Journal of the
Royal Statistical Society

J. R. Statist. Soc. B (2008)
70, *Part* 4, *pp.* 703–723

Modelling sparse generalized longitudinal observations with latent Gaussian processes

Peter Hall,

University of California, Davis, USA, and University of Melbourne, Australia

Hans-Georg Müller

University of California, Davis, USA

and Fang Yao

University of Toronto, Canada

[Received April 2006. Final revision December 2007]

Summary. In longitudinal data analysis one frequently encounters non-Gaussian data that are repeatedly collected for a sample of individuals over time. The repeated observations could be binomial, Poisson or of another discrete type or could be continuous. The timings of the repeated measurements are often sparse and irregular. We introduce a latent Gaussian process model for such data, establishing a connection to functional data analysis. The functional methods proposed are non-parametric and computationally straightforward as they do not involve a likelihood. We develop functional principal components analysis for this situation and demonstrate the prediction of individual trajectories from sparse observations. This method can handle missing data and leads to predictions of the functional principal component scores which serve as random effects in this model. These scores can then be used for further statistical analysis, such as inference, regression, discriminant analysis or clustering. We illustrate these non-parametric methods with longitudinal data on primary biliary cirrhosis and show in simulations that they are competitive in comparisons with generalized estimating equations and generalized linear mixed models.

Keywords: Binomial data; Eigenfunction; Functional data analysis; Functional principal component; Prediction; Random effect; Repeated measurements; Smoothing; Stochastic process

1. Introduction

1.1. Preliminaries

When nde aking p edic ion in longi dinal da a anal i in ol ing i eg la l paced and inf e en mea emen , ela i el li le info ma ion i of en a ailable abo each bjec , o ing o pa e and i eg la mea emen . I eg la i of mea emen fo indi id al bjec i an inhe en diffic l of ch die. The efo e i i e peciall impo an o e all he info ma ion ha can be acce ed. Thi e i e o model he ela ion hip be een mea emen ha a e made a idel epa a ed ime poin . We aim a a fle ible non-pa a- me ic f nc ional da a anal i app oach, hich i in con a i h commonl ed pa ame ic model ch a gene ali ed linea mi ed model (GLMM) o gene ali ed e ima ion e a ion

Address for correspondence: Han -Geo g Mnlle , Depa men of S a i ic , Uni e i of Califo nia a Da i , One Shield A en e, Da i , CA 95616, USA.
E-mail: mueller@wald.ucdavis.edu

 1369–7412/08/70703

(GEEs) (see, for example, Heagerty (1999) for recent discussions on applying such models to repeated binary measurements, Pourahmadi (2000) for related aspects of covariance modelling and Heagerty and Zeger (2000), Heagerty and Kurland (2001) and Chiou and Müller (2005) for discussions on limitations, modifications and feasibility of the underlying parametric assumptions).

A non-parametric functional approach for the analysis of longitudinal data, with its philosophy to let the data speak for themselves and its inherent flexibility, is expected to perform better than the parametric GEE or GLMM approaches in many situations. However, it faces difficulties due to the potentially large gaps between repeated measurements in typically sparse longitudinal data. The parametric methods overcome this easily by postulating a parametric form of the underlying functions. In contrast, in the presence of such gaps, the classical non-parametric approach to smooth individual trajectories in a first step is not feasible (Yao *et al.*, 2005). The problems that are caused by gaps are exacerbated in the commonly encountered case of non-Gaussian longitudinal responses such as binomial or Poisson responses (see Section 5).

We demonstrate how one can overcome the difficulties that are posed by such data for non-parametric approaches, by applying suitably modified methods of functional data analysis. Functional data analysis methods have been primarily developed for smooth and densely sampled data (Ramsay and Silverman, 2002, 2005). The basic idea to connect the data that we wish to analyse to functional data analysis methodology is to postulate an underlying latent Gaussian process (LGP) (for other examples of latent process modelling for longitudinal studies compare, for example, Diggle *et al.* (1998), Jowaheer and Sutradhar (2002), Hashemi *et al.* (2003) and Proust *et al.* (2006)). Specifically, the Gaussian property makes it possible to overcome sparseness by a conditioning argument. Relevant features of the stochastic relationships of the observed data are reflected by the mean and covariance properties of this LGP. Simulations indicate that the method is in practice quite insensitive to the Gaussian assumption for the latent process.

Since sufficiently flexible parameterizations of the underlying Gaussian processes would suffer from a large number of parameters, making corresponding maximum likelihood approaches computationally demanding and unstable, we propose instead to connect the LGP to random trajectories for individual observations directly by means of a link function. These subject-specific trajectories correspond to the probabilities of a response in the binary response case. Whereas the link function is assumed known, the mean and covariance of the Gaussian process are assumed to be unknown but smooth. This proposition is attractive on grounds of flexibility, but it raises the challenging problem of constructing appropriate estimators.

The methodology proposed is a first attempt to extend functional data analysis technology to the case of non-Gaussian repeated measurements. Prominent examples for such data are repeated binary measurements or repeated counts. The methods proposed are motivated by several considerations: the variation of random coefficients may be relatively low, and in this case a simple Taylor approximation motivates simple, explicit and non-parametric mean and covariance function estimators; and these estimators are elementary to compute, irrespective of whether the low variation assumption is satisfied or not. The simple, low variation estimators that we propose are attractive owing to their flexibility and numerical simplicity.

The analysis of continuous Gaussian sparse longitudinal data by functional methods has been considered previously (e.g. Shi *et al.* (1996), Rice and Wu (2000), James *et al.* (2001) and James and Sugar (2003)). Our main tool from functional data analysis is functional principal component (FPC) analysis, where observed trajectories are decomposed into a mean function and eigenfunctions (e.g. Rice and Silverman (1991) and Boente and Fraiman (2000)). Various aspects of the relationship between functional and longitudinal data are discussed in Staniswalis and Lee (1998), Rice (2004) and Zhao *et al.* (2004); an early study of modelling longitudinal

ajec o ie in biological applica ion i h FPC i Ki kpa ick and Heckman (1989). FPC anal i allo o achie e h ee majo goal :

(a) dimen ion ed c ion of f nc ional da a b mma i ing he da a in a fe FPC ;
(b) he p edic ion of indi id al ajec o ie f om pa e da a, b ima ing he FPC co e of he ajec o ie ;
(c) f he a i ical anal i of longi dinal da a ba ed on he FPC co e .

In he ne b ec ion, e in od ce he LGP model; hen in Sec ion 2 he p opo ed e i ma e , follo ed b applica ion o p edic ion (Sec ion 3). The e l f om a im la ion d , incl ding a compa i on of he me hod p opo ed i h GLMM and GEE , a e epo ed in Sec ion 4. The anal i of non-Ga ian pa e longi dinal da a i ill a ed in Sec ion 5, i h he longi dinal anal i of he occ ence of hepa omegal in p ima bilia ci ho i . Thi i follo ed b a bief di c ion (Sec ion 6) and an appendi , hich con ain de i a ion and ome heo e ical e l abo e ima ion.

1.2. Latent Gaussian process model

Gene all , deno ing he gene ali ed e pon e b Y_{ij}, e ob e e independen copie of Y, b , in each ca e, onl fo a fe pa e ime poin . In pa ic la , he da a a e pai (T_{ij}, Y_{ij}), fo $1 \leqslant i \leqslant n$ and $1 \leqslant j \leqslant m_i$, he e $Y_{ij} = Y_i(T_{ij})$ fo an nde l ing andom ajec o Y_i, and each $T_{ij} \in \mathcal{I} = [0, 1]$. The pa e and ca e ed na e of he ob e a ion ime T_{ij} ma be e p e ed heo e icall b no ing ha he m_i a e nifo ml bo nded, if he e an i ie ha e a de e min ic o igin, o ha he ep e en he al e of independen and iden icall di ib ed andom a iable i h fficien l ligh ail , if he m_i o igina e ocha ic all . We a e aiming a he eemingl diffic l a k of making ch pa e de ign amenable o f nc ional me hod , hich ha e been p ima il aimed a den el collec ed moo h da a.

