J. R. Statist. Soc. B (2018)
80, *Part* 2, *pp.* 343–364

Unified empirical likelihood ratio tests for functional concurrent linear models and the phase transition from sparse to dense functional data

Honglang Wang,

Indiana University–Purdue University, Indianapolis, USA

Ping-Shou Zhong and Yuehua Cui

Michigan State University, East Lansing, USA

and Yehua Li

Iowa State University, Ames, USA

[Received November 2015. Final revision June 2017]

Summary. We consider the problem of testing functional constraints in a class of functional concurrent linear models where both the predictors and the response are functional data measured at discrete time points. We propose test procedures based on the empirical likelihood with bias-corrected estimating equations to conduct both pointwise and simultaneous inferences. The asymptotic distributions of the test statistics are derived under the null and local alternative hypotheses, where sparse and dense functional data are considered in a unified framework. We find a phase transition in the asymptotic null distributions and the orders of detectable alternatives from sparse to dense functional data. Specifically, the tests proposed can detect alternatives of $\sqrt{n}$-order when the number of repeated measurements per curve is of an order larger than n^{η_0} with n being the number of curves. The transition points η_0 for pointwise and simultaneous tests are different and both are smaller than the transition point in the estimation problem. Simulation studies and real data analyses are conducted to demonstrate the methods proposed.

Keywords: Empirical likelihood; Functional analysis of variance; Non-parametric hypothesis testing; Unified inference

1. Introduction

We consider statistical inference problems under a general functional concurrent linear (FCL) model (Ramsay and Silverman, 2005), where both the response $Y(t)$ and the p-dimensional covariate $\mathbf{X}(t) = \{X^{(1)}(t), \ldots, X^{(p)}(t)\}^{\mathrm{T}}$ are defined continuously on a time interval $[a, b]$. The relationship between $Y(t)$ and $\mathbf{X}(t)$ is given by

$$Y(t) = \boldsymbol{\beta}_0^{\mathrm{T}}(t)\mathbf{X}(t) + \epsilon(t), \tag{1.1}$$

where $\boldsymbol{\beta}_0(t) = (\beta_{10}(t), \ldots, \beta_{p0}(t))^{\mathrm{T}}$ is a p-dimensional vector of unknown functions and $\epsilon(t)$ is a zero-mean error process independent of $\mathbf{X}$, with a covariance function $\Omega(s, t) = \mathrm{cov}\{\epsilon(s), \epsilon(t)\}$. Without loss of generality, we allow $\mathbf{X}(t)$ to be a multivariate random process with mean function

Address for correspondence: Ping-Shou Zhong, Department of Statistics and Probability, Michigan State University, East Lansing, MI 48824, USA.
E-mail: pszhong@stt.msu.edu

 1369–7412/18/80343

$\boldsymbol{\mu}(t) = E\{\mathbf{X}(t)\}$ and covariance function $\boldsymbol{\Gamma}(s,t) = \mathrm{cov}\{\mathbf{X}(s), \mathbf{X}(t)\}$. Recent literature on the FCL model includes Fan and Zhang (2000), Faraway (1997), Zhang and Chen (2007), Zhang *et al.* (2010) and Zhang (2011).

Let $\{Y_i(t), \mathbf{X}_i(t)\}$, $i = 1, \ldots, n$, be independent realizations of $\{Y(t), \mathbf{X}(t)\}$. Instead of observing the entire trajectories, one can observe $Y_i(t)$ and $\mathbf{X}_i(t)$ at only discrete time points $\{t_{ij}, j = 1, \ldots, m_i\}$ (Hall and Van Keilegom, 2007). For convenience, denote $Y_{ij} = Y_i(t_{ij})$ and $X_{ij}^{(k)} = X_i^{(k)}(t_{ij})$, and assume that m_is ($1 \leqslant i \leqslant n$) are all of the same order as $m = n^{\eta}$ for some $\eta \geqslant 0$, i.e. m_i/m are bounded below and above by some positive constants. Functional data are considered to be sparse or dense depending on the order of m (Hall *et al.*, 2006; Li and Hsing, 2010). Data with bounded m, or $\eta = 0$, are called sparse functional data; those satisfying $\eta \geqslant \eta_0$, where η_0 is a transition point to be specified below, are referred to as dense functional data. The scenarios with $\eta \in (0, \eta_0)$ are in a grey zone in the literature and we refer to them as 'moderately dense' in this paper.

Historically, sparse and dense functional data were analysed with different methods. For dense functional data, one can smooth each curve separately and proceed with further estimation and inference based on the presmoothed curves. A partial list of recent literature on dense functional data includes Castro *et al.* (1986), Rice and Silverman (1991), Zhang and Chen (2007), Eubank and Hsing (2008) and Benko *et al.* (2009). For sparse functional data, the presmoothing approach is not applicable and, instead, one needs to pool all data together to borrow strength from individual curves (Yao *et al.*, 2005a, b). Hall *et al.* (2006) investigated the theoretical properties of functional principal component analysis based on local linear smoothers. They found that, for dense functional data with $\eta \geqslant \frac{1}{4}$, the presmoothing errors are asymptotically negligible and quantities such as the mean, covariance and eigenfunctions can be estimated with a parametric $\sqrt{n}$-rate, whereas these quantities can only be estimated with a non-parametric convergence rate for sparse functional data with $\eta = 0$. However, sparse and dense functional data are asymptotic concepts and are difficult to distinguish in reality. Li and Hsing (2010) proposed an estimation procedure treating all types of functional data under a unified framework. More recently, Kim and Zhao (2013) proposed a unified, self-normalizing approach to construct pointwise confidence intervals for the mean function of functional data. Both Li and Hsing (2010) and Kim and Zhao (2013) established $\eta_0 = \frac{1}{4}$ as the transition point to a parametric convergence rate.

In contrast with estimation, less is known about inferences for functional data, with a few exceptions such as Zhang and Chen (2007) and Kim and Zhao (2013). The focus of this paper is on proposing pointwise (at a specific t) and simultaneous test (for all t in $[a, b]$) procedures for the following hypothesis under model (1.1):

$$H_0 : H\{\boldsymbol{\beta}_0(t)\} = 0 \quad \textit{versus} \quad H_1 : H\{\boldsymbol{\beta}_0(t)\} \neq 0, \tag{1.2}$$

where $H(\mathbf{z})$ is a q-dimensional function of $\mathbf{z} = (z_1, \ldots, z_p)^{\mathrm{T}} \in \mathbb{R}^p$ such that $C(\mathbf{z}) := \partial H(\mathbf{z})/\partial \mathbf{z}^{\mathrm{T}}$ is a $q \times p$ full rank matrix ($q \leqslant p$) for all $\mathbf{z}$. The test problem (1.2) is very broad, including many interesting hypotheses as special cases. For instance, if $H(\mathbf{z}) = \mathbf{z}$, the null hypothesis is equivalent to $H_0 : \beta_{k0}(\cdot) = 0$ for all k. If $H(\mathbf{z}) = (z_1 - z_2, z_2 - z_3, \ldots, z_{p-1} - z_p)^{\mathrm{T}}$, then hypothesis (1.2) is an analysis-of-variance type of hypothesis on the coefficient functions $\beta_{k0}(\cdot)$. If $H(\mathbf{z}) = \boldsymbol{\Lambda}\mathbf{z} - \mathbf{c}_0$ for a $q \times p$ known constant matrix $\boldsymbol{\Lambda}$ and a known vector $\mathbf{c}_0$, then hypothesis (1.2) becomes $H_0 : \boldsymbol{\Lambda}\boldsymbol{\beta}_0(\cdot) = \mathbf{c}_0$, which is a test for linear constraints on $\boldsymbol{\beta}_0(\cdot)$. Zhang and Chen (2007) and Zhang (2011) studied similar linear constraints test problems for dense functional data with $\eta > \frac{5}{4}$. Besides linear constraints, non-linear constraints are also special cases of hypothesis (1.2), which have broad applications in econometrics and neuroimaging studies. See Phillips and Park (1988), Critchley *et al.* (1996) and Ashby (2011) for some explicit examples of non-linear hypotheses.

In this paper, we propose both pointwise and simultaneous tests for the non-parametric hypothesis (1.2) based on empirical likelihood (EL). We show that EL-based tests enjoy a nice self-normalizing property such that we can treat both sparse and dense functional data under a unified framework. EL is a non-parametric likelihood which was introduced by Owen (1988, 1990), which maintains two key properties of a parametric likelihood: the Wilks property (Owen, 1990, 2001) and the Bartlett correction property (DiCiccio *et al.*, 1991; Chen and Cui, 2006). An overview of EL methods can be found in Owen (2001) and Chen and Van Keilegom (2009). There has been some work on EL methods for sparse functional data with $\eta = 0$, such as Xue and Zhu (2007), Chen and Zhong (2010) and Tang and Leng (2011). However, to the best of our knowledge, EL methods for dense functional data with $\eta > 0$ have not been investigated.

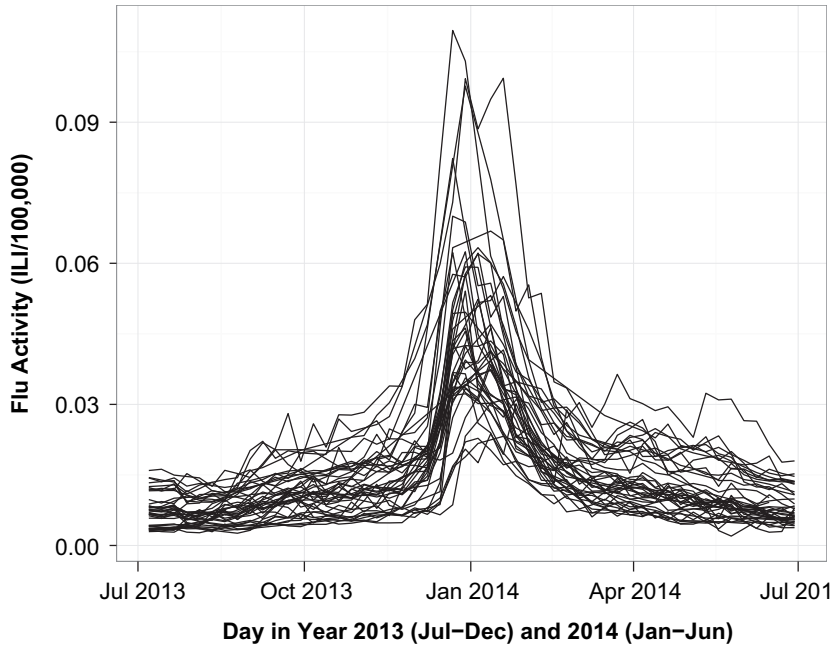
To investigate the power of the tests, we consider the local alternatives

$$H_{1n} : H\{\beta_0(t)\} = b_n \mathbf{d}(t), \tag{1.3}$$

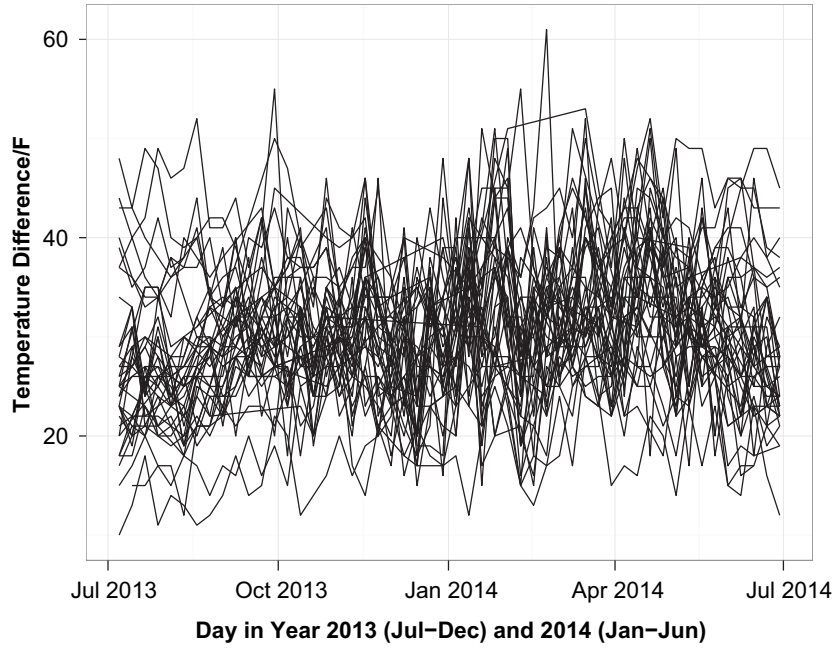
where b_n is a sequence of numbers converging to 0 at a rate to be specified later and $\mathbf{d}(t) \neq 0$ is a q-dimensional function. For any fixed non-zero $\mathbf{d}(\cdot)$, let b_n^* be the smallest order of the local alternatives such that a test has a non-trivial power (i.e. the power of a test is larger than its nominal level). Here b_n^* quantifies the order of signals that a test can detect. For sparse data with $\eta = 0$, Chen and Zhong (2010) proved that the EL method using a global bandwidth h can detect alternatives of order $b_n^* = (nh)^{-1/2}$ and $b_n^* = n^{-1/2}h^{-1/4}$ for pointwise and simultaneous tests respectively. Because $h \to 0$ in a typical non-parametric regression setting, both orders are larger than the parametric rate $n^{-1/2}$. However, for dense data with $\eta > 0$, the detectable order b_n^* is still largely unknown. One key interest in this paper is to understand the effect of η on b_n^*. Under some mild conditions and with a properly chosen bandwidth, we find that, for the pointwise EL test, b_n^* is larger than $n^{-1/2}$ for $\eta \leqslant \frac{1}{8}$ and equals $n^{-1/2}$ for $\eta > \frac{1}{8}$. For the simultaneous EL test, b_n^* is larger than $n^{-1/2}$ for $\eta \leqslant \frac{1}{16}$ and equals $n^{-1/2}$ for $\eta > \frac{1}{16}$. The transition points $\frac{1}{8}$ and $\frac{1}{16}$ will be respectively referred to as η_0 for the pointwise and simultaneous EL tests. This phase transition result echoes similar phenomena discovered by Li and Hsing (2010) for estimation problems.

