

Longitudinal semiparametric transition models with unknown link and variance functions

HUAZHEN LIN AND PETER X.-K. SONG*

We present an extension to the conventional transition mean model by adding a conditional variance model and assuming unknown link and variance functions. This extension gives rise to great flexibility of addressing not only the transitional relationship between the response and covariates but also the heteroscedastic mechanism of the underlying measurement process. We propose a kernel-based nonparametric estimation and inference for the regression parameters. Our estimation procedure for the regression coefficients detours the unknown link and variance functions, and hence its implementation is rather straightforward. The simulation studies show that the proposed methodology is particularly useful to extract the mean signals when they are heavily masked by strong variation. Both root- n parametric rate consistency and asymptotic normality of the proposed estimators are established. Numerical illustrations include also an analysis of longitudinal data on the length of women’s menstrual cycles.

KEYWORDS AND PHRASES: Heteroscedasticity, Kernel, Longitudinal data, Regression.

1. INTRODUCTION

Consider longitudinal outcomes, y_{ij} , measured at time j for subject i , $j = 1, \dots, n_i$; $i = 1, \dots, K$. Associated with each time series $\{y_{ij}, j = 1, \dots, n_i\}$ are covariate vectors \mathbf{x}_{ij} . In the analysis of such longitudinal data, transition models are used to describe the conditional distribution of response y_{ij} as a function of past responses, $y_{i,j-1}, \dots, y_{i1}$, and covariate \mathbf{x}_{ij} . In the framework of generalized linear models, the conditional mean, $\mu_{ij} = E(y_{ij}|\mathcal{H}_{ij}, \mathbf{x}_{ij})$, and the conditional variance, $\nu_{ij} = \text{var}(Y_{ij}|\mathcal{H}_{ij}, \mathbf{x}_{ij})$, are the central profiles of the conditional distribution for modelling. Here $\mathcal{H}_{ij} = \{y_{i1}, \dots, y_{i,j-1}\}$ represents the past responses for subject i . The classical generalized linear transition model explicitly expresses the influences of the past values $y_{i1}, \dots, y_{i,j-1}$ on the present observation, y_{ij} , hence it is useful to predict the future of the measurement process using the past and the present observations as well as the covariate information.

The transition model of order p takes the form (e.g. Diggle et al., 2002, Chapter 10), $h(\mu_{ij}) = \mathbf{x}'_{ij}\boldsymbol{\beta} + \sum_{k=1}^p \phi_k y_{i,j-k}$,

and $\nu_{ij} = \varphi V(\mu_{ij})$, where h is a known canonical link function and V is a known variance function determined fully from an assumed density function for the data. Here φ is the dispersion parameter. The interest for the utility of this model lies mostly in understanding the mean function of the measurement process, in which the conditional variance is completely specified by the mean model via the known variance function. However, in some practical studies, it is rather restrictive to require the conditional variance in a bond with the mean through a fixed manner. In other words, the order p in the mean model does not necessarily suit for the variance mechanism. The misspecification of the variance may lead to a great loss of efficiency for the mean parameter estimation as well as to inadequacy of commonly used goodness-of-fit statistics (Chiou and Müller, 1998; Wang and Carey, 2003; Ye and Pan, 2006). In addition, in some applications such as image restoration, microarray data analysis and financial time series analysis, modelling of the variance itself is of interest; see for example, Diggle and Verbyla (1998) and Ruppert et al. (1997), to name but a few. Hence, many researchers have advocated to relax either the dependence of the variance on covariates through only the mean function (e.g. Crowder, 1995 and 2001; Carroll and Ruppert, 1988; Paik 1992; Carroll, 2003) or the fixed form of the variance function $V(\cdot)$. For the latter case, various kernel-based estimators were suggested, such as Müller and Stadtmüller (1987), Hall and Carroll (1989), Hall and Marron (1990), and Neumann (1994); local polynomial based estimators were considered by Fan and Yao (1998) and Ruppert et al. (1997), Chiou and Müller (1998), and Fan et al. (2007). However, simultaneously relaxing both restrictions has been little studied in the literature. The paper is intended to make a contribution in this regard.

We consider a semiparametric transition model of orders p and q , denoted by STM(p, q), given by

$$(1) \quad \mu_{ij} = m \left(\mathbf{x}'_{ij}\boldsymbol{\beta} + \sum_{k=1}^p \phi_k y_{i,j-k} \right),$$

$$(2) \quad \nu_{ij} = v^2 \left(\mu_{ij}, \mathbf{z}'_{ij}\boldsymbol{\alpha} + \sum_{l=1}^q \psi_l y_{i,j-l} \right),$$

where m is an unknown smooth link function and v^2 is an unknown smooth variance function, both of which are to be estimated, and \mathbf{z}_{ij} is a vector of covariates, which may

*Corresponding author.

be a subset of \mathbf{x}_{ij} . Compared to the traditional transition model, the proposed model does not require the specification of the link and variance functions, in the hope that the nonparametric functions m and v^2 can provide, respectively, data-driven functional relationships of the present observation on past observations and covariates. In addition, our models are free of the assumption that the variance depends on covariates only through the mean.

Throughout the paper, subscripts m and v denote any terms related to the mean and variance models, respectively. Also, subscript mv indexes a joint vector of both mean and variance model components. For instance, $\boldsymbol{\theta}_m = (\boldsymbol{\beta}', \phi_1, \dots, \phi_p)'$ represents the vector of regression coefficients in the mean model, while $\boldsymbol{\theta}_v = (\boldsymbol{\alpha}', \psi_1, \dots, \psi_q)'$ represents the vector of regression coefficients in the variance model. Then, $\boldsymbol{\theta}_{mv} = (\boldsymbol{\theta}'_m, \boldsymbol{\theta}'_v)'$. Similarly, the covariate vector of the mean model is $\mathbf{w}_{m,ij} = (\mathbf{x}'_{ij}, y_{i,j-1}, \dots, y_{i,j-p})'$, and that of the variance model is $\mathbf{w}_{v,ij} = (\mathbf{z}'_{ij}, y_{i,j-1}, \dots, y_{i,j-q})'$. Then, the grand vector of covariates is $\mathbf{w}_{mv,ij} = (\mathbf{w}'_{m,ij}, \mathbf{w}'_{v,ij})'$.

We propose a kernel-based estimating equation approach to the estimation of regression coefficients $\boldsymbol{\theta}_{mv}$. A novel methodological contribution of this paper is that the $\boldsymbol{\theta}_{mv}$ is estimated directly, rather than iteratively, with no need of estimating the two nonparametric functions $m(\cdot)$ and $v^2(\cdot)$. Then, these two functions are estimated using any familiar nonparametric regression method such as that given in Rupert et al. (1997) or Fan and Yao (1998). We show that the estimator of $\boldsymbol{\theta}_{mv}$ is of parametric rate root- n consistency and asymptotically normally distributed.

The paper is organized as follows. Section 2 presents the estimation method for the regression coefficients. The large sample properties of the estimators are discussed in Section 3. Section 4 concerns a bandwidth selection procedure. Details of simulation studies are reported in Section 5. In Section 6 the proposed model is illustrated by a real data analysis. Some concluding remarks are given in Section 7. All technical details are listed in the two appendices.

2. ESTIMATION

To deal with the identifiability issue between the nonparametric function and the regression coefficients, we invoke a set-to-zero constraint; that is, the first components of $\boldsymbol{\theta}_m$ and $\boldsymbol{\theta}_v$ are fixed at certain known values c_m and c_v , respectively. Therefore, the actual number of parameters to be estimated in $\boldsymbol{\theta}_m$ or in $\boldsymbol{\theta}_v$ is reduced by one. For the ease of exposition, in this paper we keep using the above notation for the subvectors of parameters to be estimated, although both their first elements need not be estimated.