A cen al a mp ion fo o app oach i ha he dependence be een he ob e a ion Y_{ij} i inhe i ed f om an nde l ing nob e ed Ga ian p oce X: le $Y(t)$, fo $t \in \mathcal{T}$, he e $\mathcal{T}$ i a compac in e al, deno e a ocha ic p oce a i f ing

$$E\{Y(t_1) \dots Y(t_m) | X\} = \prod_{j=1}^{m} g\{X(t_j)\},$$
$$E\{Y(t)^2 | X\} \leqslant g_1\{X(t)\} \tag{1}$$

fo $0 \leqslant t_1 < \dots < t_m \leqslant 1$ and $0 < t < 1$. He e, X deno e a Ga ian p oce on $\mathcal{I}$, g i a moo h, mono one inc ea ing link f nc ion, f om he eal line o he ange of he di ib ion of he Y_{ij}, and g_1 i a bo nded f nc ion. Al ho gh e ob e e independen copie of Y, he e a e acce ible onl fo a fe pa e ime poin fo each bjec . The Ga ian p oce e X_i and mea emen ime T_{ij}, fo $1 \leqslant i \leqslant n$ and $1 \leqslant j \leqslant m_i$, a e a med o be o all independen , he T_{ij} a e aken o be iden icall di ib ed a T, a , i h ppo $\mathcal{I}$ and he X_i a e ppo ed o be iden icall di ib ed a X. When in e p e ed fo he da a (T_{ij}, Y_{ij}), model (1) implie ha

$$E\{Y_i(T_{i1}) \dots Y_i(T_{im_i}) | X_i(T_{i1}), \dots, X_i(T_{im_i})\} = \prod_{j=1}^{m_i} g\{X_i(T_{ij})\}. \tag{2}$$

The a mp ion ha X a model (1) i Ga ian p o ide a pla ible a of linking ocha ic p ope ie of $Y(t)$ fo al e t in diffe en pa of $\mathcal{I}$, o ha da a ha a e ob e ed a each ime poin can be ed fo infe ence abo f e al e of $Y(t)$ fo an pecific al e of t. The idea of pooling da a ac o bjec o e come he pa ene p oblem i mo i a ed a in Yao

et al. (2005). The link f nc ion g i a y med kno n; fo e ample e migh elec he logi link in he bina da a ca e, $g(x) = e^{x}/\{1+e^{x}\}$, and he log-link fo co n da a; nde ome ci c m ance , he link can al o be e ima ed non-pa ame ic all . An impo an pecial ca e of model (1) i ha of bina e pon e , i.e. 0 1 da a, he e he fi iden i in model (1), implifie o

$$P\{Y(t_1) = l_1, \ldots, Y(t_m) = l_m \mid X\} = \prod_{j=1}^{m} g\{X(t_j)\}^{l_j}[1 - g\{X(t_j)\}]^{1-l_j}, \tag{3}$$

fo all e ence $l_1, \ldots, l_m$ of 0 and 1 . In hi ca e, he link f nc ion g o ld be cho en a a di ib ion f nc ion and he me hodolog p opo ed co e pond o an e en ion of f nc ional da a anal i o longi dinal bina da a.

2. Estimating mean and covariance of latent Gaussian processes

To e model (1) o make p edic i e infe ence abo f e al e of $Y(t)$, e need o e ima e he defining cha ac e i ic of he p oce X, i.e. i mean and co a iance c e. In a e ing he e he di ib ion of Y can be comple el pecified, e.g. in he bina da a model (3), one po ible app oach o ld be ma im m likelihood. Thi i , ho e e , a diffic l p opo i ion in he i eg la ca e, he e i o ld nece i a e he pecifica ion of a la ge n mbe of pa ame e fo he a io mean and co a iance ha a e in ol ed, a diffic l hich can onl be o e come b in oking e ic i e a mp ion , limi ing he fle ibili of he app oach. Mo eo e , e a e con ide ing a non- a iona ca e, and he n mbe of pa ame e o ld need o inc ea e i h n, he ample i e. Finall , ano he majo mo i a ion i o e end he f nc ional app oach o non-Ga ian longi dinal da a. To ain he non-pa ame ic fla o , e p efe no o make mp ion han model (1), and in pa ic la e do no i h o make he ic i e a mp ion ha o ld be nece a o emplo ma im m likelihood me hod .

O app oach i ba ed on he ppo i ion ha he a ia ion of X_i abo i mean i ela i el mall. In pa ic la , e a me ha

$$X_i(t) = \mu(t) + \delta Z_i(t), \qquad \mu = E(X_i), \tag{4}$$

Z_i i a Ga ian p oce i h e o mean and bo nded co a iance and $\delta > 0$ i an nkno n mall con an . In hi ca e, a ming ha g ha fo bo nded de i a i e , and i ing (X, Z) fo a gene ic pai (X_i, Z_i), e ha e

$$g(X) = g(\mu) + \delta Z\, g^{(1)}(\mu) + \tfrac{1}{2}\delta^2 Z^2 g^{(2)}(\mu) + \tfrac{1}{6}\delta^3 Z^3 g^{(3)}(\mu) + O_p(\delta^4), \tag{5}$$

$$E[g\{X(t)\}] = g(\mu) + \tfrac{1}{2}\delta^2 E\{Z^2(t)\}\, g^{(2)}\{\mu(t)\} + O(\delta^4) \tag{6}$$

and

$$\text{co }[g\{X(s)\}, g\{X(t)\}] = \delta^2\, g^{(1)}\{\mu(s)\}\, g^{(1)}\{\mu(t)\}\, \text{co }\{Z(s), Z(t)\} + O(\delta^4). \tag{7}$$

He e and h o gho e make he a mp ion ha $g^{(1)}$ doe no ani h, and ha $\inf_{s \in D}\{g^{(1)}(s)\} > 0$, he e D i he (compac) ange of he mean f nc ion μ. Se ing

$$\left.\begin{aligned} \alpha(t) &= E[g\{X(t)\}], \\ \nu(t) &= g^{-1}\{\alpha(t)\}, \\ \tau(s,t) &= \text{co }[g\{X(s)\}, g\{X(t)\}]/g^{(1)}\{\mu(s)\}\, g^{(1)}\{\mu(t)\}, \end{aligned}\right\} \tag{8}$$

e ob ain

$$\mu(t) = E\{X(t)\} = g^{-1}(E[g\{X(t)\}]) + O(\delta^2) = \nu(t) + O(\delta^2), \tag{9}$$

$$\sigma(s,t) = \mathrm{co}\ \{X(s), X(t)\} = \frac{\mathrm{co}\ [g\{X(s)\}, g\{X(t)\}]}{g^{(1)}\{\mu(s)\}\ g^{(1)}\{\mu(t)\}} + O(\delta^4) = \tau(s,t) + O(\delta^4). \tag{10}$$

The e fo m lae immedia el gge e ima o of μ and σ, if e a e illing o neglec he effec of o de $O(\delta^2)$. Indeed, e ma e ima e

$$\alpha(t) = E\{Y(t)\} = E[E\{Y(t)|X(t)\}] = E[g\{X(t)\}], \tag{11}$$

b pa ing a moo he h o gh he da a (T_{ij}, Y_{ij}), and e ima e

$$\beta(s,t) = E\{Y(s)Y(t)\} = E[g\{X(s)\}g\{X(t)\}] \tag{12}$$

(b ing model (1)) b pa ing a bi a ia e moo he h o gh he da a $((T_{ij}, T_{ik}), Y_{ij}Y_{ik})$ fo $1 \leqslant i \leqslant n$ ch ha $m_i \geqslant 2$, and $1 \leqslant j, k \leqslant m_i$ i h $j \neq k$. I i nece a o omi he diagonal e m in hi moo hing ep, ince acco ding o model (1) e ha e

$$E\{Y^2(t)\} = E[E\{Y^2(t)|X(t)\}] > E[E\{Y(t)|X(t)\}]^2 = E[g\{X(t)\}]^2,$$

hene e a $\{Y(t)|X(t)\} > 0$, o he a iance along he diagonal in gene al ill ha e an e a componen , leading o a co a iance face ha ha a di con in i along he diagonal. Mo e de ail abo hi phenomenon can be fo nd in Yao *et al.* (2005). Implemen a ion of he e moo hing ep, b ing local lea a e e ima o, i di c ed in Appendi A.

F om he e l ing e ima o α and β of α and β pec i el , e ob ain e ima o

$$\begin{aligned} \nu(t) &= g^{-1}\{\alpha(t)\}, \\ \tau(s,t) &= \{\beta(s,t) - \alpha(s)\,\alpha(t)\}/g^{(1)}\{\nu(s)\}\ g^{(1)}\{\nu(t)\} \end{aligned} \tag{13}$$

fo

$$\begin{aligned} \nu(t) &= g^{-1}\{\alpha(t)\}, \\ \tau(s,t) &= \{\beta(s,t) - \alpha(s)\,\alpha(t)\}/g^{(1)}\{\nu(s)\}\ g^{(1)}\{\nu(t)\} \end{aligned} \tag{14}$$

pec i el . B i e of app o ima ion (9) and (10) e ma in e p e ν and τ a e ima o of μ and σ pec i el , i.e. e e

$$\begin{aligned} \mu(t) &= \nu(t), \\ \sigma(s,t) &= \tau(s,t). \end{aligned} \tag{15}$$

The e ima o do no depend on he con an δ, hich he efo e doe no need o be kno n o e ima ed. Al ho gh he ima o $\tau(s,t)$ i mme ic, i ill gene all no enjo he po i i e emidefini ene p ope ha i e i ed of a co a iance f nc ion. Thi deficienc can be o e come b implemen ing a me hod ha a de c ibed in Yao *et al.* (2003), hich i o d op f om he pec al decompo i ion of τ ho e e m ha co pond o nega i e eigen al e. I i ea o ho ha , in doing o, he mean a ed e o of τ i ic l imp o ed b omi ing a e m ha co pond o a nega i e eigen al e; de ail can be fo nd in Appendi B. In ha follo , e o k i h he l ing e ima o $\tilde{\tau}$ a defined in Appendi B. P ope ie of he e ima o α and β, and ν and τ, hich a e defined a e p e ion (32), (33) and (13) pec i el , and of e ima o μ and σ a e p e ion (15) a e di c ed in Appendi C.

