regions for mean functions have been proposed in the literature (see, e.g., Wang et al. 2016, and references therein). However, there are frequently range and shape restrictions of these target functions. For instance, a set of monotone random functions should have monotone $\mu(t)$. Another instance is the case of nonnegative random functions, where we would expect the confidence regions for $\mu(t)$ to stay nonnegative. Such range and shape

restriction is known to be difficult in NA-based inference, but EL provides a way to handle this naturally, as is seen in Section 3.2.

Example 5 (Functional ANOVA). In the setting of k independent samples for some positive integer $k > 1$, it is natural to conduct k-sample testing. In terms of the functional means, it is of interest to test $H_0 : \mu_1(\cdot) = \cdots = \mu_k(\cdot) \equiv \mu_0(\cdot)$ versus the omnibus alternative H_1. Here, we use the same notation as Example 4 except with a further subscript j indicating the jth sample, $j = 1, \ldots, k$. These functional means $\mu_1(\cdot), \ldots, \mu_k(\cdot)$ are infinite-dimensional features of interest, and such functional ANOVA testing has been considered by many authors (see, e.g., Zhang et al. 2019, Yuan et al. 2020). Again, an EL-based hypothesis test is called for to respect the range and monotonicity restriction in the functional data, if there is any. In addition, EL can provide better accuracy and power than less-than-optimal Wald-type tests.

Example 6 (Functional regression). A popular functional regression framework is the functional concurrent linear (FCL) model:

$$Y(t) = \boldsymbol{\beta}^\top(t)\boldsymbol{X}(t) + \varepsilon(t). \tag{10.}$$

In this setting, $Y(t)$ is a functional response, $\boldsymbol{X}(t)$ is a multivariate functional covariate, $\boldsymbol{\beta}(t)$ is a p-dimensional vector of unknown functions, and $\varepsilon(t)$ is a zero-mean error process independent of $\boldsymbol{X}(t)$. The FCL model is of great interest, for example, when exploring the link between psychological factors over time and concurrent physical activity in observational studies (see, e.g., Presseau et al. 2013). Although the regression coefficient function of interest, $\boldsymbol{\beta}(t)$, is infinite-dimensional, pointwise confidence intervals and testing may be of interest for $\boldsymbol{\beta}(t_0)$ at some given grid point $t_0 \in I$, as is often done as a preliminary step before studying simultaneous confidence regions/tests (see, e.g., Wu et al. 1998). EL is called for here to avoid estimation of bias and correlation terms and a predetermined symmetric shape of confidence regions in NA-based approaches. However, simultaneous inference for $\boldsymbol{\beta}(t)$ in Equation 10 has been proposed in the literature (see, e.g., Zhang & Chen 2007). In both cases, EL can be utilized to handle both sparse and dense functional data under a unified framework, and it enjoys optimality properties (see Section 1) that other such unified approaches do not have.

A typical starting point for relaxing the linear relationship between the response and covariate into a relationship characterized by an unknown index function $g(\cdot)$ is to employ a single-index model. An example of generalizing Equation 10 into such a single-index regression model is

$$Y(t) = g(\boldsymbol{\beta}^\top(t)\boldsymbol{X}) + \varepsilon(t), \tag{11.}$$

where the special case of nonfunctional covariate $\boldsymbol{X}$ is considered. Such a model is of interest in studying the association between imaging responses and covariates such as age and gender (Reiss et al. 2010). Under more complicated settings, such as in quantile regression and when $Y(t)$ are density functions, EL has been utilized to accommodate the within-curve correlation structure (X. Zhou, S. Ding, J. Wang, R. Liu, L. Kong & C. Huang, submitted manuscript).

Further relaxing the linearity assumption of the $\boldsymbol{\beta}^\top(t)\boldsymbol{X}(t)$ term in Equation 10, we arrive at the functional nonparametric regression model:

$$Y = r(\boldsymbol{X}) + \varepsilon, \tag{12.}$$

where $\boldsymbol{X} = \{\boldsymbol{X}(t), t \in I\}$, and Y and ε are scalar response and independent random error as in Equation 2. Extending Equation 12 to handle mixed-type predictors, the semifunctional

partially linear model is specified as

$$Y = \boldsymbol{\beta}^\top \boldsymbol{Z} + r(X) + \varepsilon, \tag{13.}$$

where $\boldsymbol{Z}$ is a p-dimensional covariate vector and $\boldsymbol{\beta}$ is a $p \times 1$ vector of unknown regression coefficients. Inference for the infinite-dimensional regression function $r(\cdot)$ is of interest, especially in Equation 12. However, mathematical challenges arise from the fact that this is a doubly infinite-dimensional situation—both the regression function and the covariate are infinite-dimensional. NA-based estimation of $r(\cdot)$ has been studied in the literature based on i.i.d. (Ferraty et al. 2007, Zhou & Lin 2016) or dependent data (Ferraty & Vieu 2004, Laib & Louani 2010, Lian 2012), with explicit expressions for the bias and variance terms available. However, these bias and variance terms involve unknown parameters that are difficult to estimate, and the resulting confidence regions can have poor empirical coverage rates (see, e.g., Lian 2012, Zhou & Lin 2016).

3.2. Empirical Likelihood Approach and Its Advantages

Let $\boldsymbol{\theta}(t)$ be an infinite-dimensional feature, and suppose we are interested in constructing a confidence region for the whole function $\boldsymbol{\theta}$, or in testing $H_0 : \boldsymbol{\theta} = \tilde{\boldsymbol{\theta}}$ versus $H_1 : \boldsymbol{\theta} \neq \tilde{\boldsymbol{\theta}}$, where $\tilde{\boldsymbol{\theta}}$ is a candidate value for $\boldsymbol{\theta}$. A powerful approach is first to construct a local EL ratio at each point $t \in I$, and then summarize the local EL statistic for each $t \in I$ in an overall test statistic. A motivation for this localization method comes from its effectiveness in non-FDA settings (see, e.g., Einmahl & McKeague 2003). Besides localization, another potential approach is to use the full nonparametric likelihood (over all t), as has been done in survival analysis (Dykstra 1982, Park et al. 2012). However, to our knowledge, this approach has not yet been adopted in FDA, whereas localization is known to be tractable (Wang et al. 2018, Chang & McKeague 2022).

