# Two-Stage Functional Mixed Models for Evaluating the Effect of Longitudinal Covariate Profiles on a Scalar Outcome

Daowen Zhang[1,*], Xihong Lin[2] and MaryFran Sowers[3]

[1] Department of Statistics, North Carolina State University, Raleigh, NC 27695

[2] Department of Biostatistics, Harvard School of Public Health, Boston, MA 02115

[3] Department of Epidemiology, University of Michigan, Ann Arbor, MI 48109

*email: zhang@stat.ncsu.edu

SUMMARY. The Daily Hormone Study, a sub-study of the Study of Women's Health Across the Nation (SWAN) consisting of more than 600 pre- and perimenopausal women, includes a scalar measure of total hip bone mineral density (BMD) together with repeated measures of creatinine-adjusted follicle stimulating hormone (FSH) assayed from daily urine samples collected over one menstrual cycle. It is of scientific interest to investigate the effect of the FSH time profile during a menstrual cycle on the total hip BMD, adjusting for age and body mass index. The statistical analysis is challenged by several features of the data: (1) The covariate FSH is measured longitudinally and its effect on the scalar outcome BMD may be complex; (2) Due to varying menstrual cycle lengths, subjects have unbalanced longitudinal measures of FSH; (3) The longitudinal measures of FSH are subject to considerable among- and within-subject variations and measurement errors. We propose a measurement error partial functional linear model, where repeated measures of FSH are modeled using a functional mixed effects model and the effect of the FSH time profile on BMD is modeled using a partial functional linear model by treating the unobserved true subject-specific FSH time profile as a functional covariate. We develop a two-stage nonparametric regression calibration method using period smoothing splines. Using the connection between smoothing splines and mixed models, we show that a key feature of our approach is that estimation at both stages can be conveniently cast into a unified mixed model framework. A simple testing

procedure for constant functional covariate effect is also proposed. The proposed methods are evaluated using simulation studies and applied to the SWAN data.

# 1 Introduction

The Daily Hormone Study is a sub-study of the Study of Women's Health Across the Nation (SWAN) involving 628 women (Sowers, et al. 2000). Total hip bone mineral density (BMD) ($g/cm^2$) was measured for each woman at a clinical visit. Each woman was also asked to collect daily urine samples during one menstrual cycle, and daily creatinine-adjusted follicle stimulating hormone (FSH) levels ($mI\mu/ml$) were measured by laboratory assay. One of the study objectives is to assess the effect of the FSH profile during a menstrual cycle on total hip BMD adjusting for age and body mass index (BMI) (Sowers, et al. 2003).

Several unique features of this data set present substantial statistical challenges. First, the primary response BMD is a scalar, but the covariate FSH is measured longitudinally. Hence standard longitudinal techniques are not applicable. Further, menstrual cycle lengths vary substantially among women, ranging from 14 to 50 days with an average of about 30 days, and result in a great deal of imbalance in the number of FSH measures per subject. Secondly, both biological understanding of reproductive hormones and empirical evidence indicate that the FSH profile exhibits a complex periodic pattern over a menstrual cycle. Figure 1 presents the FSH concentrations in a menstrual cycle for 4 randomly chosen women, where their cycle lengths were standardized by 28 days (Zhang, et al, 1998). We also superimpose in Figure 1 the estimated periodic subject-specific profiles and the population profile obtained using the estimation procedure to be discussed in Section 3. It clearly shows large between-woman and within-woman variation in FSH levels and individual FSH profiles. We are hence challenged by modeling the effects of periodic complex longitudinal FSH profiles on the scalar outcome BMD, with the longitudinal FSH measures subject to considerable imbalance and variation.

When a functional covariate is measured without error at equally spaced time points, Ramsay and Silverman (1997) proposed a functional linear model to model its relationship with a continuous scalar response variable characterized by an integrated smooth function, which is estimated using traditional smoothing techniques. When the functional covariate

is measured with error, a naive approach estimates the covariate profile for each individual separately using smoothing techniques, such as kernels or splines, and then plugs the estimators of the individual covariate profiles in the functional linear model and estimates the coefficient function following Ramsay and Silverman (1997). However, since the observed individual longitudinal covariate values measure the true individual covariate profiles with error, as demonstrated in the measurement error literature (Carroll, et al. 1995), such a naive plug-in estimator ignores measurement error and often leads to biased estimators of regression coefficients. Similar bias results are shown for mixed effects models when naive individual trajectories are fit, especially when the data are variable and the number of repeated measures is small (Fitzmaurice, et al, 2004).

James (2002) recently proposed an extension of functional linear models for discrete and survival data, where the functional covariate is measured with error. Similar to Ramsay and Silverman (1997), he expressed the true subject-specific functional covariate as a linear combination of a pre-selected fixed number of cubic spline basis functions with normally distributed subject-specific random coefficients, in a similar spirit to Rice and Wu (2001), and parameterized the fixed functional covariate effect using a similar basis function approach. The model then reduces to a special mixed model with parameters in the design matrix, and the EM algorithm can be used for maximum likelihood estimation. However, as shown in his simulation studies and our simulation study (Section 6), the nonparametric functional covariate effect estimate is sensitive to the choice of the number of knots and can be subject to considerable bias. This regression spline approach also suffers from numerical difficulties. As the number of the basis functions increases, the number of parameters grows quickly causing numerical as well as inferential problems.

In this paper, we propose a two-stage functional mixed effect model, and estimate the effect of the periodic functional covariate, e.g., the FSH profile, on a continuous scalar response variable, e.g., BMD, using a two-stage nonparametric regression calibration (TS-

NRC) method. At the first stage, a periodic nonparametric mixed model is used to estimate the periodic subject-specific profiles using periodic cubic smoothing splines. At the second stage, the estimated subject-specific profiles are then plugged in the functional linear model, and the functional covariate effect is estimated by a periodic cubic smoothing spline through a linear mixed model formulation. Our approach is an extension of regression calibration in traditional parametric measurement error models (Carroll, et al. 1995) to nonparametric settings. It allows for measurement error and imbalance in repeated measures of FSH levels, and models the within- and between- subject variability in FSH profiles. A key feature of our approach is that mixed model representations are used in a unified convenient way at both stages to estimate all the model parameters simultaneously, including nonparametric functions, variance components, and smoothing parameters. This is in contrast to the regression spline approach of James (2002), where one has to choose the number and the locations of knots, which can be difficult in practice and is likely to cause overfitting and yield artificial curvatures mainly driven by the assumed parametric form instead of the data (Carroll, et al. 2004). Our smoothing spline approach uses all data points as knots and flexibly controls the smoothness and goodness-of-fit of individual covariate profiles and the functional covariate effects through penalties, and estimates smoothing parameters conveniently as variance components within a unified mixed model framework. It also eliminates the inferential and numerical problems associated with the increasing number of parameters of the regression spline method. Previous results show that smoothing splines perform better than regression splines for longitudinal data (Carroll, et al, 2004). Our simulation study further shows a similar phenomenon in our functional regression settings.

The rest of the paper is organized as follows. We present the two-stage functional mixed model in Section 2. We describe in Section 3 the two-stage nonparametric regression calibration (TS-NRC) estimation procedure, and a test for whether or not the functional covariate effect is a constant. We provide theoretical justifications for the proposed TS-NRC method

in Section (4). We apply in Section 5 the proposed methods to analyze the SWAN data. In Section 6 we conduct an extensive simulation study to evaluate the performance of proposed estimation and testing procedures, and compare our estimation procedure to James' (2002) regression spline approach. Concluding remarks are given in Section 7.

# 2    The Periodic Partial Functional Linear Model with Measurement Error

Suppose that the data are collected from $m$ subjects. For subject $i$ $(i = 1, 2, ..., m)$, let $Y_i$ be a scalar response variable, $\boldsymbol{s}_i$ be a $p \times 1$ vector of scalar covariates, e.g., age and BMI in the Daily Hormone Study, and $W_{ij}$ be observed error-prone longitudinal measures of the true unobserved smooth periodic functional covariate $x_i(t)$ at time point $t_{ij}$ $(j = 1, ..., n_i)$, where $0 \leq t_{ij} \leq T$ and $T$ is the period length such as the standardized cycle length in the Daily Hormone Study. The response variable $Y_i$ is assumed to relate to the unobserved functional covariate $x_i(t)$ through a partial functional linear model (Ramsay and Silverman, 1997)

$$Y_i = \alpha + \boldsymbol{s}_i^T \boldsymbol{\delta} + \int_0^T x_i(t)\gamma(t)dt + \epsilon_i, \tag{1}$$

where $\alpha$ is the intercept, $\boldsymbol{\delta}$ is a $p \times 1$ vector of regression coefficients of $\boldsymbol{s}_i$, $\gamma(t)$ is an unknown periodic smooth function with period $T$ modeling the effect of $x(t)$ on $Y$, and the errors $\epsilon_i$ are assumed to be independent and identically distributed as $\mathrm{N}(0, \sigma_\epsilon^2)$.

The model is completed by assuming the observed longitudinal covariates $W_{ij}$ are related to $x_i(t)$ by an additive measurement error model as follows

$$W_{ij} = x_i(t_{ij}) + e_{ij}, \tag{2}$$

where $e_{ij}$'s are measurement errors assumed to be independent and identically distributed as $\mathrm{N}(0, \sigma_e^2)$. Our interest lies in using the data $(Y_i, \boldsymbol{s}_i, W_{ij})$ to estimate the regression coefficients $\alpha$, $\boldsymbol{\delta}$, the nonparametric periodic function $\gamma(t)$ and the residual variances $\sigma_\epsilon^2$ and $\sigma_e^2$.

When analyzing the Daily Hormone Study data, we can assess the potential effect of individual cycle length $T_i$ on $Y_i$ by including $T_i$ in the covariate vector $\boldsymbol{s}_i$ in model (1). Alternatively, we may model the effect of $T_i$ implicitly by replacing the integral in (1) with $\int_0^{T_i} \tilde{x}_i(t_i)\gamma(t)dt_i$, where $\tilde{x}_i(t_i)$ is the true hormone profile for woman $i$ in the actual time scale, and $t = Tt_i/T_i$ is the standardized day. This integral can be shown to be $(T_i/T)\int_0^T x_i(t)\gamma(t)dt$ with $x_i(t) = \tilde{x}_i(T_it/T)$ being individual hormone profile in the standardized cycle length. Therefore, the methods to be developed for model (1) can be easily modified to fit this new model. See Section 5 for more discussion.