It is crucial to build up an estimation procedure that is numerically stable for the estimation of $\boldsymbol{\theta}_v$ in the variance model. If following Chiou and Müller's (1998) three-stage approach, one may use the residuals to set up an estimating

equation for $\boldsymbol{\theta}_v$ as follows:

$$(3) \quad \frac{1}{N_v} \sum_{i=1}^K \sum_{j=p \vee q+1}^{n_i} D(\mathbf{w}_{mv,ij}, \boldsymbol{\theta}_v) \times \{\hat{e}_{m,ij} - \hat{v}^2(\mu_{ij}, \mathbf{w}'_{v,ij} \boldsymbol{\theta}_v)\} \mathbf{w}_{v,ij} = 0,$$

where N_v is the number of terms in (3), D is a weight function and $\hat{e}_{m,ij} = \{y_{ij} - \hat{m}(\mathbf{w}'_{m,ij} \boldsymbol{\theta}_m)\}^2$ are the squared residuals from the mean model. $p \vee q + 1 = \max\{p, q + 1\}$. However, two issues may affect the performance of this method. One is the analytic difficulty of establishing large sample properties at a parametric rate $K^{-1/2}$, owing to the involvement of parametric estimate $\hat{\boldsymbol{\theta}}_m$ and the nonparametric estimate $\hat{v}^2(\cdot)$. The other is the numerical stability, as squares in the $\hat{e}_{m,ij}$'s can easily create a number of unduly large values, which will impair the stability of estimation, unless a certain robustness treatment is imposed to downweight the influence of outliers. Refer to some numerical evidence reported from our simulation studies in Section 5.

To de-tangle the estimation of $\boldsymbol{\theta}_{mv}$ from the estimation of the nonparametric functions, for the purpose of better numerical stability, we propose kernel-based estimating equations. To proceed, first note that the first and second conditional moments of y_{ij} depends on covariates \mathbf{x}_{ij} and past observations \mathcal{H}_{ij} , respectively, only through index $d_{m,ij} = \mathbf{w}'_{m,ij} \boldsymbol{\theta}_m$ and through joint indices $d_{m,ij}$ and $d_{v,ij} = \mathbf{w}'_{v,ij} \boldsymbol{\theta}_v$. Second, note that these two conditional moments can be estimated nonparametrically by the well-known Nadaraya-Watson kernel method. Let \mathcal{K}_m and \mathcal{K}_v be two bounded and symmetric kernels with support $[-1, 1]$ and of orders r_m and r_v , respectively. With given bandwidths h_m and h_v , the two moment estimates are given by

$$(4) \quad E_K(\delta_m) = \frac{\sum_{k=1}^K \sum_{l=p+1}^{n_k} y_{kl} \mathcal{K}_m\left(\frac{d_{m,kl} - \delta_m}{h_m}\right)}{\sum_{k=1}^K \sum_{l=p+1}^{n_k} \mathcal{K}_m\left(\frac{d_{m,kl} - \delta_m}{h_m}\right)},$$

$$(5) \quad E_{2K}(\delta_m, \delta_v) = \frac{\sum_{k=1}^K \sum_{l=p \vee q+1}^{n_k} y_{kl}^2 \mathcal{K}_m\left(\frac{d_{m,kl} - \delta_m}{h_m}\right) \mathcal{K}_v\left(\frac{d_{v,kl} - \delta_v}{h_v}\right)}{\sum_{k=1}^K \sum_{l=p \vee q+1}^{n_k} \mathcal{K}_m\left(\frac{d_{m,kl} - \delta_m}{h_m}\right) \mathcal{K}_v\left(\frac{d_{v,kl} - \delta_v}{h_v}\right)}.$$

Then, the least squares estimates of the regression coefficients $\boldsymbol{\theta}_m$ and $\boldsymbol{\theta}_v$ are obtained as the solutions to the two following kernel estimating equations:

$$(6) \quad \mathbf{s}_m(\boldsymbol{\theta}_m) = \frac{1}{N_m} \sum_{i=1}^K \sum_{j=p+1}^{n_i} \mathbf{w}_{m,ij} \times \{y_{ij} - E_K(\mathbf{w}'_{m,ij} \boldsymbol{\theta}_m)\} = 0,$$

$$(7) \quad \mathbf{s}_v(\boldsymbol{\theta}_m, \boldsymbol{\theta}_v) = \frac{1}{N_v} \sum_{i=1}^K \sum_{j=p \vee q+1}^{n_i} \mathbf{w}_{v,ij} \times \{y_{ij}^2 - E_{2K}(\mathbf{w}'_{m,ij} \boldsymbol{\theta}_m, \mathbf{w}'_{v,ij} \boldsymbol{\theta}_v)\} = 0,$$

where N_m is the numbers of terms in (6). The estimators $\hat{\boldsymbol{\theta}}_m$ and $\hat{\boldsymbol{\theta}}_v$ are the solution to the joint equations (6) and (7). Asymptotic properties of the proposed estimators above are presented in Section 3, including \sqrt{K} -consistency and asymptotic normality. But, all the technical details such as proofs are listed in the two appendices. We make three remarks.

1. A difference of (7) from the existing method given by (3) is that the former uses directly the squared observations, y_{ij}^2 , rather than the squared residuals $\hat{e}_{m,ij}$. This treatment enables us to avoid the unknown link function $m(\cdot)$ and the unknown variance function $v^2(\cdot)$ in the estimation of $\boldsymbol{\theta}_v$.
2. The formulation of (6) is helpful to gain numerical stability because the estimation of $\boldsymbol{\theta}_m$ is little interfered by the estimation with (7). In other words, the estimating function of (6) is $\{\boldsymbol{\theta}_v, v^2(\cdot)\}$ -insensitive (Jørgensen and Knudsen, 2004), which is a desirable property in handling the estimation of nuisance parameters.
3. Another advantage for the utility of kernels in the construction of the estimating equations (6) and (7), especially (7), is the robustness against unduly large observations that are easily created by the squares of observations. Kernels essentially use local observations and hence hamper the propagation of local abnormality into a global effect.

Throughout the rest of the paper, denote the $(k_1 + k_2 + \dots + k_r)$ -th order partial derivative of any function $f(x_1, x_2, \dots, x_r)$ by $f^{(k_1 k_2 \dots k_r)}(x_1, x_2, \dots, x_r)$; that is,

$$f^{(k_1 k_2 \dots k_r)}(x_1, x_2, \dots, x_r) = \frac{\partial^{(k_1 + k_2 + \dots + k_r)} f(x_1, x_2, \dots, x_r)}{\partial x_1^{k_1} \partial x_2^{k_2} \dots \partial x_r^{k_r}}.$$

For $r = 0, 1, i = 1, \dots, K$ and $j = 1, \dots, n_i$, let

$$\begin{aligned} G_{ij,r}(\boldsymbol{\theta}_m) &= \frac{1}{N_m h_m} \sum_{k=1}^K \sum_{l=p+1}^{n_k} y_{kl}^r \\ &\quad \times \mathcal{K}_m \left(\frac{\mathbf{w}'_{m,kl} \boldsymbol{\theta}_m - \mathbf{w}'_{m,ij} \boldsymbol{\theta}_m}{h_m} \right), \\ F_{ij,r}(\boldsymbol{\theta}_m, \boldsymbol{\theta}_v) &= \frac{1}{N_v h_m h_v} \sum_{k=1}^K \sum_{l=p \vee q+1}^{n_k} y_{kl}^{2r} \\ &\quad \times \mathcal{K}_m \left(\frac{\mathbf{w}'_{m,kl} \boldsymbol{\theta}_m - \mathbf{w}'_{m,ij} \boldsymbol{\theta}_m}{h_m} \right) \\ &\quad \times \mathcal{K}_v \left(\frac{\mathbf{w}'_{v,kl} \boldsymbol{\theta}_v - \mathbf{w}'_{v,ij} \boldsymbol{\theta}_v}{h_v} \right). \end{aligned}$$

The Newton-Raphson algorithm is used to solve the above equations (6) and (7). The Hessian matrix of $\mathbf{s}_{mv}(\boldsymbol{\theta}_m, \boldsymbol{\theta}_v) = (\mathbf{s}_m(\boldsymbol{\theta}_m)', \mathbf{s}_v(\boldsymbol{\theta}_m, \boldsymbol{\theta}_v)')'$ is a 2×2 block matrix of the following form:

$$\Delta_{mv}(\boldsymbol{\theta}_m, \boldsymbol{\theta}_v) = \begin{pmatrix} \Delta_{11}(\boldsymbol{\theta}_m) & 0 \\ \Delta_{21}(\boldsymbol{\theta}_m, \boldsymbol{\theta}_v) & \Delta_{22}(\boldsymbol{\theta}_m, \boldsymbol{\theta}_v) \end{pmatrix},$$