For each given $t \in I$, the localization approach essentially treats $\boldsymbol{\theta}(t)$ in the same way as the EL-based inference for finite-dimensional features, as in Section 2.2. More specifically, EL handles the case when the truth $\boldsymbol{\theta}_0(t)$ satisfies an estimating equation $\mathrm{E}[\boldsymbol{M}(\boldsymbol{X}, \boldsymbol{\theta}_0)(t)] = 0$, where $\boldsymbol{M}(\boldsymbol{X}, \boldsymbol{\theta})(t)$ is a smooth function of the functional data $\boldsymbol{X} = \boldsymbol{X}(t)$ and the feature, such as $\boldsymbol{M}(\boldsymbol{X}, \boldsymbol{\theta})(t) = \boldsymbol{X}(t) - \boldsymbol{\theta}(t)$ for inference about the mean. EL-based local hypotheses are $H_0^t : \boldsymbol{\theta}(t) = \tilde{\boldsymbol{\theta}}(t)$ versus $H_1^t : \boldsymbol{\theta}(t) \neq \tilde{\boldsymbol{\theta}}(t)$. The local EL test relies on the local EL ratio statistic $-2 \log R(\tilde{\boldsymbol{\theta}}, \boldsymbol{X}_{1:n})(t)$ (see the last paragraph of this section for more details), a nonparametric version of the likelihood ratio statistic for testing the local hypotheses, where (recall from Section 2.2) $\boldsymbol{X}_{1:n} = \{\boldsymbol{X}_1(t), \ldots, \boldsymbol{X}_n(t), t \in I\}$ denotes the functional data at hand. This local test statistic is always nonnegative (because a likelihood ratio is bounded between 0 and 1), and it evaluates how far the candidate value $\tilde{\boldsymbol{\theta}}(t)$ of the feature deviates from the local truth $\boldsymbol{\theta}(t)$. For pointwise EL-based inference (existing methods are reviewed in Section 3.2.1), the local test statistic is compared with some cutoff value q_t as the upper α-quantile of some calibration/reference distribution, where (again, recall from Section 2.2) $\alpha \in [0, 1]$ is a user-chosen type I error rate. This reference can be obtained from the large-sample limit of $-2 \log R(\tilde{\boldsymbol{\theta}}, \boldsymbol{X}_{1:n})(t)$, which is typically chi-squared, or from repeated bootstrap samples from the original data (see the first paragraph of Section 2.2 for more details).

The EL-based pointwise confidence set [a set of possible $\boldsymbol{\theta}(t)$], $C(\boldsymbol{X}_{1:n})(t)$, is constructed like a localized confidence set/conformal set that contrasts a local score $s(\tilde{\boldsymbol{\theta}}, \boldsymbol{X}_{1:n})(t)$ with a threshold q_t':

$$C(\boldsymbol{X}_{1:n})(t) = \{\tilde{\boldsymbol{\theta}}(t) : s(\tilde{\boldsymbol{\theta}}, \boldsymbol{X}_{1:n})(t) \leq q_t'\}. \tag{14.}$$

For EL, we set the score $s(\tilde{\boldsymbol{\theta}}, \boldsymbol{X}_{1:n})(t)$ in Equation 14 to be the local EL ratio statistic $-2 \log R(\tilde{\boldsymbol{\theta}}, \boldsymbol{X}_{1:n})(t)$. Setting q_t' as q_t typically leads to an asymptotic level $1 - \alpha$ guarantee as

$\mathrm{P}(\boldsymbol{\theta}_0(t) \in C(\boldsymbol{X}_{1:n})(t)) \to 1-\alpha$; that is, the probability that the pointwise confidence set contains the true local value of the feature is (asymptotically) $1-\alpha$. Note that the EL-based pointwise confidence set can be used to test H_0^t versus H_1^t, too, by using the rejection region $\{\boldsymbol{\theta}_0(t) \notin C(\boldsymbol{X}_{1:n})(t)\}$.

As for simultaneous EL-based inference (existing methods reviewed in Section 3.2.2), the local statistics $-2\log R(\tilde{\boldsymbol{\theta}}, \boldsymbol{X}_{1:n})(t)$ across $t \in I$ are combined into the following maximally selected or integrated EL statistics:

$$K_n = \sup_{t\in I}\left[-2\log R(\tilde{\boldsymbol{\theta}}, \boldsymbol{X}_{1:n})(t)\right], \quad I_n = \int_{t\in I}\left[-2\log R(\tilde{\boldsymbol{\theta}}, \boldsymbol{X}_{1:n})(t)\right]\mathrm{d}\nu(t). \tag{15.}$$

Here, ν is some measure that controls the weighting for $-2\log R(\tilde{\boldsymbol{\theta}}, \boldsymbol{X}_{1:n})(t)$ at each $t \in I$; we recommend avoiding ad hoc choices of $\nu(t)$ by setting $\nu(t) = \hat{F}(t)$ that corresponds to equal weight for data with no ties, where $\hat{F}(t)$ is the empirical cumulative distribution function of $\boldsymbol{X}_{1:n}(t)$. This is because the local EL statistic implicitly provides optimal (i.e., nonparametric-likelihood-optimized) weighting for the target feature (Chang et al. 2021). The K_n statistic is also known as a maximal-deviation-type or Kolmogorov–Smirnov-type statistic (like the commonly used Kolmogorov–Smirnov statistic for cumulative distribution distributions) and is sensitive to any local difference between the target function and $\tilde{\boldsymbol{\theta}}(\cdot)$, with a larger value indicating larger local difference. In contrast, the I_n statistic is an integral-type statistic and can detect differences [between the target function and $\tilde{\boldsymbol{\theta}}(\cdot)$] dispersed over the range of indices, with a larger value indicating a larger dispersed difference. The choice of which statistic to use should be made a priori, guided by prior knowledge and practical considerations. In particular, if prior knowledge does not suggest the presence of local differences, we recommend I_n for general use. Otherwise, K_n can be implemented to exploit the additional knowledge of the existence of a local difference. For simultaneous EL tests, the test statistics K_n/I_n are compared with some cutoff value q_K/q_I as the upper α-quantile of some calibration/reference distribution. This distribution can be obtained from the large-sample limit of K_n and I_n, or from bootstrap.

The EL-based simultaneous confidence region [a set of possible $\boldsymbol{\theta}(\cdot)$], $C(\boldsymbol{X}_{1:n})$, can be constructed based on K_n (not I_n). In detail, it is a collection of the localized confidence set/conformal set that contrasts the same local score $s(\tilde{\boldsymbol{\theta}}, \boldsymbol{X}_{1:n})(t)$ with a less stringent threshold q' than the q'_t for pointwise inference to account for multiple type I errors that arise when there is more than one $t \in I$:

$$C(\boldsymbol{X}_{1:n}) = \{(t, \tilde{\boldsymbol{\theta}}(t)) : \sup_{t\in I} s(\tilde{\boldsymbol{\theta}}, \boldsymbol{X}_{1:n})(t) \le q'\}. \tag{16.}$$

Setting q' as q_K typically leads to an asymptotic level $1-\alpha$ guarantee as $\mathrm{P}((t, \tilde{\boldsymbol{\theta}}(t)) \in C(\boldsymbol{X}_{1:n})$ for all $t \in I) \to 1-\alpha$; that is, the probability that the confidence set contains all true local values of the feature is (asymptotically) $1-\alpha$. Note that the EL-based simultaneous confidence set can be trivially extended to simultaneous testing for H_0 versus H_1, by using the rejection region $\{(t, \tilde{\boldsymbol{\theta}}(t)) \notin C(\boldsymbol{X}_{1:n})$ for some $t \in I\}$. Note the distinction between this and the K_n test mentioned earlier in this paragraph, with the rejection region $\{K_n > q_K\}$. While the latter is easier to construct, it does not readily produce the aforementioned EL-based simultaneous confidence region. This is because the construction/computation of a confidence set is more challenging than that of the corresponding EL test. The difference here is that the construction involves solving for the roots $\boldsymbol{\theta}_B(t)$ of $-2\log R(\boldsymbol{\theta}_B, \boldsymbol{X}_{1:n})(t) = q_K$ when finding the boundary of the confidence region.