Our study is motivated by inference problems from two real data applications where one involves dense functional data and another involves sparse functional data. In both data examples, the interest is to study the dynamic time varying effects of functional covariates on the functional response. In the Google flu trend data, the response is weekly flu activity from 42 states in the USA for an entire year. One of the covariates of interest is the maximum daily temperature variation (MDTV) in these states, which is also obtained weekly over the same time period. As shown in Fig. 1, both flu activity and temperature variation curves are dense functional data. In a data set collected by the Alzheimer's disease (AD) neuroimaging initiative at the University of Southern California, the mini-mental state examination (MMSE) score and the volume of the hippocampus of the brain are measured for each enrolled patient during clinic visits. These variables are repeatedly measured 3–10 times during a 1-year follow-up, which yields sparse functional data. More details of these data are provided in Section 7.

The rest of the paper is organized as follows. In Section 2, we present a bias-corrected estimator and some preliminary results. We propose the unified pointwise EL test in Section 3 and the simultaneous EL test in Section 4, where we investigate the asymptotic distributions of the test statistics under both the null hypothesis and local alternatives and obtain the transition phase for b_n^*. In Section 5, we address implementation issues such as bandwidth selection and covariance estimation. Simulation studies are presented in Section 6, followed by analysis of the two data examples in Section 7. All technical details are included in the on-line supplemental material.



(a)



(b)

Fig. 1. Google flu trend data: (a) flu activity (percentage of influenza-like illness cases per 100000 physician visits) and (b) the MDTV for 42 states of the USA (both variables are measured weekly during the 2013–2014 flu season; each curve represent one state)

The data that are analysed in the paper and the programs that were used to analyse them can be obtained form

`http://wileyonlinelibrary.com/journal/rss-datasets`

2. A bias-corrected estimator and some preliminary results

We first introduce a bias-corrected estimator $\check{\boldsymbol{\beta}}(t)$ based on an initial local linear estimator $\hat{\boldsymbol{\beta}}(t)$ (Fan and Gijbels, 1996) for $\boldsymbol{\beta}_0(t)$ and then provide some preliminary results for $\check{\boldsymbol{\beta}}(t)$.

2.1. A bias-corrected estimator

Let $K(\cdot)$ be a symmetric probability density function that we use as a kernel and h be a bandwidth, and denote $K_h(\cdot) = K(\cdot/h)/h$. For any t in a neighbourhood of t_0, $\beta_{k0}(t)$ can be approximated by $\beta_{k0}(t) \approx \beta_{k0}(t_0) + \{\partial\beta_{k0}(t_0)/\partial t\}(t - t_0) := a_k + b_k(t - t_0)$ for $k = 1, 2, \ldots, p$. Denote $\boldsymbol{\vartheta} = (a_1, \ldots, a_p, hb_1, \ldots, hb_p)^{\mathrm{T}}$ and $\mathbf{D}_{ij}(t) = (\mathbf{X}_{ij}^{\mathrm{T}}, (t_{ij} - t)\mathbf{X}_{ij}^{\mathrm{T}}/h)^{\mathrm{T}}$. For each $1 \leqslant i \leqslant n$, put $\mathbf{Y}_i = (Y_{i1}, Y_{i2}, \ldots, Y_{im_i})^{\mathrm{T}}$, $\mathbf{D}_i(t) = (\mathbf{D}_{i1}(t), \mathbf{D}_{i2}(t), \ldots, \mathbf{D}_{im_i}(t))^{\mathrm{T}}$ and $\mathbf{W}_i(t) = \mathrm{diag}\{K_h(t_{i1} - t), \ldots, K_h(t_{im_i} - t)\}/m_i$. Further, let $\mathbf{Y} = (\mathbf{Y}_1^{\mathrm{T}}, \mathbf{Y}_2^{\mathrm{T}}, \ldots, \mathbf{Y}_n^{\mathrm{T}})^{\mathrm{T}}$, $\mathbf{D}(t) = (\mathbf{D}_1^{\mathrm{T}}(t), \mathbf{D}_2^{\mathrm{T}}(t), \ldots, \mathbf{D}_n^{\mathrm{T}}(t))^{\mathrm{T}}$ and $\mathbf{W}(t) = \mathrm{diag}\{\mathbf{W}_1(t), \mathbf{W}_2(t), \ldots, \mathbf{W}_n(t)\}$. An estimator for $\boldsymbol{\vartheta}$ is obtained as

$$\hat{\boldsymbol{\vartheta}} = \{\mathbf{D}^{\mathrm{T}}(t_0)\mathbf{W}(t_0)\mathbf{D}(t_0)\}^{-1}\mathbf{D}^{\mathrm{T}}(t_0)\mathbf{W}(t_0)\mathbf{Y}.$$

Thus the local linear estimator for $\boldsymbol{\beta}_0(t_0)$ is

$$\hat{\boldsymbol{\beta}}(t_0) = (\mathbf{I}_p, \mathbf{0}_p)\hat{\boldsymbol{\vartheta}} = (\mathbf{I}_p, \mathbf{0}_p)\{\mathbf{D}^{\mathrm{T}}(t_0)\mathbf{W}(t_0)\mathbf{D}(t_0)\}^{-1}\mathbf{D}^{\mathrm{T}}(t_0)\mathbf{W}(t_0)\mathbf{Y}, \tag{2.1}$$

where $\mathbf{I}_p$ is a $p \times p$ identity matrix and $\mathbf{0}_p$ is a $p \times p$ zero matrix. It is shown in lemma 1 in the on-line supplemental material that

$$\sup_{t\in[a,b]} |\hat{\boldsymbol{\beta}}(t) - \boldsymbol{\beta}_0(t)| = O\left[h^2 + \left\{\frac{\log(n)}{n} + \frac{\log(n)}{nmh}\right\}^{1/2}\right] \qquad \text{almost surely.} \tag{2.2}$$

The $O(h^2)$ term in result (2.2) is the order for the bias of $\hat{\boldsymbol{\beta}}(t)$. To eliminate the influence of bias, undersmoothing is typically needed for an unbiased testing procedure (Xue and Zhu, 2007). Instead of performing artificial undersmoothing, we propose to reduce the bias in $\hat{\boldsymbol{\beta}}(t)$ by introducing a bias-corrected estimator $\check{\boldsymbol{\beta}}(t)$ (Xue and Zhu, 2007) as the solution of the following residual-adjusted estimating equation for $\boldsymbol{\beta}(t)$:

$$\bar{g}_n\{\boldsymbol{\beta}(t)\} := \frac{1}{n}\sum_{i=1}^{n} g_i\{\boldsymbol{\beta}(t)\} = 0, \tag{2.3}$$

with $g_i\{\boldsymbol{\beta}(t)\} = (1/m_i)\Sigma_{j=1}^{m_i}\{Y_{ij} - \boldsymbol{\beta}^{\mathrm{T}}(t)\mathbf{X}_{ij} - (\hat{\boldsymbol{\beta}}(t_{ij}) - \hat{\boldsymbol{\beta}}(t))^{\mathrm{T}}\mathbf{X}_{ij}\}\mathbf{X}_{ij}K_h(t_{ij} - t)$, where $\hat{\boldsymbol{\beta}}(t)$ is the initial local linear estimator in equation (2.1). Remark S.1 in the on-line supplemental material provides some motivation for the above bias correction.

There are several existing methods to estimate $\boldsymbol{\beta}(t)$ in the FCL literature, including the Fourier basis approach of Faraway (1997) and Ramsay and Silverman (2005), the local polynomial methods in Fan and Zhang (2000), Zhang and Chen (2007) and Zhang (2011), the wavelet approach of Zhang *et al.* (2010) and the smoothing spline method in Eggermont *et al.* (2010). Our purpose in introducing the bias-corrected estimating equation approach is not only to reduce the bias in the estimator but also to use $g_i\{\boldsymbol{\beta}(t)\}$ in the construction of our empirical likelihood (3.1). One can develop other estimators for $\boldsymbol{\beta}(t)$, such as the local cubic polynomial estimator (Fan and Gijbels, 1996), that yield a similar bias of $O(h^4)$ as in $\check{\boldsymbol{\beta}}(t)$; however, those

estimators cannot be naturally extended to empirical likelihood. For illustration, we also provide a simulation study in the on-line supplemental material to compare the numerical performance of the bias-corrected estimator $\check{\boldsymbol{\beta}}(t)$ with that of the local cubic estimator. We find that the local cubic estimator has a slightly smaller bias than $\check{\boldsymbol{\beta}}(t)$ but higher variance and mean-squared error. This perhaps is because the local cubic estimation method needs to estimate more nuisance parameters such as the higher order derivatives of $\boldsymbol{\beta}(t)$.

2.2. Regularity conditions and preliminary results

We now present some preliminary results regarding the asymptotics of $\check{\boldsymbol{\beta}}(t)$. Assume that t_{ij} are independent and identically distributed random variables following a probability density function $f(t)$. For convenience, define $\boldsymbol{\Gamma}(t)=\boldsymbol{\Gamma}(t,t)$, $\Omega(t)=\Omega(t,t)$ and $\mathbf{C}(t)=C\{\boldsymbol{\beta}_0(t)\}$. We shall also use $\tilde{o}_p$ and $\tilde{O}_p$ to represent respectively that o_p and O_p hold uniformly for all $t\in[a,b]$. The following conditions are needed for our asymptotic results.

Condition 1. The kernel function $K(\cdot)$ is a symmetric probability density function with a bounded support in $[-1,1]$.

Condition 2. Assume that $\mathbb{E}\{\sup_{t\in[a,b]}\|\mathbf{X}(t)\|^{\lambda_1}\}<\infty$ and $\mathbb{E}\{\sup_{t\in[a,b]}|\epsilon(t)|^{\lambda_2}\}<\infty$ where $\lambda_1,\ \lambda_2\geqslant 5$ and $\|\cdot\|$ is the L_2-norm of a vector.

Condition 3. Both $f(t)$ and $\boldsymbol{\Gamma}(t)$ are twice continuously differentiable, $\boldsymbol{\beta}_0(t)$ is three-times continuously differentiable and $\mathbf{C}(t)$ is uniformly bounded on $[a,b]$.

Condition 4. Define $\lambda=\min(\lambda_1,\lambda_2)$ and let $h\asymp n^{-\alpha_0}$ with $\alpha_0\in(0,1)$ the order of the bandwidth. Assume that

(a) $\alpha_0<1-\eta-2/\lambda$ if $\eta\in[0,\frac{1}{8}]$ and $\alpha_0<\frac{1}{2}-1/\lambda$ if $\eta>\frac{1}{8}$;
(b) $(1+\eta)/9<\alpha_0$ if $\eta\in[0,\frac{1}{8}]$ and $\frac{1}{8}<\alpha_0<\eta$ if $\eta>\frac{1}{8}$.

Conditions 1 and 3 are commonly used regularity conditions in non-parametric regressions. Condition 2 is similar to that in Li and Hsing (2010). The upper bounds on the bandwidth h in condition 4, part (a), are adapted from Li and Hsing (2010). A detailed explanation on the restrictions of h in condition 4, part (b), will be given in remark 2 after proposition 2. Selecting a bandwidth that satisfies condition 4 will be discussed in Section 5.

The following proposition provides an asymptotic expansion for $\check{\boldsymbol{\beta}}(t)$, the proof of which is provided in the on-line supplemental paper.