When the functional coefficient $\gamma(t) = \gamma$ is a constant, model (1) reduces to the familiar case of using the "area under the curve" $A_i = \int_0^T x_i(t)dt$ as a covariate. One can use $\overline{W}_i = \sum_{j=1}^{n_i} W_{ij}(t_{ij} - t_{ij-1})$ as an error-prone measure of $A_i$ and apply the conventional linear model measurement error approach (Carroll, et al., 1995). However, since the whole profile $x_i(t)$ is unknown, with finite $n_i$, $\overline{W}_i$ is often a biased measure of $A_i$. Results in the standard measurement error literature mainly concern parametric regression and do not directly apply here even in this special case (Carroll et al., 1995). When the repeated measures of $W_{ij}$ are observed at the same time points without error, see Ramsay and Silverman (1997).

# 3 The Two-stage Nonparametric Regression Calibration Method

Both the periodic nonparametric function $\gamma(t)$ and the true periodic subject-specific covariate profiles $x_i(t)$ in the functional linear model (1) are infinitely dimensional parameters. We estimate them by periodic smoothing splines. Specifically, denote by $\mathbf{t}^0$ the $q \times 1$ vector of the distinct values of the time points $t_{ij}$'s (0 and $T$ are considered as the same time point). We assume that $\{\gamma(t), x_i(t)\} \in W_{ps} =$ {periodic cubic smoothing splines in $[0, T]$ with knots $\mathbf{t}^0$}.

For any function $d(t) \in W_{ps}$, it is easy to show using the results in Green and Silver-

man (1994) that there exist $q$ piece-wise periodic cubic polynomial basis functions $\mathbf{c}(t) = \{c_1(t), ..., c_q(t)\}^T$ such that $d(t) = \mathbf{d}^T \mathbf{c}(t)$, where $\mathbf{d} = d(\mathbf{t}^0)$ is the vector formed by evaluating $d(t)$ at knot $\mathbf{t}^0$. Each function $c_j(t)$ itself is also a periodic cubic smoothing spline uniquely determined by its values at knot $\mathbf{t}^0$ assumed as $(0, \cdots, 0, 1, 0, \cdots, 0)$ with the $j$th element being 1 and the rest being 0. This gives periodic cubic smoothing spline basis function expressions for the functional covariate effect $\gamma(t)$ and the subject-specific profile $x_i(t)$ as

$$\gamma(t) = \boldsymbol{\gamma}^T \mathbf{c}(t), \quad x_i(t) = \mathbf{x}_i^T \mathbf{c}(t),$$

where $\boldsymbol{\gamma} = \gamma(\mathbf{t}^0)$ and $\mathbf{x}_i = x_i(\mathbf{t}^0)$. Using these expressions, the measurement error partial functional linear model becomes

$$
\begin{aligned}
Y_i &= \alpha + \boldsymbol{s}_i^T \boldsymbol{\delta} + \mathbf{x}_i^T \mathbf{C} \boldsymbol{\gamma} + \epsilon_i, \qquad (3) \\
\mathbf{W}_i &= \mathbf{N}_i \mathbf{x}_i + \mathbf{e}_i
\end{aligned}
$$

where $\mathbf{C} = \int_0^T \mathbf{c}(t) \mathbf{c}^T(t) dt$ and the integrals are evaluated element-wise, $\mathbf{W}_i = (W_{i1}, ..., W_{in_i})^T$, $\mathbf{N}_i$ is the incidence matrix mapping $(t_{i1}, ..., t_{in_i})^T$ to $\mathbf{t}^0$ such that the $(j, l)$th element of $\mathbf{N}_i$ is 1 if $t_{ij} = t_l^0$ and 0 otherwise $(j = 1, \ldots, n_i, l = 1, \ldots, q)$, and $\mathbf{e}_i = (e_{i1}, ..., e_{in_i})^T$. Since each $c_j(t)$ is a piece-wise periodic cubic polynomial, calculation of the matrix $\mathbf{C}$ is straightforward.

Model (3) takes the same form as the standard linear measurement error model. However, unlike the traditional linear measurement error model, the dimensions of $\gamma(t)$ and $x_i(t)$ are infinite so they have to be estimated nonparametrically. Motivated by the parametric regression calibration method in traditional parametric measurement error models (Carroll, et al., 1995), we propose a nonparametric regression calibration approach. At stage I, we calculate the subject-specific smoothing spline estimator $\hat{x}_i(t)$, equivalently $\hat{\mathbf{x}}_i$, under (2) using the functional mixed model formulation. At stage II, we calculate the smoothing spline estimator of $\gamma(t)$, equivalently $\boldsymbol{\gamma}$, by fitting the nonparametric regression calibration model

$$Y_i = \alpha + \boldsymbol{s}_i^T \boldsymbol{\delta} + \hat{\mathbf{x}}_i^T \mathbf{C} \boldsymbol{\gamma} + \epsilon_i^*, \qquad (4)$$

via a linear mixed model smoothing spline formulation and treating the new residual error $\epsilon_i^*$ to be independent and identically distributed. We describe in Sections 3.1 and 3.2 the two-stage regression calibration estimation procedure. In Section 3.3, we propose a score test for testing $\gamma(t)$ is a constant, which corresponds to the simple linear regression model with the area under the curve $x_i(t)$ as a covariate.

## 3.1 Stage I: Estimation of the individual profile $x_i(t)$

Decompose the subject-specific profile $x_i(t)$ as

$$x_i(t) = f(t) + g_i(t),$$

where $f(t)$ is a periodic cubic smoothing spline modeling the population profile, and $g_i(t)$ is a random periodic cubic smoothing spline characterizing the subject-specific deviation from the population profile. Under this decomposition, the model for longitudinal measurements $W_{ij}$ becomes

$$W_{ij} = f(t_{ij}) + g_i(t_{ij}) + e_{ij}, \tag{5}$$

where $g_i(t)$ is a mean zero Gaussian stochastic process whose covariance matrix is given later in this section. Equivalently, the vector $\mathbf{W}_i$ can be re-written as

$$\mathbf{W}_i = \mathbf{N}_i \mathbf{f} + \mathbf{N}_i \mathbf{g}_i + \mathbf{e}_i, \tag{6}$$

where $\mathbf{f} = f(\mathbf{t}^0)$ and $\mathbf{g}_i = g_i(\mathbf{t}^0)$. It is sufficient to estimate $\mathbf{f}$ and $\mathbf{g}_i$ in order to estimate $\mathbf{x}_i$, and hence the process $x_i(t)$.

The smoothing spline estimator of $f(t)$ maximizes the penalized likelihood function

$$\ell_p(\mathbf{f}; \mathbf{W}) = \sum_{i=1}^{m} \ell(\mathbf{f}; \mathbf{W}_i) - \frac{\lambda_f}{2} \int_0^T \{f''(t)\}^2 dt = \sum_{i=1}^{m} \ell(\mathbf{f}; \mathbf{W}_i) - \frac{\lambda_f}{2} \mathbf{f}^T \mathbf{K} \mathbf{f},$$

where $\mathbf{W} = (\mathbf{W}_1^T, ..., \mathbf{W}_m^T)^T$, $\ell(\mathbf{f}; \mathbf{W}_i)$ is the log-likelihood of $\mathbf{f}$ under (6), $\lambda_f > 0$ is a smoothing parameter that controls the goodness-of-fit of the model and the roughness of $f(t)$, and $\mathbf{K}$ is the $q \times q$ smoothing matrix for a periodic smoothing spline and has rank $r = q - 1$

9

and satisfies $\mathbf{K1}_q = 0$, with $\mathbf{1}_q$ being a $q-$dimensional vector of ones. Following Zhang, et al. (1998, 2000), the smoothing spline estimator of $\mathbf{f}$ has a mixed model representation. Specifically, decompose $\mathbf{K}$ as $\mathbf{K} = \mathbf{LL}^T$, where $\mathbf{L}$ is a $q \times r$ full rank matrix. Then $\mathbf{f}$ has a mixed effect representation as $\mathbf{f} = \mathbf{1}_q\eta + \mathbf{Ba}$, where $\eta$ is a fixed effect, $\mathbf{B} = \mathbf{L}(\mathbf{L}^T\mathbf{L})^{-1}$ and $\mathbf{a}$ are random effects distributed as $\mathbf{a} \sim \mathrm{N}(0, \tau_f\mathbf{I}_{r\times r})$ with $\tau_f = \lambda_f^{-1}$ and identity matrix $\mathbf{I}_{r\times r}$. Hence the inverse of the smoothing parameter can be treated as a variance component.

Since $\mathbf{g}_i = g(\mathbf{t}^0)$ and $g_i(t)$ models the random subject-specific deviation from the population profile, the mixed effect representation of $\mathbf{f}$ motivates us to model $\mathbf{g}_i$ by $\mathbf{g}_i = \mathbf{1}_q b_i + \mathbf{Ba}_i$, where $b_i \sim \mathrm{N}(0, \sigma_b^2)$ and $\mathbf{a}_i \sim \mathrm{N}(0, \tau_g\mathbf{I}_{r\times r})$ are two independent random variables. Substituting the expressions of $\mathbf{f}$ and $\mathbf{g}_i$ in model (6), we get the following smoothing spline mixed model representation of model (6)

$$\mathbf{W}_i = \mathbf{1}_{n_i}\eta + \mathbf{N}_i\mathbf{Ba} + \mathbf{1}_{n_i}b_i + \mathbf{N}_i\mathbf{Ba}_i + \mathbf{e}_i, \tag{7}$$

where $\eta$ is the fixed effect, $\mathbf{a} \sim \mathrm{N}(0, \tau_f\mathbf{I}_{r\times r})$, $b_i \sim \mathrm{N}(0, \sigma_b^2)$ and $\mathbf{a}_i \sim \mathrm{N}(0, \tau_g\mathbf{I}_{r\times r})$ are independent random effects, and $\mathbf{e}_i$ is a vector of measurement errors.

Following Zhang et al. (1998), we treat $\tau_f$ as an extra variance component and estimate it with other variance components $(\sigma_b^2, \tau_g\ \sigma_e^2)$ simultaneously using the REML approach under the linear mixed model (7). We then estimate the population profile $\mathbf{f}$ by $\widehat{\mathbf{f}} = \mathbf{1}_q\widehat{\eta} + \mathbf{B}\widehat{\mathbf{a}}$ and subject-specific deviation $\mathbf{g}_i$ by $\widehat{\mathbf{g}}_i = \mathbf{1}_q\widehat{b}_i + \mathbf{B}\widehat{\mathbf{a}}_i$, where $\widehat{\eta}$, $\widehat{\mathbf{a}}, \widehat{b}_i$ and $\widehat{\mathbf{a}}_i$ are the BLUPs. This gives the estimate of subject-specific profile $\widehat{\mathbf{x}}_i = \widehat{\mathbf{f}} + \widehat{\mathbf{g}}_i$, which is the BLUP of $\mathbf{x}_i$ under the linear mixed model (7). Hence the first stage provides a nonparametric subject-specific periodic smoothing estimate of $\mathbf{x}_i$, equivalently, the subject-specific smoothing spline estimate $\hat{x}_i(t) = \hat{\mathbf{x}}_i^T\mathbf{c}(t)$.