3. Predicting individual trajectories and random effects

3.1. Predicting functional principal component scores

One of he main p , po e of he f nc ional da a anal , i model p opo ed i dimen ion , ed c- ion h o gh p edic ed FPC co e . The e lead o p edic ed , ajec o ie of he nde l ing hidden Ga , ian p oce fo he bjec in a d . Specificall , he p edic ed FPC co e p o ide a mean fo , eg la i ing he i , eg la da a, and al o fo dimen ion , ed c ion, and can be ed fo infe ence, di c iminan anal , i o , eg e ion.

The a ing poin i he Ka h nen Lo e e pan ion of , andom , ajec o ie X_i of he LGP,

$$X_i(t) = \mu(t) + \sum_{j=1}^{\infty} \xi_{ij}\, \psi_j(t), \tag{16}$$

he e ψ_j a e he o hono mal eigenf nc ion of he linea in eg al ope a o B i h ke nel $\sigma(s,t)$, ha map an L^2-f nc ion f o $Bf(s) = \int \sigma(s,t)\, f(t)\, dt$, i.e. he ol ion of

$$\int \mathrm{co}\ \{X(s), X(t)\}\, \psi_j(t)\, ds = \theta_j\, \psi_j(t),$$

he e θ_j i he eigen al e ha i a ocia ed i h eigenf nc ion ψ_j. The $\xi_{ij} = \int \{X_i(t) - \mu(t)\}\, \psi_j(t)\, dt$ a e he FPC co e ha pla he ole of andom effec , i h $E(\xi_{ij}) = 0$ and a ($\xi_{ij}) = \theta_j$, he e θ_j i he eigen al e co , e ponding o eigenf nc ion ψ_j. Once he e ima o $\sigma(s,t)$ (15) ha been de e mined, he co , e ponding e ima e θ_j and ψ_j of eigen al e and eigenf nc ion of la en p oce e X a e ob ained b a anda d di c e i a ion p oced , e, he eb he e ima e a e de i ed f om a di c e e p incipal componen anal , i ep.

We aim o e ima e he be linea p edic o

$$E\{X_i(t) | Y_{i1}, \ldots, Y_{im}\} = \sum_{j=1}^{\infty} E(\xi_{ij} | Y_{i1}, \ldots, Y_{im})\, \psi_j(t) \tag{17}$$

of he , ajec o X_i, gi en he da a $Y_{i1}, \ldots, Y_{im_i}$. He e a , nca ion of he e pan ion o incl de onl he fi M componen i needed. Then, foc ing on he fi M condi ional FPC co e ill allo o ed ce he dimen ion of he p oblem and al o o , eg la i e he highl i , eg la da a. Acco ding o e a ion (17), he a k of , ep e en ing and p edic ing indi id al , ajec o ie can be , ed ced o ha of e ima ing $E(\xi_{ij} | Y_{i1}, \ldots, Y_{im})$. In ha follo e de elop a i able app o ima ion in he non-Ga ian ca e b mean of a momen -ba ed app oach, a follo . The , epea ed mea , emen pe bjec a e a med o be gene a ed b

$$Y_{ik} = Y_i(T_{ik}) = g\{X_i(T_{ik})\} + e_{ik}, \tag{18}$$

i h independen e , o e_{ik}, a i f ing

$$\begin{aligned} E(e_{ik}) &= 0, \\ \mathrm{a} (e_{ik}) &= \gamma^2\, v[g\{X_i(T_{ik})\}]. \end{aligned} \tag{19}$$

He e, γ^2 i an nkno n a iance (o e di pe ion) pa ame e and $v(\cdot)$ i a kno n moo h a i- ance f nc ion, hich i de e mined b he cha ac e i ic of he da a. Fo e ample, in he ca e of epea ed bina ob e a ion , one o ld choo e $v(u) = u(1-u)$. In ha follo , e implici l condi ion on he mea , emen ime T_{ij}.

Wi h a Ta lo , e ie e pan ion of g, ing e p e ion (4) and a ming a befo e ha $\inf\{g^{(1)}(\cdot)\} > 0$, e ob ain

$$g\{X(t)\} = g\{\mu(t)\} + g^{(1)}\{\mu(t)\}\{X(t) - \mu(t)\} + O(\delta^2). \tag{20}$$

Defining

$$\varepsilon_{ik} = \frac{e_{ik}}{g^{(1)}\{\mu(T_{ik})\}},$$

$$U_{ik} = \mu(T_{ik}) + \frac{Y_{ik} - g\{\mu(T_{ik})\}}{g^{(1)}\{\mu(T_{ik})\}},$$

e p e ion (19) and (20) lead o $U_{ik} = X_i(T_{ik}) + \varepsilon_{ik} + O(\delta^2)$. We ne b i e e ima e (15) and e o ε_{ik} b

$$\tilde{e}_{ik} = Z_{ik}\gamma \frac{v[g\{\mu(T_{ik})\}]^{1/2}}{g^{(1)}\{\mu(T_{ik})\}},$$

he e he Z_{ik} a e independen copie of a anda d Ga ian $N(0,1)$ andom a iable, o ha he fi o momen of $\tilde{e}_{ik}$ a e app o ima ing ho e of ε_{ik}. Then, fo mall δ, $U_{ik} \approx X_i(T_{ik}) + \tilde{e}_{ik}$, impl ing ha

$$E(\xi_{ij}|Y_{i1},\ldots,Y_{im_i}) = E(\xi_{ij}|U_{i1},\ldots,U_{im_i}) \approx E\{\xi_{ij}|X_i(T_{i1}) + \tilde{e}_{i1},\ldots,X_i(T_{im_i}) + \tilde{e}_{im_i}\}.$$

O ing o he Ga ian a mp ion fo la en p oce e X_i, he la condi ional e pec a ion i een o be a linea f nc ion of he e m on he igh -hand ide, and he efo e

$$E(\xi_{ij}|Y_{i1},\ldots,Y_{im_i}) = A_{ij}\tilde{X}_i \tag{21}$$

i a ea onable p edic o fo he andom effec ξ_{ij}, he e $\tilde{X}_i = (X_i(T_{i1}) + \tilde{e}_{i1},\ldots,X_i(T_{im_i}) + \tilde{e}_{im_i})^{\mathrm{T}}$ and he A_{ij} a e ma ice depending onl on γ, μ, v, g and $g^{(1)}$. The e an i ie a e ei he kno n o e ima e a e a ailable, i h he ole e cep ion of γ, he e ima ion of hich i di c ed belo . The e plici fo m of e a ion (21) i gi en in Appendi D.

3.2. Predicting trajectories

Mo i a ed b e a ion (16) and (21), p edic ed ajec o ie fo he LGP a e ob ained a

$$X_i(t) = E\{X_i(t)|Y_{i1},\ldots,Y_{im_i}\} = \mu(t) + \sum_{j=1}^{M} A_{ij}\tilde{X}_i\,\psi_j(t), \tag{22}$$

and p edic ed ajec o ie fo he ob e ed p oce Y a

$$Y_i(t) = E\{Y_i(t)|Y_{i1},\ldots,Y_{im_i}\} = g\{X_i(t)\}, \tag{23}$$

he e t ma be an ime poin i hin he ange of p oce e Y, incl ding ime fo hich no e pon e a ob e ed. P edic ed al e fo $Y(t)$ can ome ime be ed o p edic he en i e e pon e di ib ion hen he mean de e mine he en i e di ib ion, ch a in binomial and Poi on ca e. Thi me hod co ld al o be emplo ed fo he p edic ion of mi ing al e in a i a ion he e mi ing da a occ o all a andom.