Next we discuss the construction of the local EL statistic, $-2\log R(\tilde{\boldsymbol{\theta}}, \boldsymbol{X}_{1:n})(t)$, which is vital in both the EL-based pointwise and simultaneous inference discussed above. We proceed by treating $\boldsymbol{\theta}(t)$ similarly to the EL-based inference for finite-dimensional features as in Section 2.2, except for the derivation of the limiting distribution. More specifically, the local EL ratio $R(\tilde{\boldsymbol{\theta}}, \boldsymbol{X}_{1:n})(t)$ is

just like a parametric likelihood ratio, with the numerator maximizing a nonparametric likelihood subject to constraint under H_0^t, and the denominator maximizing the nonparametric likelihood subject to constraint defined by the union of the spaces under H_0^t and H_1^t. That is, the EL ratio for $\boldsymbol{\theta}(t)$ at a given $t \in I$ is

$$\mathcal{R}(\tilde{\boldsymbol{\theta}}, \boldsymbol{X}_{1:n})(t) = \frac{\sup\{L(F_t) : E_{F_t}[\boldsymbol{M}(\boldsymbol{X}, \tilde{\boldsymbol{\theta}})(t)] = 0, F_t \in \mathcal{F}_t\}}{\sup\{L(F_t) : F_t \in \mathcal{F}_t\}}, \tag{17.}$$

where $F_t(\cdot)$ is a candidate for the cumulative distribution function of $\boldsymbol{X}(t)$, $\mathcal{F}_t$ is the set of distributions supported on the data $\boldsymbol{X}_{1:n}(t)$,

$$L(F_t) = \prod_{i=1}^{n} p_i(t)$$

is the nonparametric likelihood, $p_i(t)$ is the point mass of $F_t(\cdot)$ at $\boldsymbol{X}_i(t)$, and again we follow the convention $\sup \emptyset = 0$. As mentioned in Section 1, Equation 17 indeed avoids assuming specific forms in relating the likelihood to the target feature by optimizing the nonparametric likelihood $L(F_t)$ subject to constraints imposed on the target feature. This flexibility may look too good to be true at first, and one may doubt its implementability. Note that the denominator of Equation 17 is again $\prod_{i=1}^{n}(1/n)$. The numerator can be shown by Lagrange's method to be $\prod_{i=1}^{n} \tilde{p}_i(t)$, where $\tilde{p}_i(t) = \{n[1 + \tilde{\boldsymbol{\lambda}}^\top(t)\tilde{\boldsymbol{M}}_i(t)]\}^{-1}$, $\tilde{\boldsymbol{M}}_i(t) = \boldsymbol{M}(\boldsymbol{X}_i, \tilde{\boldsymbol{\theta}})(t)$, and $\tilde{\lambda}(t)$ satisfies the estimating equation $\sum_{i=1}^{n} \tilde{p}_i(t)\tilde{\boldsymbol{M}}_i(t) = 0$. It follows that

$$-2 \log \mathcal{R}(\tilde{\boldsymbol{\theta}}, \boldsymbol{X}_{1:n})(t) = 2 \sum_{i=1}^{n} \log\left[1 + \tilde{\boldsymbol{\lambda}}^\top(t)\tilde{\boldsymbol{M}}_i(t)\right].$$

By the same reasoning as in EL for a multivariate mean (Owen 2001, p. 70), any feasible solution for $\tilde{\boldsymbol{\theta}}(t)$ in computing $\mathcal{R}(\tilde{\boldsymbol{\theta}}, \boldsymbol{X}_{1:n})(t)$ lies in the convex hull of $\{\tilde{\boldsymbol{M}}_1(t), \ldots, \tilde{\boldsymbol{M}}_n(t)\}$; this is what we mean by EL respecting the range restrictions inherent in the data, as mentioned in Section 1. The limiting distribution of $-2 \log R(\tilde{\boldsymbol{\theta}}, \boldsymbol{X}_{1:n})(t)$ can typically be expressed in terms of a zero-mean Gaussian process (that is marginally chi-squared). Remarkably, this pointwise nonparametric analogue of Wilks's theorem holds across various kinds of problems (some examples are provided in the next two sections), rendering the calibration of the pointwise confidence set in Equation 14 quite standard—just take q_t to be the upper α-quantile of chi-squared. Note, however, that there are exceptions to this pointwise chi-squared behavior, similar to those discussed near the end of the first paragraph in Section 2.2. As for the simultaneous confidence set, bootstrap methods have been applied to approximate the entire limit of $\{-2 \log R(\tilde{\boldsymbol{\theta}}, \boldsymbol{X}_{1:n})(t), t \in I\}$ (Wang et al. 2018, Chang & McKeague 2022). The latter reference even does so without having to estimate the covariance function of the limiting Gaussian process.

3.2.1. Pointwise empirical likelihood–based inference. The above EL-based inference does not involve variance estimation, as mentioned in Section 2.1. To avoid estimating the bias and correlation terms involved in NA-based inference in Equation 10, Xue & Zhu (2007) establish EL-based pointwise confidence intervals regarding $\boldsymbol{\beta}(t_0)$ in Equation 10 for some given grid point $t_0 \in I$, in the case of sparse functional data. Wang et al. (2018) consider EL-based pointwise testing and confidence intervals that extend the work of Xue & Zhu (2007) to the case of dense functional data and handle both sparse and dense functional data under a unified EL framework, thanks to the self-normalization property of EL. Note that both Xue & Zhu (2007) and Wang et al. (2018) construct simultaneous inference after these pointwise inferential considerations, but discussion of these more general results is deferred to Section 3.2.2. Lian (2012) develops EL-based pointwise

confidence intervals for the regression function r in Equations 12 or 13 at a fixed covariate value x_0, to avoid estimating the bias and variance involved in NA-based inference in both models, while allowing for weakly dependent data. Xiong & Lin (2013) also study EL-based pointwise confidence intervals for the regression function r in Equation 12 at a fixed covariate value x_0, but for a different type of dependent data, to avoid the poor convergence properties of NA-based confidence regions and the use of bootstrap that needs additional theoretical justification. Zhou & Lin (2016) construct another such EL-based pointwise confidence interval, but using a better type of smoother (see Section 4), to avoid estimating the unknown parameters involved in NA-based inference. All the references in this paragraph deal with the presence of nuisance parameters (including bias correction terms for smoothing approaches) in $\boldsymbol{M}(\boldsymbol{X}, \cdot)$ by plug-in estimates, as mentioned in the first paragraph of Section 2.2.

The EL-based confidence regions in Section 3.2 have data-driven shapes that are not predetermined like estimate ± error in NA-based confidence intervals. This can be seen in the work of Lian (2012, figure 2), where NA-based confidence regions are simply ellipses, while the EL-based regions have more flexible shapes (determined by the data). This advantage has been emphasized by Xue & Zhu (2007), Lian (2012), and Zhou & Lin (2016). Lian (2012) attributes the better empirical performance of EL-based confidence intervals over NA-based ones to the data-dependent shape of the EL confidence regions.

As for the optimality and improved accuracy of EL, Xue & Zhu (2007) show in simulation studies that EL-based pointwise confidence intervals for $\boldsymbol{\beta}(t_0)$ in Equation 10, for some given grid point $t_0 \in I$, are shorter and achieve higher empirical coverage than NA-based and bootstrap methods. On the other hand, Lian (2012) and Zhou & Lin (2016) both find that the EL-based confidence intervals for the nonparametric regression function r at a fixed covariate value x_0 are shorter (i.e., more efficient) and have better coverage (i.e., more accurate) compared with NA-based methods. Note that such results from Wang et al. (2018) are in terms of simultaneous inference and are discussed in the next section.