where

$$\begin{aligned} \Delta_{11}(\boldsymbol{\theta}_m) &= -\frac{1}{N_m} \sum_{i=1}^K \sum_{j=p+1}^{n_i} \mathbf{w}_{m,ij} \\ &\quad \times \left\{ \frac{G_{ij,1}^{(1)}(\boldsymbol{\theta}_m)' G_{ij,0}(\boldsymbol{\theta}_m) - G_{ij,1}(\boldsymbol{\theta}_m) G_{ij,0}^{(1)}(\boldsymbol{\theta}_m)'}{G_{ij,0}^2(\boldsymbol{\theta}_m)} \right\}, \\ \Delta_{21}(\boldsymbol{\theta}_m, \boldsymbol{\theta}_v) &= -\frac{1}{N_v} \sum_{i=1}^K \sum_{j=p \vee q+1}^{n_i} \mathbf{w}_{v,ij} \\ &\quad \times \left\{ \frac{F_{ij,1}^{(10)}(\boldsymbol{\theta}_m, \boldsymbol{\theta}_v)' F_{ij,0}(\boldsymbol{\theta}_m, \boldsymbol{\theta}_v) - F_{ij,1}(\boldsymbol{\theta}_m, \boldsymbol{\theta}_v) F_{ij,0}^{(10)}(\boldsymbol{\theta}_m, \boldsymbol{\theta}_v)'}{F_{ij,0}^2(\boldsymbol{\theta}_m, \boldsymbol{\theta}_v)} \right\}, \\ \Delta_{22}(\boldsymbol{\theta}_m, \boldsymbol{\theta}_v) &= -\frac{1}{N_v} \sum_{i=1}^K \sum_{j=p \vee q+1}^{n_i} \mathbf{w}_{v,ij} \\ &\quad \times \left\{ \frac{F_{ij,1}^{(01)}(\boldsymbol{\theta}_m, \boldsymbol{\theta}_v)' F_{ij,0}(\boldsymbol{\theta}_m, \boldsymbol{\theta}_v) - F_{ij,1}(\boldsymbol{\theta}_m, \boldsymbol{\theta}_v) F_{ij,0}^{(01)}(\boldsymbol{\theta}_m, \boldsymbol{\theta}_v)'}{F_{ij,0}^2(\boldsymbol{\theta}_m, \boldsymbol{\theta}_v)} \right\}. \end{aligned}$$

An iterative update scheme is used to search the solution of equation $\mathbf{s}_{mv}(\boldsymbol{\theta}_m, \boldsymbol{\theta}_v) = 0$. Since the Hessian matrix is lower block-triangular, it is numerically stable to iteratively update the values of $\boldsymbol{\theta}_m$ and $\boldsymbol{\theta}_v$ as follows,

$$\begin{aligned} \boldsymbol{\theta}_m^{k+1} &= \boldsymbol{\theta}_m^k - \left\{ \Delta_{11}(\boldsymbol{\theta}_m^k) \right\}^{-1} \mathbf{s}_m(\boldsymbol{\theta}_m^k), \\ \boldsymbol{\theta}_v^{k+1} &= \boldsymbol{\theta}_v^k - \left\{ \Delta_{22}(\boldsymbol{\theta}_m^{k+1}, \boldsymbol{\theta}_v^k) \right\}^{-1} \mathbf{s}_v(\boldsymbol{\theta}_m^{k+1}, \boldsymbol{\theta}_v^k). \end{aligned}$$

The starting values of the search may be specified from a GLM fit, where both $m(\cdot)$ and $v^2(\cdot)$ are approximated with high-order polynomial functions.

3. LARGE SAMPLE PROPERTIES

Suppose the distributions of y_{ij} , $\mathbf{w}_{m,ij}$, $\mathbf{w}_{v,ij}$, $d_{m,ij}$ and $d_{v,ij}$ to be the same as those of y , \mathbf{w}_m , \mathbf{w}_v , d_m and d_v , respectively. Denote the true value of $\boldsymbol{\theta}_{mv}$ by $\boldsymbol{\theta}_{mv,0}$. In the rest of the paper, let $d_m = \mathbf{w}'_m \boldsymbol{\theta}_{m,0}$, $\mathbf{d}_{mv} = (\mathbf{w}'_m \boldsymbol{\theta}_{m,0}, \mathbf{w}'_v \boldsymbol{\theta}_{v,0})'$, $p_m(\cdot)$ and $p_{mv}(\cdot)$ be the density functions of d_m and \mathbf{d}_{mv} . Denote $M(\delta_m, \delta_v) = m^2(\delta_m) + v^2(\delta_m, \delta_v)$, $\varepsilon_{m,ij} = y_{ij} - m(d_{m,ij})$, $\varepsilon_{v,ij} = y_{ij}^2 - M(d_{m,ij}, d_{v,ij})$,

$$\begin{aligned} \tau_{mv}(\delta_m, \delta_v) &= \mathbb{E}\{I(\mathbf{w}_m \in \boldsymbol{\Omega}_m) \mathbf{w}_m | \mathbf{d}_{mv} = (\delta_m, \delta_v)'\}, \\ \tau_m(\delta_m) &= \mathbb{E}\{I(\mathbf{w}_m \in \boldsymbol{\Omega}_m) \mathbf{w}_m | d_m = \delta_m\}, \\ \tau_v(\delta_m, \delta_v) &= \mathbb{E}\{I(\mathbf{w}_{mv} \in \boldsymbol{\Omega}_{mv}) \mathbf{w}_v | \mathbf{d}_{mv} = (\delta_m, \delta_v)'\}, \end{aligned}$$

where $\boldsymbol{\Omega}_m$ and $\boldsymbol{\Omega}_{mv}$ are two compact sets satisfying conditions (13) and (14) given in Appendix 1, respectively. Here $I(C)$ is an indicator of set C .

Theorem 1. (Consistency) Under Condition List A in Appendix 1, $\hat{\boldsymbol{\theta}}_{mv}$ converges in probability to $\boldsymbol{\theta}_{mv,0}$ as $K \rightarrow \infty$.

Theorem 2. (*Asymptotic Normality*) Under both Condition Lists A and B in Appendix 1, as $K \rightarrow \infty$, $h_m \rightarrow 0$ and $h_v \rightarrow 0$,

$$(8) \quad \sqrt{K} \left(\hat{\boldsymbol{\theta}}_m - \boldsymbol{\theta}_{m,0} \right) \rightarrow N \left\{ 0, \left(\mathbf{A}_1 \mathbf{V}_m^{-1} \mathbf{A}_1' \right)^{-1} \right\},$$

where the sensitivity matrix is

$$\mathbf{A}_1 = E \left[I(\mathbf{w}_m \in \boldsymbol{\Omega}_m) m^{(1)}(d_m) \mathbf{w}_m \{ \mathbf{w}_m - \tau_m(d_m) \}' \right]$$

and the variability matrix is

$$\mathbf{V}_m = E \left[\sum_{j=p+1}^{n_i} \{ \mathbf{w}_{m,ij} - \tau_m(d_{m,ij}) \} \varepsilon_{m,ij} \right]^{\otimes 2};$$

and

$$(9) \quad \sqrt{K} \left(\hat{\boldsymbol{\theta}}_v - \boldsymbol{\theta}_{v,0} \right) \rightarrow N \left\{ 0, \left(\mathbf{B}_2 \mathbf{V}_v^{-1} \mathbf{B}_2' \right)^{-1} \right\},$$

where the sensitivity matrix is

$$\mathbf{B}_2 = E \left[I(\mathbf{w}_{mv} \in \boldsymbol{\Omega}_{mv}) M^{(01)}(\mathbf{d}_{mv}) \mathbf{w}_v \{ \mathbf{w}_v - \tau_v(\mathbf{d}_{mv}) \}' \right]$$

and the variability matrix is

$$\mathbf{V}_v = E \left[\sum_{j=p \vee q+1}^{n_i} \{ \mathbf{w}_{v,ij} - \tau_v(d_{mv,ij}) \} \varepsilon_{v,ij} - \mathbf{B}_1 \mathbf{A}_1^{-1} \sum_{j=p+1}^{n_i} \{ \mathbf{w}_{m,ij} - \tau_m(d_{m,ij}) \} \varepsilon_{m,ij} \right]^{\otimes 2},$$

with

$$\mathbf{B}_1 = E \left[I(\mathbf{w}_{mv} \in \boldsymbol{\Omega}_{mv}) M^{(10)}(\mathbf{d}_{mv}) \mathbf{w}_m \{ \mathbf{w}_m - \tau_{mv}(\mathbf{d}_{mv}) \}' \right].$$

Here the Kronecker product is $\mathbf{a}^{\otimes 2} = \mathbf{a} \mathbf{a}'$ for any vector \mathbf{a} .

Directly evaluating the above sandwich asymptotic covariance matrices is numerically challenging, because it involves estimating the derivatives of the nonparametric functions, which appears very tedious. To overcome this difficulty in the numerical implementation, we adopt the bootstrap method to compute the standard errors of the estimates.