## 3.2 Stage II: Estimation of $\gamma(t)$

The second stage of the nonparametric regression calibration approach is to estimate the nonparametric function $\gamma(t)$ in the functional model (1) using the nonparametric regression

calibration model (4) by holding the subject-specific profiles $\mathbf{x}_i$ fixed as the nonparametric BLUP estimates $\hat{\mathbf{x}}_i$. We estimate $\gamma(t)$, equivalently $\boldsymbol{\gamma}$, using a smoothing spline by developing a mixed model representation of the nonparametric regression calibration model (4).

Specifically, we jointly estimate $\boldsymbol{\gamma}$ and the parametric regression coefficients $\alpha$ and $\boldsymbol{\delta}$ by maximizing the following penalized pseudo-likelihood function

$$\ell_p(\alpha, \boldsymbol{\delta}, \boldsymbol{\gamma}; \mathbf{Y}, \boldsymbol{s}, \widehat{\mathbf{x}}) = \sum_{i=1}^{m} \ell(\alpha, \boldsymbol{\delta}, \boldsymbol{\gamma}; Y_i, \boldsymbol{s}_i, \widehat{\mathbf{x}}_i) - \frac{\lambda_\gamma}{2} \int_0^T \{\gamma''(t)\}^2 dt = \sum_{i=1}^{m} \ell(\alpha, \boldsymbol{\delta}, \boldsymbol{\gamma}; Y_i, \boldsymbol{s}_i, \widehat{\mathbf{x}}_i) - \frac{\lambda_\gamma}{2} \boldsymbol{\gamma}^T \mathbf{K} \boldsymbol{\gamma},$$

where $\mathbf{Y} = (Y_1, ..., Y_m)^T$, $\ell(\alpha, \boldsymbol{\delta}, \boldsymbol{\gamma}; Y_i, \boldsymbol{s}_i, \widehat{\mathbf{x}}_i)$ is the pseudo log-likelihood of $(\alpha, \boldsymbol{\delta}, \boldsymbol{\gamma})$ contributed by subject $i$ by treating $\epsilon_i^*$ in model (4) as independent and identically distributed as $N(0, \sigma_{\epsilon^*}^2)$, $\lambda_\gamma$ is a smoothing parameter controlling goodness-of-fit of the model to the data and roughness of $\gamma(t)$, and $\mathbf{K}$ is the non-negative definite matrix given in Section 3.1. The resulting penalized pseudo-likelihood estimator of $\gamma(t)$ is a smoothing spline.

This penalized pseudo-likelihood has essentially the same form as the one for $\mathbf{f}$ given in Section 3.1. Therefore, the estimator of $\boldsymbol{\gamma}$ has a similar mixed effect representation

$$\boldsymbol{\gamma} = \mathbf{1}_q \xi + \mathbf{B} \mathbf{u}, \tag{8}$$

where $\xi$ is a fixed effect, $\mathbf{u} \sim N(\mathbf{0}, \tau_\gamma \mathbf{I}_{r \times r})$ and $\tau_\gamma = \lambda_\gamma^{-1}$. Substituting this mixed effect representation of $\boldsymbol{\gamma}$ into model (4), we get the following working linear mixed model

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u} + \boldsymbol{\epsilon}^*, \tag{9}$$

where $\mathbf{X}$ is the new design matrix whose $i$th row is $(1, \boldsymbol{s}_i^T, \widehat{\mathbf{x}}_i^T \mathbf{C} \mathbf{1}_q)$, $\boldsymbol{\beta} = (\alpha, \boldsymbol{\delta}^T, \xi)^T$ are new fixed effects, $\mathbf{Z}$ is the new design matrix for random effects $\mathbf{u}$ whose $i$th row is $\widehat{\mathbf{x}}_i^T \mathbf{C} \mathbf{B}$ and $\boldsymbol{\epsilon}^* = (\epsilon_1^*, ..., \epsilon_m^*)^T$ is a $m \times 1$ residual error vector independent of $\mathbf{u}$ and its components are assumed to be independent and follow $N(0, \sigma_{\epsilon^*}^2)$. Note that the design matrix $\mathbf{Z}$ for random effects $\mathbf{u}$ is not block-diagonal. One can show that given $\tau_\gamma$ and $\sigma_{\epsilon^*}^2$, the BLUP estimators $\widehat{\boldsymbol{\beta}}$ and $\widehat{\boldsymbol{\gamma}} = \mathbf{1}_q \widehat{\xi} + \mathbf{B} \widehat{\mathbf{u}}$ under the linear mixed model (9) is the maximizer of $\ell_p(\alpha, \boldsymbol{\delta}, \boldsymbol{\gamma}; \mathbf{Y}, \boldsymbol{s}, \widehat{\mathbf{x}})$. The smoothing parameter $\tau_\gamma$ and the residual variance $\sigma_{\epsilon^*}^2$ can be estimated using the REML estimators under the linear mixed model (9) by treating both as variance components.

The covariance of $(\hat{\boldsymbol{\beta}}, \hat{\mathbf{u}})$ calculated under linear mixed model (9) leads to an approximate estimator of the covariance of $(\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\delta}}, \hat{\boldsymbol{\gamma}})$. The REML estimator $\hat{\sigma}_{\epsilon*}^2$ provides an estimator of the residual variance $\sigma_\epsilon^2$. The inverse of the information matrix of $(\tau_\gamma, \sigma_{\epsilon*}^2)$ from the REML likelihood provides an approximate variance estimator of the REML estimator of $\sigma_\epsilon^2$.

It is well recognized in literature that the smoothing parameters $\lambda_f$ and $\lambda_\gamma$ play important roles in estimating nonparametric functions $f(t)$ and $\gamma(t)$ and thus have to be well estimated. Here we proposed to estimate them simultaneously with other model parameters in a unified and convenient mixed model framework (model (7) for $\lambda_f$ and working model (9) for $\lambda_\gamma$). Simulation results presented in Section 6 indicate that this estimation strategy works well in estimating underlying true nonparametric functions $f(t)$ and $\gamma(t)$.

## 3.3   Hypothesis Testing for $\gamma(t)$

It is often of interest to test whether the nonparametric functional effect $\gamma(t)$ of the process $x(t)$ is a constant in the functional linear model (1), i.e., $H_0 : \gamma(t) = \gamma$. Under this null hypothesis model (1) reduces to the simple linear regression model in which the area under the process $x(t)$ is a covariate with $\gamma(t) = \gamma$ being its effect. In this section, we propose a simple score test for $H_0 : \gamma(t) = $ constant vs $H_1 : \gamma(t)$ is a smoothing spline.

The mixed effect representation of $\gamma(t)$ in (8) suggests that $H_0 : \gamma(t) = $ constant is equivalent to $H_0 :$ variance component $\tau_\gamma = 0$. Hence we can perform a score test for $\tau_\gamma$ in the mixed model (9) in the same spirit as Lin (1998) and Zhang and Lin (2003).

The score of $\tau_\gamma$ under the linear mixed model (9) is $U_{\tau_\gamma} = (\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}})^T \mathbf{Z}\mathbf{Z}^T(\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}})$, where $\hat{\boldsymbol{\beta}} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{Y}$ is the MLE of $\boldsymbol{\beta}$ under $H_0$. It is easy to show that $U_\tau$ has a distribution the same as that of weighted chi-squared random variables $\sum_{i=1}^q \psi_i \chi_{1i}^2$, where $\chi_{1i}^2$ are independent chi-squared random variables with one degree of freedom and $\psi_i$ are ordered non-zero eigen-values of $\mathbf{Z}^T\mathbf{P}\boldsymbol{\Sigma}_{\epsilon*}\mathbf{P}\mathbf{Z}$ and decay rapidly to zero, where $\mathbf{P} = \mathbf{I} - \mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T$ is the projection matrix under $H_0$. Calculations of $\psi_i$ are often computationally expensive

and the probability distribution of the associated weighted chi-squared random variables is prohibitive to calculate. Following Zhang and Lin (2003), we approximate the distribution of $U_{\tau_\gamma}$ by the distribution of a scaled chi-squared random variable $\kappa\chi^2_\nu$ by matching their first two moments. One can easily see that the mean and variance of $U_{\tau_\gamma}$ are $\mu = \mathrm{tr}(\mathbf{Z}^T\mathbf{P}\mathbf{\Sigma}_{\epsilon^*}\mathbf{P}\mathbf{Z})$ and $\theta = 2\mathrm{tr}\{(\mathbf{Z}^T\mathbf{P}\mathbf{\Sigma}_{\epsilon^*}\mathbf{P}\mathbf{Z})^2\}$ respectively. Matching the first two moments of $U_{\tau_\gamma}(\mathbf{Y};\sigma^2_\epsilon)$ and those of $\kappa\chi^2_\nu$ gives $\kappa = \theta/(2\mu)$ and $\nu = 2\mu^2/\theta$. Then a test statistic is calculated as $S = U_{\tau_\gamma}/\kappa$ and the evidence against $H_0$ is given by the p-value $= P[\chi^2_\nu \geq S]$.

Note that the mixed model (9) is conditional on $\mathbf{W}$, the above argument indicates that our testing procedure has approximately right statistical properties such as a correct type I error probability conditional on $\mathbf{W}$. Double expectation theorem then implies that our testing procedure should also have approximately right statistical properties unconditionally. This is verified by the simulation study given in Section 6.2.

# 4 Justifications for the Two-stage Nonparametric Regression Calibration Method

In this section, we provide some justifications for the two-stage nonparametric regression calibration (TS-NRC) method described in Sections 3.1 and 3.2. We also propose a bias-corrected estimator of $\sigma^2_\epsilon$ and variance estimators of the parameter estimators in model (1), including $\widehat{\gamma}(t)$. The TS-NRC method can be justified by showing that if $x_i(t)$ and $\gamma(t)$ are periodic smoothing splines, the true $Y_i|\mathbf{W}_i$ model is identical to the nonparametric regression calibration model (4) except that the errors $\epsilon^*_i$ are correlated.