3.2.2. Simultaneous empirical likelihood–based inference. While EL-based simultaneous inference may involve variance estimation, as explained in Section 3.1, it still enjoys the advantage of avoiding a predetermined shape for the confidence region. In addition, EL-based simultaneous testing has been shown to adapt well to heteroscedasticity and imbalance in the sample sizes if there are multiple groups (Chang & McKeague 2022). Using the Bonferroni adjustment, Xue & Zhu (2007) established EL-based simultaneous confidence bands regarding $\boldsymbol{\beta}(t)$ in Equation 10, in the case of sparse functional data. Without using Bonferroni adjustment, Wang et al. (2018) consider EL-based simultaneous testing based on integrating the local EL statistics via a Cramér–von Mises-type formulation, for both sparse and dense functional data. Chang & McKeague (2022) introduce EL-based simultaneous confidence bands and k-sample simultaneous testing for functional means, in the presence of potential nonsmoothness in functional data/means/covariances; Chang (2022) provides an R package for implementing the methods proposed by Chang & McKeague (2022). They emphasize the advantage of data-driven shapes of EL-based confidence regions, in terms of respecting range and monotonicity constraints inherent in the data. This means EL is well suited to the task of analyzing shape-constrained functional data, such as bounded and nonincreasing data (see **Figure 2** for an example). Zhou et al. (X. Zhou, S. Ding, J. Wang, R. Liu, L. Kong & C. Huang, submitted manuscript) consider an EL-based estimator of $\boldsymbol{\beta}(t)$ in a density-on-scalar single-index quantile regression model, generalizing Equation 11 into quantile regression with $Y(t)$ as density functions. However, these authors only derive a Wald-type simultaneous confidence region for $\boldsymbol{\beta}(t)$, failing to utilize the benefit of data-driven shapes of EL-based confidence sets. Except for Chang & McKeague (2022), all the other references in this paragraph

deal with the presence of nuisance parameters (including bias correction terms for smoothing approaches) in $\boldsymbol{M}(\boldsymbol{X}, \cdot)$ by plug-in estimates.

As for the optimality and improved accuracy of EL shown in simulation studies, Xue & Zhu (2007) find that EL-based confidence bands for $\boldsymbol{\beta}(t)$ in Equation 10 are narrower than those found using NA-based and bootstrap methods. However, since these bands are all based on Bonferroni adjustment that is known to be conservative, they yield high coverage probabilities that are close to 1 for all the methods. Wang et al. (2018) find that their proposed EL-based simultaneous test for linear hypotheses regarding $\boldsymbol{\beta}(t)$ in Equation 10 has similar empirical accuracy but better power than an alternative non-EL-based bootstrap method (Zhang & Chen 2007) in the case of dense functional data. Chang & McKeague (2022) show that their proposed simultaneous test maintains empirical accuracy while being more powerful than other Wald-type functional ANOVA tests, and their proposed confidence band has more accurate coverage while being narrower than alternative Wald-type approaches.

4. SMOOTH VERSUS NONSMOOTH FRAMEWORKS IN FUNCTIONAL DATA ANALYSIS

FDA posits that the data, if fully observed, are d-dimensional random functions $\{X_1(t), \ldots, X_n(t), t \in I\}$. However, in practice, instead of fully observed trajectories, we can only observe the $X_i(t)$ on a (not necessarily equispaced) finite grid of points in I. For FDA under fixed design, the grid is common across the study units and is denoted $\mathbf{G}_n$. In contrast, random design allows the grid to be different across study units, and the grid is denoted $\mathbf{G}_{i,n}$. Note that the dependence of $\mathbf{G}_n$ or $\mathbf{G}_{i,n}$ on the sample size n is to facilitate asymptotic analysis. A common condition in FDA is that the mesh of $\mathbf{G}_n$ or $\mathbf{G}_{i,n}$ (the maximal distance between adjacent grid points) approaches zero as more data become available ($n \to \infty$). For simplicity of exposition, we consider $d = 1$ and fixed design in the following; generalization to $d > 1$ and random design is straightforward.

For reasons explained in the next paragraph, there is a need to reconstruct $X_i(t)$ for all $t \in I$ using the $X_i(t)$ observed only at $t \in \mathbf{G}_n$. The reconstruction is conducted based on smooth or nonsmooth frameworks in FDA, which differ in that the resulting $\{X_i(t), t \in I\}$ typically looks smoother in smooth frameworks than in nonsmooth ones; see **Figure 4**, which contrasts the two methods based on the EEG data analyzed by Sang et al. (2019, section 5.2). The collection of rebuilt data is then treated as the functional data $\boldsymbol{X}_{1:n}$ in computing the EL ratio in Sections 2.2 and 3.2. Note that there are some methods that do not seem to reconstruct the data explicitly, which is accomplished by working with the full trajectories directly in the theoretical development. These methods, however, typically still need a reconstruction step in implementation (see, e.g., Cuevas et al. 2004). Therefore, we suggest incorporating this step into theory to assess the quality of and the underlying assumptions needed for the (chosen) reconstruction (see, e.g., Zhang et al. 2019).

4.1. Smoothing Approach

Many FDA methods use smoothing for solving the following two major issues. First, since treating each measurement of a study unit (i.e., $X_i(t)$ at a $t \in \mathbf{G}_n$) as an isolate random variable leads to high dimensionality, there is a need to pool information from neighborhood measurements for dimension reduction (Wang et al. 2016). Second, it is sometimes desirable to dampen fluctuations of the observations across the indices, in particular measurement errors. More specifically, each (random) function $\{X_i(t), t \in I\}$ is modeled as a linear combination of basis functions:

$$X_i(t) = \mu(t) + \sum_{k=1}^{\infty} A_{ik}\phi_k(t), \qquad 18.$$

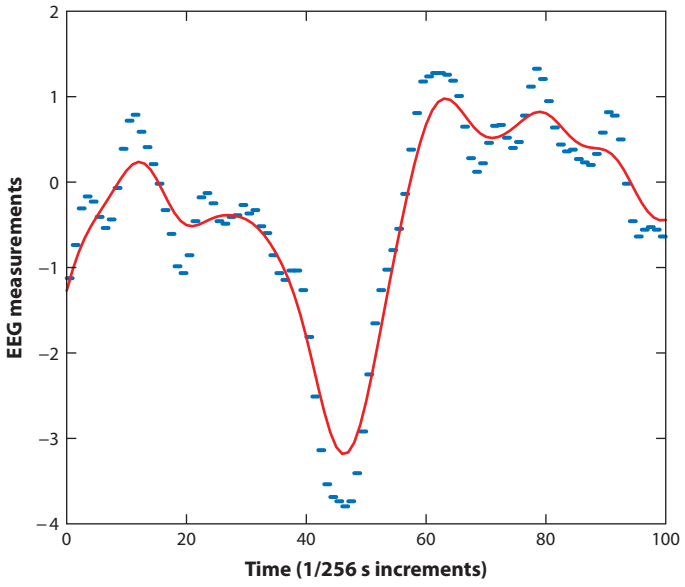

Figure 4

Example of smooth (*red*) versus nonsmooth (*blue*) reconstruction of functional data based on 256 electroencephalogram (EEG) measurements of the C5 electrode from the first alcoholic subject from the study introduced in Section 2.2.

where A_{ik} are the random coefficients, and $\{\phi_k(t), k = 1, \ldots\}$ is a set of smooth basis functions of the function space $L^2(I)$. Equation 18 is often approximated by

$$X_i(t) = \mu(t) + \sum_{k=1}^{K} A_{ik}\phi_k(t) \qquad 19.$$

for large enough K, where K is determined by some model selection criterion. This way, the function $\{X_i(t), t \in I\}$ can be expressed as a finite vector of uncorrelated random variables $A_{i1}, \ldots, A_{iK}$, thereby achieving the purpose of dimension reduction. Analogous dimension reduction can be conducted on the (nonrandom) coefficient functions in Equations 2, 3, 9, and 10, but in this case the corresponding scores are nonrandom. These smoothed terms will typically be included in the resulting EL ratio construction.