Proposition 1. Let $\mathbf{A}(t)=\boldsymbol{\Gamma}(t)f(t)$. Under conditions 1–3 and 4, part (a),

$$\check{\boldsymbol{\beta}}(t)-\boldsymbol{\beta}_0(t)=-\mathbf{A}^{-1}(t)\bar{\boldsymbol{\xi}}_n(t)\{1+\tilde{o}_p(1)\}+\tilde{O}_p(h^4), \tag{2.4}$$

where $\bar{\boldsymbol{\xi}}_n(t)=n^{-1}\Sigma_{i=1}^n\xi_i(t)$ and $\xi_i(t)=m_i^{-1}\Sigma_{j=1}^{m_i}\mathbf{X}_{ij}\epsilon_{ij}K_h(t_{ij}-t)$. Let $\mu_{ts}=\int u^sK^t(u)\mathrm{d}u$ and $\bar{r}=\lim_{n\to\infty}n^{-1}\Sigma_{i=1}^n m/m_i$; then

$$\mathrm{var}\{\bar{\boldsymbol{\xi}}_n(t)\}=\boldsymbol{\Gamma}(t)\Omega(t)f(t)\left\{\frac{\bar{r}}{mnh}\mu_{20}+\frac{m-\bar{r}}{nm}f(t)\right\}\{1+\tilde{o}(1)\}. \tag{2.5}$$

Remark 1. The genuine bias and variance of $\check{\boldsymbol{\beta}}(t)$ cannot be easily obtained, but the $\tilde{O}_p(h^4)$ remainder term and the variance of the leading order term in the asymptotic expansion (2.4) can be considered respectively as substitutes for the bias and variance of $\check{\boldsymbol{\beta}}(t)$. Similarly, throughout the rest of this paper, the bias and variance of an estimator are not exact but those of the leading terms in the asymptotic expansion. Our calculation of asymptotic mean-squared error

(AMSE) is also based on these 'pseudo'asymptotic bias and variance. Proposition 1 shows that the AMSE of $\check{\boldsymbol{\beta}}(t)$ is

$$\mathrm{AMSE}\{\check{\boldsymbol{\beta}}(t)\} = O\left(h^8 + \frac{1}{mnh} + \frac{1}{n}\right).$$

Hence the optimal bandwidth h_{opt} that minimizes the AMSE of $\check{\boldsymbol{\beta}}(t)$ is $h_{\mathrm{opt}} \asymp (mn)^{-1/9} = n^{-(1+\eta)/9}$. It follows that $\check{\boldsymbol{\beta}}(t) - \boldsymbol{\beta}_0(t) = O_p\{h_{\mathrm{opt}}^4 + (mnh_{\mathrm{opt}})^{-1/2} + n^{-1/2}\} = O_p\{n^{-1/2} + n^{-4(1+\eta)/9}\}$. The optimal convergence rate of $\check{\boldsymbol{\beta}}(t)$ is then of order $n^{-4(1+\eta)/9}$ if $\eta \leqslant \frac{1}{8}$ and of order $n^{-1/2}$ if $\eta > \frac{1}{8}$. Thus, $\eta_0 = \frac{1}{8}$ is the transition point for the convergence rate of $\check{\boldsymbol{\beta}}(t)$. When $\eta > \eta_0$, $\check{\boldsymbol{\beta}}(t)$ is no longer sensitive to the choice of h and its convergence rate remains at $O_p(n^{-1/2})$ as long as $h = O(n^{-1/8})$ and $h \gg m^{-1} = n^{-\eta}$.

The next proposition provides the asymptotic distribution of $\check{\boldsymbol{\beta}}(t)$.

Proposition 2. Suppose that $mh \to \kappa_0 \in [0, \infty]$; define

$$C_{n,\alpha_0,\eta} = \begin{cases} \{n/(mh)\}^{1/2}, & \text{if } 0 \leqslant \kappa_0 < \infty, \\ n^{1/2}, & \text{if } \kappa_0 = \infty \end{cases} \tag{2.6}$$

and $\mathbf{B}(t) = \boldsymbol{\Gamma}(t)\Omega(t)f(t)[\{\bar{r}\mu_{20} + \kappa_0 f(t)\}I(\kappa_0 < \infty) + f(t)I(\kappa_0 = \infty)]$. Under conditions 1–4, we have

$$nC_{n,\alpha_0,\eta}^{-1}\{\check{\boldsymbol{\beta}}(t) - \boldsymbol{\beta}_0(t)\} \xrightarrow{\mathrm{d}} N\{\mathbf{0}, \mathbf{V}(t)\}, \tag{2.7}$$

where $\mathbf{V}(t) = \mathbf{A}^{-1}(t)\mathbf{B}(t)\mathbf{A}^{-1}(t)$ and $\mathbf{A}(t)$ was defined in proposition 1.

Remark 2. Following proposition 1, the bias in $nC_{n,\alpha_0,\eta}^{-1}\{\check{\boldsymbol{\beta}}(t) - \boldsymbol{\beta}_0(t)\}$ is of order $O(nh^4 \times C^{-1}{}_{n,\alpha_0,\eta})$. Condition 4, part (b), warrants that this bias is asymptotically negligible. Specifically, when $\eta \leqslant \eta_0 = \frac{1}{8}$, the condition $\alpha_0 > (1+\eta)/9$ ensures that $mh < \infty$. Hence $nh^4/C_{n,\alpha_0,\eta} = n^{1/2}m^{1/2}h^{9/2} \asymp n^{(1+\eta-9\alpha_0)/2} = o(1)$; when $\eta > \eta_0$, the condition that $\frac{1}{8} < \alpha_0 < \eta$ implies $mh \to \infty$ and $nh^4/C_{n,\alpha_0,\eta} = n^{1/2}h^4 \asymp n^{1/2-4\alpha_0} \to 0$. Being asymptotically unbiased is particularly important for constructing a valid unbiased test.

Remark 3. By proposition 2 and the delta method, we can show that, under H_0 in expression (1.2), $nC_{n,\alpha_0,\eta}^{-1}H\{\check{\boldsymbol{\beta}}(t)\} \to^{\mathrm{d}} N\{\mathbf{0}, \mathbf{R}^{-1}(t)\}$, where $\mathbf{R}(t) = \{\mathbf{C}(t)\mathbf{V}(t)\mathbf{C}(t)^{\mathrm{T}}\}^{-1}$. Note that the asymptotic variances of $H\{\check{\boldsymbol{\beta}}(t)\}$ are different under sparse and dense cases.

A Wald-type test statistic may be constructed by using remark 3 if an appropriate estimator for the variance of $H\{\check{\boldsymbol{\beta}}(t)\}$ can be obtained. We shall not pursue this direction because the estimation of the asymptotic variance involves several non-parametric functions, e.g. $\boldsymbol{\Gamma}(t)$, $\Omega(t)$ and $f(t)$, which requires properly selecting several bandwidths. Instead, we propose a self-normalizing EL method in the next section which avoids estimating the asymptotic variance explicitly. Some recent reviews of self-normalizing methods and theories can be found in Peña *et al.* (2008), Shao (2010) and Shao and Wang (2013).

3. A unified pointwise test

To test H_0 in expression (1.2) at any fixed time t, we propose a test based on the EL ratio statistic. Following Owen (1990), the EL for $\boldsymbol{\beta}(t)$ is given by

$$L\{\boldsymbol{\beta}(t)\} = \max_{p_1, p_2, \ldots, p_n} \left[\prod_{i=1}^{n} p_i : \sum_{i=1}^{n} p_i = 1, p_i \geqslant 0, \sum_{i=1}^{n} p_i g_i\{\boldsymbol{\beta}(t)\} = 0\right]. \tag{3.1}$$

Applying the method of Lagrange multipliers, the log-EL function becomes

$$l\{\boldsymbol{\beta}(t)\} := \log[L\{\boldsymbol{\beta}(t)\}] = -\sum \log[1+\boldsymbol{\gamma}^{\mathrm{T}}(t) g_i\{\boldsymbol{\beta}(t)\}] - n\log(n),$$

where $\boldsymbol{\gamma}(t)$ is a solution to the equation

$$Q_{1n}\{\boldsymbol{\beta}(t), \boldsymbol{\gamma}(t)\} := \frac{1}{n}\sum_{i=1}^{n} \frac{g_i\{\boldsymbol{\beta}(t)\}}{1+\boldsymbol{\gamma}^{\mathrm{T}}(t) g_i\{\boldsymbol{\beta}(t)\}} = 0. \tag{3.2}$$

The maximum log-EL without any constraint is $l\{\breve{\boldsymbol{\beta}}(t)\} = -n\log(n)$. It follows that the negative log-EL ratio for testing $H_0 : H\{\boldsymbol{\beta}_0(t)\} = 0$ is

$$l(t) := \min_{H\{\boldsymbol{\beta}(t)\}=0} l_0\{\boldsymbol{\beta}(t)\}, \tag{3.3}$$

where $l_0\{\boldsymbol{\beta}(t)\} = \Sigma_{i=1}^{n} \log[1+\boldsymbol{\gamma}^{\mathrm{T}}(t) g_i\{\boldsymbol{\beta}(t)\}]$. To solve equation (3.3), we minimize the following objective function (Qin and Lawless, 1995):

$$M\{\boldsymbol{\beta}(t), \boldsymbol{\nu}(t)\} = \frac{1}{n} l_0\{\boldsymbol{\beta}(t)\} + \boldsymbol{\nu}^{\mathrm{T}}(t) H\{\boldsymbol{\beta}(t)\},$$

where $\boldsymbol{\nu}(t)$ is a $q \times 1$ vector of Lagrange multipliers. Differentiating $M(\cdot,\cdot)$ with respect to $\boldsymbol{\beta}$ and $\boldsymbol{\nu}$ and setting them to 0, we have

$$Q_{2n}\{\boldsymbol{\beta}(t), \boldsymbol{\gamma}(t), \boldsymbol{\nu}(t)\} := \frac{1}{n}\frac{\partial l_0\{\boldsymbol{\beta}(t)\}}{\partial \boldsymbol{\beta}^{\mathrm{T}}(t)} + C^{\mathrm{T}}\{\boldsymbol{\beta}(t)\}\boldsymbol{\nu}(t) = 0,$$
$$H\{\boldsymbol{\beta}(t)\} = 0.$$

Combining equation (3.2) for $\boldsymbol{\gamma}(t)$, the constrained minimization problem (3.3) is equivalent to solving the following estimating equation system:

$$\left.\begin{array}{c} Q_{1n}\{\boldsymbol{\beta}(t), \boldsymbol{\gamma}(t)\} = 0, \\ Q_{2n}\{\boldsymbol{\beta}(t), \boldsymbol{\gamma}(t), \boldsymbol{\nu}(t)\} = 0, \\ H\{\boldsymbol{\beta}(t)\} = 0. \end{array}\right\} \tag{3.4}$$

The existence of a consistent solution to system (3.4) follows from similar arguments to those given by Qin and Lawless (1995), the proof of which is omitted but available from the authors on request. Denote the solution to system (3.4) as $\{\tilde{\boldsymbol{\beta}}(t), \tilde{\boldsymbol{\gamma}}(t), \tilde{\boldsymbol{\nu}}(t)\}$ and $\tilde{\boldsymbol{\beta}}(t)$ as the restricted maximum EL estimator. Then the test statistic (3.3) becomes $l(t) = l_0\{\tilde{\boldsymbol{\beta}}(t)\}$. Several existing non-linear optimization algorithms implemented in the C package '`NLopt`' are applicable to obtain the solution. The C package `NLopt` was developed by S. Johnson (Johnson, 2010) and is available from `http://ab-initio.mit.edu/nlopt`.

The following proposition provides an asymptotic expansion of $2l(t)$.

Proposition 3. Under H_0 at $t \in [a, b]$ and conditions 1–4,

$$2l(t) = \mathbf{U}_n(t)^{\mathrm{T}} \mathbf{U}_n(t) + O_p(nh^4/C_{n,\alpha_0,\eta}), \tag{3.5}$$

where $\mathbf{U}_n(t) = nC_{n,\alpha_0,\eta}^{-1} \mathbf{G}(t)\bar{\boldsymbol{\xi}}_n(t)$, $\mathbf{G}(t) = \mathbf{R}^{1/2}(t)\mathbf{C}(t)\mathbf{A}^{-1}(t)$ and $\mathbf{R}(t)$ is defined in remark 3.

The asymptotic expansion in equation (3.5) makes a connection between $2l(t)$ and the bias-corrected estimator $\breve{\boldsymbol{\beta}}(t)$ that was described in Section 2. Using Taylor series expansion, one can show that $H\{\breve{\boldsymbol{\beta}}(t)\} = H\{\boldsymbol{\beta}_0(t)\} + \mathbf{C}(t)\{\breve{\boldsymbol{\beta}}(t) - \boldsymbol{\beta}_0(t)\} + o_p(C_{n,\alpha_0,\eta}/n + h^4)$. Then by proposition 1, we have $\mathbf{U}_n(t) = nC_{n,\alpha_0,\eta}^{-1} \mathbf{R}^{1/2}(t) H\{\breve{\boldsymbol{\beta}}(t)\} + o_p(1)$ under H_0 where $H\{\boldsymbol{\beta}_0(t)\} = 0$. Applying results in remark 3, it can be shown that $\mathbf{U}_n(t)$ asymptotically follows a q-dimensional multivariate

standard normal distribution. Naturally, $2l(t) \to^{\mathrm{d}} \chi^2_q$ under the null hypothesis. The fact that the asymptotic distribution of $2l(t)$ does not depend on m (or η) proves that it is a self-normalized test statistic. This is a very appealing property because the test procedure is the same for all types of functional data and solving system (3.4) does not require estimating the variance of $H\{\check{\boldsymbol{\beta}}(t)\}$.

The following theorem summarizes the asymptotic distribution of $2l(t)$ under both the local alternative (1.3) and the null hypothesis H_0.

Theorem 1. Under conditions 1–4 and under the alternative H_{1n}, $H\{\boldsymbol{\beta}_0(t)\} = b_n \mathbf{d}(t)$ for $t \in [a, b]$, where $b_n = n^{-1} C_{n,\alpha_0,\eta}$ and $\mathbf{d}(t)$ is any fixed real vector of functions, we have $2l(t) \to^{\mathrm{d}} \chi^2_q\{\mathbf{d}^{\mathrm{T}}(t)\mathbf{R}(t)\mathbf{d}(t)\}$, where $\mathbf{d}^{\mathrm{T}}(t)\mathbf{R}(t)\mathbf{d}(t)$ is the non-centrality parameter. In particular, under H_0 and $\mathbf{d}(t) = 0$, $2l(t)$ follows a χ^2_q-distribution asymptotically.