Note that to make the above asymptotic properties valid both \mathcal{K}_m and \mathcal{K}_v are required to be fourth-order kernels, in which the corresponding bandwidths h_m and h_v are of order $n^{-\eta}$, $1/8 < \eta < 1/5$. Fourth-order kernels can be found in Müller (1984). It is known in the literature (e.g. Hastie and Tibshirani, 1990; Carroll et al., 1997) that to achieve consistency at the root- n parametric rate, the nonparametric components need to be under-smoothed, resulting in the requirement of higher-order kernels than the familiar second-order kernels.

4. BANDWIDTH SELECTION

The proposed method needs to choose suitable bandwidths h_m and h_v . We followed Chiou and Müller's (1998) method of bandwidth selection, as it is found to fit well into our setting. To proceed, let the estimated moments in (4) and (5) by $\hat{E}_K(\delta_m)$ and $\hat{E}_{2K}(\delta_m, \delta_v)$, respectively, resulted from the terms of $d_{m,kl}$ and $d_{v,kl}$ being replaced with their corresponding estimates, $\hat{d}_{m,kl} = \mathbf{w}'_{m,kl} \hat{\boldsymbol{\theta}}_m$ and $\hat{d}_{v,kl} = \mathbf{w}'_{v,kl} \hat{\boldsymbol{\theta}}_v$. Define an indicator, for $j = 1, \dots, n_i$, $i = 1, \dots, K$,

$$a_{ij} = \begin{cases} 1, & \text{if } \hat{E}_{2K}(\hat{d}_{m,ij}, \hat{d}_{v,ij}) > \left\{ \hat{E}_K(\hat{d}_{m,ij}) \right\}^2; \\ 0, & \text{otherwise.} \end{cases}$$

A Pearson-type goodness-of-fit measure is given by

$$\chi^2(h_m, h_v) = \sum_{i=1}^K \sum_{j=p \vee q+1}^{n_i} \frac{a_{ij} \left\{ y_{ij} - \hat{E}_K(\hat{d}_{m,ij}) \right\}^2}{\hat{E}_{2K}(\hat{d}_{m,ij}, \hat{d}_{v,ij}) - \left\{ \hat{E}_K(\hat{d}_{m,ij}) \right\}^2}.$$

To find the optimal solution, in this paper we considered a "greedy" search. That is, we first specified a sequence of paired grid points of (h_m, h_v) , and at each point we then computed $Q(h_m, h_v)$. The optimal bandwidths h_m and h_v can be easily obtained as the pair giving the minimum of the following objective function:

$$(10) \quad Q(h_m, h_v) = |\chi^2(h_m, h_v) / (N - k) - 1|,$$

with $N = \sum_{i=1}^K \sum_{j=p \vee q+1}^{n_i} a_{ij}$, and $k = \dim(\mathbf{w}_{mv,ij})$.

5. SIMULATION EXPERIMENTS

To examine the performance of the proposed model and inference method, we conduct two simulation experiments in this section. One experiment is designed to have strong signals in the mean model but weak variability in the variance model, while the other is set to have strong variability in the variance model that badly masks the signals in the mean model. In both simulation studies, the true link and variance functions are known. So, we compare the estimates obtained from our approach with unknown link and variance functions to those obtained from the approach with known link and variance functions. According to Hall and Severini (1998), we consider the competing estimates obtained from the following two estimating equations,

$$(11) \quad \tilde{\mathbf{s}}_m(\boldsymbol{\theta}_m, \boldsymbol{\theta}_v) = \sum_{i=1}^K \sum_{j=p+1}^{n_i} m^{(1)}(\mathbf{w}'_{m,ij} \boldsymbol{\theta}_m) \times \{ v^2(\mathbf{w}'_{v,ij} \boldsymbol{\theta}_v) \}^{-1} \{ y_{ij} - m(\mathbf{w}'_{m,ij} \boldsymbol{\theta}_m) \} \mathbf{w}_{m,ij} = 0,$$

$$(12) \quad \tilde{\mathbf{s}}_{mv}(\boldsymbol{\theta}_m, \boldsymbol{\theta}_v) = \sum_{i=1}^K \sum_{j=q+1}^{n_i} v^{2(1)}(\mathbf{w}'_{v,ij} \boldsymbol{\theta}_v) \times \left[\{ y_{ij} - m(\mathbf{w}'_{m,ij} \boldsymbol{\theta}_m) \}^2 - v^2(\mathbf{w}'_{v,ij} \boldsymbol{\theta}_v) \right] \mathbf{w}_{v,ij} = 0,$$

where both m and v^2 are specified as their true functions. The Newton-Raphson algorithm is implemented to solve the above equations (11) and (12). The resulting estimates are referred to as the estimates with CLVF (correct link and variance functions).

We simulate longitudinal data from a transition model of orders $p = 2$ and $q = 1$, where the covariate vector for the mean model is $\mathbf{x}_{ij} = (x_{ij1}, x_{ij2}, x_{ij3})'$ with

$$\begin{aligned} x_{ij1} &= x_{i1} \sim \text{Uniform}(0, 3), \\ x_{ij2} &= x_{i2} \sim \text{Bernoulli}(0.5), \\ x_{ij3} &= 1 + j/n_i, \quad j = 1, \dots, n_i. \end{aligned}$$

Hypothetically, x_1 represents an exposure over time per subject, x_2 represents a treatment over time per subject, and x_3 represents a time-dependent trend. The covariate vector for the variance model is $\mathbf{z}_{ij} = (x_{ij1}, x_{ij3})'$. The initial observation is generated by

$$\begin{aligned} y_{ij} &= m(\mathbf{x}'_{ij}\boldsymbol{\beta}) + \sqrt{v^2(\mathbf{z}'_{ij}\boldsymbol{\alpha})}\epsilon_{ij}, \\ \epsilon_{ij} &\sim N(0, \sigma^2), \quad j = 1, 2, \end{aligned}$$

and the other observations are generated sequentially from the following transition model,

$$\begin{aligned} y_{ij} &= m(\mathbf{x}'_{ij}\boldsymbol{\beta} + \phi_1 y_{i,j-2} + \phi_2 y_{i,j-1}) \\ &\quad + \sqrt{v^2(\mathbf{z}'_{ij}\boldsymbol{\alpha} + \psi_1 y_{i,j-1})}\epsilon_{ij} \\ \epsilon_{ij} &\stackrel{iid}{\sim} N(0, \sigma^2), \quad j = 3, \dots, n_i. \end{aligned}$$

In the two simulation studies, we always fix parameters $\boldsymbol{\alpha} = (-0.5, 2.5)'$, $\psi_1 = 0.6$, or $\boldsymbol{\theta}_{v,0} = (-0.5, 2.5, 0.6)'$, $v^2(u) = u^2$, $n_i = 5$, and $K = 200$. But, we vary regression parameters $\boldsymbol{\beta} = (1, -4, 4) \times a_0$, $\phi_1 = -0.5 \times a_0$, $\phi_2 = 0.5 \times a_0$, or $\boldsymbol{\theta}_{m,0} = (1, -4, 4, -0.5, 0.5) \times a_0$, by a factor a_0 in the mean model, and the variance parameter σ^2 . The link functions are $m(u) = \exp(u)$ in the first simulation and $m(u) = \sin^2(u\pi/2)$ in the second simulation, respectively. Varying the mean function, $\boldsymbol{\theta}_{m,0}$ and σ^2 will lead to different levels of the so-called signal-to-noise ratio (SNR) defined as follows:

$$\text{SNR} = \frac{\text{SD} \{m(\mathbf{w}'_{m,ij}\boldsymbol{\theta}_{m,0})\}}{\text{SD} \left\{ \sqrt{v^2(\mathbf{w}'_{v,ij}\boldsymbol{\theta}_{v,0})\sigma^2} \right\}},$$

where SD stands for the standard deviation. Clearly, the SNR measures the severity that the variance masks the mean signals. The sample version of the SNR is calculated by the substitution of the corresponding sample standard deviations.

It is worth noting that in the first design of simulation, compared to the square variance function $v^2(u) = u^2$, the exponential link function, $m(u) = \exp(u)$, clearly presents a dominant trend, so the corresponding SNR is large. In contrast, the second design of simulation specifies a periodic link

function, $m(u) = \sin^2(u\pi/2)$, which is confined within $[0, 1]$ and presents a weaker trend than the square variance function. So, the mean signals will be largely buried in strong noise, and the resulting SNR is small. Alternatively, we vary the factor a_0 in the mean model and σ in the variance model to change the strength of the mean and variance signal, which also results in different values of the SNR.