Various choices of the basis functions, including spline, Fourier, or wavelet bases, have been proposed in the literature. Ideally, one should choose the smoother to reflect features of the functional data of interest (Ramsay & Silverman 2005), such as Fourier smoothers for cyclical or periodic functional data, splines for noncyclical nonperiodic data, and wavelet bases for data displaying rapid changes in behavior (see, e.g., Ullah & Finch 2013, and references therein). An inappropriate choice would require a large number of basis functions to adequately model the data, leading to an unnecessary dimensionality increase (see, e.g., Ratcliffe et al. 2002).

4.2. Nonsmoothing Approach

The two issues above can be solved by carefully devised nonsmooth methods, where by nonsmooth approaches we mean interpolation (see, e.g., Cai & Yuan 2011), where the interpolated data at $t \in \mathbf{G}_n$ remain the same as the original $X_i(t)$. For the first issue, there are ways to pool information via piecewise-constant, linear (see, e.g., Yuan et al. 2020), or other interpolation approaches instead

of smoothing. The benefits of these nonsmoothing approaches include a more flexible framework, weaker assumptions, and elimination of smoothing-related bias correction terms, compared with existing approaches in the literature. For example, the piecewise-constant interpolation of Chang & McKeague (2022) allows discontinuities in the functional means and covariances while accounting for dense discretization of the observed trajectories, in contrast to the existing literature (see, e.g., Degras 2011, Cao et al. 2012, Choi & Reimherr 2018, Zhang et al. 2019). For the second issue, if the functional data at hand are smooth already and an explicit modeling of measurement error is not necessarily implemented in the relevant subject matter area, smoothing may not be necessary. Not implementing smoothing does not mean measurement errors cannot be taken into account, but just that they are incorporated implicitly into the covariance function of the limiting distributions of the relevant statistics. More specifically, instead of assuming a model for each (random) function $\{X_i(t), t \in I\}$ as in Equation 18, each observation is viewed as a discretized version of the fully observed trajectory, denoted as $f_n(X_i)$. Here, $f_n(g)$ is the discretization of a function $g : [\alpha_1, \alpha_2] \to \mathbb{R}^d$ defined by

$$f_n(g)(a) = \begin{cases} g(a), & a \in \mathbf{G}_n \\ h_a(g(b_1(a)), \ldots, g(b_m(a))), & a \in [\alpha_1, \alpha_2] \setminus \mathbf{G}_n, \end{cases} \quad 20.$$

where $b_\ell(a)$ is the ℓth point based on which we interpolate the value at a, m is the number of such points, and $h_a(\cdot)$ is a function that determines how we interpolate. For example, piecewise-constant interpolation from the right would involve $m = 1$, $h_a(x) = x$, and $b(a)$ as the closest point on the grid to the right of a. Linear interpolation would involve $m = 2$, $b_1(a)/b_2(a)$ being the closest points on the grid to the left/right of a, and $h_a(g(b_1(a)), g(b_2(a))) = g(b_1(a)) + [a - b_1(a)][g(b_2(a)) - g(b_1(a))]/[b_2(a) - b_1(a)]$. Since Equation 20 does not involve dimension reduction as in Equation 19, there is no loss of information from smoothing/dimension reduction (Dette et al. 2020).

Analogous discretization can be conducted on the estimated infinite-dimensional features in Section 3.1, the relevant terms inside the EL ratio, and even the EL ratio itself as a function of t in Equation 17 to characterize the discretized version of the fully observed trajectory of the estimated features and the EL ratio. Choices of b_a depend on the type of interpolation in use (for more examples, see the last paragraph of Section 4.3). As with issues involved in choosing an appropriate smoother, here one should choose the interpolation method that reflects features of the functional data being studied. For example, if the functional data include points of discontinuity from the left, then piecewise-constant interpolation from the right would be preferable to linear interpolation.

Nonsmooth FDA methods differ from smoothing approaches in FDA based on wavelet bases in the following two aspects. First, the former do not involve a tuning parameter (and hence its selection) for controlling the degree of smoothness in the resulting estimator. Second, although wavelet methods are known to capture rapid changes in a function, such a guarantee may still hinge on a certain degree of smoothness, such as continuity or Lipschitz continuity (see, e.g., Antoniadis et al. 1994). In contrast, nonsmooth FDA methods can completely do without a continuity assumption on the target function (see, e.g., Chang & McKeague 2022).

Although nonsmooth methods can accommodate discontinuities, they differ from the approach used in change point problems. The reasons are as follows. First, nonsmooth methods are not designed for estimating/selecting the change points per se (see, e.g., Bai 2010). The goal is to conduct inference for the target function (be it the mean function or coefficient function) of interest. Even if an attempt is made to modify a change point method to do this, it would be necessary to treat those change points as plug-in estimates. Such an approach would lead to a much more complicated theory. Second, nonsmooth methods typically involve assumptions that are different from certain assumptions, such as parametric models (see, e.g., Bélisle et al. 1998), smooth conditional covariances (see, e.g., Zhu et al. 2014), or stationarity (see, e.g., Xi & Pang 2021), that are used in

the change point literature. Last but not least, nonsmooth methods address an important feature of FDA that treats the observed trajectories as discretized versions of an underlying trajectory and is able to make inference about the complete trajectory. In contrast, the change point literature is not known to do that.

4.3. Smoothing and Nonsmoothing Approaches Used in Sections 2 and 3

Many of the articles cited in Section 2 employ smoothing. For the problem of dynamical correlation, Dubin & Müller (2005) use a linear smoother (including kernel and local linear smoothers) and Sang et al. (2019) employ local linear regression to smooth each random function, which is assumed to be twice differentiable with certain equi-continuity assumptions on the second derivatives. The advantage of using local linear smoothing has been shown in finite-dimensional data to include high minimax efficiency (Fan 1993), bias reduction from kernel smoothing, and adaptation to edge effects (Fan 1992). For inference about PFLMs, Shin (2009) and Hu et al. (2018) expand the covariate function $X(\cdot)$ and the coefficient function $\gamma(\cdot)$ of Equation 3 in the same eigenbasis, assuming the covariance function of $X(\cdot)$ is continuous. The use of such an expansion is called functional principal components analysis (FPCA). The advantage is that among all basis expansions that use K components for a fixed K, the FPCA expansion explains most of the variation in $X(\cdot)$ in the L^2 sense. Hu & Liang (2022) use B-spline approximation for the slope function $\gamma(t)$ in Equation 9, assuming continuity of at least the second derivatives of $\gamma(t)$. For FCL regression in Equation 10, Xue & Zhu (2007) focus their theoretical development on inferring $\boldsymbol{\beta}(t_0)$ for some given grid point $t_0 \in I$ using kernel smoothing, assuming commonly used smoothness conditions such as that the covariance function of $\boldsymbol{X}(t)$ has continuous first derivatives at t_0, and that $\boldsymbol{\beta}(t)$ has continuous second derivatives at t_0. Li et al. (2021) handle Equation 3 with Equation 4's error structure using the same smoothing approach for PFLM as that of Shin (2009). The same smoothing approach and smoothness assumption on the covariance function of $X(\cdot)$ are also used by Yang et al. (2023) when generalizing the model considered by Li et al. (2021).