Remark 4. An asymptotic α-level test is given by rejecting H_0 at a fixed point t if $2l(t) > \chi^2_{q,\alpha}$ where $\chi^2_{q,\alpha}$ is the upper α-quantile of χ^2_q. By taking a special function $H(\boldsymbol{\beta}) = \beta_j(t)$, we can also construct a $100(1-\alpha)\%$ confidence interval for $\beta_j(t)$ as $\mathrm{CI}_\alpha = \{\beta_j(t) : 2l(t) < \chi^2_{1,\alpha}\}$, which can be computed numerically. This provides an alternative self-normalized confidence interval to those proposed by Kim and Zhao (2013) who considered a mean model rather than the FCL model (1.1). Another distinction is that Kim and Zhao (2013) used a higher order kernel to reduce the bias whereas our method is based on bias-corrected estimating equations.

We define the detectable order b_n^* as the smallest order b_n in the alternative (1.3) that can be detected by the test proposed. For a given level of significance α,

$$\begin{aligned} b_n^* = \min_h b_n \text{ subject to} \\ \text{(a) type I error} \leqslant \alpha \text{ under } H_0 \text{ and} \\ \text{(b) the power is non-trivial under } H_{1n}. \end{aligned} \tag{3.6}$$

Theorem 1 guarantees that the test proposed controls the type I error at the nominal level asymptotically. For sparse and moderately dense functional data ($\eta \leqslant \frac{1}{8}$), condition 4 implies that $mh \to 0$ and hence $b_n = (nmh)^{-1/2}$ by theorem 1. In this case,

$$b_n^* = \min_h b_n = \min_h (nmh)^{-1/2} \qquad \text{subject to condition 4 on } h.$$

The optimal h that solves the minimization problem above is $h_* = n^{-(1+\eta+\delta)/9}$ for an arbitrarily small $\delta > 0$. This implies that the optimal b_n is $n^{-4(1+\eta)/9+\delta/18}$, which results in $b_n^* = n^{-4(1+\eta)/9}$ by letting $\delta \to 0$. For dense data ($\eta > \frac{1}{8}$), condition 4 leads to $mh \to \infty$. Theorem 1 implies that the test proposed has a non-trivial power under a local alternative of size $b_n^* = n^{-1/2}$, which is the detectable order of a parametric test.

4. A unified simultaneous test

We now consider a simultaneous test on H_0 in expression (1.2) for all $t \in [a, b]$. Intuitively, $2l(t)$ measures the distance between $H\{\boldsymbol{\beta}_0(t)\}$ and 0 at any $t \in [a, b]$. To test hypothesis (1.2) for all t, we propose a Cramér–von Mises-type test statistic

$$T_n = \int_a^b 2l(t) w(t) \mathrm{d}t, \tag{4.1}$$

where $w(\cdot)$ is a known probability density function. The construction of T_n enables us to borrow information across the time domain and yields a more powerful test than the pointwise test. Similar constructions were used by Härdle and Mammen (1993) and Chen and Zhong (2010).

4.1. Null distribution and local power

By the asymptotic decomposition of $2l(t)$ in proposition 3, we need first to understand the covariance structure of the process $\mathbf{U}_n(t)$ to investigate the distribution of T_n.

Proposition 4. Under conditions 1–4 and the null hypothesis H_0, we have $\mathrm{cov}\{\mathbf{U}_n(s), \mathbf{U}_n(t)\} = \boldsymbol{\Sigma}_n(s,t)\{1+o(1)\}$ where

$$\boldsymbol{\Sigma}_n(s,t) = \begin{cases} \mu_{20}^{-1} K^{(2)}\{(s-t)/h\}\mathbf{I}_q, & \text{if } m^2h \to 0, \\ \mathbf{I}_q I(s=t) + mh\boldsymbol{\Sigma}_0(s,t) I(s \neq t), & \text{if } m^2h \to \infty \text{ and } mh \to 0, \\ \boldsymbol{\Sigma}_0(s,t), & \text{if } mh \to \infty, \end{cases}$$

$K^{(2)}(x) = \int K(y)K(x-y)\,\mathrm{d}y$ and $\boldsymbol{\Sigma}_0(s,t) = \mathbf{G}(s)\boldsymbol{\Gamma}(s,t)\mathbf{G}^{\mathrm{T}}(t)\Omega(s,t)f(s)f(t)$.

The leading term in the covariance of $\mathbf{U}_n(t)$ is different under different asymptotic scenarios. In the second case, the $\mathbf{I}_q I(s=t)$ term in the expression for $\boldsymbol{\Sigma}_n(s,t)$ seems to dominate but is only non-zero in an area with Lebesgue measure 0; the $mh\boldsymbol{\Sigma}_0(s,t)I(s \neq t)$ term, in contrast, is the leading term almost everywhere.

Suppose that the covariance function $\boldsymbol{\Sigma}_n(s,t)$ has the spectral decomposition (Balakrishnan, 1960)

$$\boldsymbol{\Sigma}_n(s,t) = \sum_{k=1}^{\infty} \gamma_{nk}\boldsymbol{\phi}_{nk}(s)\boldsymbol{\phi}_{nk}^{\mathrm{T}}(t) \qquad \text{for any } s,t \in [a,b],$$

where $\gamma_{n1} \geqslant \gamma_{n2} \geqslant \ldots \geqslant 0$ are the ordered eigenvalues and $\boldsymbol{\phi}_{n1}(t), \boldsymbol{\phi}_{n2}(t), \ldots$ are the associated eigenfunctions. The eigenfunctions are vector-valued orthonormal functions satisfying $\int_a^b \boldsymbol{\phi}_{nk}^{\mathrm{T}}(t)\boldsymbol{\phi}_{nl}(t)w(t)\,\mathrm{d}t = \delta_k^l$ where $\delta_k^l = 1$ if $k = l$ and $\delta_k^l = 0$ otherwise. Even though the eigenvalues γ_{nk} change under different asymptotic scenarios, it is easy to verify that $\sum_{k=1}^{\infty}\gamma_{nk} = \mathrm{tr}\{\int \boldsymbol{\Sigma}_n(t,t)w(t)\mathrm{d}t\} = q$ for all cases in proposition 4. Also note that, in the third case of proposition 4, $\boldsymbol{\Sigma}_n = \boldsymbol{\Sigma}_0$ does not depend on n and therefore $\gamma_{nk} \equiv \gamma_k$ and $\boldsymbol{\phi}_{nk}(t) \equiv \boldsymbol{\phi}_k(t)$ for all k.

To establish the asymptotic distribution of T_n, we need to replace condition 4, part (b), with the following condition.

Condition 4.

(b′) $2(1+\eta)/17 < \alpha_0$ if $\eta \in [0, \frac{1}{8}]$ and $\frac{1}{8} < \alpha_0 < \eta$ if $\eta > \frac{1}{8}$.

Under the null hypothesis, we can show that $\mathbf{U}_n(t)$ is an asymptotically q-dimensional Gaussian process with mean $\mathbf{0}$ and covariance $\mathrm{cov}\{\mathbf{U}_n(s), \mathbf{U}_n(t)\} = \boldsymbol{\Sigma}_n(s,t)$. We shall show that the limiting distribution of T_n is the same as that of $Z_n = \int_a^b \mathbf{U}_n^{\mathrm{T}}(t)\mathbf{U}_n(t)w(t)\,\mathrm{d}t$, which follows a χ^2-mixture distribution. This result is described in the following theorem, the proof of which is provided in the on-line supplemental material.

Theorem 2. Under H_0 in hypothesis (1.2) and conditions 1–3, 4, parts (a) and (b′), $T_n =^{\mathrm{d}} Z_n\{1 + o_p(1)\}$, where $Z_n =^{\mathrm{d}} \sum_{k=1}^{\infty}\gamma_{nk}\chi_{1,k}^2$ and $\chi_{1,k}^2$, $k = 1, 2, \ldots$, are independent χ^2 random variables with 1 degree of freedom.

Remark 5. The asymptotic χ^2-mixture distribution in theorem 2 is quite different from the asymptotic normal distribution for classic EL ratio tests for independent data, time series or sparse longitudinal data (Chen *et al.*, 2003; Chen and Zhong, 2010). In fact, for dense functional data, our calculation shows that $\mathbb{E}[\{T_n - \mathbb{E}(T_n)\}^4] \neq 3\mathrm{var}^2(T_n)$, and hence T_n can behave quite differently from a Gaussian variable. However, for sparse or moderately dense functional data with $\eta \leqslant \frac{1}{16}$, the χ^2-mixture is also asymptotically normal. This result is given in the following corollary.

Corollary 1. Under the same conditions as those in theorem 2, if $\eta \leqslant \frac{1}{16}$, we have $h^{-1/2}(T_n - q) \to^{\mathrm{d}} N(0, q\sigma_0^2)$, where $\sigma_0^2 = 2\mu_{20}^{-2} \int_a^b w^2(t)\,\mathrm{d}t \int_{-2}^{2} K^{(2)}(u)^2 \mathrm{d}u$.

Corollary 1 makes a connection between the general results in theorem 2 and the classic results. The null distribution of T_n is different under different asymptotic scenarios and may depend on some unknown quantities such as γ_{nk}, which makes it difficult to use in practice. In the next subsection, we shall propose a bootstrap method that is unanimously applicable to all types of functional data to estimate this null distribution.

Next, we study the power of the simultaneous test under the local alternatives.

Theorem 3. Suppose that the local alternative hypothesis (1.3) holds and conditions 1–3, 4, part (a), and 4, part (b′), are satisfied.

(a) If $\eta \leqslant \frac{1}{16}$ and $b_n = n^{-1/2}(m^2h)^{-1/4}$, then $h^{-1/2}(T_n - q) \to^{\mathrm{d}} N(\mu_0, q\sigma_0^2)$, where

$$\mu_0 = \int_a^b \mathbf{d}^{\mathrm{T}}(t)\mathbf{R}(t)\mathbf{d}(t)w(t)\,\mathrm{d}t$$

and σ_0^2 is defined in corollary 1.

(b) If $\frac{1}{16} < \eta \leqslant \frac{1}{8}$, $\alpha_0 < 2\eta$ and $b_n = n^{-1/2+\epsilon}$ for an arbitrarily small $\epsilon > 0$, then $\sigma_1^{-1}(T_n - q - nb_n^2 mh\mu_0) \to^{\mathrm{d}} N(0,1)$ where $\sigma_1^2 = 4nb_n^2(mh)^2\mu_1$ and

$$\mu_1 = \int_a^b \int_a^b \mathbf{d}^{\mathrm{T}}(t)\mathbf{R}^{1/2}(t)\boldsymbol{\Sigma}_0(t,s)\mathbf{R}^{1/2}(s)\mathbf{d}(s)w(t)w(s)\,\mathrm{d}t\mathrm{d}s.$$

(c) Let $u_k = \int_a^b (\mathbf{R}^{1/2}(t)\mathbf{d}(t))^{\mathrm{T}}\phi_k(t)w(t)\mathrm{d}t$. If $\eta > \frac{1}{8}$ and $b_n = n^{-1/2}$, then $T_n \to^{\mathrm{d}} \Sigma_{k=1}^{\infty}\{\gamma_k \times \chi_{1,k}^2(u_k^2/\gamma_k)\}$.

We can use theorem 3 to examine the power and size of detectable signals of the simultaneous test under various scenarios. We use the principle (3.6) to determine the optimal detectable order b_n^*. When $\eta \leqslant \frac{1}{16}$, following part (a) in theorem 3, the asymptotic power of the test is $\mathcal{B}(\mathbf{d}) = \Phi(-z_\alpha + \mu_0/\sigma_0\sqrt{q})$ where μ_0 and σ_0 are defined in theorem 3 and $\Phi(\cdot)$ is the cumulative distribution function of the standard normal distribution. The test has non-trivial powers for signals of size $b_n = n^{-1/2}(m^2h)^{-1/4}$. Under constraints 4, part (a), and 4, part (b′), on h, b_n attains its minimum at $h_* = n^{-2(1+\eta+\delta)/17}$ for any arbitrary small $\delta > 0$ such that $b_n = n^{-8(1+\eta)/17+\delta/34}$. By letting $\delta \to 0$, the optimal detectable order is $b_n^* = n^{-8(1+\eta)/17}$.

When $\frac{1}{16} < \eta \leqslant \frac{1}{8}$, by our calculations in proposition 4 and theorem 2 the null distribution of T_n is a χ^2-mixture with mean $(\Sigma_{k=1}^{\infty}\gamma_{nk})\{1+o(1)\} = q\{1+o(1)\}$ and variance $(2\Sigma_k\gamma_{nk}^2)\{1+o(1)\} = \mathrm{tr}\{\int\int \boldsymbol{\Sigma}_n^2(s,t)\,w(s)\,w(t)\,\mathrm{d}s\,\mathrm{d}t\}\{1+o(1)\} = O(mh)$. Therefore, the threshold for an α-level test is of the form $q + c_{n,\alpha}$, where $c_{n,\alpha} \leqslant (2\Sigma_k\gamma_{nk}^2/\alpha)^{1/2} = O(mh)$ by Chebyshev's inequality. By part (b) of theorem 3, the asymptotic power is

$$\mathcal{B}(\mathbf{d}) = \Phi\left(-\frac{c_{n,\alpha}}{2\sqrt{nb_n}mh\sqrt{\mu_1}} + \frac{\mu_0}{2\sqrt{\mu_1}}\sqrt{nb_n}\right) \to 1,$$

for $b_n = n^{-1/2+\epsilon}$ with an arbitrarily small $\epsilon > 0$. This also means that the test has non-trivial powers for signals of size $b_n^* = n^{-1/2}$.