Specifically, since $x_i(t)$ is a period smoothing spline, using the results at the beginning of Section 3, $x_i(t)$ is fully determined by its values at the knots $\mathbf{x}_i$. The smoothing spline mixed model representation of $(\mathbf{W}_i, \mathbf{x}_i)$ in Section 3.1 can be written as

$$\mathbf{W}_i = \mathbf{N}_i\mathbf{x}_i + \mathbf{e}_i, \quad \mathbf{x}_i = \mathbf{1}_q\eta + \mathbf{B}\mathbf{a} + \mathbf{1}_q b_i + \mathbf{B}\mathbf{a}_i,$$

where $\eta$ is a fixed effect parameter, $\mathbf{a} \sim \mathrm{N}(\mathbf{0}, \tau_f\mathbf{I}_{r\times r})$, $b_i \sim \mathrm{N}(\mathbf{0}, \sigma^2_b)$, $\mathbf{a}_i \sim \mathrm{N}(\mathbf{0}, \tau_g\mathbf{I}_{r\times r})$ and

$\mathbf{e}_i \sim \mathrm{N}(\mathbf{0}, \sigma_e^2 I_{n_i \times n_i})$ are independent. Denote by $\mathbf{x} = (\mathbf{x}_1^T, ..., \mathbf{x}_m^T)^T$, $\mathbf{N} = \mathrm{diag}\{\mathbf{N}_1, ..., \mathbf{N}_m\}$, $\mathbf{E}_x = \sigma_b^2 \mathbf{1}_q \mathbf{1}_q^T + \tau_g \mathbf{B} \mathbf{B}^T$, $\mathbf{E}_W = \mathbf{E}_x + \sigma_e^2 \mathbf{I}_{r \times r}$ and $\widetilde{\mathbf{B}} = \tau_f^{1/2}[\mathbf{B}^T, ..., \mathbf{B}^T]^T$. Then under the mixed model representation, $\mathbf{x} \sim N(\eta \mathbf{1}, \boldsymbol{\Sigma}_x)$, $\mathbf{W} \sim N(\eta \mathbf{1}, \boldsymbol{\Sigma}_W)$, where $\boldsymbol{\Sigma}_x = \mathrm{cov}(\mathbf{x}) = \mathrm{diag}\{\mathbf{E}_x, ..., \mathbf{E}_x\} + \widetilde{\mathbf{B}}\widetilde{\mathbf{B}}^T$, and $\boldsymbol{\Sigma}_W = \mathrm{cov}(\mathbf{W}) = \mathbf{N}\mathrm{diag}\{\mathbf{E}_W, ..., \mathbf{E}_W\}\mathbf{N}^T + \mathbf{N}\widetilde{\mathbf{B}}\widetilde{\mathbf{B}}^T\mathbf{N}^T$. Since $\mathbf{W}$ and $\mathbf{x}$ are jointly normal, $\mathbf{x}|\mathbf{W}$ follows a normal distribution with mean $\mathrm{E}(\mathbf{x}|\mathbf{W}) = \widehat{\mathbf{x}}$ and covariance $\mathrm{cov}(\mathbf{x}|\mathbf{W})$, which can be equivalently written as

$$\mathbf{x}_i = \widehat{\mathbf{x}}_i + \mathbf{U}_i,$$

where $\widehat{\mathbf{x}}_i$ is the estimator of $\mathbf{x}_i$ given in the first stage, $\mathbf{U} = (\mathbf{U}_1, \cdots, \mathbf{U}_m)^T \sim \mathrm{N}(\mathbf{0}, \boldsymbol{\Sigma}_U)$ is independent of $\hat{\mathbf{x}}_i$, and $\boldsymbol{\Sigma}_U = \mathrm{cov}(\mathbf{x}|\mathbf{W}) = \boldsymbol{\Sigma}_x - \boldsymbol{\Sigma}_x \mathbf{N}^T \boldsymbol{\Sigma}_W^{-1} \mathbf{N} \boldsymbol{\Sigma}_x$. Note that $\boldsymbol{\Sigma}_U$ is not a block diagonal matrix and hence $\mathbf{U}_i$ and $\mathbf{U}_{i'}$ $(i \neq i')$ are correlated. Also note that the direct inverse of $\boldsymbol{\Sigma}_W$ is prohibitive because of its usually large dimension (i.e., $\sum_{i=1}^m n_i$). However, we can exploit the special structure of $\boldsymbol{\Sigma}_W$, and calculate $\boldsymbol{\Sigma}_W^{-1}$ by only inverting a matrix with the same dimension as $\mathbf{B}$ and the $n_i \times n_i$ matrices $\mathbf{N}_i \mathbf{E}_W \mathbf{N}_i^T$.

Since $x_i(t)$ is a periodic cubic smoothing spline, we have

$$x_i(t) = \mathbf{x}_i \mathbf{c}(t) = \widehat{\mathbf{x}}_i \mathbf{c}(t) + \mathbf{U}_i \mathbf{c}(t) = \widehat{\mathbf{x}}_i(t) + \mathbf{U}_i(t), \tag{10}$$

where $\mathbf{U}_i(t)$ is a mean zero stochastic process whose covariance is determined by $\mathbf{c}^T(t)\boldsymbol{\Sigma}_U\mathbf{c}(t)$. Note that $\mathbf{U}_i(t)$ and $\mathbf{U}_{i'}(t)$ $(i \neq i')$ are correlated stochastic processes. Substituting expression (10) into model (1), we have the true $\mathbf{Y}|\mathbf{W}$ model:

$$Y_i = \alpha + \boldsymbol{s}_i^T \boldsymbol{\delta} + \widehat{\mathbf{x}}_i^T \mathbf{C} \boldsymbol{\gamma} + \boldsymbol{\gamma}^T \mathbf{C} \mathbf{U}_i + \epsilon_i = \alpha + \boldsymbol{s}_i^T \boldsymbol{\delta} + \widehat{x}_i \mathbf{C} \boldsymbol{\gamma} + \epsilon_i^*, \tag{11}$$

where $\epsilon_i^* = \boldsymbol{\gamma}^T \mathbf{C} \mathbf{U}_i + \epsilon_i$, $\boldsymbol{\epsilon}^* = (\epsilon_1^*, ..., \epsilon_m^*)^T$, $\boldsymbol{\epsilon}^* \sim N(0, \boldsymbol{\Sigma}_{\epsilon^*})$, and $\boldsymbol{\Sigma}_{\epsilon^*} = \mathbf{G}\boldsymbol{\Sigma}_U\mathbf{G}^T + \sigma_\epsilon^2 \mathbf{I}_{m \times m}$ and $\mathbf{G} = \mathrm{diag}\{\boldsymbol{\gamma}^T \mathbf{C}, ..., \boldsymbol{\gamma}^T \mathbf{C}\}$.

A comparison of the true $\mathbf{Y}|\mathbf{W}$ model (11) with the nonparametric regression calibration model (4) shows that the nonparametric regression calibration model at the second stage correctly specifies the mean structure of the $\mathbf{Y}|\mathbf{W}$ model, but assumes the errors $\epsilon_i^*$ are

independent across subjects. In other words, it estimates the smoothing spline estimator of $\boldsymbol{\gamma}$ by maximizing the following penalized pseudo-likelihood function by treating $\epsilon_i^* \sim \mathrm{N}(0, \sigma_{\epsilon*}^2)$ as independent errors

$$-\frac{1}{2\sigma_{\epsilon*}^2} \sum_{i=1}^{m} (Y_i - \alpha - \boldsymbol{s}_i^T \delta - \widehat{\mathbf{x}}_i^T \mathbf{C}\boldsymbol{\gamma})^2 - \frac{\lambda_\gamma}{2} \boldsymbol{\gamma}^T \mathbf{K}\boldsymbol{\gamma}.$$

The results of Lin et al. (2003) shows that such a working independent smoothing spline estimator of $\gamma(t)$ is consistent. However the estimator $\hat{\sigma}_{\epsilon*}^2$ calculated under the linear mixed model assuming working independence is a biased estimator of $\sigma_\epsilon^2$. Examination of equation (11) shows that a simple biased corrected moment estimator of $\sigma_\epsilon^2$ can be constructed as

$$\hat{\sigma}_\epsilon^2 = \hat{\sigma}_{\epsilon*}^2 - \frac{1}{m} \sum_{i=1}^{m} \widehat{\boldsymbol{\gamma}}^T \mathbf{C} \boldsymbol{\Sigma}_{U_i} \mathbf{C} \widehat{\boldsymbol{\gamma}}, \tag{12}$$

where $\widehat{\mathbf{G}} = \mathrm{diag}\{\widehat{\boldsymbol{\gamma}}^T \mathbf{C}, ..., \widehat{\boldsymbol{\gamma}}^T \mathbf{C}\}$ and $\boldsymbol{\Sigma}_{U_i} = \mathrm{cov}(\mathbf{U}_i)$.

The naive variance estimators of $\widehat{\boldsymbol{\beta}}$ and $\widehat{\boldsymbol{\gamma}}$ calculated under the linear mixed model (9) assuming working independence ignores the correlation in $\epsilon^*$ and may not capture the true variation in $\widehat{\boldsymbol{\beta}}$ and $\widehat{\boldsymbol{\gamma}}$. Write $\widehat{\boldsymbol{\beta}} = \mathbf{A}_\beta \mathbf{Y}$ and $\widehat{\boldsymbol{\gamma}} = \mathbf{A}_\gamma \mathbf{Y}$. Then their covariance matrices can be estimated as $\widehat{\mathrm{cov}}(\widehat{\boldsymbol{\beta}}) = \mathbf{A}_\beta \widehat{\boldsymbol{\Sigma}}_{\epsilon*} \mathbf{A}_\beta^T$ and $\widehat{\mathrm{cov}}(\widehat{\boldsymbol{\gamma}}) = \mathbf{A}_\gamma \widehat{\boldsymbol{\Sigma}}_{\epsilon*} \mathbf{A}_\gamma^T$, where $\widehat{\boldsymbol{\Sigma}}_{\epsilon*} = \widehat{\mathbf{G}} \boldsymbol{\Sigma}_U \widehat{\mathbf{G}}^T + \hat{\sigma}_\epsilon^2 \mathbf{I}_{m \times m}$. Note that such covariance estimators of $\widehat{\boldsymbol{\beta}}$ and $\widehat{\gamma}(t)$ accounts for the variability of the nonparametric regression calibration estimators $\hat{x}_i(t)$ resulted from the first stage estimation. Simulation studies conducted in Section 6 show that the corrected estimator $\hat{\sigma}_\epsilon^2$ in (12) is virtually unbiased and the above covariance estimates of $\widehat{\boldsymbol{\beta}}$ and $\widehat{\boldsymbol{\gamma}}$ estimate the true variances of $\widehat{\boldsymbol{\beta}}$ and $\widehat{\gamma}(t)$ well.

# 5  Application to the Hormone Data

In this section, we use the two-stage nonparametric regression calibration method developed in Sections 3 and (4) to analyze the SWAN data introduced in Section 1. Denote by $Y_i$ total hip BMD (g/cm$^2$), $\mathrm{AGE}_i$ age in years, $\mathrm{BMI}_i$ body mass index (kg/m$^2$), $x_i(t)$ the true FSH profile and $W_{ij}$ the observed FSH level at the $j$th time point during a standardized

menstrual cycle for woman $i$ ($i=1, \cdots, m=628$). Cycle lengths varied considerably among women, with the mean equal to 30 days and a range between 14 and 50 days. Following the common practice in hormone studies (Zhang, et al., 1998), we standardized the cycle length for each woman to 28 days. The average age of the participants was 47 years, the average hip BMD was 0.94 (g/cm$^2$) and the average FSH level was 35.8(mIU/ml).