Many of the papers cited in Section 3 also employ smoothing. Wang et al. (2018) build their theory on the residual-adjusted estimating equations proposed by Xue & Zhu (2007) based on kernel smoothing, with stronger smoothness assumptions for handling $\boldsymbol{\beta}(t)$ uniformly over $t \in I$: The covariance function of $\boldsymbol{X}(t)$ is twice continuously differentiable, and $\boldsymbol{\beta}(t)$ is three-times continuously differentiable. For weakly dependent functional data, Lian (2012) applies kernel smoothing to estimating equations for inferring the regression function r in Equations 12 or 13 at a fixed covariate value x_0, assuming that $r(\cdot)$ is Lipschitz continuous and the noise variance function $\sigma^2(X) = \mathrm{Var}(\varepsilon|X)$ is continuous in a neighborhood of x_0. For functional stationary ergodic data, Xiong & Lin (2013) apply kernel smoothing to estimating equations for inferring the regression function r in Equation 12 at a fixed covariate value x_0, assuming that $r(\cdot)$ is Hölder continuous and that the noise variance function and the conditional fourth moment $\mathrm{E}(\varepsilon^4|X = x)$ is continuous in a neighborhood of x_0. Zhou & Lin (2016) use local linear smoothing to handle the same nonparametric regression problem as Xiong & Lin (2013) with a smaller bias term, using similar continuity assumptions above. Zhou et al. (X. Zhou, S. Ding, J. Wang, R. Liu, L. Kong & C. Huang, submitted manuscript) use a boundary-corrected kernel density estimator for the density-function-valued response $Y(t)$, which has the advantage of converging to the true density function uniformly (Petersen & Müller 2016) compared with traditional kernel density estimators.

A few of the references in Section 3 examine nonsmooth methods. Cai & Yuan (2011) consider discretely observed functional data using spline interpolation (rather than piecewise-constant interpolation) to reconstruct the data between fixed common design points, and they show that the minimax convergence rate (for the mean) does not improve with the additional smoothing. Other nonsmooth methods for dense functional data include linear interpolation for

two-sample tests (Yuan et al. 2020) and direct discretization for ANOVA tests (Zhang et al. 2019). However, these methods require the assumption of smooth covariance functions, which is not needed in the following approach. Concerning functional means, Chang & McKeague (2022) use piecewise-constant interpolation and allow discontinuities in the underlying functional means and covariances, in contrast to the existing literature (see, e.g., Degras 2011, Cao et al. 2012, Choi & Reimherr 2018, Zhang et al. 2019) that deals with dense designs. Their simulations show that when there is a discontinuity in the functional mean and covariance, the EL band maintains accuracy, whereas alternative bands based on smoothing can have severe undercoverage.

5. A LOOK TO THE FUTURE

While EL has improved FDA in various aspects, such as achieving pointwise optimality, circumventing the estimation of complicated bias/variance terms in NA-based inference, and providing data-driven shapes of confidence regions, it has been applied to only limited examples (see Examples 1–6) resembling those illustrated in this article. This is perhaps due to the perceived computational issues involved in using EL on FDA, as mentioned in the third to last paragraph of Section 1. However, as shown in Section 3.2, even in the most demanding case of simultaneous inference of infinite-dimensional features, the computation of the local EL statistics is essentially the same as in the case of finite-dimensional features, and the reference distribution for simultaneous inference can be obtained via a bootstrap. This implies that the only additional effort needed to apply EL in FDA is in deriving the theoretical limiting distribution for simultaneous inference. Since this derivation typically involves steps that are quite standard in EL, we conclude that EL in FDA is not much more challenging than EL for traditional univariate/multivariate data or than ordinary FDA methods. Therefore, we advocate the use of EL in FDA to exploit its many benefits, thereby improving the accuracy and efficiency of the statistical inference.

The major advantage of EL in building more efficient confidence regions and hypothesis tests can be apparent in high-risk settings, such as medical diagnostics, where parameter/model estimation is insufficient, whereas prediction with uncertainty quantification plays a more essential role. This aligns with current needs in statistics and machine learning. While the literature has focused on EL-based inference, prediction has been less studied. In fact, prediction naturally fits into the EL framework. Considering the case of finite-dimensional features (see Section 2.2; similar arguments work for EL for infinite-dimensional features, as in Section 3.2.2), the predictive EL statistic can be constructed by modifying Equation 7 into

$$-2\log R((\boldsymbol{X}_{n+1},\boldsymbol{y});(\boldsymbol{X}_1,\boldsymbol{Y}_1),\ldots,(\boldsymbol{X}_n,\boldsymbol{Y}_n)) \qquad 21.$$

for testing $H_0:\boldsymbol{Y}_{n+1}=\boldsymbol{y}$ (see, e.g., Owen 2001, section 4.9), where the term before the semicolon denotes the variables relevant to prediction, and the term after the semicolon indicates the training dataset. To our knowledge, however, no work has been done in EL-based prediction for FDA. One promising future direction is to integrate EL-based prediction for FDA with conformal prediction (see, e.g., Lei et al. 2018). For such EL-based conformal prediction, one can potentially use a version of the above predictive EL statistic based on leaving out one unit of the augmented functional dataset $(\boldsymbol{X}_1,\boldsymbol{Y}_1),\ldots,(\boldsymbol{X}_n,\boldsymbol{Y}_n),(\boldsymbol{X}_{n+1},\boldsymbol{y})$ (after the semicolon of Equation 21) for the prediction tasks on the left-out functional observation $(\boldsymbol{X}_i,\boldsymbol{Y}_i)$ for $i=1,\ldots,n$ or $(\boldsymbol{X}_{n+1},\boldsymbol{y})$ (before the semicolon of Equation 21). The resulting EL ratios can be used as calibration scores $s_{y,i}$ for $i=1,\ldots,n$ and $s_{y,n+1}$, respectively, in the context of conformal inference. Since Equation 21 is a function symmetric in its last n arguments, finite-sample (nonasymptotic) validity may be obtained as in typical conformal prediction. Further research is needed to see whether predictive EL continues to enjoy the many optimality properties of EL (as mentioned in Section 1), thereby outperforming competing choices of conformal scores.