Tables 1 and 2 report the summary results of the regression parameter estimation, including the summary statistics of bias, empirical standard deviation (ESD) and mean squared error (MSE) based on 200 replications. The reported SNR in the tables is the median of the 200 sample SNR values, each from one simulated data set. In the tables, $a[i]$ means the i -th component of vector a .

From the simulation studies, we learned:

- (i) The method with CLVF performs better than the proposed STM when the SNR is around 1, indicating a balance between the signals from both mean and variance models. See the top block of Table 1. This is not surprising, because the CLVF estimation is carried out under the true link and variance functions, whereas the STM method assumes these two functions unknown. One exception, though, is the estimation of parameter ϕ_1 , the lag-2 autoregressive effect in the mean model, where the bias and ESD of the proposed STM estimation are smaller. Also, we found in our simulation studies that the CLVF is very sensitive to the SNR level. From some additional simulation experiments, we observed that with a slight increase in the SNR, the bias and ESD of CLVF's estimates of mean parameters become smaller than those of the STM's, whereas the bias and ESD of CLVF's estimates of variance parameters tend to become larger than those of the STM. In contrast, the proposed STM is relatively robust against the change of SNR level.
- (ii) The CLVF method has been greatly challenged when the SNR balance is broken. When the mean signal is very strong, the case presented in the bottom block of Table 1, the CLVF method appears to have great difficulty of capturing the variance structure properly, resulting in about 74% of runs failing to converge. This is due largely to that in the CLVF method, the estimation of the mean and variance model parameters are deeply involved each other. In contrast, the proposed method works well, which is beneficial from the fact that equation (6) is $\{\boldsymbol{\theta}_v, v^2(\cdot)\}$ -insensitive. In addition, the kernel-based estimation utilizes essentially local observations, and unduly large observations created by y_{ij}^2 in the variance model can be automatically downweighted or deleted from a local fit. Thus, the performance of the STM method appears quite stable.
- (iii) The performance of the STM appears steadily better than the CLVF method when the variance model dominates over the mean model, as indicated by the

Table 1. Simulation results of the regression parameter estimation by the proposed STM method and the method with CLVF under $m(u) = \exp(u)$ and $n = 200$. The summaries are based on 200 replicates. The results from the method with CLVF marked by * are based only on 52 replications, because the Newton-Raphson algorithm failed to converge in 148 out of the 200 simulation runs. Cautious interpretations are needed for the results marked by *

(σ, a_0)	Median SNR			Bias	ESD	$\sqrt{\text{MSE}}$		
(1/3, 1/5)	1.2367	$\beta_m[2]$	STM	-0.0201	0.0835	0.0859		
			CLVF	-0.0071	0.0584	0.0589		
		$\beta_m[3]$	STM	0.0130	0.0999	0.1007		
			CLVF	-0.0010	0.0348	0.0348		
		ϕ_1	STM	-0.0011	0.0092	0.0093		
			CLVF	0.0018	0.0154	0.0155		
		ϕ_2	STM	0.0017	0.0105	0.0106		
			CLVF	-0.0021	0.0100	0.0102		
		$\alpha[2]$	STM	-0.0311	0.5796	0.5804		
			CLVF	0.0544	0.3280	0.3324		
		ψ_1	STM	-0.0119	0.1447	0.1452		
			CLVF	-0.0464	0.1250	0.1333		
		(1/3, 1/3.5)	6×10^6	$\beta_m[2]$	STM	-0.0063	0.0452	0.0457
					CLVF*	-0.0042	0.0212	0.0216
$\beta_m[3]$	STM			0.0100	0.0806	0.0813		
	CLVF*			0.0003	0.0021	0.0022		
ϕ_1	STM			0.0000	0.0075	0.0075		
	CLVF*			0.0000	0.0003	0.0003		
ϕ_2	STM			0.0007	0.0088	0.0088		
	CLVF*			0.0000	0.0000	0.0000		
$\alpha[2]$	STM			-0.0659	0.6045	0.6081		
	CLVF*			1.6706	1.3321	2.1367		
ψ_1	STM			-0.0165	0.1357	0.1367		
	CLVF*			-0.2628	0.0967	0.2801		

results in Table 2. In such a situation, strong variability severely masks the mean signals. So, even if the link and variance functions are correctly specified, the CLVF method does not work well. The top block of Table 2 presents the summary based only on selected 95 replicates that have the smallest squared errors, $\|\hat{\theta}_{mv} - \theta_{mv,0}\|^2$. Take one case among the other 105 runs as an example. The CLVF estimates of the parameters in the mean model are $\hat{\theta}_m = (938.377, 620.207, -884.402, -180.466, 28.747)'$, which are very biased in comparison to the true values $\theta_{m,0} = (1, -4, 4, -0.5, 0.5)'/3$ (with the factor $a_0 = 1/3$); the estimates of the parameters in the variance model are $\hat{\theta}_v = (-0.419, 2.581, 0.357)'$, which are reasonably close to the true values $\theta_{v,0} = (-0.5, 2.5, 0.6)'$. A possible explanation for such a poor performance is that the large variance can lead to distortion or cancellation of the mean effects in the estimating equation (11), described as follows: for any θ_{mv} in a neighborhood around the $\theta_{mv,0}$,

$$\frac{\{y_{ij} - m(\mathbf{w}'_{m,ij}\theta_m)\}^2}{v^2(\mathbf{w}'_{v,ij}\theta_v)} \approx \frac{y_{ij}^2}{v^2(\mathbf{w}'_{v,ij}\theta_v)},$$

where both ratios $m(\mathbf{w}'_{m,ij}\theta_m)/v^2(\mathbf{w}'_{v,ij}\theta_v)$ and

$m^2(\mathbf{w}'_{m,ij}\theta_m)/v^2(\mathbf{w}'_{v,ij}\theta_v)$ are so small that the mean function $m(\mathbf{w}'_{m,ij}\theta_m)$ literally has no power to determine the root of the equation with respect to parameter θ_m . As a result, properly estimating the $\theta_{m,0}$ becomes very difficult.

- (iv) In the bottom block of Table 2, when the dominance of the variance model decreases by lowering σ^2 from 1 to 1/4, because again of the weak trend pattern of $m(u) = \sin^2(u\pi/2)$, the performance of the kernel-based estimation remains still better than the CLVF method, although the concern on convergence of the latter is alleviated.

With regard to the computing time, the proposed procedure converged on average with 10 iterations, and each iteration took approximately 1.5 CPU seconds on an average PC that we used to run the simulation studies.

6. A DATA EXAMPLE

To illustrate the application of the proposed methodology, we re-analyzed the data from a 1-year longitudinal study on the length of menstrual cycles in a sample of 157 college women (Harlow and Matanoski, 1991; Lin et al., 1997). In this study, a total of 1158 standard cycles were observed with each woman contributing from $n_i = 1$ to

Table 2. Simulation results of the regression parameter estimation by the proposed STM method and the method with CLVF under $m(u) = \sin^2(x\pi/2)$ and $n = 200$. The summaries are based on 200 replicates, 95 of which are used to produce the results marked by *. The Newton-Raphson algorithm failed to converge in 75 out of the 200 simulation runs, and among the remaining 125 convergent cases about 30% of them converged to the values far away from the true parameter values

(σ, a_0)	Median SNR			Bias	ESD	$\sqrt{\text{MSE}}$		
(1, 1/3)	0.0748	$\beta_m[2]$	STM	-0.0054	0.0870	0.0872		
			CLVF*	0.0841	3.2543	3.2554		
		$\beta_m[3]$	STM	0.0114	0.1542	0.1547		
			CLVF*	0.5449	4.2442	4.2790		
		ϕ_1	STM	-0.0020	0.0157	0.0158		
			CLVF*	0.0952	0.5953	0.6029		
		ϕ_2	STM	0.0020	0.0138	0.0140		
			CLVF*	0.2578	1.1351	1.1640		
		$\alpha[2]$	STM	0.0276	0.5691	0.5697		
			CLVF	0.0455	0.2643	0.2681		
		ψ_1	STM	-0.0135	0.1381	0.1388		
			CLVF	-0.0780	0.0920	0.1206		
		(1/2, 1/3)	0.1634	$\beta_m[2]$	STM	-0.0207	0.1003	0.1024
					CLVF	1.9801	19.5615	19.6615
$\beta_m[3]$	STM			0.0191	0.1266	0.1280		
	CLVF			-0.5633	15.0120	15.0225		
ϕ_1	STM			-0.0014	0.0185	0.0186		
	CLVF			0.2194	7.2234	7.2268		
ϕ_2	STM			0.0020	0.0131	0.0132		
	CLVF			-0.0623	4.3761	4.3766		
$\alpha[2]$	STM			-0.0605	0.5935	0.5966		
	CLVF			-0.0101	0.1773	0.1776		
ψ_1	STM			-0.0045	0.1573	0.1574		
	CLVF			-0.0300	0.1200	0.1237		

$n_i = 15$ cycles. The median value of n_i was nine. Three covariates considered in Harlow and Matanoski (1991) include: (i) history of an extreme cycle (greater than 43 days, yes= 1/no= 0), (ii) dieting status (yes= 1/no= 0), and (iii) living at home or on campus (yes= 1/no= 0).