We considered the following partial functional linear model to examine the effect of the FSH profile on BMD, adjusting for age and body mass index

$$Y_i = \alpha + \text{AGE}_i \delta_1 + \text{BMI}_i \delta_2 + \int_0^{28} x_i(t)\gamma(t)dt + \epsilon_i, \tag{13}$$

where $\alpha, \delta_1, \delta_2$ are regression coefficients, $\gamma(t)$ is a periodic smooth function in $[0, 28]$ characterizing the effect of the FSH profile on BMD, and $\epsilon_i \sim \text{N}(0, \sigma_\epsilon^2)$ are independent errors. The subject-specific observed FSH values $W_{ij}$ were obtained at irregular time points and were assumed to relate to the subject-specific true profile $x_i(t)$ by equation (5).

We used the first stage functional mixed model estimation procedure discussed in Section 3.1 to obtain subject-specific smoothing spline estimates of the $x_i(t)$. For numerical reasons, BMD was multiplied by 100 and FSH was divided by 100. Figure 1 presents the estimated subject-specific profiles $\hat{x}_i(t)$ for 4 randomly chosen women, plotted together with their raw FSH values and the estimated population profile. Figure 1 shows that the observed individual raw data have considerable between- and within-woman variability. The estimated subject-specific profiles tracked each woman's raw data very well and lied between the estimated population profile and the individual raw data. The estimated population profile is consistent with the biological knowledge of FSH functioning. During menstruation, FSH levels increase and promote growth and selection of follicles. FSH begins to decrease followed by a secondary rise as a dominant follicle is selected to continue to develop. After surging with luteinizing hormone (LH) about one day before ovulation, FSH decreases to the lowest level. It then gradually increases at the end of a cycle, with the initiation of the next cycle.

Figure 1 shows that there is complicated and tremendous between-woman variation in

FSH profiles. The variances of $\mathbf{b}_i$ and $\mathbf{a}_i$ in the mixed model representation (7) of $x_i(t)$ were estimated as $\hat{\sigma}_b^2 = 0.12$ and $\hat{\tau}_g = 0.80$, and the average of the $\text{var}\{x_i(t)\}$ was estimated as 0.14. The measurement error variance $\sigma_e^2$ in model (7) was estimated as $\hat{\sigma}_e^2 = 0.06$, which measured the departure of the FSH measurements from the subject-specific curves. These results indicate that the between-woman variation in the hormone profile dominates the total FSH variation.

At the second stage of the proposed TS-NRC method, we replaced $x_i(t)$ in the functional linear model (13) by its subject-specific estimate $\hat{x}_i(t)$ obtained from the first stage. We then calculated the maximum pseudo-likelihood estimate of $\boldsymbol{\beta}$ and the periodic smoothing spline estimate $\gamma(t)$ using the linear mixed model representation (11). We estimate the standard errors of $\hat{\boldsymbol{\delta}}$ and $\hat{\gamma}(t)$ using the method described at the end of (4). Figure 2(a) presents the estimated periodic cubic smoothing spline $\hat{\gamma}(t)$ and its 95% point-wise confidence intervals. The estimated curve $\hat{\gamma}(t)$ was negative in most of the interval $[0, 28]$, implying the overall negative association between BMD and the FSH profile. This result was consistent with the finding of negative association between BMD and serum FSH from a cross-sectional study (Sowers et al. 2003).

To compare our smoothing spline regression calibration method with James' (2002) regression spline approach, we estimated $\gamma(t)$ using his approach with 2, 3 and 4 equally spaced interior knots assuming periodic cubic regression splines. The results are given in Figure 2(b). Although the estimates with 2 interior knots had an overall similar shape to the two-stage estimate $\hat{\gamma}(t)$, the values of the regression spline estimates were much greater in some regions. The estimates of $\gamma(t)$ with 3 and 4 interior knots exhibited an additional mode and was dramatically different from the other estimates. These results indicate that regression splines estimates are sensitive to the choice of the number of knots. James (2002) did not provide a data driven method to estimate the number and the locations of the knots. See Section 7 for more discussions. In Section 6, we use simulations to more systematically

compare the performance of the smoothing spline method and the regression spline method.

The estimated age coefficient was $\widehat{\delta}_1 = 0.16$ (SE=0.18), and the estimated BMI coefficient was $\widehat{\delta}_2 = 1.37$ (SE=0.06). These results indicated that there was no significant age effect on BMD, and the women with higher BMI values had higher BMD values, characteristics well recognized in the epidemiological literature. The uncorrected estimate of the residual variance of $\sigma_\epsilon^2$ was 116.51 and the corrected estimate in (12) was 116.50. Because $\widehat{\gamma}(t)$ was much smaller than $\widehat{\sigma}_\epsilon^2$, the correction given in equation (12) did not have much impact on the estimation of $\sigma_\epsilon^2$. However, the simulation results in Section 6 show that when the measurement error becomes larger, the uncorrected estimate can be quite biased, while the corrected estimator still performs well.

The relatively flat 95% point-wise confidence intervals of the nonparametric regression calibration estimate $\widehat{\gamma}(t)$ given in Figure 2(a) motivated testing the null hypothesis $H_0 : \gamma(t)$ is a constant, i.e., $\gamma(t)=\gamma$. We applied the score test developed in Section 3.3. The test statistic was $S = 2.57$ with 1.80 degrees of freedom (p-value = 0.24). Therefore, it is reasonable to assume that $\gamma(t)$ in model (13) is a constant $\gamma$. Under this assumption, we re-fit model (13) with the area under the curve $\int_0^{28} x_i(t)dt$ as a true unobserved covariate, and obtained $\widehat{\gamma} = -0.091$ with the estimated standard error 0.046. This indicates a strong negative relationship between BMD and cumulative FSH hormone concentrations characterized by the area under the profile. The estimates of the other parameters remained almost unchanged. The estimated $R^2$ for both models were very close (0.4334 vs. 0.4320), and the residual plots (not shown) did not reveal any unusual patterns, indicating a reasonable fit of either model to the data.

We also included actual individual cycle length $T_i$ in model (13) to assess the potential effect of cycle length on BMD. The estimated regression coefficient was -0.004 with estimated standard error 0.068 (p-value=0.95), indicating that it may be reasonable to use the standardized cycle length in model (13).

18

We then conducted a naive analysis where the area under the profile $x_i(t)$ was estimated using raw FSH observations from woman $i$ with simple interpolation described in the last paragraph of Section 2. The analysis gave almost identical results to those from the two-stage procedure and hence was not shown. This is because the measurement error variance $\sigma_e^2$ in FSH measurements is much smaller than the between-woman variance of the FSH profiles, implying negligible measurement error in the naive estimates of the areas under the profiles. In fact, the Pearson correlation coefficient between the two areas under the profile using naive approach and two-stage approach was almost equal to 1. The simulation results show that if the measurement error is more prominent, the bias of the naive method would become significant.

# 6    Simulation Studies

## 6.1    Simulation Study for the Estimation Procedure

In this section, we present results from simulation studies designed to evaluate the finite sample performance of the two-stage nonparametric regression calibration method developed in Sections 3.1 and 3.2. Two sample sizes $m = 200$ and $m = 400$ were considered.

Let the $t_j$'s be 40 equally spaced time points in $[0, 10)$. For each subject $i = 1, ..., m$, longitudinal covariate data $W_{ij}$ were generated from

$$W_{ij} = f(t_j) + g_i(t_j) + e_{ij}, \quad j = 1, 2, ..., 40,$$

where $f(t) = cos(2\pi t/10)$ is a periodic population profile with the period length equal to $T = 10$, $g_i(t)$ is a periodic cubic smoothing spline characterizing subject-specific deviations from $f(t)$ whose values at the distinct knots $\{t_j\}$ are given by $\mathbf{1}_{40}b_i + \mathbf{B}\mathbf{a}_i$, where $\mathbf{B}$ is the matrix defined in Section 3.1, $b_i \sim \mathrm{N}(0, 0.8^2)$, $\mathbf{a}_i \sim \mathrm{N}(0, 0.5^2\mathbf{I}_{39\times39})$, and $e_{ij} \sim \mathrm{N}(0, 0.7^2)$ are the measurement errors. The values of the variance components were chosen in such a way that the measurement error variance was comparable to the between-subject variance of the

subject-specific profiles $x_i(t) = cos(2\pi t/10) + g_i(t)$. The primary response $Y_i$ was generated from the following functional linear model

$$Y_i = \alpha + \int_0^T x_i(t)\gamma(t)dt + \epsilon_i, \tag{14}$$

where $\alpha = 2$, $\gamma(t) = sin(2\pi t/10)$ and $\epsilon_i \sim \text{N}(0, \sigma_\epsilon^2 = 4)$. Since $g_i(t)$ is assumed to be a periodic cubic smoothing spline, the integral $\int_0^T x_i(t)\gamma(t)dt$ can be evaluated in a closed form once the values of $g_i(t)$ at the distinct knots are given.

One hundred data sets were generated, and the TS-NRC method developed in Section 3 was applied to each data set to estimate $\gamma(t)$ in $[0, T = 10]$. To compare the two-stage procedure with James' (2002) regression spline approach, these generated data sets were re-analyzed using his approach with 3 and 5 equally spaced interior knots.

Figure 3 presents the true $\gamma(t)$, the average of the estimates $\hat{\gamma}(t)$, and their biases, based on 100 simulation runs for two sample sizes $m = 200$ and $m = 400$, using the TS-NRC method and James' (2002) periodic cubic regression spline approach with 3 and 5 interior knots. For fair visual comparison, all biases were plotted on the same scale. The TS-NRC method yielded an almost unbiased estimate of the underlying true function, and the bias was smaller than that of the regression spline method. James' periodic cubic regression spline approach with 3 interior knots estimated $\gamma(t)$ reasonably well when the sample size was 400. The estimates in other settings had considerably higher biases. The results indicate that the regression spline method is sensitive to the choice of the number of knots and their allocations. Due to the assumed parametric forms, regression splines are more likely to result in over-fitting and uncover artificial bumps and valleys. Carroll, et al (2004) also found smoothing splines performed better than regression splines in other settings. Another reason regression splines perform worse in our setting is numerical. For example, with 5 interior knots, a $6 \times 6$ unstructured variance matrix for the random coefficients of the periodic cubic basis functions has to be updated in each EM iteration. These findings of regression splines are consistent with the simulation study results reported in James (2002).