Other possible future directions include developing EL methods for more complicated functional data, such as those involving random objects in metric spaces with little structure (Petersen & Müller 2019) or involving censoring (Yang et al. 2021). We anticipate EL will provide better inferential procedures in these cases, because as we have seen in Section 2.1, the advantage of EL is even more apparent in settings where the bias/variance estimation in NA-based inference is problematic. Another possible direction is to investigate optimality results for EL-based simultaneous inference in FDA. Such a study would further promote the use of EL in FDA, although the optimality properties of EL mentioned in the second paragraph of Section 1 are already applicable to EL-based pointwise inference in FDA. Still another important direction is to develop more efficient computational techniques and software packages for EL in FDA. This research direction can leverage existing EL-based methods for univariate/multivariate data or non-EL-based methods for functional data, which are relatively more developed.

Although varying-coefficient models contain functional parameters and are closely related to Equation 10, these are not covered in this review. The reason is that the data involved per study unit are typically multivariate rather than functional (see, e.g., Ramsay & Silverman 2005, p. 259). Note that there are varying-coefficient situations where each study unit contains functional data, such as the functional varying-coefficient single-index model considered in Li et al. (2017). These can be considered as special cases or extensions of Equation 10 discussed in this article.

DISCLOSURE STATEMENT

The authors are not aware of any affiliations, memberships, funding, or financial holdings that might be perceived as affecting the objectivity of this review.

ACKNOWLEDGMENTS

We thank Chi-Ru Fang for help with literature search and figure preparation. H.-w.C. was supported by Academia Sinica under award AS-CDA-113-M03 and by the National Science and Technology Council of Taiwan under grant 12-2118-M-001-006 and 113-2118-M-001-014-MY2. I.W.M. was supported by the National Institutes of Health (NIH) under award 1R01 AG062401 and by the National Science Foundation (NSF) under award DMS-2112938.

LITERATURE CITED

Anastasiou A, Reinert G. 2020. Bounds for the asymptotic distribution of the likelihood ratio. *Ann. Appl. Probab.* 30(2):608–43

Antoniadis A, Gregoire G, McKeague IW. 1994. Wavelet methods for curve estimation. *J. Am. Stat. Assoc.* 89(428):1340–53

Bahadur RR. 1967. Rates of convergence of estimates and test statistics. *Ann. Math. Stat.* 38(2):303–24

Bai J. 2010. Common breaks in means and variances for panel data. *J. Econom.* 157(1):78–92

Bélisle P, Joseph L, MacGibbon B, Wolfson DB, du Berger R. 1998. Change-point analysis of neuron spike train data. *Biometrics* 54(1):113–23

Bravo F. 2003. Second-order power comparisons for a class of nonparametric likelihood-based tests. *Biometrika* 90(4):881–90

Cai TT, Yuan M. 2011. Optimal estimation of the mean function based on discretely sampled functional data: phase transition. *Ann. Stat.* 39(5):2330–55

Cao G, Yang L, Todem D. 2012. Simultaneous inference for the mean function based on dense functional data. *J. Nonparametric Stat.* 24(2):359–77

Cao J. 2018. WEL: weighted empirical likelihood inference for dynamical correlations. *R Package*. **https://github.com/caojiguo/WEL**

Chang H. 2020. survELtest: Comparing multiple survival functions with crossing hazards. *R Package*, version 2.0.1. **https://CRAN.R-project.org/package=survELtest**
Chang H. 2022. fdEL. *R Package*, version 0.0.0.9000. **https://github.com/news11/fdEL**
Chang H, McKeague IW. 2022. Empirical likelihood based inference for functional means with application to wearable device data. *J. R. Stat. Soc. Ser. B* 84(5):1947–68
Chang H, Tsai PY, Kao JT, Lan GY. 2021. Comparing multiple survival functions with crossing hazards in R. *R J.* 12(2):20–42
Chen J, Huang Y. 2013. Finite-sample properties of the adjusted empirical likelihood. *J. Nonparametric Stat.* 25(1):147–59
Chen SX, Van Keilegom I. 2009. A review on empirical likelihood methods for regression. *Test* 18:415–47
Choi H, Reimherr M. 2018. A geometric approach to confidence regions and bands for functional parameters. *J. R. Stat. Soc. Ser. B* 80(1):239–60
Cuevas A, Febrero M, Fraiman R. 2004. An ANOVA test for functional data. *Comput. Stat. Data Anal.* 47(1):111–22
Dabo-Niang S, Guillas S. 2010. Functional semiparametric partially linear model with autoregressive errors. *J. Multivar. Anal.* 101(2):307–15
Degras DA. 2011. Simultaneous confidence bands for nonparametric regression with functional data. *Stat. Sin.* 21(4):1735–65
Dette H, Kokot K, Aue A. 2020. Functional data analysis in the Banach space of continuous functions. *Ann. Stat.* 48(2):1168–92
Dubin JA, Müller HG. 2005. Dynamical correlation for multivariate longitudinal data. *J. Am. Stat. Assoc.* 100(471):872–81
Durbin J, Watson GS. 1950. Testing for serial correlation in least squares regression: I. *Biometrika* 37(3/4):409–28
Dykstra RL. 1982. Maximum likelihood estimation of the survival functions of stochastically ordered random variables. *J. Am. Stat. Assoc.* 77(379):621–28
Einmahl JHJ, McKeague IW. 2003. Empirical likelihood based hypothesis testing. *Bernoulli* 9(2):267–90
Fan J. 1992. Design-adaptive nonparametric regression. *J. Am. Stat. Assoc.* 87(420):998–1004
Fan J. 1993. Local linear regression smoothers and their minimax efficiencies. *Ann. Stat.* 21:196–216
Ferraty F, Mas A, Vieu P. 2007. Nonparametric regression on functional data: inference and practical aspects. *Aust. N. Z. J. Stat.* 49(3):267–86
Ferraty F, Vieu P. 2004. Nonparametric models for functional data, with application in regression, time series prediction and curve discrimination. *Nonparametric Stat.* 16(1–2):111–25
Fontana M, Zeni G, Vantini S. 2023. Conformal prediction: a unified review of theory and new challenges. *Bernoulli* 29(1):1–23
Hall P, La Scala B. 1990. Methodology and algorithms of empirical likelihood. *Int. Stat. Rev./Rev. Int. Stat.* 58(2):109–27
Hjort NL, McKeague IW, Van Keilegom I. 2009. Extending the scope of empirical likelihood. *Ann. Stat.* 37(3):1079–111
Hu L, Huang T, You J. 2021. Robust inference in varying-coefficient additive models for longitudinal/functional data. *Stat. Sin.* 31(2):773–96
Hu Y, Xue L, Feng S. 2018. Empirical likelihood inference for partial functional linear model with missing responses. *Commun. Stat. Theory Methods* 47(19):4673–91
Hu YP, Liang HY. 2022. Empirical likelihood in single-index partially functional linear model with missing observations. *Commun. Stat. Theory Methods* 53(3):882–908
Kim E, MacEachern SN, Peruggia M. 2024. melt: Multiple empirical likelihood tests in R. *J. Stat. Softw.* 108(5):1–33
Kim JS, Staicu AM, Maity A, Carroll RJ, Ruppert D. 2018. Additive function-on-function regression. *J. Comput. Graph. Stat.* 27(1):234–44
Kitamura Y. 2007. Empirical likelihood methods in econometrics: theory and practice. In *Advances in Economics and Econometrics: Theory and Applications, Ninth World Congress*, Vol. III, ed. R Blundell, W Newey, T Persson, pp. 174–237. Cambridge, UK: Cambridge Univ. Press