The previous studies have indicated that the menstrual cycle is highly variable; see for example, Figure 1 in Harlow and Zeger (1991) and Figure 2 in Lin et al. (1997). Such strong within-subject heterogeneity at the variance level motivates Lin et al. to extend Harlow and Zeger's analysis based on the standard linear mixed-effects model. In this paper, using the proposed semiparametric transition model, our analysis will provide a different approach to modelling and interpreting this strong heterogeneity. Thus, we present some new understanding about the transitional behaviour of ovarian function. It is also useful to forecast the transitional mechanism of the menstrual cycle based on the historic cycle information and the baseline covariates.

Denote by y_{ij} the j -th menstrual cycle length for the i -th women ($i = 1, \dots, 157, j = 1, \dots, n_i$). The following model STM(1, 1) is applied to fit the data,

$$y_{ij} = m(\beta_1 \text{HIS}_i + \beta_2 \text{CAMP}_{ij} + \beta_3 \text{DIET}_{ij} + \phi_1 y_{i,j-1}) + \sqrt{v^2(\mu_{ij}, \alpha_1 \text{HIS}_i + \alpha_2 \text{DIET}_{ij} + \psi_1 y_{i,j-1})} \epsilon_{ij},$$

where the errors $\epsilon_{ij}, i = 1, \dots, K, j = 1, \dots, n_i$ are *iid* with mean zero and variance $\sigma^2 = 1$. $m(\cdot)$ and $v^2(\cdot)$ are unknown link and variance functions, respectively. $\theta_m = (\beta_1, \beta_2, \beta_3, \phi_1)'$ and $\theta_v = (\alpha_1, \alpha_2, \psi_1)'$ are the unknown regression parameters. Following Lin et al. (1997), we do not include CAMP in the variance model.

We compare the STM(1,1) based analysis to that given by a simple linear transition model (LTM) of order 1:

$$y_{ij} = b_0 + b_1 \text{HIS}_i + b_2 \text{CAMP}_{ij} + b_3 \text{DIET}_{ij} + \psi_1 y_{i,j-1} + \varsigma_{ij},$$

where ς_{ij} are *iid* errors with mean zero and variance σ^2 . The simple LTM analysis assumes the identity link function, $m(u) = u$, and does not model the variance of the measurement process. Using the bandwidth selection method given in (10), we found the bandwidths equal to $h_m = 2.7$ and $h_v = 0.15$, and the resulting estimates of the regression coefficients and their standard errors are listed in Table 3. The calculation of the standard errors was done via the resampling method with 300 bootstrap samples, in which each subject is treated as a resampling unit in order to preserve the inherent transitional feature of the data. The choice of 300 sample size was determined by monitoring the stability of the standard errors; we found that when the bootstrap sample size is between 200 and 300, the resulting standard errors got stabilized and the difference was only marginal.

Table 3. The estimation results from the analysis of the menstrual cycle length data using the proposed STM(1,1) and the simple linear transition model (LTM) of order 1

	STM(1,1)				LTM(1)	
	$\theta_{m,0}$	(std. err.)	θ_v	(std. err.)	$\theta_{m,0}$	(std. err.)
HIS(x_1)	1.198	(-)	4.235	(-)	1.198	(0.307)
CAMP(x_2)	-0.856	(0.4605)	-	(-)	-0.668	(0.306)
DIET(x_3)	-3.601	(1.3906)	-7.415	(1.5390)	-2.896	(0.455)
LAG-1 OBS.(y_{j-1})	0.101	(0.0501)	0.346	(0.0809)	0.087	(0.022)

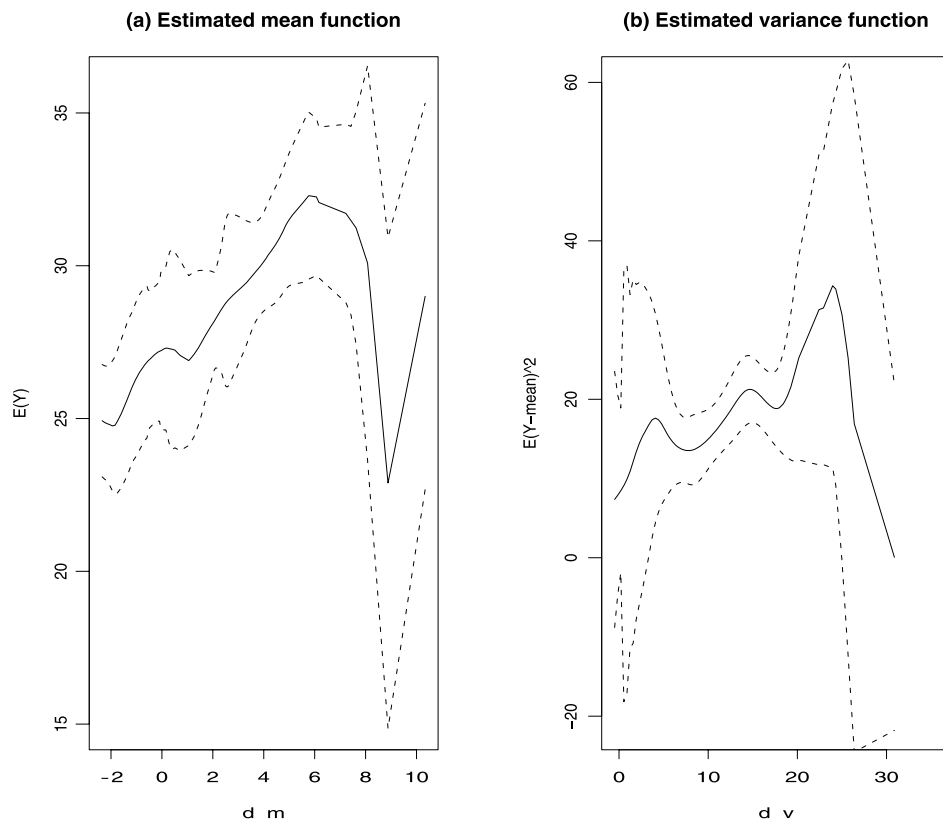


Figure 1. The estimated mean function (the left panel) and the estimated variance function (the right panel) are obtained by the standard kernel-based nonparametric regression function in the semiparametric transition model of order (1, 1).

The estimates obtained from the simple LTM(1) model are also included in Table 3 for comparison.

Our analysis concludes that the diet has a significant effect on both the mean length of menstrual cycle and the variance of the cycle. This finding is in agreement with the previous two analyses by Harlow and Matanoski (1991) and Lin et al. (1997). What is new from our analysis is that the variance of the cycle has a significantly positive relation with the length of cycle at the previous time; that is, the longer the previous cycle takes, the more variable the current cycle length is.

Figure 1 displays the estimated mean and variance functions of the STM(1,1) model, with 95% bootstrap confidence bands. For the variance function, we fix the mean index d_m at the sample mean (equal to 2.7) of $\hat{d}_{m,ij}$'s. These

estimated functions are obtained by the standard kernel-based nonparametric regression approach, after the estimation of the regression coefficients in both the mean and variance models is done. Clearly, the two estimated functions appear to be highly non-linear. The mean function begins to decrease when the mean index d_m is around 6. Considering the edge effect of kernel smoothing technique, our conclusion of pattern change would be drawn cautiously for the range of $6 < d_m < 8$. This means that for example, for an average college woman (with d_m around the average 2.7) dieting will reduce the length of her menstrual cycle, but this effect may be reversed for a woman with approximately $6 < d_m < 8$. Similarly, the variance function exhibits an overall increasing pattern with a few waves. Given these findings of non-linearity, one should be cautious to interpret

the results given in the LTM(1) based analysis, where the three covariates and the lag-1 observation are all found statistically significant. In particular, the covariate CAMP is only marginally significant from the LTM analysis, but not significant from the STM analysis.