Figure 4(a) presents the estimated and empirical standard errors of the smoothing spline estimates of $\gamma(t)$ for sample size $m = 200$ and $m = 400$ using our two-stage estimation procedure. The estimated SEs were calculated using the procedure described at the end of Section (4). The estimated standard errors agreed very well with their empirical counterparts, and the agreement became better as the sample size increased. This indicated our standard error estimation method works well. Figure 4(b) presents the empirical standard errors of the estimates of $\gamma(t)$ using James' (2002) regression spline approach. It is clear from Figure (3) and Figures 4(a) and 4(b) that our smoothing spline estimate not only has smaller bias but also smaller variance compared to James' (2002) regression spline estimate. Figure 4(c) presents the empirical coverage probabilities of 95% point-wise confidence intervals of $\gamma(t)$ for sample sizes $m = 200$ and $m = 400$. They were overall very close to the nominal level (95%). Although the coverage probabilities for sample size $m = 200$ fluctuated at times, they became more stable and closer to the nominal level (95%) for a larger sample size ($m = 400$).

Table 1 presents the results of the estimates of the other model parameters in (14) using the TS-NRC method. The estimates of the regression coefficients ($\alpha$) were virtually unbiased. As expected, the uncorrected estimate of $\hat{\sigma}_\epsilon^2$ in the second stage over-estimated the true value. However its corrected estimate using (12) was almost unbiased. The estimated standard errors agreed well with their empirical counterparts. The empirical coverage probabilities of the 95% confidence intervals of the regression coefficient estimates and the corrected estimate of $\hat{\sigma}_\epsilon^2$ were close to the nominal level, especially when the sample size is $m = 400$.

We repeated the simulation by reducing the number of repeated measures of $W_{ij}$ to a half of the current setting, i.e., $n_i = 20$ instead of 40 and doubling the variance of their measurement errors $e_{ij}$'s, i.e., $\sigma_e^2 = 1$. The results are similar and are omitted.

## 6.2    Simulation Study of the Testing Procedure

In this section, we report a simulation study to evaluate the performance of the testing procedure developed in Section 3.3 for $H_0 : \gamma(t) = \gamma$, including its empirical sizes and powers. The simulation design was similar to that used in Section 6.1, except that we only considered the case where the sample size was $m = 200$ due to computational consideration. We considered a series of functions $\gamma_d(t) = 1 + d \times sin(2\pi t/10)/20$ indexed by $d = 0, 1, 2, 3, 4$ and 5. Five hundred simulation runs were done for $d = 0$ to more accurately evaluate the empirical sizes and 100 runs were done for $d > 0$ to evaluate the empirical powers.

Figure 5(a) presents the empirical distribution (histogram) of the test statistic $S$ obtained under $d = 0$, superimposed with the estimated chi-squared density. The estimated density matched the empirical distribution reasonably well. This finding was further supported by the nearly uniform distribution of the p-value for testing $H_0 : \gamma(t) = $ constant under $d = 0$ (Figure 5(b)). Table 2 gives the empirical sizes and powers of the testing procedure for three commonly used nominal levels 0.01, 0.05 and 0.1. The empirical sizes were very close to the corresponding nominal levels, re-assuring the evidence in Figure 5 (b). The empirical powers quickly increased to about 1 when $d$ increased from 0 to 5, indicating that the test had a good statistical power to detect a non-constant functional covariate effect. Note that the non-constant function $\gamma_5(t)$ is equal to only one-fourth of $\gamma(t)$ used in Section 6.1.

When $d = 0$, $\gamma_0(t) = \gamma = 1$ is the covariate effect of the area under the true profile $x_i(t)$. Table 3 compares the estimates of the regression coefficients and the residual variance using the TS-NRC method with the naive method, where the area under the true profile $x_i(t)$ was calculated using the error-prone observations $W_{ij}$ with simple interpolation discussed in the last paragraph of Section 2. The regression coefficient estimates and the corrected residual variance estimate using the TS-NRC method were all unbiased. The estimated standard errors agreed well with their empirical counterparts, and the 95% confidence intervals had right coverage probabilities. Although the estimated intercept using the naive method per-

formed well, its estimate of the key regression coefficient $\gamma$ had a downward bias as large as its standard error, resulting in lower than the nominal coverage probability of the confidence interval. This is because the naive estimate of the area under the true profile $x_i(t)$ calculated using the error-prone observations $W_{ij}$ can itself be viewed as a covariate measured with error, and gives an attenuated regression coefficient estimate of the true covariate. This also explains the upward bias of the naive estimate of the residual variance $\sigma_\epsilon^2$.

Similarly, we repeated the simulation for the testing procedure using $n_i = 20$ and $\sigma_e^2 = 1$. The results are similar and hence are omitted.

# 7 Discussion

In this paper we proposed a two-stage nonparametric regression calibration method in semiparametric functional linear models. In the first stage, we used periodic cubic smoothing splines to estimate the subject-specific functional covariates nonparametrically. At the second stage, we fitted the semiparametric functional linear model using a smoothing spline under working independence by plugging in the nonparametric regression calibration estimates of the subject-specific functional covariates obtained from the first stage. A key feature of our approach is that we cast nonparametric estimation at both stages in a unified mixed model framework and estimate all model parameters including smoothing parameters simultaneously at each stage. Unlike the traditional functional data analysis approaches, our approach allows for unbalanced designs of longitudinal covariates. Theoretical justifications were given for the validity of the two-stage nonparametric regression calibration method. Compared to the existing methods, our approach does not require estimating a large number of parameters and is quite robust for estimating subject-specific profiles in the first stage and the functional covariate effect in the second stage. We also proposed a score test for testing whether or not the functional linear model can be reduced to the popular model with the "area under the curve" as a covariate. Our simulation results indicated that the

23

proposed procedures performed well.

Our simulation results show that the proposed TS-NRC method using smoothing spline mixed models perform better than the existing regression spline method. James (2002) did not provide a data drive method to estimate the number of knots in $x_i(t)$ and $\gamma(t)$. It would be of future research interest to develop a data-driven regression spline knot selection procedure and compare the performance of such an adaptive regression spline method to the proposed smoothing spline TS-NRC method. However, for the regression spline method, since the optimal number of knots needed for estimating $x_i(t)$ might differ from that for estimating $\gamma(t)$, the numbers of knots for estimating $x_i(t)$ and $\gamma(t)$ need to be allowed to be different and selected differently. A unified selection criterion such as CV or GCV might be difficult to develop in this context. The results of Carroll, et al (2004) found in simple settings, even with the number of knots estimated using CV or GCV for the regression spline method, it still underperforms the smoothing spline method. An alternative approach is to develop a penalized spline (P-spline) method (Wand, et al, 2003) to fit the proposed functional linear mixed model. This would be of future research interest as well.

The proposed functional linear mixed model is applicable to many other disciplines besides hormone research. For example, we may be interested in assessing the effect of the longitudinal profile of a biomarker on a scalar clinical endpoint. Furthermore, the clinical endpoint might not be continuous. It is hence of future interest to extend the proposed method to discrete outcomes.

# ACKNOWLEDGMENTS

# REFERENCES

Carroll, R. J., Ruppert, D. and Stefanski, L. A. (1995). *Measurement Error in Nonlinear Models.* London: Chapman and Hall.

Carroll, R. J., Hall, P., Apanasovich, T. V., and Lin, X. (2004). Histospline method in nonparametric regression models with application to longitudinal/clustered data. *Statistical Sinica* **14**, 649-674.

Fitzmaurice, G. M., Laird, N. M. and Ware, J. H. (2004). *Applied Longitudinal Analysis.* New Jersey: John Wiley & Sons.

Guo, W. (2002). Functional Mixed Effects Models. *Biometrics* **58**, 121-128.

Green, P. J. and Silverman B. W. (1994). *Nonparametric Regression and Generalized Linear Models.* London: Chapman and Hall.

James, G. M. (2002). Generalized Linear Models with Functional Predictors. *Journal of Royal Statistical Society, B* **64**, 411-432.

Lin, X. (1997). Variance Components Testing in Generalized Linear Models with Random Effects. *Biometrika* **84**, 309-326.

Lin, X., Wang, N., Welsh, A. H. and Carroll, R. J. (2004). Equivalent Kernels of Smoothing Splines in Nonparametric Regression for Clustered/Longitudinal Data. *Biometrika* **91**, 177-193.

Ramsay, J. O. and Silverman, B. W. (1997). *Functional Data Analysis.* New York, Springer.

Rice, J. A. and Wu, C. O. (2001). Nonparametric Mixed Effects Models for Unequally Sampled Noisy Curves. *Biometrics* **57**, 253-259.

Sowers, M. R., Crawford, S., Sternfeld B., Morgenstein, D., Gold, E., Greendale G., Evans, D., Neer, R., Matters, K., Sherman, S., Lo, A., Weiss, G. and Kelsey, J. (2000).

Design, Survey Sampling and Recruitment Methods of SWAN: A Multi-center, Multi-ethnic, Community-based Cohort Study of Women and the Menopausal Transition. In: *Menopause: Biology and Pathobiology.* Eds: Wren, J., Lobo, R. A., Kelsey, J. and Marcus, R. Academic Press.

Sowers MF, Finkelstein J, Ettinger B, Bondarenko I, Neer R, Cauley J, Sherman S, Greendale G. (2003). The Association of Endogenous Hormone Concentrations and Bone Mineral Density Measures in Pre- and Perimenopausal Women of Four Ethnic Groups: SWAN. *Osteporos International* **14**, 44-52.

Zhang, D., Lin, X., Raz, J. and Sowers, M. F. (1998). Semiparametric Stochastic Mixed Models for Longitudinal Data. *Journal of the American Statistical Association* **93**, 710-719.

Zhang, D., Lin, X. and Sowers, M. F. (2000). Semiparametric Regression for Periodic Longitudinal Hormone Data from Multiple Menstrual Cycles. *Biometrics* **56**, 31-39.

Zhang, D. and Lin, X. (2003). Hypothesis Testing in Semiparametric Additive Mixed Models. *Biostatistics* **4**, 57-74.