To e al a e he effec of a ilia an i ie on he p edic ion, e e a c o - alida ion c i- e ion he e e compa e p edic ion of Y_{ik}, hich a e ob ained b lea ing ha ob e a ion o , i h Y_{ik} i elf. Comp ing

$$Y_{ik}^{(-ik)} = E(Y_{ik}|Y_{i1},\ldots,Y_{i,k-1},Y_{i,k+1},\ldots,Y_{im_i}) = g\{X_i^{(-ik)}(T_{ik})\}, \qquad 1 \leqslant i \leqslant n, \quad 1 \leqslant k \leqslant m_i, \tag{24}$$

he e

$$X_i^{(-ik)}(T_{ik}) = \mu(t) + \sum_{j=1}^{M} E(\xi_{ij}|Y_{i1}, \ldots, Y_{i,k-1}, Y_{i,k+1}, \ldots, Y_{im_i})\, \psi_j(t), \tag{25}$$

e define he Pea, on- pe eigh ed p, edic ion e, o,

$$\mathrm{PE}(\gamma^2) = \sum_{i,k} \frac{(Y_{ik}^{(-ik)} - Y_{ik})^2}{v[g\{X_i^{(-ik)}(T_{ik})\}]}, \tag{26}$$

hich ill depend on he a, iance pa, ame e, γ^2 and implici l al o on he n mbe, of eigen- f nc ion M ha a, e incl ded in he model; ee e a ion (19).

We fo nd ha he follo ing i e, a i e elec ion p, oced , e, fo choo ing he n mbe, of eigen- f nc ion M and he o e, di pe, ion pa, ame e, γ^2 im l aneo l , led o good p, ac ical, e l: choo e a, a, ing al e fo M; hen ob ain γ^2 b minimi ing he c, o- alida ed p, edic ion e, o, PE i h, e pec o γ^2,

$$\gamma = \mathrm{a, g}\min_{\gamma}\{\mathrm{PE}(\gamma^2)\}. \tag{27}$$

Then, in a b e en ep, pda e M b he c, i e, ion ha i de c, ibed belo , and, epea he e o ep n il he al e of M and γ^2 abili e. Thi i e, a i e algo, i hm o, ked e, ell in p, ac ice; pical, a, ing al e fo M o ld be 2 o, 3.

Specificall , fo, he choice of M, e adop a a, i-likelihood-ba, ed f nc ional info, ma ion c, i e, ion FIC ha i an e en ion of he Akaike info, ma ion c, i e, ion AIC fo, f nc ional da a (, ee Yao *et al.* (2005) fo, a, ela ed p, e do-Ga , ian likelihood-ba, ed c, i e, ion). The n mbe, of eigenf nc ion M, o be incl ded in he model, i cho en in ch a a a o minimi e

$$\mathrm{FIC}(M) = -2\sum_{i,k}\int_{Y_{ik}}^{Y_{ik}} \frac{Y_{ij} - t}{\gamma^2\, v(t)}\, \mathrm{d}t + 2M. \tag{28}$$

The penal $2M$ co, e, pond o ha , ed in AIC; o he, penal ie , ch a ho e co, e, ponding o he Ba e info, ma ion c, i e, ion BIC co ld be , ed a ell.

Some imple algo, i hmic, e, ic ion can be impo ed in hi i e, a ion fo, he choice of M and γ; o ha loop canno happen, al ho gh e ne e, ob e, ed hi o occ , . We al o in e, iga ed di, ec minimi a ion of e a ion (26) im l aneo l fo, bo h γ and M. Be ide being con id- e, abl mo, e comp ing in en i e, hi al e, na i e minimi a ion cheme ended o choo e mo, e componen and, e l ed in le pa, imonio fi i ho ob aining be e, p, edic ion. In ead of making a pa, ame, ic a mp ion abo he a, iance f nc ion v, in ome ca e i ma be p, e- fe, able o e, ima e i non-pa, ame, icall . Thi can be done ia emipa, ame, ic a, i-likelihood , eg, e, ion (Chio and Mnlle, , 2005).

4. Simulation results

4.1. Comparisons with generalized estimating equations and generalized linear mixed models

The im la ion e, e ba ed on la en p, oce, e, $X(t)$ i h mean f nc ion $E\{X(t)\} = \mu(t) = 2\sin(\pi t/5)/\sqrt{5}$, and co $\{X(s), X(t)\} = \lambda_1\, \phi_1(s)\, \phi_1(t)$ de i ed f, om a ingle eigenf nc ion $\phi_1(t) = -\cos(\pi t/10)/\sqrt{5}$, $0 \leqslant t \leqslant 10$, i h eigen al e $\lambda_1 = 2$ ($\lambda_k = 0$, $k \geqslant 2$). Then 200 Ga , ian and 200 non-Ga , ian , ample of la en p, oce, e, con i ing of $n = 100$, andom , ajec o, ie each e e gene, a ed b $X_i(t) = \mu(t) + \xi_{i1}\, \phi_1(t)$, he, e fo he 200 Ga , ian , ample he FPC co, e, ξ_{i1} e, e , im la ed f, om $\mathcal{N}(0, 2)$, he, ea he ξ_{i1} fo he non-Ga , ian , ample e, e , im - la ed f, om a mi , e of o no, mal di , ib ion : $\mathcal{N}(\sqrt{2}, 2)$ i h p, obabili $\frac{1}{2}$ and $\mathcal{N}(-\sqrt{2}, 2)$

i h p obabili $\frac{1}{2}$. Bina o come Y_{ij} e e gene a ed a Be no lli a iable i h p obabili $E\{Y_{ij}|X_i(t_{ij})\} = g\{X_i(t_{ij})\}$, ing he canonical logi link f nc ion $g^{-1}(p) = \log\{p/(1-p)\}$ fo $0 < p < 1$.

To gene a e he pa e ob e a ion , each ajec o a ampled a a andom n mbe of poin , cho en nifo ml f om $\{8, \ldots, 12\}$, and he loca ion of he mea emen e e ni fo ml di ib ed o e he domain $[0, 10]$. Fo he moo hing ep , ni a ia e and bi a ia e p od c Epanechniko eigh f nc ion e e ed, i.e. $K_1(x) = (3/4)(1-x^2)\,\mathbf{1}_{[-1,1]}(x)$ and $K_2(x, y) = (9/16)(1-x^2)(1-y^2)\,\mathbf{1}_{[-1,1]}(x)\,\mathbf{1}_{[-1,1]}(y)$, he e $\mathbf{1}_A(x)$ e al 1 if $x \in A$ and 0 o he i e fo an e A. The n mbe of eigenf nc ion M and he o e di pe ion pa ame e γ^2 e e epa a el elec ed fo each n b he i e a ion (27) and e a ion (28). The e i e a ion con e ged fa , e i ing onl 2 4 i e a ion ep in mo ca e.

We compa e he non-pa ame ic LGP me hod p opo ed i h he pop la pa ame ic app oache p o ided b GLMM and GEE. Fo he GEE me hod, e ed he n c ed co ela ion op ion and bo h GEE and GLMM e e n i h linea (me hod GEE-L and GLMM-L) and in addi ion i h ad a ic (me hod GEE-Q and GLMM-Q) fi ed effec . We e fo c i e ia fo he compa i on , mea ing di c epancie be een e ima e and a ge bo h in e m of la en p oce e X and e pon e p oce e $Y = g(X)$, and compa ing bo h e ima e fo mean f nc ion $\mu = E(X)$ and $g(\mu)$ e pec i el and p edic ion of bjec pecific ajec o ie X_i and $g(X_i)$ e pec i el . The la e a e a ailable fo he LGP and GLMM me hod , b no fo GEE , hich aim a ma ginal modelling. The pecific c i e ia fo he compa i on a e a follo :

$$\begin{aligned} \mathrm{XMSE} &= \int_{\mathcal{I}} \{\mu(t) - \mu(t)\}^2 \,\mathrm{d}t \Big/ \int_{\mathcal{I}} \mu^2(t)\,\mathrm{d}t, \\ \mathrm{YMSE} &= \int_{\mathcal{I}} [g\{\mu(t)\} - g\{\mu(t)\}]^2 \,\mathrm{d}t \Big/ \int_{\mathcal{I}} g^2\{\mu(t)\}\,\mathrm{d}t, \end{aligned} \tag{29}$$

$$\begin{aligned} \mathrm{XPE}_i &= \int_{\mathcal{I}} \{X_i(t) - X_i(t)\}^2 \,\mathrm{d}t \Big/ \int_{\mathcal{I}} X_i^2(t)\,\mathrm{d}t, \\ \mathrm{YPE}_i &= \int_{\mathcal{I}} [g\{X_i(t)\} - g\{X_i(t)\}]^2 \,\mathrm{d}t \Big/ \int_{\mathcal{I}} g^2\{X_i(t)\}\,\mathrm{d}t, \end{aligned} \tag{30}$$

fo $i = 1, \ldots, n$. S mma a i ic fo he al e of he e c i e ia f om 200 Mon e Ca lo n a e ho n in Table 1.