7. CONCLUDING REMARKS

There are several types of models developed in the literature for the analysis of longitudinal data. The class of transition models provides a unique perspective to address the serial dependence mechanism for the longitudinal data. Transitional patterns are of primary interest in dealing with serially correlated data and forecasting. Besides the applications shown in the paper, our proposed model can be applied to analyze time-course microarray data and panel data from finance and economics. In these subject-matters fields, signals at the mean level are typically masked by strong noise, so that it is hard to extract the mean signals effectively and appropriately. This paper presents a useful approach to overcoming this difficulty.

A striking experience gained from our simulation studies is the importance of numerical stability. The standard transition model with correctly specified link and variance functions can easily run into the problem of non-convergence, and hence numerically very unstable. Practically, this is a big concern, which can stop practitioners from using the transition model. In contrast, the proposed method performs remarkably stable, due to the use of local observations in the kernel-based estimating equations and the exclusion of the variance parameters from the estimating equation related to the mean model. Our approach may result in lower estimation efficiency in the case when both link and variance functions are known and the signal-to-noise ratio is around 1. However, in practice this efficiency loss is hardly avoided because the true link and variance functions are never known.

With no doubt, to fit the proposed model well, in which the unknowns include two nonparametric functions and some regression coefficients, the sample size cannot be too small. As suggested by a reviewer, we repeated the simulation studies on STM(2,1) model in Section 5 with sample size $K = 100$. We found that the proposed method failed to converge for 27.5% of the samples, and the CLVF performed even worse. However, using the convergent cases we obtained findings similar to those from Tables 1 and 2.

One limitation of the proposed methodology is that it is not applicable to data with categorical response variables. For the example of binary responses, equation (7) is invalid because $y_{ij}^2 = y_{ij}$. A new approach is needed to overcome this limitation. Another remaining issue is the selection of orders p and q in the application of the proposed STM methodology. This is a non-trivial issue as in the context of estimating functions, the lack of objective functions gives rise to the difficulty of proposing suitable model selection

criteria, such as analogs of the AIC or BIC in the likelihood based inference. One possible direction is in the line of Pan (2001), but further exploration is needed.

As for other possible future directions of research in the area of transition models for longitudinal data, although being very useful, such types of models are currently limited to fit only regularly balanced time series. Some extensions are needed to cope with (i) data with irregular visit times, (ii) subjects who do not have sufficient number of observations for the transition models of orders p and/or q , and (iii) data with intermittently missing observations. Methodologies such as continuous-time Markov models and missing data imputation are of possibility, which have been listed as part of our future projects.

ACKNOWLEDGEMENTS

The authors are thankful to the Associate Editor and an anonymous reviewer for their constructive comments and suggestions on the paper. The authors' research is supported in part by the Chinese Natural Science Foundation (10771148) to the first author and by the USA National Science Foundation Grant (DMS 0904177) to the second author. The authors are also grateful to Dr. Xihong Lin for the Women's menstrual cycle data analyzed in the paper and to Dr. Eddy Lam for his valuable comments on the manuscript.

APPENDIX 1: CONDITIONS

To establish the consistency of the proposed estimators, the following *Condition List A* is required.

- A1: Kernels \mathcal{K}_m and \mathcal{K}_v are one-dimensional bounded symmetric density functions around zero with compact supports. Without loss of generality, the supports are $[-1, 1]$. In addition, the \mathcal{K}_m and \mathcal{K}_v have bounded and continuous second order derivatives that satisfy Lipschitz conditions, namely, $|\mathcal{K}_m^{(2)}(x_1) - \mathcal{K}_m^{(2)}(x_2)| \leq M|x_1 - x_2|$ and $|\mathcal{K}_v^{(2)}(x_1) - \mathcal{K}_v^{(2)}(x_2)| \leq M|x_1 - x_2|$ for some constant $M < \infty$.
- A2: $\sum_{k=1}^{\infty} \exp\{-\gamma k h_m^2 h_v^2\} < \infty$ for all $\gamma > 0$.
- A3: $p_m(\varpi'_m \boldsymbol{\theta}_m)$ and $E(y|d_m = \varpi'_m \boldsymbol{\theta}_m)$ are uniformly continuous in ϖ_m and $\boldsymbol{\theta}_m$; $p_{mv}(\varpi'_m \boldsymbol{\theta}_m, \varpi'_v \boldsymbol{\theta}_v)$ and $E(y^2|\mathbf{d}_{mv} = (\varpi'_m \boldsymbol{\theta}_m, \varpi'_v \boldsymbol{\theta}_v)')$ are uniformly continuous in $\varpi_m, \varpi_v, \boldsymbol{\theta}_m$ and $\boldsymbol{\theta}_v$.
- A4: There exist compact sets Θ_m and Θ_v such that $\boldsymbol{\theta}_m \in \Theta_m$ and $\boldsymbol{\theta}_v \in \Theta_v$ and

$$\inf_{i,j,\boldsymbol{\theta}_m \in \Theta_m, \boldsymbol{\theta}_v \in \Theta_v} p_{mv}(\omega'_{m,ij} \boldsymbol{\theta}_m, \omega'_{v,ij} \boldsymbol{\theta}_v) > 0.$$

- A5: Let $g_v(\boldsymbol{\theta}_{mv}) = E\{\mathbf{w}_v E(y^2|\mathbf{d}_{mv} = (\mathbf{w}'_m \boldsymbol{\theta}_m, \mathbf{w}'_v \boldsymbol{\theta}_v)')\} - E(\mathbf{w}_v y^2)$, and $g_m(\boldsymbol{\theta}_m) = E\{\mathbf{w}_m E(y|d_m = \mathbf{w}'_m \boldsymbol{\theta}_m)\} - E[\mathbf{w}_m y]$. There exists a unique solution to the system of two equations, $g_m(\boldsymbol{\theta}_m) = 0$ and $g_v(\boldsymbol{\theta}_{mv}) = 0$.

To establish asymptotic normality for the proposed estimators, some additional assumptions are given in the *Condition List B* below.

B1: There exist compact sets Ω_m and Ω_v , with $\Omega_{mv} = \Omega_m \times \Omega_v$, such that

$$(13) \quad \sum_{i=1}^K \sum_{j=p+1}^{n_i} I(\mathbf{w}_{m,ij} \notin \Omega_m) |\mathbf{w}_{m,ij}| = o_p(K^{1/2}),$$

and

$$(14) \quad \sum_{i=1}^K \sum_{j=p \vee q+1}^{n_i} I(\mathbf{w}_{mv,ij} \notin \Omega_{mv}) |\mathbf{w}_{v,ij}| = o_p(K^{1/2}).$$

B2: Derivatives $p_m^{(k_1)}(\delta_m)$, $p_{mv}^{(k_1, k_2)}(\delta_m, \delta_v)$, $m^{(k_1)}(\delta_m)$, $v^{2(0, k_2)}(\delta_m, \delta_v)$, $\tau_{mv}^{(k_1, k_2)}(\delta_m, \delta_v)$, $\tau_m^{(k_1)}(\delta_m)$ and $\tau_v^{(k_1, k_2)}(\delta_m, \delta_v)$, $k_1 = 1, \dots, r_m + 1$, $k_2 = 1, \dots, r_v + 1$, exist and are bounded, where r_m and r_v are the orders of \mathcal{K}_m and \mathcal{K}_v , respectively.

B3: $K h_m^4 \rightarrow \infty$, $K h_v^4 \rightarrow \infty$, $K^3 h_m^2 h_v^8 \rightarrow \infty$, $K^3 h_m^8 h_v^2 \rightarrow \infty$, $\frac{\log K}{\sqrt{K h_m^3 h_v}} \rightarrow 0$, $\frac{\log K}{\sqrt{K h_m h_v^3}} \rightarrow 0$, $K h_m^{2r_m} \rightarrow 0$ and $K h_v^{2r_v} \rightarrow 0$.