Similarly, the power of the test under case (c) is $\mathcal{B}(\mathbf{d}) = P\{\Sigma_{k=1}^{\infty}\gamma_k\chi_{1,k}^2(u_k^2/\gamma_k) > q + c_\alpha\}$, where $q + c_\alpha$ is the αth quantile of $\Sigma_{k=1}^{\infty}\gamma_k\chi_{1,k}^2$. In this case, $\mathcal{B}(\mathbf{d})$ is a constant as long as $\mathbf{d}(t)$ is a fixed non-zero function, which implies that the test has a non-trivial power if $b_n = n^{-1/2}$. Combining parts (b) and (c), the optimal detectable order of the simultaneous test is $b_n^* = n^{-1/2}$ when $\eta > \frac{1}{16}$.

Note that the optimal detectable order for the simultaneous test is smaller than that of the pointwise test when $\eta \leqslant \frac{1}{8}$. This is understandable because the simultaneous test borrows

information over the entire time domain and is more powerful. Both the pointwise and the simultaneous tests can detect signals of $\sqrt{n}$-order for dense functional data with $\eta > \frac{1}{8}$.

An interesting question that was raised by one referee is about the possibility of selecting the bandwidth and the weight function $w(t)$ through maximizing the power function of the test. Following the discussion above, we find that the asymptotic power depends not only on h and $w(t)$, but also on an unknown function $\mathbf{d}(t)$, which is the deviation of the truth from the null hypothesis. Moreover, the expression of the power depends on the asymptotic regime. As we discussed earlier, sparse and dense functional data are asymptotic concepts, and are difficult to distinguish in practice. Therefore, choosing the optimal bandwidth and weight function through maximizing the power remains a challenging open question. For now, we leave $w(t)$ as a subjective choice of the practitioner. The most commonly used weight function is a uniform density to put equal weights on all points. Another natural choice of $w(t)$ is the density function of t_{ij}, which puts higher weights on the interval with more data information. In addition, if there is prior knowledge on the importance of a particular subinterval, one can change $w(t)$ to put more weights on the important subinterval. Bandwidth selection is further addressed in Section 5.

4.2. Wild bootstrap procedure

For the simultaneous test, the asymptotic distributions of T_n are different for sparse and dense functional data. But the boundary between different scenarios is defined only in the asymptotic sense, making different asymptotic scenarios difficult to distinguish in practice. To unify the inference procedure, we propose a wild bootstrap procedure (Mammen, 1993). Some residual-based bootstrap procedures have also been proposed in Faraway (1997) and Zhang and Chen (2007) for dense functional data, but the consistency of such procedures was not investigated.

The bootstrap procedure proposed consists of the following steps.

Step 1: generate bootstrapped samples $(Y_{ij}^*, t_{ij}, \mathbf{X}_{ij})$ according to the model $Y_{ij}^* = \tilde{\boldsymbol{\beta}}^{\mathrm{T}}(t_{ij})\mathbf{X}_{ij} + \epsilon_{ij}^*$, where $\tilde{\boldsymbol{\beta}}(t_{ij})$ is the solution of the estimating equations (3.4). The residual vector $\boldsymbol{\epsilon}_i^* = (\epsilon_{i1}^*, \ldots, \epsilon_{im_i}^*)^{\mathrm{T}}$ is generated from an m_i-dimensional multivariate normal distribution with mean $\mathbf{0}$ and covariance $\hat{\boldsymbol{\Omega}}_i = \{\hat{\Omega}(t_{ij}, t_{ik})\}_{j,k=1}^{m_i}$ where $\hat{\Omega}(t,s)$ is a consistent estimator of $\Omega(t,s)$ described in Section 5.2.

Step 2: on the basis of the bth bootstrapped sample, compute a bootstrapped version of T_n, denoted as $T_n^{*(b)}$.

Step 3: repeat steps 1 and 2 B times to obtain bootstrap values $\{T_n^{*(b)}\}_{b=1}^B$ and let $\hat{t}_\alpha$ be the $100(1-\alpha)\%$ quantile of $\{T_n^{*(b)}\}_{b=1}^B$. Reject the null hypothesis if $T_n > \hat{t}_\alpha$.

The following theorem justifies the above bootstrap procedure.

Theorem 4. Let $\mathcal{X}_n = \{(Y_{ij}, X_{ij}, t_{ij}),\ j = 1, \ldots, m_i,\ i = 1, \ldots, n\}$ denote the original data and $\mathcal{L}(T_n)$ be the asymptotic distribution of T_n under the null hypothesis. Under the same conditions as theorem 2 and supposing that $\hat{\Omega}(s,t)$ is a consistent covariance estimator, the conditional distribution of T_n^* given $\mathcal{X}_n$, $\mathcal{L}(T_n^* | \mathcal{X}_n)$, converges to $\mathcal{L}(T_n)$ almost surely.

5. Implementation issues

5.1. Bandwidth selection

The performance of the estimation and test procedures depends on the choice of the bandwidth h. Our asymptotic theory relies on h falling in the range defined in condition 4. For longitudinal data (sparse functional data) where subjects are assumed to be independent, one may apply a 'leave-one-out' cross-validation strategy (Rice and Silverman, 1991) to choose the

bandwidth. However, cross-validation is time consuming and, in general, its performance for dense functional data is unknown.

We propose to select the bandwidth through minimizing the conditional integrated asymptotic mean-squared error (IAMSE) of the local polynomial estimator $\hat{\boldsymbol{\beta}}(t)$. By expression (2.2), the bandwidth h that minimizes the IAMSE of $\hat{\boldsymbol{\beta}}(t)$ is of the order of $n^{-(1+\eta)/5}$, which satisfies condition 4 for both sparse and dense cases. Let $\mathcal{D}=\{(t_{ij},\mathbf{X}_{ij}),\ j=1,2,\ldots,m_i,\ i=1,2,\ldots,n\}$. It is not difficult to show that, for any fixed t, $\mathrm{AMSE}\{\hat{\boldsymbol{\beta}}(t)|\mathcal{D}\}=\mathbf{b}^{\mathrm{T}}(t)\mathbf{b}(t)+\mathrm{tr}[\mathrm{cov}\{\hat{\boldsymbol{\beta}}(t)|\mathcal{D}\}]$, where $\mathbf{b}(t)=\mathrm{bias}\{\hat{\boldsymbol{\beta}}(t)|\mathcal{D}\}$. The IAMSE is defined as

$$\mathrm{IAMSE}\{\hat{\boldsymbol{\beta}}(\cdot)|\mathcal{D}\}=\int_a^b \mathrm{AMSE}\{\hat{\boldsymbol{\beta}}(t)|\mathcal{D}\}\varpi(t)f(t)\mathrm{d}t,$$

where $\varpi(t)$ is a known weight function and $f(t)$ is the probability density function of t_{ij}. The conditional bias is $\mathbf{b}(t)=(\mathbf{I},\mathbf{0})\{\mathbf{D}^{\mathrm{T}}(t)\mathbf{W}(t)\mathbf{D}(t)\}^{-1}\mathbf{D}^{\mathrm{T}}(t)\mathbf{W}(t)\mathbf{l}(t)$, where $\mathbf{l}(t)=(l_{11}(t),\ldots,l_{1m_1}(t),$ $l_{21}(t),\ldots,l_{nm_n}(t))^{\mathrm{T}}$ with $l_{ij}(t)\approx\mathbf{X}_{ij}^{\mathrm{T}}\boldsymbol{\beta}^{(2)}(t)(t_{ij}-t)^2/2$, and $\boldsymbol{\beta}^{(s)}(t)=\{\beta_1^{(s)}(t),\ldots,\beta_p^{(s)}(t)\}^{\mathrm{T}}, s=1,2,$ is the sth derivative of $\boldsymbol{\beta}_0(t)$. The conditional covariance is

$$\mathrm{cov}\{\hat{\boldsymbol{\beta}}(t)|\mathcal{D}\}=(\mathbf{I},\mathbf{0})\{\mathbf{D}^{\mathrm{T}}(t)\mathbf{W}(t)\mathbf{D}(t)\}^{-1}\mathbf{D}^{\mathrm{T}}(t)\mathbf{W}(t)\boldsymbol{\Omega}\mathbf{W}(t)\mathbf{D}(t)\{\mathbf{D}^{\mathrm{T}}(t)\mathbf{W}(t)\mathbf{D}(t)\}^{-1}$$

where $\boldsymbol{\Omega}=\mathrm{cov}(\mathbf{Y}|\mathcal{D})=\mathrm{diag}(\boldsymbol{\Omega}_1,\boldsymbol{\Omega}_2,\ldots,\boldsymbol{\Omega}_n)$ and $\boldsymbol{\Omega}_i=\{\Omega(t_{ij},t_{ik})\}_{j,k=1}^{m_i}$.

An estimator of the covariance $\Omega(s,t)$ is described in Section 5.2. To estimate $\boldsymbol{\beta}^{(2)}(t)$, we use a higher order local polynomial estimator of $\boldsymbol{\beta}_0(t)$ with a pilot bandwidth h_*. The pilot bandwidth is obtained by minimizing the residual squares criterion in Zhang and Lee (2000). Additional details are given in the on-line supplemental material. By replacing $\boldsymbol{\beta}^{(2)}(t)$ and $\boldsymbol{\Omega}$ with their estimators $\widehat{\boldsymbol{\beta}^{(2)}}(t)$ and $\hat{\boldsymbol{\Omega}}$, we obtain estimators of the conditional mean and covariance, $\hat{\mathbf{b}}(t)$ and $\widehat{\mathrm{cov}}\{\hat{\boldsymbol{\beta}}(t)|\mathcal{D}\}$. Then the bandwidth h is chosen by minimizing the empirical IAMSE

$$\hat{h}=\arg\min_h \frac{1}{N}\sum_{i=1}^{n}\sum_{j=1}^{m_i}\widehat{\mathrm{AMSE}}\{\hat{\boldsymbol{\beta}}(t_{ij})|\mathcal{D}\}\varpi(t_{ij}),$$

where $N=\Sigma_{i=1}^n m_i$ and $\widehat{\mathrm{AMSE}}\{\hat{\boldsymbol{\beta}}(t)|\mathcal{D}\}=\hat{\mathbf{b}}^{\mathrm{T}}(t)\hat{\mathbf{b}}(t)+\mathrm{tr}[\widehat{\mathrm{cov}}\{\hat{\boldsymbol{\beta}}(t)|\mathcal{D}\}]$. Note that we only guarantee that the order of $\hat{h}$ falls in the right range and selecting the optimal bandwidth for testing remains an unsolved problem. Our extensive simulation study showed that multiplying $\hat{h}$ by a constant can significantly improve the numerical performance, and any constant between 0.25 and 0.75 produced similar results. We used $0.25\hat{h}$ as the bandwidth in our numerical studies.

5.2. Covariance estimation

The covariance function $\Omega(\cdot,\cdot)$ can be estimated by the non-parametric kernel estimator of Yao *et al.* (2005a), which is uniformly consistent (Li and Hsing, 2010). However, the non-parametric covariance estimator is not necessarily positive semidefinite. Instead, we adopt the semiparametric covariance estimator of Fan *et al.* (2007). Suppose that the covariance function can be decomposed as $\Omega(s,t)=\sigma(s)\rho(s,t)\sigma(t)$. We model the variance function $\sigma^2(t)$ non-parametrically and the correlation function $\rho(s,t)$ parametrically. For estimation, we first apply the non-parametric kernel estimators of $\rho(s,t)$ and $\sigma^2(t)$ (Yao *et al.*, 2005a) to obtain information about the parametric structure of $\rho(s,t)$. Then we can fit a parametric model to $\rho(s,t)$ by using least squares or the quasi-maximum-likelihood estimator of Fan *et al.* (2007). The parametric structure guarantees the positive semidefiniteness of the estimated correlation function. For more details of the implementation, see Section 6 and section B in the on-line supplemental material.

6. Simulation studies

Simulation studies were conducted to evaluate the performance of the proposed unified inference procedures. We generated data from the model

$$Y_i(t_{ij}) = \beta_1(t_{ij})X_i^{(1)}(t_{ij}) + \beta_2(t_{ij})X_i^{(2)}(t_{ij}) + \epsilon_i(t_{ij}), \tag{6.1}$$

for $i = 1, 2, \ldots, n$ and $j = 1, 2, \ldots, m$ where the t_{ij}s are independent and identically distributed Unif(0,1), $X_i^{(1)}(t_{ij}) = 1 + 2\exp(t_{ij}) + v_{ij}$ and $X_i^{(2)}(t_{ij}) = 3 - 4t_{ij}^2 + u_{ij}$. Here, u_{ij} and v_{ij} are independent and identically distributed $N(0, 1)$ random variables, which are independent of t_{ij} and $\epsilon_i(t_{ij})$. The random error $\epsilon_i(t_{ij})$ was generated from a zero-mean auto-regressive AR(1) process such that $\mathrm{var}\{\epsilon(t)\} = 1$ and $\mathrm{cov}\{\epsilon(t), \epsilon(t-s)\} = \rho^{10s}$ for some $\rho \in (0, 1)$. To evaluate the proposed methods for both sparse and dense data, we set $m = 5, 10, 50$. The sample sizes were chosen to be 100 and 200. The Epanechnikov kernel $K(x) = \frac{3}{4}(1 - x^2)_+$ was used for estimation and hypotheses testing, where $(a)_+ = \max(a, 0)$. Bandwidth selection was conducted for every simulated data set by using the method that was described in Section 5. In our numeric studies, we used the uniform density function on (0,1) as the weight function $w(t)$ if not particularly specified, which assumes no preference for any particular intervals.