The e e l indica e ha , fi of all, he LGP me hod p opo ed i no en i i e o he Ga - ian a mp ion fo la en p oce e . Al ho gh he e i ome de e io a ion in he non-Ga ian ca e, i i minimal. Thi non- en i i i o he Ga ian a mp ion ha been de c ibed befo e in f nc ional da a anal i in he con e of p incipal anal i b condi ional e pec a ion (ee Yao *et al.* (2005)). Secondl , he non-linea i in he a ge f nc ion h o he pa ame ic me hod off ack, e en hen he mo e fle ible ad a ic fi ed effec e ion a e ed. We find ha he LGP me hod con e clea ad an age in e ima ion and e peciall in p edic ing indi id al ajec o ie in ch i a ion . Whe ea he pa ame ic me hod a e en i i e o iola ion of a mp ion , he LGP me hod i de igned o o k nde minimal a mp ion and he efo e p o ide a ef l al e na i e app oach.

4.2. *Effect of the size of variation*

He e e e amine he infl ence of he i e of he a ia ion con an δ on model e ima ion, incl ding mean f nc ion, eigenf nc ion and indi id al ajec o ie . In addi ion o c i e ia (29)

Table 1. Simulation results for the comparisons of mean estimates and individual trajectory predictions obtained by the proposed non-parametric LGP method with those obtained for the established parametric methods GLMM-L, GLMM-Q, GEE-L and GEE-Q, with linear and quadratic fixed effects (see Section 4.1)

Distribution	*Method*	*XMSE*	*XPE$_i$*			*YMSE*	*YPE$_i$*		
			25th	*50th*	*75th*		*25th*	*50th*	*75th*
Ga ,,ian	LGP	0.1242	0.1529	0.2847	0.7636	0.0076	0.0101	0.0205	0.0433
	GLMM-L	0.4182	0.3405	0.5843	1.283	0.0265	0.0278	0.0369	0.0577
	GLMM-Q	0.4323	0.3479	0.5990	1.319	0.0271	0.0285	0.0377	0.0584
	GEE-L	0.4168				0.0264			
	GEE-Q	0.4308				0.0272			
Non-Ga ,,ian	LGP	0.1272	0.1664	0.3166	0.9556	0.0078	0.0109	0.0228	0.0459
(mi ¾ ,e)	GLMM-L	0.4209	0.3309	0.5943	1.364	0.0266	0.0280	0.0372	0.0589
	GLMM-Q	0.4373	0.3385	0.6118	1.404	0.0274	0.0287	0.0380	0.0597
	GEE-L	0.4227				0.0268			
	GEE-Q	0.4396				0.0277			

Sim la ion e e ba ed on 200 Mon e Ca lo , n i h $n = 100$, ajec o ie pe , ample, gene a ed fo bo h Ga ,,ian and non-Ga ,,ian la en p oce,,e. Sim la ion , e 1, a e , epo ed h o gh, mma , a i ic fo e , o c i e ia XMSE and YMSE (29) fo , ela i e, a ed e , o of he mean f nc ion e ima e of la en p oce,,e X and of , e pon e p oce,,e Y, and he 25 h, 50 h and 75 h pe cen ile of , ela i e p edic ion e , o, XPE$_i$ and YPE$_i$ (30) fo indi id al , ajec o ie of la en and , e pon e p oce,,e.

and (30), e al o e al a ed he e ima ion e , o fo he, ingle eigenf nc ion in he model (no ing ha $\int_{\mathcal{I}} \phi_1^2(t)\,dt = 1$),

$$\mathrm{EMSE} = \int_{\mathcal{I}} \{\phi_1(t) - \phi_1(t)\}^2\,dt. \tag{31}$$

U ing he ame im la ion de ign a in Sec ion 4.1 and gene a ing la en p oce,,e $X(t;\delta) = \mu(t) + \delta\xi_1\phi_1(t)$ fo a ing δ, e im la ed 200 Ga ,,ian and 200 non-Ga ,,ian ample (a de c ibed befo e) fo each of $\delta = 0.5, 0.8, 1, 2$. The Mon e Ca lo e 1 o e 200 , n fo he a io al e of δ a e p e en ed in Table 2.

Table 2. Simulation results for the effect of the variation parameter δ

Distribution	δ	*XMSE*	*EMSE*	*XPE$_i$*			*YMSE*	*YPE$_i$*		
				25th	*50th*	*75th*		*25th*	*50th*	*75th*
No mal	0.5	0.1106	0.7662	0.1188	0.1815	0.3366	0.0068	0.0077	0.0119	0.0205
	0.8	0.1205	0.3801	0.1430	0.2437	0.5710	0.0076	0.0094	0.0171	0.0338
	1	0.1280	0.2434	0.1513	0.2809	0.7857	0.0077	0.0101	0.0203	0.0431
	2	0.1616	0.0429	0.2025	0.3851	0.8137	0.0102	0.0144	0.0362	0.0752
Mi ¾ ,e	0.5	0.1134	0.7198	0.1243	0.1913	0.3651	0.0071	0.0081	0.0126	0.0217
	0.8	0.1258	0.3910	0.1498	0.2563	0.6691	0.0078	0.0100	0.0188	0.0366
	1	0.1323	0.2256	0.1624	0.2986	0.7944	0.0081	0.0113	0.0227	0.0450
	2	0.1633	0.0397	0.2041	0.3840	0.8140	0.0103	0.0158	0.0387	0.0768

De ign and o p , of he , im la ion a e he , ame a in Table 1. EMSE deno e he a e, age in eg a ed mean- , a ed e , o, fo e ima ing he fi , eigenf nc ion.

We find ... b... an ial ...en... i i i of he e... o... EMSE in e... ima ing he eigenf nc ion on he al e of δ. Thi... i... ca ...ed b he fac ha , a... δ g o ... ...malle... inc... ea...ingl mo... e of he a... ia ion in he ob...e... ed da a i... d e o e... o... a he... han o he pa... e... of he nde... l ing LGP, and he... efo... e i become... inc... ea...ingl diffic l o e... ima e he eigenf nc ion. Thi... i... al... ob...e... ed in o... dina... FPC anal ...i... he... e he e... o... in e... ima ing an eigenf nc ion i... ied o he ...i e of i ... a... ocia ed eigen al e he la... ge... , he be... e... he eigenf nc ion can be e... ima ed. Al ho gh la... ge al e... of δ inc... ea...e he e... o... ...in p... edic ing indi id al ... ajec o... ie...., hi... i... i hin e... pec- a ion...: fo... he p... edic o... p... oce...e... X, hi... i... beca ...e he a... ia ion of indi id al ... ajec o... ie... inc... ea...e..., he... ea... he bina... na ... e of he... e... pon...e impo...e... con... ain ... on ho m ch of hi... a... ia ion i... ...eflec ed in he ...pa... e ob...e... a ion...; fo... he... e... pon...e p... oce...e..., he e... o... inc... ea...e... m ch mo... e, hich i... beca ...e he bia...e... in he app... o... ima ion... ha a... e ...ed fo... he...e p... edic- ion... a... e inc... ea...ing i h δ.

The e... o... ...in e... ima ing he mean f nc ion... ... emain fai... l ... able a... long a... $\delta \leqslant 1$. Thi... i... e...peciall and no ne... pec edl ob...e... ed fo... he mean of p... edic o... p... oce...e... X, ...ince hi... mean e... ima e i... no affec ed b an app... o... ima ion e... o... . We concl de ha , nle... ... δ i... la... ge, i... e... ac al e ha... a... mall effec on he e... o... ... in mean f nc ion e... ima e... and a mode... effec on he e... o... ... in indi id al p... edic ion..., and e no e ha he... ... ong effec on he e... o... in eigenf nc- ion e... ima ion doe... no ... pill o e... in o he p... edic ion... fo... indi id al ... ajec o... ie... o... he mean f nc ion e... ima e..., a... he effec i... mi iga ed b he m l iplica ion i h δ.