**Table 1**

*Simulation results for model (14): Mean and EMP-SE are the average of the parameter estimates and the empirical SEs; EST SE is the average of model-based estimated standard errors; CP is the empirical coverage probability of the 95% confidence intervals*

| Par. | True | Method | $m = 200$ | | | | $m = 400$ | | | |
|------|------|--------|------|--------|--------|------|------|--------|--------|------|
| | | | Mean | EMP SE | EST SE | CP | Mean | EMP SE | EST SE | CP |
| $\alpha$ | 2.00 | | 2.00 | 0.31 | 0.32 | 0.92 | 2.02 | 0.23 | 0.22 | 0.96 |
| $\sigma_\epsilon^2$ | 4.00 | Uncorrected | 4.53 | 0.50 | 0.46 | 0.82 | 4.59 | 0.33 | 0.33 | 0.57 |
| | | Corrected | 3.96 | 0.51 | 0.46 | 0.88 | 4.02 | 0.33 | 0.33 | 0.96 |

**Table 2**

*Empirical sizes and powers of the testing procedure from the simulation study*

| Nominal level | Empirical size | | | Power | | |
|------|------|------|------|------|------|------|
| | | $d = 1$ | $d = 2$ | $d = 3$ | $d = 4$ | $d = 5$ |
| 0.01 | 0.01 | 0.05 | 0.09 | 0.28 | 0.66 | 0.94 |
| 0.05 | 0.06 | 0.10 | 0.26 | 0.56 | 0.87 | 0.97 |
| 0.1 | 0.11 | 0.19 | 0.40 | 0.73 | 0.93 | 0.98 |

Empirical size was based on 500 simulation runs.

Empirical powers were based on 100 simulation runs

**Table 3**

*Simulation results under the area under the curve model assuming a constant $\gamma$: Mean and EMP-SE are the average of the parameter estimates and the empirical SEs; EST SE is the average of model-based estimated standard errors; CP is the empirical coverage probability of the 95% confidence intervals*

| Par. | True | Method | Naive | | | | Two-stage | | | |
|------|------|--------|------|--------|--------|------|------|--------|--------|------|
| | | | Mean | EMP SE | EST SE | CP | Mean | EMP SE | EST SE | CP |
| $\alpha$ | 2.00 | | 2.00 | 0.16 | 0.16 | 0.95 | 2.00 | 0.16 | 0.16 | 0.95 |
| $\gamma$ | 1.00 | | 0.98 | 0.02 | 0.02 | 0.87 | 1.00 | 0.02 | 0.02 | 0.96 |
| $\sigma_\epsilon^2$ | 4.00 | Uncorrected | 5.23 | 0.50 | | | 5.20 | 0.50 | 0.52 | 0.33 |
| | | Corrected | | | | | 4.00 | 0.50 | 0.52 | 0.94 |

# LIST OF FIGURES

**Figure 1**. FSH levels (divided by 100) from four randomly selected women plotted against days in a standardized menstrual cycle, superimposed with the estimated population and individual FSH profiles: —— population profile; **- - - -** individual profile.

**Figure 2.** (a) Estimated nonparametric function $\widehat{\gamma}(t)$ in model (13) and its 95% confidence intervals: —— estimated $\widehat{\gamma}(t)$; **- - - -** confidence intervals

(b) Estimated function $\widehat{\gamma}(t)$ using periodic cubic regression splines: —— estimate with 2 interior knots; **- - - -** estimate with 3 interior knots; $----$ estimate with 4 interior knots;

**Figure 3**. The true $\gamma(t)$ and the average of the estimated nonparametric functions $\widehat{\gamma}(t)$ (Figure (a)), and their biases (Figure (b)) for two sample sizes from the simulation study: ——, true; **- - - -**, $m = 200$; $----$, $m = 400$. The first row is based on the two-stage NRC method, the second and third rows are based on the periodic regression spline approach with 3 and 5 interior knots.

**Figure 4**.

(a) Empirical SEs and the average estimated standard errors of the smoothing spline estimates $\widehat{\gamma}(t)$s for two sample sizes from the simulation study: ——, empirical SE; **- - - -**, estimated SE;

(b): Empirical SEs of the regression spline estimates $\widehat{\gamma}(t)$s with 3 and 5 knots for two sample sizes from the simulation study: ——, 3 knots; **- - - -**, 5 knots;

(c) Empirical coverage probabilities of the 95% confidence intervals of $\gamma(t)$ for two sample sizes from the simulation study: ——, nominal level; **- - - -**, $m = 200$; $----$, $m = 400$

**Figure 5**. (a) Empirical distribution of the score test statistic $S$ in Section 6.2 for testing $H_0 : \gamma(t)$=constant, from the simulation study, superimposed with its estimated density function.

(b) Empirical distribution of the P-value of the test statistic $S$ in Section 6.2 for testing $H_0 : \gamma(t)$ =constant, from the simulation study.
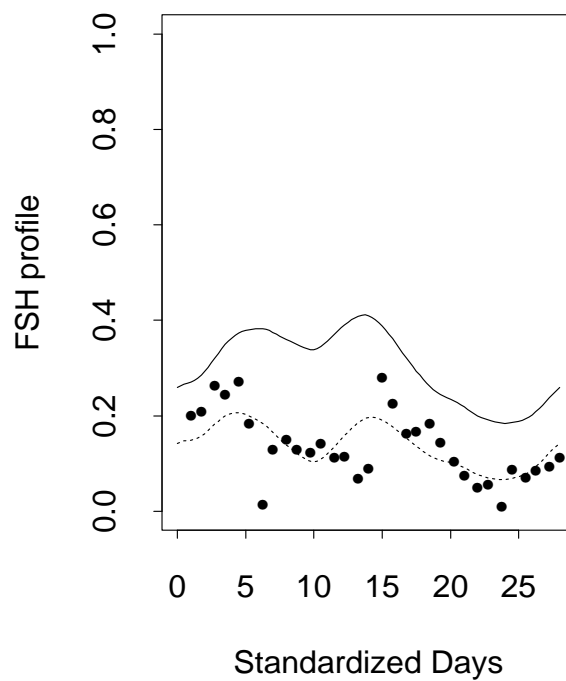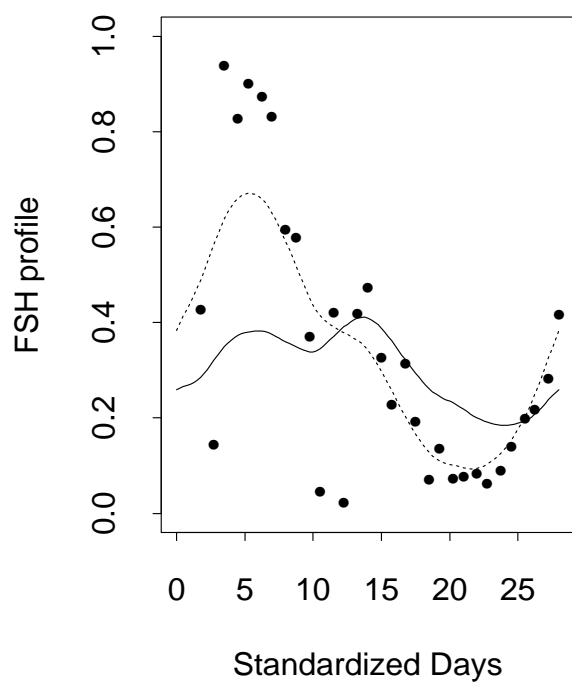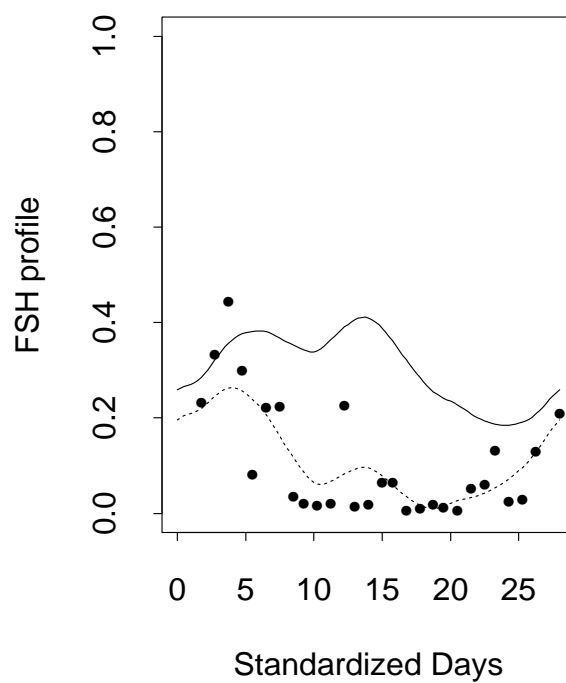
Figure 1(a)
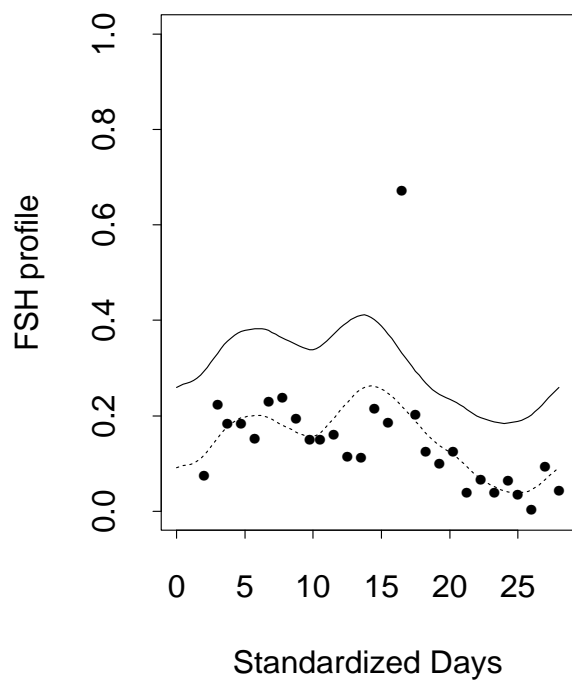
Figure 1(b)

Figure 1(c)

Figure 1(d)