We first set $\beta_1(t) = \frac{1}{2}\sin(t)$ and $\beta_2(t) = 2\sin(t + 0.5)$ in model (6.1) and applied the procedure in Section 3 to construct pointwise confidence intervals for $\beta_1(t)$. Table 1 summarizes the empirical coverage probability as percentages and the average length of the confidence intervals (in parentheses) for $\beta_1(t)$ at $t = 0.3, 0.5, 0.7$ based on 1000 simulation replicates. As we can see from Table 1, the coverage probabilities are close to the nominal level 95% in both sparse and dense cases and the average lengths are shorter under a larger sample size. In addition, the average lengths improve as m increases from 5 to 50.

Next, we considered unified simultaneous inference. We considered three scenarios A, B and C, corresponding to three hypotheses on $\boldsymbol{\beta}(t)$. In scenario A, we used $H\{(z_1, z_2)^{\mathrm{T}}\} = z_1 - z_2$ to test $H_{0\mathrm{A}} : \beta_1(\cdot) = \beta_2(\cdot)$ *versus* $H_{1A} : \beta_1(\cdot) \neq \beta_2(\cdot)$, where we set $\beta_1(t) = \frac{1}{2}\sin(t)$ and $\beta_2(t) = (\frac{1}{2} + a)\sin(t)$ for $a = 0, 0.1, 0.2, 0.3, 0.4$ in model (6.1) to evaluate the empirical size (when $a = 0$) and powers (when $a > 0$). In scenario B, we set $H\{(z_1, z_2)^{\mathrm{T}}\} = z_2$ to test $H_{0\mathrm{B}} : \beta_2(\cdot) = 0$ *versus* $H_{1\mathrm{B}} : \beta_2(\cdot) \neq 0$, where we chose $\beta_1(t) = \frac{1}{2}\sin(t)$ and $\beta_2(t) = c$ for $c = 0, 0.02, 0.04, \ldots, 0.14$. In scenario C, we considered a non-linear functional constraint $H\{(z_1, z_2)^{\mathrm{T}}\} = z_1 z_2 - 1$ to test $H_{0\mathrm{C}} : \beta_1(\cdot)\beta_2(\cdot) = 1$ *versus* $H_{1\mathrm{C}} : \beta_1(\cdot)\beta_2(\cdot) \neq 1$, where $\beta_1(t) = 1 + t$ and $\beta_2(t) = (1 + d)/(1 + t)$ for $d = 0, 0.05, 0.1, 0.5$. A similar non-linear hypothesis to that in scenario C was considered by Dufour (1989). In the construction of the test statistic T_n, we chose the weight function $w(t) = 1$ for $t \in (0, 1)$ and $w(t) = 0$ otherwise. The covariance function was estimated by the quasi-maximum-likelihood method of Fan *et al.* (2007). All simulation results below were based on 500 simulation replicates and the critical value of the test was estimated by 500 bootstrap samples in each simulation run. We performed the same bandwidth selection procedure in each bootstrap sample to take into account the extra variation in the test caused by bandwidth selection.

Table 2 summarizes the empirical sizes and powers for hypothesis $H_{0\mathrm{A}}$ at the 5% nominal level. It can be seen that the empirical sizes are reasonably controlled around the nominal level. As we expected, the empirical power increases as the increase in the sample size n and the number of repeated measurements m, which confirms our theoretical results in Section 4. In addition, the correlation ρ does not have a clear effect on the power, indicating that the procedure proposed is robust with respect to the covariance structure of the random error.

For comparison, we also implemented the method that was proposed by Zhang and Chen (2007), which was proposed to test linear hypotheses for dense functional data. We applied the bootstrap test procedure of Zhang and Chen (2007) to scenario A with $m = 50$ and summarized

Table 1. Empirical coverage probability and average length of pointwise confidence intervals (in parentheses) for $\beta_1(t)$ at $t = 0.3, 0.5, 0.7$

t	n	*Results (%) for* $m=5$		*Results (%) for* $m=10$		*Results (%) for* $m=50$	
		$\rho=0.2$	$\rho=0.5$	$\rho=0.2$	$\rho=0.5$	$\rho=0.2$	$\rho=0.5$
0.3	100	92.3 (0.531)	92.4 (0.537)	94.2 (0.393)	92.8 (0.394)	95.4 (0.205)	95.0 (0.206)
	200	94.0 (0.395)	95.3 (0.391)	94.8 (0.301)	96.5 (0.298)	95.6 (0.157)	95.2 (0.160)
0.5	100	94.0 (0.524)	93.0 (0.524)	93.9 (0.394)	95.0 (0.404)	95.6 (0.212)	94.2 (0.217)
	200	95.4 (0.399)	94.4 (0.397)	94.8 (0.299)	94.7 (0.297)	95.7 (0.159)	95.4 (0.162)
0.7	100	93.6 (0.528)	93.3 (0.525)	94.8 (0.394)	94.0 (0.397)	95.0 (0.222)	95.0 (0.219)
	200	94.2 (0.401)	94.7 (0.409)	94.2 (0.302)	94.6 (0.307)	94.1 (0.160)	95.3 (0.160)

Table 2. Empirical size and power for testing $H_{0A} : \beta_1(\cdot) = \beta_2(\cdot)$ under scenario A using the proposed method†

a	n	*Results for the method proposed*						*Results for the method of Zhang and Chen (2007)*	
		$m=5$		$m=10$		$m=50$		$m=50$	
		$\rho=0.2$	$\rho=0.5$	$\rho=0.2$	$\rho=0.5$	$\rho=0.2$	$\rho=0.5$	$\rho=0.2$	$\rho=0.5$
0.0	100	0.052	0.062	0.048	0.050	0.040	0.052	0.062	0.042
	200	0.052	0.050	0.056	0.040	0.038	0.038	0.066	0.046
0.1	100	0.168	0.152	0.290	0.248	0.828	0.800	0.348	0.316
	200	0.222	0.264	0.444	0.450	0.982	0.980	0.654	0.684
0.2	100	0.490	0.444	0.762	0.740	1.000	1.000	0.982	0.968
	200	0.710	0.738	0.956	0.976	1.000	1.000	1.000	1.000
0.3	100	0.810	0.804	0.984	0.978	1.000	1.000	1.000	1.000
	200	0.980	0.980	1.000	1.000	1.000	1.000	1.000	1.000
0.4	100	0.964	0.954	0.998	1.000	1.000	1.000	1.000	1.000
	200	0.998	1.000	1.000	1.000	1.000	1.000	1.000	1.000

†For the dense case with $m = 50$, we also compared it with the method proposed by Zhang and Chen (2007).

the empirical sizes and powers in the last two columns of Table 2. As we observe from Table 2, Zhang and Chen's method controlled the empirical sizes to the nominal level but had slightly lower power than our method.

The simulation results for scenario B are illustrated in Figs 2(a) and 2(b). The results under $n = 100$ and $n = 200$ are represented by full and broken lines respectively. We observed a very similar pattern to that under scenario A. The size is well controlled at the 5% nominal level and the power increases as the value of c increases. At each value of c, the power increases as we increase n or m. For the non-linear hypothesis testing that was considered in scenario C, the results are shown in Figs 2(c) and 2(d). We observe very similar results in scenario C to those for linear hypotheses in scenarios A and B.

To demonstrate further the performance of the proposed bandwidth selection method in Section 5, we show in Fig. 3 the boxplots of $\hat{h}$ selected for model (6.1) with $\beta_1(t) = \frac{1}{2}\sin(\pi t)$ and $\beta_2(t) = 2\sin(\pi t + 0.5)$ based on 500 replicates. Both the median and the spread of $\hat{h}$ decreased as n and m increased and the correlation ρ had little effect on the bandwidth selection result.

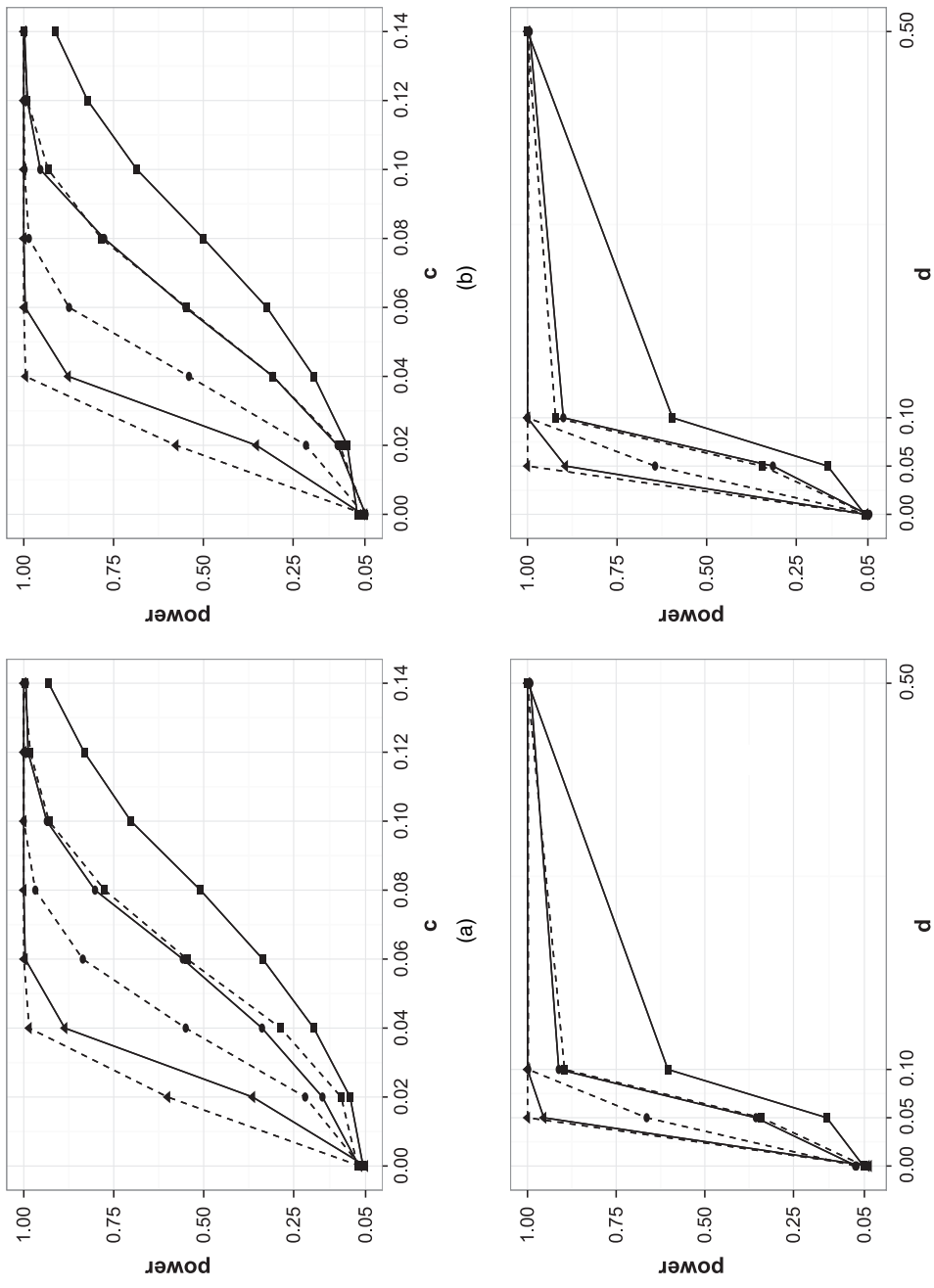


Fig. 2. (a), (b) Empirical size and power for testing $H_{0B}: \beta_2(\cdot) = 0$ at the 5% nominal level under scenario B and (*c*), (d) empirical size and power for testing $H_{0C}: \beta_1(\cdot)\beta_2(\cdot) = 1$ at the 5% nominal level under scenario C (■, $m = 5$; ●, $m = 10$; ▲, $m = 50$; ———, $n = 100$; – – –, $n = 200$): (a), (c) $\rho = 0.2$; (b), (d) $\rho = 0.5$

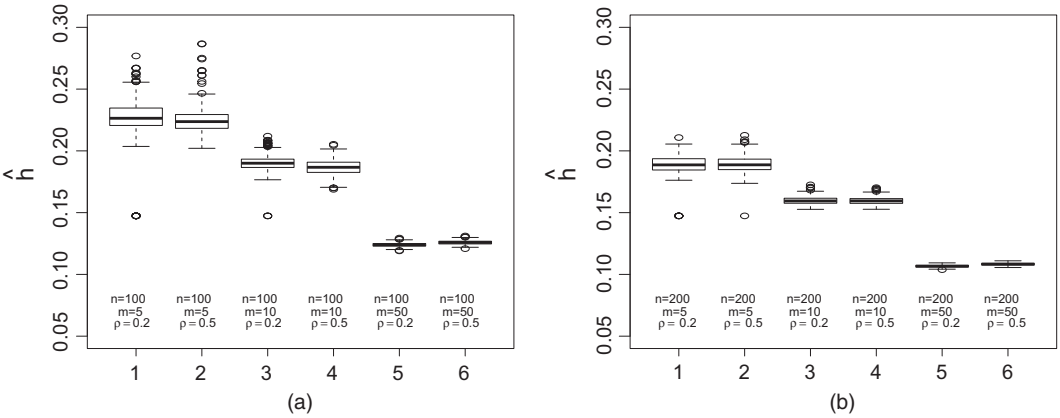


Fig. 3. Boxplots for bandwidths selected for model (6.1) with $\beta_1(t) = \frac{1}{2}\sin(\pi t)$ and $\beta_2(t) = 2\sin(\pi t + 0.5)$ by using the bandwidth selection method proposed: (a) $n = 100$; (b) $n = 200$

These plots also show that our bandwidth selection procedure is very stable as there are very few outliers in each case.