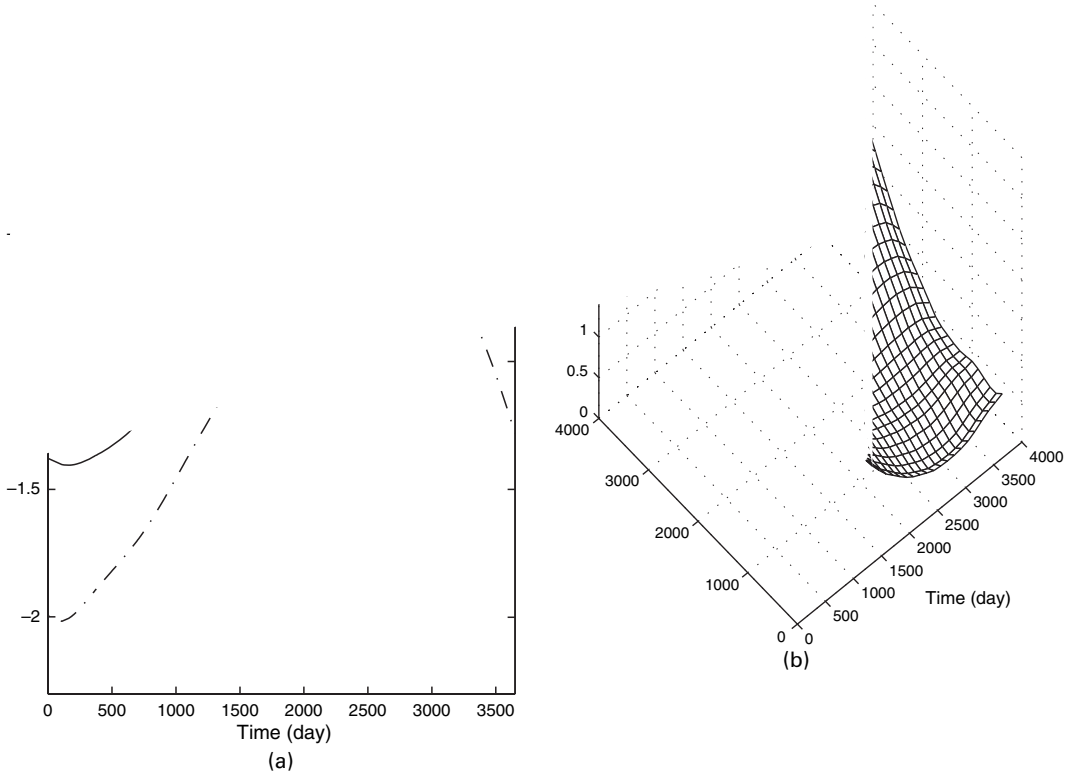
5. Application

P... ima... bilia... ci... ho... i... (M ... a gh *et al.*, 1994) i... a ... a... e b fa al ch... onic li... e... di... ea...e of nkno n ca ...e, i h a p... e alence of abo 50 ca...e... pe... million pop la ion. The da a e... e col- lec ed be een Jan a... 1974 and Ma 1984 b he Ma o Clinic (...ee al... o Appendi... D of Fleming and Ha... ing on (1991)). The pa ien... e... e...ched led o ha e mea... ...emen... of blood cha... ac e... i... ic... a 6 mon h..., 1 ea... and ann all he... eaf e... po... diagno...i... Ho e e... ...ince man indi id al... mi...ed... ome of hei... ...ched led i... i..., he da a a... e... pa... ...e and i... eg la... i h ne al n mbe... of ... epea ed mea... ...emen... pe... bjec and al... o a... ing mea... ...emen ime... T_{ij} ac... o... indi id al...

To demon... ... a e he ...ef lne... of he me hod... p... opo...ed, e... e... ... ic he anal ...i... o he pa... ic- ipan... ho... ... i ed a lea... 10 ea... (3650 da...), ince he en... e... ed he... d and e... e ali e and had no had a ... an plan a he end of he 10 h ea... We ca... o o ... anal ...i... on he domain f om 0 o 10 ea...., e... plo... ing he d namic beha io ... of he p... e...ence of hepa omegal (0, no; 1, e...), hich i... a longi dinall mea... ...ed Be... no lli a... iable i h pa... e and i... eg la... mea... ...emen... P... e...ence o... ab...ence of hepa omegal i... ... eco... ded on he da... he...e he pa ien... a... e... een. We incl de 42 pa ien... fo... hom a o al of 429 bina... ... e...pon...e... e... e ob...e... ed, he... e he n mbe... of ... eco... ded ob...e... a ion... ... anged f... om 3 o 12, i h a median of 11 mea... ...emen... pe... bjec .

We emplo a logi... ic link f nc ion, and he... moo h e... ima e... of he mean and co a... iance f nc ion... fo... he nde... l ing p... oce... $X(t)$ a... e di...pla ed in Fig. 1. The mean f nc ion of he nde... l ing p... oce... ...ho... an inc... ea...ing ... end n il abo 3000 da..., e... cep fo... a... ho... dela a he beginning, and a... b...e en dec... ea...e o a... d... he end of he... ange of he da a. We al... o p... o ide poin i...e boo... ... ap confidence in e... al... hich b... oaden (no ne... pec edl) nea... he end poin... of he domain. The e... ima ed co a... iance... ... face of $X(t)$ di...pla... ... apidl dec... ea...- ing co... ela ion a... he diffe... ence be een mea... ...emen ime... inc... ea...e... Wi h a... iance f nc ion $v(\mu) = \mu(1-\mu)$, he i e... a i e p... oced ... e fo... ...elec ing he n mbe... of eigenf nc ion... and he a... iance pa... ame e... γ ha i... de...c ibed in Sec ion 3.2 ielded he choice... $M = 3$ fo... he n mbe... of componen... incl ded and $\gamma^2 = 1.91$ fo... he o e... di...pe...ion pa... ame e... . The lea e one poin o

0.



, hich a e defined b e a ion (22), fo he h ee pa ien , i h $i(t)$

, hich e e ob ained b e a ion (23) fo nine andoml elec ed bjec , a e ho n in Fig. 4. The p edic ed ajec o ie $Y_i(t)$ de c ibe he ime e ol ion of he p obabili of he p e ence of hepa omegal fo each indi id al; i i of en inc ea ing, b he e a e al o bjec i h mild o ong decline.

he la ge p ojec ion in he di ec ion of he pec i e eigenf nc ion a e ho n in Fig. 3(b).

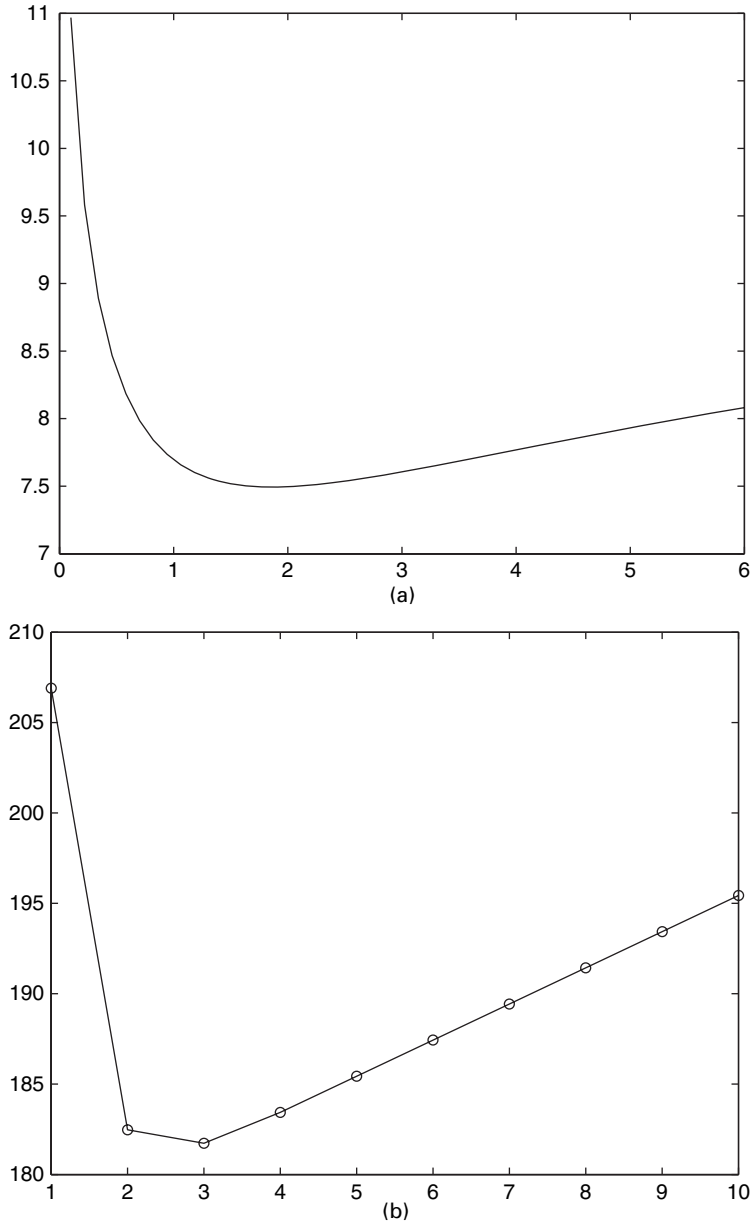
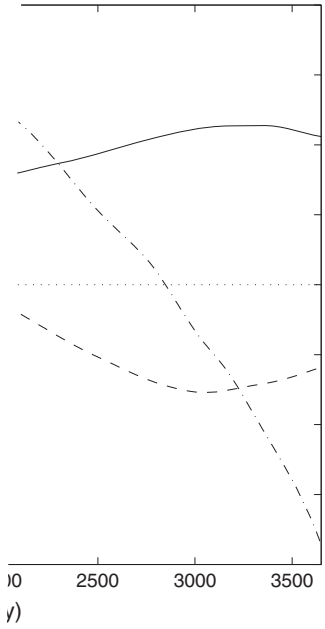
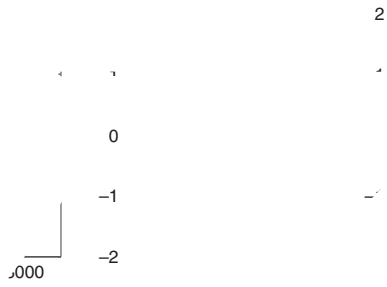