Kitamura Y, Santos A, Shaikh AM. 2012. On the asymptotic optimality of empirical likelihood for testing moment restrictions. *Econometrica* 80(1):413–23
Koner S, Staicu AM. 2023. Second-generation functional data. *Annu. Rev. Stat. Appl.* 10:547–72
Laib N, Louani D. 2010. Nonparametric kernel regression estimation for functional stationary ergodic data: asymptotic properties. *J. Multivar. Anal.* 101(10):2266–81
Lazar NA. 2021. A review of empirical likelihood. *Annu. Rev. Stat. Appl.* 8:329–44
Lei J, G'Sell M, Rinaldo A, Tibshirani RJ, Wasserman L. 2018. Distribution-free predictive inference for regression. *J. Am. Stat. Assoc.* 113(523):1094–111
Li J, Huang C, Zhu H. 2017. A functional varying-coefficient single-index model for functional response data. *J. Am. Stat. Assoc.* 112(519):1169–81
Li Q, Tan X, Wang L. 2021. Testing for error correlation in partially functional linear regression models. *Commun. Stat. Theory Methods* 50(3):747–61
Lian H. 2012. Empirical likelihood confidence intervals for nonparametric functional data analysis. *J. Stat. Plan. Inference* 142(7):1669–77
Ljung GM, Box GE. 1978. On a measure of lack of fit in time series models. *Biometrika* 65(2):297–303
Matthews CE, Chen KY, Freedson PS, Buchowski MS, Beech BM, et al. 2008. Amount of time spent in sedentary behaviors in the United States, 2003–2004. *Am. J. Epidemiol.* 167(7):875–81
Morris JS. 2015. Functional regression. *Annu. Rev. Stat. Appl.* 2:321–59
Mukerjee R. 1994. Comparison of tests in their original forms. *Sankhyā Ser. A* 56(1):118–27
Nordman DJ, Lahiri SN. 2014. A review of empirical likelihood methods for time series. *J. Stat. Plan. Inference* 155:1–18
Otsu T. 2010. On Bahadur efficiency of empirical likelihood. *J. Econom.* 157(2):248–56
Owen AB. 1988. Empirical likelihood ratio confidence intervals for a single functional. *Biometrika* 75(2):237–49
Owen AB. 2001. *Empirical Likelihood*. Boca Raton, FL: Chapman and Hall/CRC
Parente PM, Smith RJ. 2014. Recent developments in empirical likelihood and related methods. *Annu. Rev. Econ.* 6:77–102
Park Y, Kalbfleisch JD, Taylor JMG. 2012. Constrained nonparametric maximum likelihood estimation of stochastically ordered survivor functions. *Can. J. Stat.* 40(1):22–39
Petersen A, Müller HG. 2016. Functional data analysis for density functions by transformation to a Hilbert space. *Ann. Stat.* 41(1):183–218
Petersen A, Müller HG. 2019. Fréchet regression for random objects with Euclidean predictors. *Ann. Stat.* 47(2):691–719
Presseau J, Tait RI, Johnston DW, Francis JJ, Sniehotta FF. 2013. Goal conflict and goal facilitation as predictors of daily accelerometer-assessed physical activity. *Health Psychol.* 32(12):1179
Qin J, Lawless J. 1994. Empirical likelihood and general estimating equations. *Ann. Stat.* 22(1):300–25
Ramsay JO, Silverman BW. 2005. *Functional Data Analysis*. New York: Springer. 2nd ed.
Ratcliffe SJ, Leader LR, Heller GZ. 2002. Functional data analysis with application to periodically stimulated foetal heart rate data. I: Functional regression. *Stat. Med.* 21(8):1103–14
Reiss PT, Huang L, Mennes M. 2010. Fast function-on-scalar regression with penalized basis expansions. *Int. J. Biostat.* 6(1):28
Sang P, Wang L, Cao J. 2019. Weighted empirical likelihood inference for dynamical correlations. *Comput. Stat. Data Anal.* 131:194–206
Shin H. 2009. Partial functional linear regression. *J. Stat. Plan. Inference* 139(10):3405–18
Ullah S, Finch CF. 2013. Applications of functional data analysis: a systematic review. *BMC Med. Res. Methodol.* 13:43
US Natl. Cent. Health Stat. 2006. *National health and nutrition examination survey data*. Dataset, US Cent. Dis. Control Prev., Atlanta, GA. **https://wwwn.cdc.gov/nchs/nhanes/continuousnhanes/default.aspx?BeginYear=2005**
van der Vaart AW. 2000. *Asymptotic Statistics*. Cambridge, UK: Cambridge Univ. Press
Wang H, Zhong PS, Cui Y, Li Y. 2018. Unified empirical likelihood ratio tests for functional concurrent linear models and the phase transition from sparse to dense functional data. *J. R. Stat. Soc. Ser. B* 80(2):343–64
Wang JL, Chiou JM, Müller HG. 2016. Functional data analysis. *Annu. Rev. Stat. Appl.* 3:257–95

Wilks SS. 1938. The large-sample distribution of the likelihood ratio for testing composite hypotheses. *Ann. Math. Stat.* 9(1):60–62

Wu CO, Chiang CT, Hoover DR. 1998. Asymptotic confidence regions for kernel smoothing of a varying-coefficient model with longitudinal data. *J. Am. Stat. Assoc.* 93(444):1388–402

Xi D, Pang T. 2021. Common breaks in means for panel data under short-range dependence. *Commun. Stat. Theory Methods* 50(2):486–505

Xiong X, Lin Z. 2013. Empirical likelihood inference for nonparametric regression functions with functional stationary ergodic data. *Commun. Stat. Theory Methods* 42(19):3421–31

Xue L, Zhu L. 2007. Empirical likelihood for a varying coefficient model with longitudinal data. *J. Am. Stat. Assoc.* 102(478):642–54

Yang B, Chen M, Zhou J. 2023. Testing for error correlation in semi-functional linear models. *J. Syst. Sci. Complex.* 36:1697–716

Yang H, Zhu H, Ahn M, Ibrahim JG. 2021. Weighted functional linear Cox regression model. *Stat. Methods Med. Res.* 30(8):1917–31

Yao F, Müller HG. 2010. Functional quadratic regression. *Biometrika* 97(1):49–64

You J, Chen G, Zhou Y. 2006. Block empirical likelihood for longitudinal partially linear regression models. *Can. J. Stat.* 34(1):79–96

Yuan A, Fang HB, Li H, Wu CO, Tan MT. 2020. Hypothesis testing for multiple mean and correlation curves with functional data. *Stat. Sin.* 30:1095–116

Zhang JT, Chen J. 2007. Statistical inferences for functional data. *Ann. Stat.* 35(3):1052–79

Zhang JT, Cheng MY, Wu HT, Zhou B. 2019. A new test for functional one-way ANOVA with applications to ischemic heart screening. *Comput. Stat. Data Anal.* 132:3–17

Zhou XC, Lin JG. 2014. Empirical likelihood inference in mixture of semiparametric varying-coefficient models for longitudinal data with non-ignorable dropout. *Statistics* 48(3):668–84

Zhou Z, Lin Z. 2016. Asymptotic normality of locally modelled regression estimator for functional data. *J. Nonparametric Stat.* 28(1):116–31

Zhu H, Fan J, Kong L. 2014. Spatially varying coefficient model for neuroimaging data with jump discontinuities. *J. Am. Stat. Assoc.* 109(507):1084–98