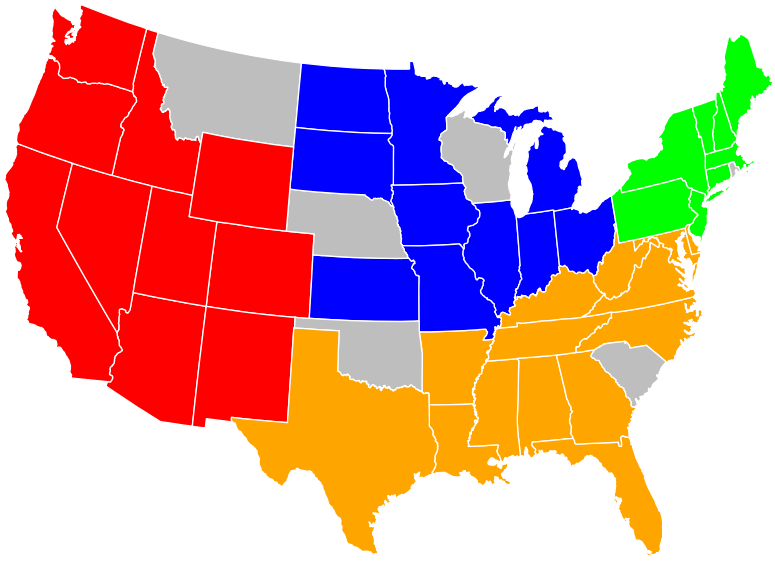
7. Real data analysis

We applied our proposed methods to the two data sets that were described in Section 1, which contain dense and sparse functional data.

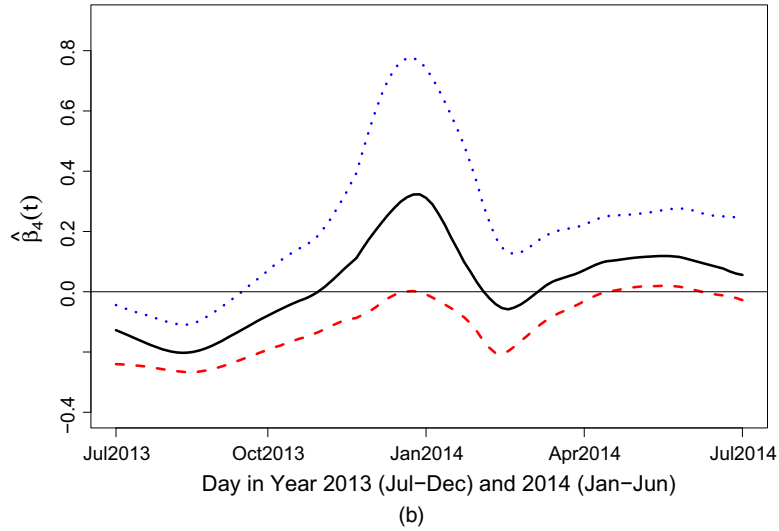
7.1. Google flu data

Google flu trend is a realtime Web service providing aggregated search-queries-based estimates of flu activity for a number of countries and regions. Google flu trend reports weekly the estimated numbers of influenza-like illness (ILI) cases per 100 000 doctor visits. Most of these estimates were found to be consistent with the estimates that are provided by the Centers for Disease Control and Prevention. We were interested in studying the relationship between flu activity and temperature fluctuation in the USA. For this, we collected state level flu activity data in the 2013–2014 flu season (July 2013–June 2014) from the Google flu trend Web site and the state level temperature data from the US historical climatology network. The US historical climate network provides daily maximum and minimum temperature averaged over weather stations within each continental state of the USA. The daily temperature variation is the difference between the daily maximum and daily minimum. We aggregated the temperature fluctuation data to the same resolution as the flu activity data by taking the MDTV within each week. Because part of the temperature records are missing for some states in the US historical climate network, the data set that we used contains ILI percentages and MDTV for 42 states, which are illustrated in Fig. 1. These 42 states can be classified into four regions ('north-east', 'midwest', 'south' and 'west') according to the US Census Bureau. For illustration, the 42 states and their regions are plotted in Fig. 4(a).

In our analysis, we standardized ILI percentage and MDTV at each time point t by dividing the variables by their root mean squares. The original dates from July 1st, 2013, to June 30th, 2014, were numbered by integers from 1 to 365. We then rescaled the time t to the [0,1] interval by dividing the numbers by 365. Let $Y(t)$ and $\mathcal{U}(t)$ be respectively the standardized ILI percentage and MDTV at time $t \in [0, 1]$. To incorporate the regional effects, we use midwest as the baseline



(a)



(b)

Fig. 4. (a) The 42 states with complete temperature records in the US historical climate network and their corresponding regions (■, north-east states; ■, midwest states; ■, south states; ■, west states; ■, states with missing values) and (b) pointwise confidence bands for $\beta_4(t)$ in the reduced model (7.2) (· · · · · ·, upper; ———, $\hat{\beta}_4(t)$; – – –, lower)

level and let dummy variables Z_1, Z_2 and Z_3 be indicators for north-east, south and west respectively. We considered the following functional concurrent linear model:

$$Y_i(t) = \beta_0(t) + \sum_{k=1}^{3} \beta_k(t) Z_{ki} + \beta_4(t)\mathcal{U}_i(t) + \sum_{k=5}^{7} \beta_k(t) Z_{(k-4)i}\mathcal{U}_i(t) + \epsilon_i(t), \tag{7.1}$$

where i is the index for state, $\beta_4(t)$ represents the main effect of MDTV, $\beta_j(t)$ ($j = 1, 2, 3$) represent the regional main effects and $\beta_j(t)$ ($j = 5, 6, 7$) are the interactions between

MDTV and the regional indicators. This FCL model helps us to assess the dynamic effect of MDTV on flu activities and to take timely action to prevent and control flu outbreaks. In a matrix form, model (7.1) can be represented as $\mathbf{Y}(t) = \mathbf{X}(t)\boldsymbol{\beta}(t) + \boldsymbol{\epsilon}(t)$ where $\mathbf{Y}(t) = (Y_1(t), \ldots, Y_n(t))^{\mathrm{T}}$, $\mathbf{X}_i(t) = (1, Z_{1i}, Z_{2i}, Z_{3i}, \mathcal{U}_i(t), Z_{1i}\mathcal{U}_i(t), Z_{2i}\mathcal{U}_i(t), Z_{3i}\mathcal{U}_i(t))^{\mathrm{T}}$, $\boldsymbol{\epsilon}(t) = (\epsilon_1(t), \ldots, \epsilon_n(t))^{\mathrm{T}}$, $\mathbf{X}(t) = (\mathbf{X}_1(t), \ldots, \mathbf{X}_n(t))^{\mathrm{T}}$ and $\boldsymbol{\beta}(t) = (\beta_0(t), \ldots, \beta_7(t))^{\mathrm{T}}$.

Since the ILI percentages were collected spatially, the random errors $\{\epsilon_i(t)\}_{i=1}^n$ in model (7.1) might be spatially correlated. To apply our test procedure, we first preprocessed the data to remove spatial correlations. For each time point t_0, we model the spatial dependence in $\epsilon(t_0)$ by a conditionally auto-regressive (CAR) model (Wall, 2004; Banerjee *et al.*, 2014). Specifically, we assume that the conditional distribution of $\epsilon_i(t_0)$ given the rest of the states $\boldsymbol{\epsilon}_{(-i)}(t_0)$ has mean $\Sigma_{j\neq i} c_{ij}\epsilon_j(t_0)$ and variance $\sigma_\epsilon^2(t_0)$. Then the joint distribution of $\boldsymbol{\epsilon}(t_0)$ has mean $\mathbf{0}$ and covariance $\{\mathbf{I} - \mathbb{C}(t_0)\}^{-1}\sigma_\epsilon^2(t_0)$. Following the standard CAR model, we assume that $\mathbb{C}(t_0) = \lambda(t_0)\mathbb{W}(t_0)$ where $\lambda(t_0)$ is a spatial auto-correlation parameter and $\mathbb{W}(t_0) = (w_{ij}(t_0))$ is an adjacency matrix, i.e. $w_{ij}(t_0) = 1$ if states i and j share a boundary; otherwise $w_{ij}(t_0) = 0$. We fitted the CAR model to the residuals of the FCL model to obtain estimates $\hat{\sigma}_\epsilon^2(t_0)$ and $\hat{\mathbb{C}}(t_0)$ at every t_0. We used the function `spautolm` in the R package `spdep` (Bivand, 2017) to fit the CAR model. Define the transformed response as $\tilde{\mathbf{Y}}(t_0) = \{\mathbf{I} - \hat{\mathbb{C}}(t_0)\}^{1/2}\mathbf{Y}(t_0)/\hat{\sigma}_\epsilon(t_0)$, and the transformed covariate and error as $\tilde{\mathbf{X}}(t_0) = \{\mathbf{I} - \hat{\mathbb{C}}(t_0)\}^{1/2}\mathbf{X}(t_0)/\hat{\sigma}_\epsilon(t_0)$ and $\tilde{\boldsymbol{\epsilon}}(t_0) = \{\mathbf{I} - \hat{\mathbb{C}}(t_0)\}^{1/2}\boldsymbol{\epsilon}(t_0)/\hat{\sigma}_\epsilon(t_0)$. The transformed model becomes $\tilde{\mathbf{Y}}(t_0) = \tilde{\mathbf{X}}(t_0)\boldsymbol{\beta}(t_0) + \tilde{\boldsymbol{\epsilon}}(t_0)$, where the coefficient functions $\boldsymbol{\beta}(t)$ remain unchanged but the errors $\tilde{\boldsymbol{\epsilon}}(t_0)$ are spatially uncorrelated. We applied the proposed method to the transformed model to make inference about $\boldsymbol{\beta}(t)$.

We first tested the significance of interactions with $H_0 : \beta_5(t) = \beta_6(t) = \beta_7(t) = 0$ and then tested the regional main effects with $H_0 : \beta_1(t) = \beta_2(t) = \beta_3(t) = 0$. The EL test p-values for the two hypotheses were 0.354 and 0.272 respectively, based on 1000 bootstrap samples. Since neither hypothesis was significant at the nominal level 0.05, we considered the following reduced model:

$$Y_i(t) = \beta_0(t) + \beta_4(t)\mathcal{U}_i(t) + \epsilon_i(t). \tag{7.2}$$

After removing the spatial correlations as described above, we applied the proposed EL method to construct the 95% pointwise confidence bands for $\beta_4(t)$ under model (7.2), which is presented in Fig. 4(b). On the basis of estimation of $\beta_4(t)$, we observe that the effect of temperature fluctuation on flu activity changes over time and reaches its peak value in the winter season around January. We further considered the hypothesis $H_0 : \beta_4(\cdot) = 0$ *versus* $H_1 : \beta_4(\cdot) \neq 0$, and the p-value of the EL test was 0.052, which was moderately significant.

7.2. Alzheimer's disease neuroimaging initiative data

AD is an irreversible, progressive brain disorder and one of the most common forms of dementia. According to the 2010 world Alzheimer's report, the disease affects about 35.6 million people around the world (Weiner *et al.*, 2012). There are four stages for the disease: cognitively normal (CN), early mild cognitive impairment (EMCI), late mild cognitive impairment (LMCI) and AD. The AD neuroimaging initiative is an on-going, multicentre longitudinal project designed to identify biomarkers for early detection and tracking of the disease, particularly focused on the use of brain imaging methods. The hippocampus is the brain region that is damaged first by AD and it is the functional region associated with memory loss and disorientation. Thus, we used the volume of hippocampus as the covariate in our analysis.