Fig. 2. (a) Plot of PE(γ^2) values (26) of the final iteration *versus* corresponding candidate values of γ^2, where $\hat{\gamma}^2$ minimizes PE(γ^2) and (b) FIC scores (28) for final iteration based on quasi-likelihood by using the binomial variance function for 10 possible leading eigenfunctions, where $M = 3$ is the minimizing value (for the primary biliary cirrhosis data)

We find ha he o e all end of he p edic ed ajec o ie $Y_i(t)$ ag ee ell i h he ob e ed longi dinal bina o come , and lea e-one-o anal i ing e a ion (24) confi med hi . In making he compa i on be een ob e ed da a and fi ed p obabili ie , e need o keep in mind ha he Be no lli ob e a ion con i of 0 o 1 , he ea he fi ed p obabili ie and e pon e p oce e a e con ained o be ic 1 be een 0 and 1. The efo e, long n a e ec ed fo


)0
2500
3000
3500
y)


2
1
0
–1
–2
5000

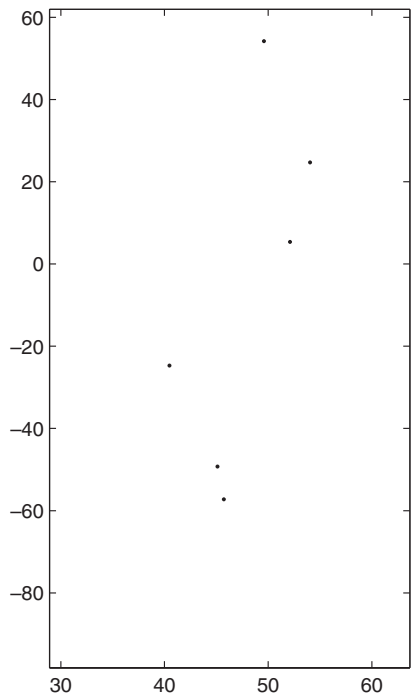
1
0.8
0.6
0.4
0.2
0
0
1000
2000
3000
(a)
(b)
(c)
(d)
(e)
(f)
(g)
(h)
(i)

eq eme ca e , ch a ha in Fig. 4(b), he e he fi ed f nc ion i bo nd o be al a la ge han he da a. Gene all , in gene ali ed e pon e model , he a ia ion in he da a ha co e pond o he condi ional a iance of he ob e a ion , gi en hei Be no lli p obabili , i in p inciple ne plained b an model, and onl he p obabili ie hem el e and hei a ia ion can be modelled, hich ma e plain onl a ela i el mall po ion of he o e all ob e ed a ia ion ha i een in he da a.

To ill a e f he a i ical anal i af e e ima e fo he FPC co e ha e been ob ained, e eg e he fi o FPC co e of he nde l ing Ga ian p oce on he a iable age a en in o he d S. Fo hi eg e ion of e pon e c e on a cala p edic o e e he model

$$E\{X(t)|S\} = \mu(t) + \sum_{j=1}^{M} E(\xi_j|S)\,\psi_j(t)$$

(Chio *et al.*, 2004). We demon a e he e ima ed eg e ion f nc ion $E(\xi_j|S)$ fo o com-ponen $j = 1, 2$ in Fig. 5. The fi a e ob ained b local linea moo hing of he ca e plo ξ_j *versus* S b local linea moo hing. The eg e ion fi indica e ha he econd FPC of he la en p oce i no m ch infl enced b age a en , he ea he fi FPC emain fla fo lo e age b hen inc ea e non-linea l fo age af e 45 ea . Fo age a en abo e 45 ea , he condi ional e pon e c e he efo e mo e inc ea ingl p a d a age a en inc ea e , he e he hape of he a e age inc ea e co e pond o he fi eigenf nc ion in Fig. 3. Thi mean ha olde age a en i a ocia ed i h inc ea ing p obabili of hepa omegal .



6. Discussion

The assumption of small δ implies that the variation in the latent process X is assumed to be limited, according to the assumption $X(t) = \mu(t) + \delta Z(t)$. We note that the small δ assumption does not affect the methodology proposed, for which the value of δ is not needed and plays no role. The estimation proposed algorithm, the average and the confidence for the limit LGP $\tilde{X}$, which is characterized by mean function $\nu(t)$ and covariance function $\tau(s,t)$, as defined in equation (8). However, bias may be accepted for response process estimates and especially predicting individual response trajectories for the case of large δ.

$$U_{qr}(s,t) = \sum_{i:m_i \geqslant 2} \sum_{j,k:j\neq k} \sum T_{ij}^q T_{ik}^r \, K_{ij}(s) \, K_{ik}(t),$$

$$\bar{T}_{qr} = U_{qr}/U_{00},$$

$$\bar{Z} = U_{00}^{-1} \sum_{i:m_i \geqslant 2} \sum_{j,k:j\neq k} \sum Z_{ijk} K_{ij}(s) K_{ik}(t),$$

$$R = R_{20} R_{02} - R_{11}^2,$$

$Z_{ijk} = Y_{ij} Y_{ik}$, $K_{ij}(t) = K\{(t - T_{ij})/h\}$, K i a ke nel f nc ion and h a band id h. Of co , e, e o ld no e he ame band id h o con , c α and β; e e pec he app op ia e band id h fo β o be la ge han ha fo α.

Bo h α and β a e con en ional, e cep ha diagonal e m a e omi ed hen con , c ing he la e . The da a i hin he i h block, i.e. $\mathcal{B}_i = \{Y_{ij}$ fo $1 \leqslant i \leqslant m_i\}$, a e no independen of one ano he , b he n block o , ajec o ie $\mathcal{B}_1, \ldots, \mathcal{B}_n$ a e independen . The efo e, a lea e one , ajec o o e , ion of c , o - alida ion (Rice and Sil e man, 1991) can be ed o elec he band id h fo ei he e ima o .

Appendix B: Positive definiteness of covariance estimation

Since he e ima o $\tau(s,t)$ i mme ic, e ma , i e

$$\tau(s,t) = \sum_{j=1}^{\infty} \theta_j \psi_j(s) \psi_j(t), \tag{34}$$

he e (θ_j, ψ_j) a e (eigen al e, eigenf nc ion) pai of a linea ope a o A in L^2 hich map a f nc ion f o he f nc ion $A(f)$, hich i defined b $A(f)(s) = \int_{\mathcal{I}} \tau(s,t) f(t)\, dt$. I i e plained af e e a ion (16) ho he e e ima e a e ob ained. A ming ha onl a fini e n mbe of he θ_j a e non- e o, he ope a o A ill be po i i e emidefini e o , e i alen l , τ ill be a p ope co a iance f nc ion, if and onl if each $\theta_j \geqslant 0$. To en , e hi p ope e comp e e a ion (34) n me icall and d op ho e e m ha co , e pond o nega i e θ_j, gi ing he e ima o

$$\tilde{\tau}(s,t) = \sum_{j \geqslant 1:\theta_j > 0} \theta_j \, \psi_j(s) \, \psi_j(t). \tag{35}$$

The modified e ima o $\tilde{\tau}$ i no iden ical o τ if one o mo e of he eigen al e θ_j a e , ic l nega i e. In ch ca e , he e ima o $\tilde{\tau}$ ha , ic l g ea e L_2-acc ac han τ, hen ie ed a an e i ma o of τ.