# Figure 2(a)



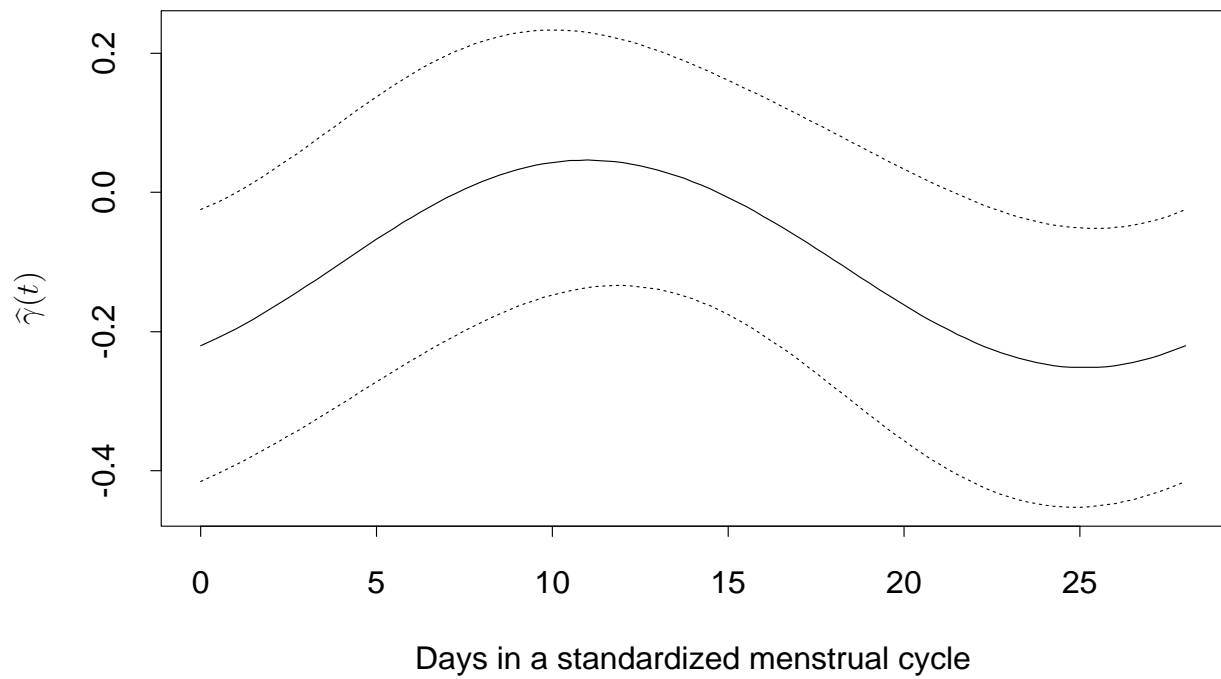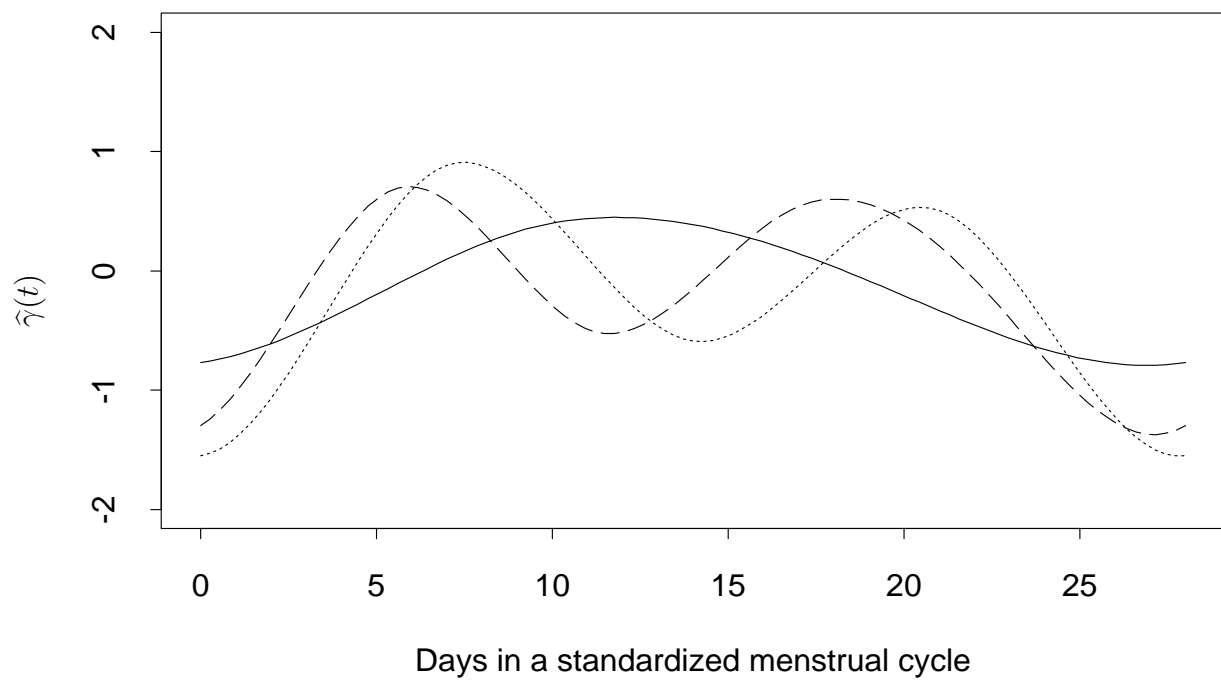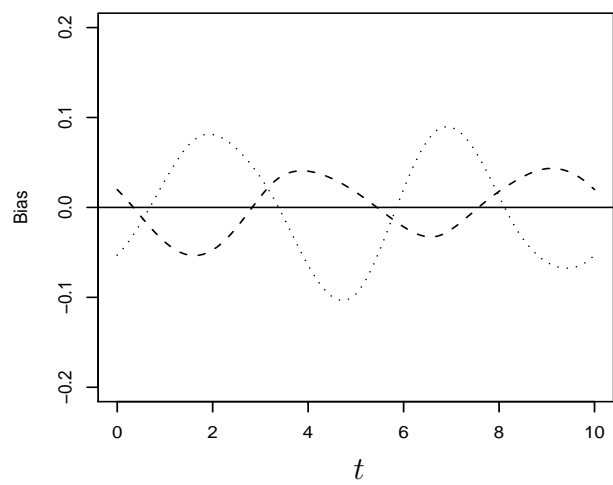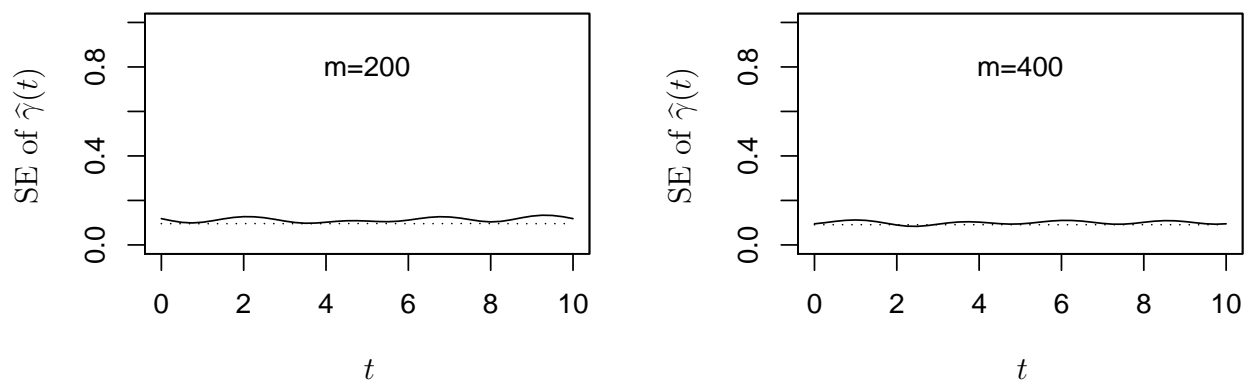$\widehat{\gamma}(t)$

Days in a standardized menstrual cycle

# Figure 2(b)



$\widehat{\gamma}(t)$

Days in a standardized menstrual cycle
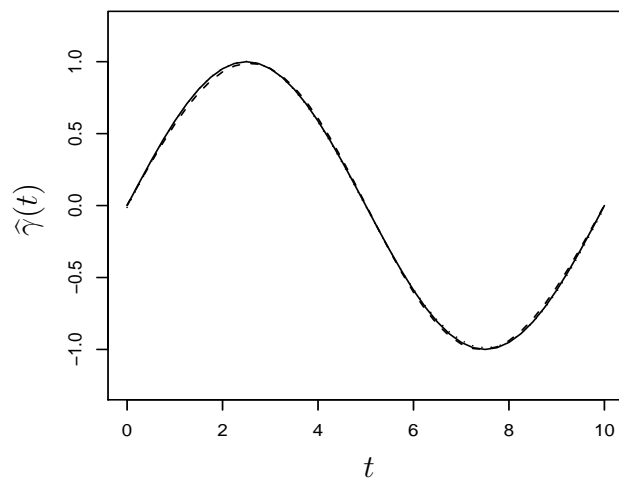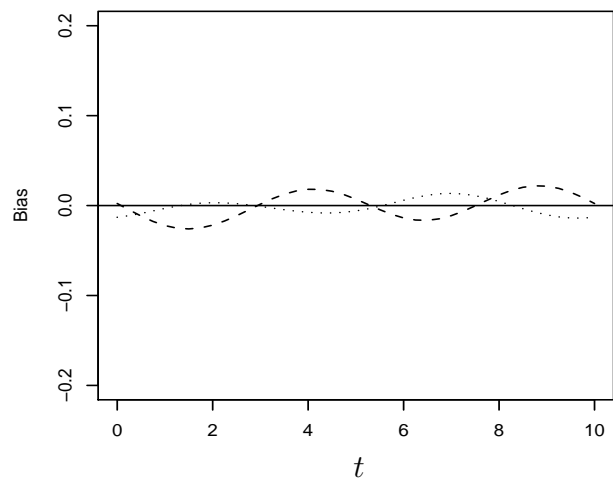
**Figure 3(a)**



**Figure 3 (b)**



**Figure 3 (c)**
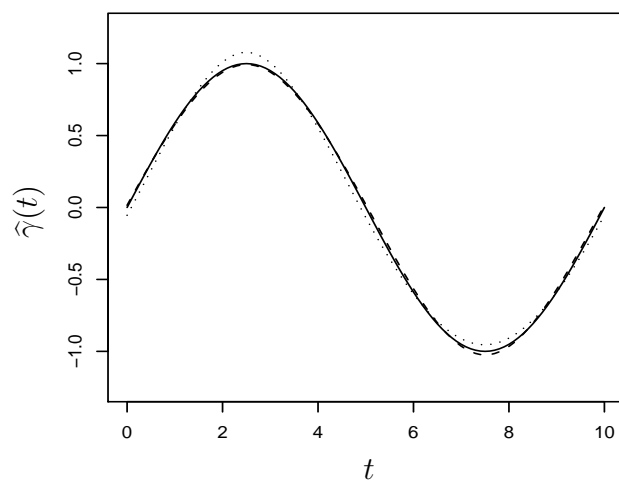


**Figure 3 (d)**



**Figure 3 (e)**



**Figure 3 (f)**



g replacements

$\widehat{\gamma}(t)$

ariance of $\widehat{\gamma}(t)$

## Figure 4(a)



PSfrag replacements

$\widehat{\gamma}(t)$

Variance of $\widehat{\gamma}(t)$

$\widehat{\gamma}(t)$

## Figure 4(b)



PSfrag replacements

$\widehat{\gamma}(t)$

Variance of $\widehat{\gamma}(t)$

$\widehat{\gamma}(t)$

## Figure 4(c)



PSfrag replacements

$\widehat{\gamma}(t)$

Variance of $\widehat{\gamma}(t)$

$\widehat{\gamma}(t)$

SE of $\widehat{\gamma}(t)$

# Figure 5(a)

m=200

Density

Test statistic S

# Figure 5(b)

m=200

Density

P−value