The data set in our analysis consists of 628 subjects at the four stages of the disease: 215 CN, 99 EMCI, 254 LMCI and 60 AD. The earliest examination date for this cohort was September 7th, 2005, and the latest date was April 23rd, 2015. Interested readers can download the data set from `http://adni.loni.usc.edu`. The study followed up most patients up to 1 year

Table 3. *p*-values for pairwise comparisons in the AD neuroimaging initiative study

Hypothesis	*p-value*	*Hypothesis*	*p-value*
CN *versus* EMCI : $\beta_1(\cdot)=0$	0.001	EMCI *versus* LMCI : $\beta_1(\cdot)=\beta_2(\cdot)$	0.000
CN *versus* LMCI : $\beta_2(\cdot)=0$	0.000	EMCI *versus* AD : $\beta_1(\cdot)=\beta_3(\cdot)$	0.000
CN *versus* AD : $\beta_3(\cdot)=0$	0.000	LMCI *versus* AD : $\beta_2(\cdot)=\beta_3(\cdot)$	0.994

and the numbers of repeated measurements ranged from 3 to 10. One of the major symptoms of AD is cognitive impairment. In the AD neuroimaging iniative studies, cognitive performance was measured by MMSE, which is a questionnaire test. The maximum MMSE score is 30 and typically the MMSE score declines as the disease progresses.

The interest of our study was to understand the relationship between the volume of hippocampus and the MMSE score at different stages. We used the real age of a patient as time, and let $Y_i(t)$ and $X_i(t)$ be the MMSE score and the volume of hippocampus region for the ith individual measured at time t. To include the effects of different stages of the disease, let Z_{1i}, Z_{2i} and Z_{3i} be indicators of EMCI, LMCI and AD respectively. We considered the model

$$Y_i(t_{ij}) = \beta_0(t_{ij}) + \sum_{k=1}^{3} \beta_k(t_{ij}) Z_{ki} + X_i(t_{ij})\beta_4(t_{ij}) + \sum_{k=5}^{7} \beta_k(t_{ij}) Z_{(k-4)i} X_i(t_{ij}) + \epsilon_i(t_{ij}).$$

The advantage of this model is that we can evaluate the time varying and stage-dependent effects of $X(t)$ on the MMSE score. To avoid overfitting, we conducted hypothesis tests to select the appropriate model. We first tested the interactions between the hippocampus volume and the stages indicators, where the hypothesis was $H_0 : \beta_5(t) = \beta_6(t) = \beta_7(t) = 0$ for all t. The simultaneous EL test yielded a p-value of 0.167 based on 1000 bootstrap replicates. Since there was no significant interaction, we considered the reduced model

$$Y_i(t_{ij}) = \beta_0(t_{ij}) + \sum_{k=1}^{3} \beta_k(t_{ij}) Z_{ki} + X_i(t_{ij})\beta_4(t_{ij}) + \epsilon_i(t_{ij}). \tag{7.3}$$

Under such a model, we tested the main effect of hippocampus volume, $H_0 : \beta_4(t) = 0$, and the mean effects of the stage indicators, $H_0 : \beta_1(t) = \beta_2(t) = \beta_3(t) = 0$. The simultaneous EL test for both hypotheses yielded p-values less than 0.001.

We further conducted pairwise comparisons between the different groups of patients, namely testing $H_{0j} : \beta_j(\cdot) = 0$ for $j = 1, 2, 3$ and $H_{0,jk} : \beta_j(\cdot) = \beta_k(\cdot)$, $j \neq k = 1, 2, 3$. The p-values are summarized in Table 3. All p-values for the pairwise comparisons are less than or equal to 0.001 except that with 0.994 when comparing LMCI with AD. This indicates that there is no significant difference between the LMCI and AD groups. These findings also stress the importance of early treatment for the disease.

8. Discussion

Non-parametric hypothesis testing has been a very active research area for the past 25 years (González-Manteiga and Crujeiras, 2013). We made a new contribution in this area by proposing EL-based procedures to make pointwise and simultaneous inferences on functional concurrent linear models, treating sparse and dense functional data in a unified framework. We showed that EL is an effective tool for unifying the inference due to the self-normalization property.

We studied the asymptotic distributions of the EL-based test statistics under the null and local alternative hypotheses for both sparse and dense functional data.

Another important contribution of this paper is on establishing the transition phase in η, the order of repeated measurements, for pointwise and simultaneous tests. The transition point η_0 was shown to be $\frac{1}{8}$ for the pointwise test and $\frac{1}{16}$ for the simultaneous test. If $\eta \leqslant \eta_0$, we showed that the method proposed can detect alternatives of order $b_n^* = n^{-4(1+\eta)/9}$ for the pointwise test and of order $b_n^* = n^{-8(1+\eta)/17}$ for the simultaneous test. For dense functional data such that $\eta > \eta_0$, we found that the tests proposed can detect alternatives of magnitude $n^{-1/2}$ both pointwise and simultaneously, which is the same order of alternative that a parametric test can detect. The transition points that we established for the hypothesis testing problems are different from those in estimation (Li and Hsing, 2010) and pointwise confidence interval (Kim and Zhao, 2013).

Moreover, we proposed a practical bandwidth selection method for functional data. Many bandwidth selection methods were proposed for independent or weakly dependent data, but bandwidth selection for functional data is still a challenging problem; see Zhang *et al.* (2013) for a recent study. Numerical experiments in this paper showed that the bandwidth selection method proposed worked well in practice.

Acknowledgements

We are grateful to the Joint Editor, Associate Editor and three referees for the constructive comments that significantly improved the paper. Zhong's research was partially supported by National Science Foundation grant DMS-1462156. Li's research was partially supported by National Science Foundation grant DMS-1317118.

References

Ashby, F. G. (2011) *Statistical Analysis of fMRI Data*. Cambridge: MIT Press.

Balakrishnan, A. (1960) Estimation and detection theory for multiple stochastic processes. *J. Math. Anal. Appl.*, **1**, 386–410.

Banerjee, S., Carlin, B. P. and Gelfand, A. E. (2014) *Hierarchical Modeling and Analysis for Spatial Data*. Boca Raton: CRC Press.

Benko, M., Härdle, W. and Kneip, A. (2009) Common functional principal components. *Ann. Statist.*, **37**, 1–34.

Bivand, R. (2017) spdep—spatial dependence: weighting schemes, statistics and models. *R Package*. (Available from `http://r-forge.r-project.org/projects/spdep/`.)

Castro, P., Lawton, W. and Sylvestre, E. (1986) Principal modes of variation for processes with continuous sample curves. *Technometrics*, **28**, 329–337.

Chen, S. X. and Cui, H. (2006) On Bartlett correction of empirical likelihood in the presence of nuisance parameters. *Biometrika*, **93**, 215–220.

Chen, S. X., Härdle, W. and Li, M. (2003) An empirical likelihood goodness-of-fit test for time series. *J. R. Statist. Soc.* B, **65**, 663–678.

Chen, S. X. and Van Keilegom, I. (2009) A review on empirical likelihood methods for regression. *Test*, **18**, 415–447.

Chen, S. X. and Zhong, P.-S. (2010) ANOVA for longitudinal data with missing values. *Ann. Statist.*, **38**, 3630–3659.

Critchley, F., Marriott, P. and Salmon, M. (1996) On the differential geometry of the Wald test with nonlinear restrictions. *Econometrica*, **64**, 1213–1222.

DiCiccio, T., Hall, P. and Romano, J. (1991) Empirical likelihood is Bartlett-correctable. *Ann. Statist.*, **19**, 1053–1061.

Dufour, J.-M. (1989) Nonlinear hypotheses, inequality restrictions, and non-nested hypotheses: exact simultaneous tests in linear regressions. *Econometrica*, **57**, 335–355.

Eggermont, P., Eubank, R. and LaRiccia, V. (2010) Convergence rates for smoothing spline estimators in varying coefficient models. *J. Statist. Planng Inf.*, **140**, 369–381.

Eubank, R. and Hsing, T. (2008) Canonical correlation for stochastic processes. *Stoch. Processes Appl.*, **118**, 1634–1661.

Fan, J. and Gijbels, I. (1996) *Local Polynomial Modelling and Its Applications*. Boca Raton: CRC Press.
Fan, J., Huang, T. and Li, R. (2007) Analysis of longitudinal data with semiparametric estimation of covariance function. *J. Am. Statist. Ass.*, **102**, 632–641.
Fan, J. and Zhang, J.-T. (2000) Two-step estimation of functional linear models with applications to longitudinal data. *J. R. Statist. Soc.* B, **62**, 303–322.
Faraway, J. J. (1997) Regression analysis for a functional response. *Technometrics*, **39**, 254–261.
González-Manteiga, W. and Crujeiras, R. M. (2013) An updated review of goodness-of-fit tests for regression models. *Test*, **22**, 361–411.
Hall, P., Müller, H.-G. and Wang, J.-L. (2006) Properties of principal component methods for functional and longitudinal data analysis. *Ann. Statist.*, **34**, 1493–1517.
Hall, P. and Van Keilegom, I. (2007) Two-sample tests in functional data analysis starting from discrete data. *Statist. Sin.*, **17**, 1511–1531.
Härdle, W. and Mammen, E. (1993) Comparing nonparametric versus parametric regression fits. *Ann. Statist.*, **21**, 1926–1947.
Johnson, S. G. (2010) The NLopt nonlinear-optimization package. Massachusetts Institute of Technology, Cambridge. (Available from `http://lab-initio.mit.edu/nlopt`.)
Kim, S. and Zhao, Z. (2013) Unified inference for sparse and dense longitudinal models. *Biometrika*, **100**, 203–212.
Li, Y. and Hsing, T. (2010) Uniform convergence rates for nonparametric regression and principal component analysis in functional/longitudinal data. *Ann. Statist.*, **38**, 3321–3351.
Mammen, E. (1993) Bootstrap and wild bootstrap for high dimensional linear models. *Ann. Statist.*, **21**, 255–285.
Owen, A. (1990) Empirical likelihood ratio confidence regions. *Ann. Statist.*, **18**, 90–120.
Owen, A. B. (1988) Empirical likelihood ratio confidence intervals for a single functional. *Biometrika*, **75**, 237–249.
Owen, A. B. (2001) *Empirical Likelihood*. Boca Raton: CRC Press.
Peña, V. H., Lai, T. L. and Shao, Q.-M. (2008) *Self-normalized Processes: Limit Theory and Statistical Applications*. Heidelberg: Springer Science and Business Media.
Phillips, P. C. and Park, J. Y. (1988) On the formulation of Wald tests of nonlinear restrictions. *Econometrica*, **56**, 1065–1083.
Qin, J. and Lawless, J. (1995) Estimating equations, empirical likelihood and constraints on parameters. *Can. J. Statist.*, **23**, 145–159.
Ramsay, J. O. and Silverman, B. W. (2005) *Functional Data Analysis*. New York: Springer.
Rice, J. A. and Silverman, B. W. (1991) Estimating the mean and covariance structure nonparametrically when the data are curves. *J. R. Statist. Soc.* B, **53**, 233–243.
Shao, X. (2010) A self-normalized approach to confidence interval construction in time series. *J. R. Statist. Soc.* B, **72**, 343–366.
Shao, Q.-M. and Wang, Q. (2013) Self-normalized limit theorems: a survey. *Probab. Surv.*, **10**, 69–93.
Tang, C. Y. and Leng, C. (2011) Empirical likelihood and quantile regression in longitudinal data analysis. *Biometrika*, **98**, 1001–1006.
Wall, M. M. (2004) A close look at the spatial structure implied by the car and sar models. *J. Statist. Planng Inf.*, **121**, 311–324.
Weiner, M. W., Veitch, D. P., Aisen, P. S., Beckett, L. A., Cairns, N. J., Green, R. C., Harvey, D., Jack, C. R., Jagust, W., Liu, E., Morris, J. C., Petersen, R. C., Saykin, A. J., Schmidt, M. E., Shaw, L., Siuciak, J. A., Soares, H., Toga, A. W., Trojanowski, J. Q. and Alzheimer's Disease Neuroimaging Initiative (2012) The Alzheimer's Disease Neuroimaging Initiative: a review of papers published since its inception. *Alz. Demen.*, **8**, suppl. 1, S1–68.
Xue, L. and Zhu, L. (2007) Empirical likelihood for a varying coefficient model with longitudinal data. *J. Am. Statist. Ass.*, **102**, 642–654.
Yao, F., Müller, H.-G. and Wang, J.-L. (2005a) Functional data analysis for sparse longitudinal data. *J. Am. Statist. Ass.*, **100**, 577–590.
Yao, F., Müller, H.-G. and Wang, J.-L. (2005b) Functional linear regression analysis for longitudinal data. *Ann. Statist.*, **33**, 2873–2903.
Zhang, J.-T. (2011) Statistical inferences for linear models with functional responses. *Statist. Sin.*, **21**, 1431–1451.
Zhang, J.-T. and Chen, J. (2007) Statistical inferences for functional data. *Ann. Statist.*, **35**, 1052–1079.
Zhang, J., Clayton, M. and Townsend, P. (2010) Functional concurrent linear regression model for spatial images. *J. Agric. Biol. Environ. Statist.*, **16**, 105–130.
Zhang, W. and Lee, S.-Y. (2000) Variable bandwidth selection in varying-coefficient models. *J. Multiv. Anal.*, **74**, 116–134.
Zhang, X., Park, B. U. and Wang, J.-L. (2013) Time-varying additive models for longitudinal data. *J. Am. Statist. Ass.*, **108**, 983–998.

Supporting information

Additional 'supporting information' may be found in the on-line version of this article:

'Supplemental material for "Unified empirical likelihood ratio tests for functional concurrent linear models and the phase transition from sparse to dense functional data"'.