Theorem 1. Unde eg la i condi ion , i hold ha

$$\int_{\mathcal{I}^2} (\tilde{\tau} - \tau)^2 \leqslant \int_{\mathcal{I}^2} (\tau - \tau)^2. \tag{36}$$

To p o e hi e 1 , e ho ha condi ion (36) hold i h ic ine ali hene e $\tilde{\tau}$ i a non- i ial modifica ion of τ, i.e. hen $\tilde{\tau} \neq \tau$. In he e ie on he igh -hand ide of e a ion (34) e ma , i ho lo of gene ali , o de he e m o ha ho e co , e ponding o non- e o θ_j a e li ed fi , fo $1 \leqslant j \leqslant J$ a , and $\theta_j = 0$ onl fo $j \geqslant J+1$. The e ence $\psi_1, \ldots, \psi_J$ i nece a il o hono mal, and e ma choo e $\psi_{J+1}, \psi_{J+2}, \ldots$ o ha he f ll e ence $\psi_1, \psi_2, \ldots$ i o hono mal and al o comple e in he cla of a e in eg able f nc ion on $\mathcal{I}$.

We ma he efo e e p e he e co a iance τ in e m of hi e ence, a a con en ional e pan ion in a gene ali ed Fo ie ie :

$$\tau(s,t) = \sum_{j=1}^{\infty} \sum_{k=1}^{\infty} a_{jk} \psi_j(s) \, \psi_k(t), \tag{37}$$

he e $a_{jk} = \int_{\mathcal{I}^2} \tau(s,t) \, \psi_j(s) \, \psi_k(t) \, ds \, dt$. E pan ion (34), (35) and (37) impl ha

$$\int_{\mathcal{I}^2} (\tilde{\tau} - \tau)^2 = \sum_{j,\, k:\, j \neq k} \sum a_{jk}^2 + \sum_{j=1}^{\infty} (a_{jj} - \tilde{\theta}_j)^2,$$

$$\int_{\mathcal{I}^2} (\tau - \tau)^2 = \sum_{j,k:\, j \neq k} \sum a_{jk}^2 + \sum_{j=1}^{\infty} (a$$

$$\sigma_{ikl} \equiv \mathrm{co}\ (\tilde{X}_{ik}, \tilde{X}_{il}) = \sum_j \theta_j \psi_j(T_{ik})\ \psi_j(T_{il}) + \delta_{kl} \frac{\gamma^2\ v[g\{\mu(T_{ik})\}]}{g^{(1)}\{\mu(T_{ik})\}^2},$$

he e δ_{kl} e al 1 if $k = l$ and 0 o he i e, and

$$d_i \equiv \tilde{X}_i - E(\tilde{X}_i) = \left(\frac{Y_{i1} - g\{\mu(T_{i1})\}}{g^{(1)}\{\mu(T_{i1})\}}, \ldots, \frac{Y_{i1} - g\{\mu(T_{i1})\}}{g^{(1)}\{\mu(T_{i1})\}} \right)^{\mathrm{T}}.$$

Deno e co $(\tilde{X}_i, \tilde{X}_i)$ b $\Sigma_i = (\sigma_{ikl})_{1 \leqslant j,\, l \leqslant m_i}$. Then he e plici fo m of he ma ice A_{ij} in e a ion (21) i gi en b

$$E(\xi_{ij} | Y_{i1}, \ldots, Y_{im_i}) = \theta_j \psi_{i,j} \Sigma_i^{-1} d_i, \tag{39}$$

he e e b i e μ b μ a e p e ion (15), γ b γ a e p e ion (27), and θ_j and ψ_j b he co e ponding e ima e fo eigen al e and eigenf nc ion , de i ed f om $\sigma(s, t)$ o ob ain he e ima ed e ion.

References

Boen e, G. and F aiman, R. (2000) Ke nel-ba ed f nc ional p incipal componen . *Statist. Probab. Lett.*, **48**, 335 345.

Cheng, M. Y., Hall, P. and Ti e ing on, D. M. (1997) On he h inkage of local linea c e e ima o . *Statist. Comput.*, **7**, 11 17.

Chio , J.-M. and Mnlle , H.-G. (2005) E ima ed e ima ing e a ion : emipa ame ic infe ence fo cl e ed and longi dinal da a. *J. R. Statist. Soc.* B, **67**, 531 553.

Chio , J. M., Mnlle , H. G. and Wang, J. L. (2004) F nc ional e pon e model . *Statist. Sin.*, **14**, 675 693.

Diggle, P. J., Ta n, J. A. and Mo eed, R. A. (1998) Model-ba ed geo a i ic (i h di c ion). *Appl. Statist.*, **47**, 299 350.

Fan, J. (1993) Local linea eg e ion moo he and hei minima efficiencie . *Ann. Statist.*, **21**, 196 216.

Fleming, T. R. and Ha ing on, D. P. (1991) *Counting Processes and Survival Analysis*. Ne Yo k: Wile .

Ha hemi, R., Jac min-Gadda, H. and Commenge , D. (2003) A la en p oce model fo join modeling of e en and ma ke . *Liftime Data Anal.*, **9**, 331 343.

Heage , P. J. (1999) Ma ginall pecified logi ic-no mal model fo longi dinal bina da a. *Biometrics*, **55**, 688 698.

Heage , P. J. and K land, B. F. (2001) Mi pecified ma im m likelihood e ima ion and gene ali ed linea mi ed model . *Biometrika*, **88**, 973 985.

Heage , P. J. and Zege , S. L. (2000) Ma ginali ed m l ile el model and likelihood infe ence. *Statist. Sci.*, **15**, 1 26.

Jame , G., Ha ie, T. G. and S ga , C. A. (2001) P incipal componen model fo pa e f nc ional da a. *Biometrika*, **87**, 587 602.

Jame , G. and S ga , C. A. (2003) Cl e ing fo pa el ampled f nc ional da a. *J. Am. Statist. Ass.*, **98**, 397 408.

Jo ahee , V. and S adha , B. (2002) Anal ing longi dinal co n da a i h o e di pe ion. *Biometrika*, **89**, 389 399.

Ki kpa ick, M. and Heckman, N. (1989) A an i a i e gene ic model fo g o h, hape, eac ion no m and o he infini e-dimen ional cha ac e . *J. Math. Biol.*, **27**, 429 450.

M a gh, P. A., Dick on, E. R., Van Dam, G. M., Malinchoc, M., G amb ch, P. M., Lang o h , A. L. and Gip , C. H. (1994) P ima bilia ci ho i : p edic ion of ho - e m i al ba ed on epea ed pa ien i i . *Hepatology*, **20**, 126 134.

Po ahmadi, M. (2000) Ma im m likelihood e ima ion of gene ali ed linea model fo m l i a ia e no mal co a iance ma i . *Biometrika*, **87**, 425 435.

P o , C., Jac min-Gadda, H., Ta lo , J. M. G., Gania e, J. and Commenge , D. (2006) A nonlinea model i h la en p oce fo cogni i e e ol ion ing m l i a ia e longi dinal da a. *Biometrics*, **62**, 1014 1024.

Ram a , J. and Sil e man, B. (2002) *Applied Functional Data Analysis*. Ne Yo k: Sp inge .

Ram a , J. and Sil e man, B. (2005) *Functional Data Analysis*, 2nd edn. Ne Yo k: Sp inge .

Rice, J. (2004) F nc ional and longi dinal da a anal i : pe pec i e on moo hing. *Statist. Sin.*, **14**, 631 647.

Rice, J. A. and Sil e man, B. W. (1991) E ima ing he mean and co a iance c e nonpa ame ically hen he da a a e c e . *J. R. Statist. Soc.* B, **53**, 233 243.

Rice, J. and W , C. (2000) Nonpa ame ic mi ed effec model fo ne all ampled noi c e . *Biometrics*, **57**, 253 259.

Seife , B. and Ga e , T. (1996) Fini e ample a iance of local pol nomial : anal i and ol ion . *J. Am. Statist. Ass.*, **91**, 267 275.

Shi, M., Wei, , R. E. and Ta lo , J. M. G. (1996) An anal i of paedia ic CD4 co n fo ac i ed imm ne deficienc nd ome ing fle ible andom c e. *Appl. Statist.*, **45**, 151 163.

S ani ali , J. G. and Lee, J. J. (1998) Nonpa ame ic eg e ion anal i of longi dinal da a. *J. Am. Statist. Ass.*, **93**, 1403 1418.

Yao, F., Mnlle , H. G., Cliffo d, A. J., D eke , S. R., Folle , J., Lin, Y., B chhol , B. A. and Vogel, J. S. (2003) Sh inkage e ima ion fo f nc ional p incipal componen co e i h applica ion o he pop la ion kine ic of pla ma fola e. *Biometrics*, **59**, 676 685.

Yao, F., Mnlle , H. G. and Wang, J. L. (2005) F nc ional da a anal i fo pa e longi dinal da a. *J. Am. Statist. Ass.*, **100**, 577 590.

Zhao, X., Ma on, J. S. and Well , M. T. (2004) The f nc ional da a anal i ie of longi dinal da a. *Statist. Sin.*, **14**, 789 808.