APPENDIX 2: PROOFS

Proof of Theorem 1. We present the proof to the consistency of $\hat{\theta}_m$, and the consistency of $\hat{\theta}_v$ can be argued in a similar manner. Clearly, evaluated at the true values, $g_m(\theta_{m,0}) = 0$. First, we show that for any θ_m in a neighborhood of the true $\theta_{m,0}$ such that $\|\theta_m - \theta_{m,0}\| < \zeta$, $s_{m,K}(\theta_m) \rightarrow g_m(\theta_m)$ with probability one. Note that

$$\begin{aligned} s_{m,K}(\theta_m) &= \frac{1}{N_m} \sum_{i=1}^K \sum_{j=p+1}^{n_i} \mathbf{w}_{m,ij} \{y_{ij} - \hat{E}_K(\mathbf{w}'_{m,ij} \theta_m)\} \\ &= \frac{1}{N_m} \sum_{i=1}^K \sum_{j=p+1}^{n_i} \mathbf{w}_{m,ij} \{y_{ij} - E(y_{ij} | \mathbf{w}_{m,ij})\} \\ &\quad - \frac{1}{N_m} \sum_{i=1}^K \sum_{j=p+1}^{n_i} \mathbf{w}_{m,ij} \{\hat{E}_K(\mathbf{w}'_{m,ij} \theta_m) \\ &\quad - E(y_{ij} | \mathbf{w}'_{m,ij} \theta_m)\} \\ &\quad - \frac{1}{N_m} \sum_{i=1}^K \sum_{j=p+1}^{n_i} \mathbf{w}_{m,ij} \{E(y_{ij} | \mathbf{w}'_{m,ij} \theta_m) \\ &\quad - E(y_{ij} | \mathbf{w}_{m,ij})\}. \end{aligned}$$

By the strong law of large numbers, the first term $\frac{1}{N_m} \sum_{i=1}^K \sum_{j=p+1}^{n_i} \mathbf{w}_{m,ij} \{y_{ij} - E(y_{ij} | \mathbf{w}_{m,ij})\} \xrightarrow{a.s.} 0$, and the third term $\frac{1}{N_m} \sum_{i=1}^K \sum_{j=p+1}^{n_i} \mathbf{w}_{m,ij} \{E(y_{ij} | \mathbf{w}'_{m,ij} \theta_m) - E(y_{ij} | \mathbf{w}_{m,ij})\} \xrightarrow{a.s.} g_m(\theta_m)$ as $K \rightarrow \infty$. By Lemma 1 of Ying and Cook (2005), under Conditions A1 to A4, the second term converges to zero in probability. Finally, given the

compactness of the set Θ_m there exists a $\theta_m^* \in \Theta_m$ such that $\theta_{m,K} = \theta_m^* + o_p(1)$. This implies that with probability one $g_m(\theta_m^*) = 0$, because $\theta_{m,K}$ is a sequence of roots such that $s_{m,K}(\theta_{m,K}) = 0$ for all K . Condition A5 of the unique root immediately leads to $\theta_m^* = \theta_{m,0}$. The consistency is proved. \square

Lemma A1. Under the Condition List A and Condition B2 and B3, for $k_1, k_2 = 0, 1$,

$$\begin{aligned} &\frac{(-1)^{k_1+k_2}}{h_m^{k_1+1} h_v^{k_2+1}} E \left\{ \mathcal{K}_m^{(k_1)} \left(\frac{d_m - \delta_m}{h_m} \right) \mathcal{K}_v^{(k_2)} \left(\frac{d_v - \delta_v}{h_v} \right) g(\mathbf{d}_{mv}) \right\} \\ &= \{g(\delta_{mv}) p_{mv}(\delta_{mv})\}^{(k_1, k_2)} + O(h_m^{r_m} + h_v^{r_v}), \end{aligned}$$

where $g(\cdot)$ is a function whose derivatives $g^{(r_m+1, r_v+1)}(\delta_{mv})$ exist and are bounded.

Proof of Lemma A1. See Horowitz (1996). \square

Lemma A2. Let $d_{t,ij} = \mathbf{w}'_{t,ij} \theta_{t,0}$, $\hat{d}_{t,ij} = \mathbf{w}'_{t,ij} \hat{\theta}_t$, $t = m, v$. For any given $\varpi_{mv} \in \Omega_{mv}$, $\delta_t = \varpi'_t \theta_{t,0}$ and $\hat{\delta}_t = \varpi'_t \hat{\theta}_t$, $t = m, v$. Denote

$$\begin{aligned} \hat{F}_{mv,r}(\delta_{mv}) &= \frac{1}{N_v h_m h_v} \sum_{i=1}^K \sum_{j=p \vee q+1}^{n_i} y_{ij}^{2r} \mathcal{K}_m \left(\frac{\hat{d}_{m,ij} - \delta_m}{h_m} \right) \\ &\quad \times \mathcal{K}_v \left(\frac{\hat{d}_{v,ij} - \delta_v}{h_v} \right), r = 0, 1, \\ F_{mv,r}(\delta_{mv}) &= \frac{1}{N_v h_m h_v} \sum_{i=1}^K \sum_{j=p \vee q+1}^{n_i} y_{ij}^{2r} \mathcal{K}_m \left(\frac{d_{m,ij} - \delta_m}{h_m} \right) \\ &\quad \times \mathcal{K}_v \left(\frac{d_{v,ij} - \delta_v}{h_v} \right), r = 0, 1. \end{aligned}$$

As $K \rightarrow \infty$,

$$\begin{aligned} \hat{F}_{mv,0}(\hat{\delta}_{mv}) &= F_{mv,0}(\delta_{mv}) + (\hat{\theta}_m - \theta_{m,0})' \Gamma_m^{(100)}(\delta_m, \delta_v, \varpi_m) \\ &\quad + (\hat{\theta}_v - \theta_{v,0})' \Gamma_v^{(010)}(\delta_m, \delta_v, \varpi_v) \\ &\quad + o_p(|\hat{\theta}_m - \theta_{m,0}|) + o_p(|\hat{\theta}_v - \theta_{v,0}|), \\ \hat{F}_{mv,1}(\hat{\delta}_{mv}) &= F_{mv,1}(\delta_{mv}) + (\hat{\theta}_m - \theta_{m,0})' \Delta_m^{(100)}(\delta_m, \delta_v, \varpi_m) \\ &\quad + (\hat{\theta}_v - \theta_{v,0})' \Delta_v^{(010)}(\delta_m, \delta_v, \varpi_v) \\ &\quad + o_p(|\hat{\theta}_m - \theta_{m,0}|) + o_p(|\hat{\theta}_v - \theta_{v,0}|), \end{aligned}$$

where $\Gamma_t(\delta_{mv}, \varpi_t) = -\{E(\mathbf{w}_t | \mathbf{d}_{mv} = \delta_{mv}) - \varpi_t\} p_{mv}(\delta_{mv})$, and

$$\begin{aligned} \Delta_t(\delta_{mv}, \varpi_t) &= -\{E(y^2 \mathbf{w}_t | \mathbf{d}_{mv} = \delta_{mv}) - M(\delta_{mv}) \varpi_t\} \\ &\quad \times p_{mv}(\delta_{mv}), t = m, v. \end{aligned}$$

Proof of Lemma A2. We prove the first result for $\hat{F}_{mv,0}(\hat{\delta}_m, \hat{\delta}_v)$ and the second can be proved in a similar way. Let the first-order Taylor expansion gives, with $r = 0$,

(15)

$$\begin{aligned}
& \widehat{F}_{mv,0}(\widehat{\delta}_m, \widehat{\delta}_v) \\
&= \frac{1}{N_v h_m h_v} \sum_{i=1}^K \sum_{j=p \vee q+1}^{n_i} \mathcal{K}_m \left(\frac{d_{m,ij} - \delta_m}{h_m} \right) \mathcal{K}_v \left(\frac{d_{v,ij} - \delta_v}{h_v} \right) \\
&+ \frac{1}{N_v h_m^2 h_v} \sum_{i=1}^K \sum_{j=p \vee q+1}^{n_i} \mathcal{K}_m^{(1)} \left(\frac{d_{m,ij} - \delta_m}{h_m} \right) \\
&\times \mathcal{K}_v \left(\frac{d_{v,ij} - \delta_v}{h_v} \right) (\mathbf{w}_{m,ij} - \varpi_m)' (\widehat{\boldsymbol{\theta}}_m - \boldsymbol{\theta}_{m,0}) \\
&+ \frac{1}{N_v h_m h_v^2} \sum_{i=1}^K \sum_{j=p \vee q+1}^{n_i} \mathcal{K}_m \left(\frac{d_{m,ij} - \delta_m}{h_m} \right) \\
&\times \mathcal{K}_v^{(1)} \left(\frac{d_{v,ij} - \delta_v}{h_v} \right) (\mathbf{w}_{v,ij} - \varpi_v)' (\widehat{\boldsymbol{\theta}}_v - \boldsymbol{\theta}_{v,0}) \\
&+ o_p(|\widehat{\boldsymbol{\theta}}_m - \boldsymbol{\theta}_{m,0}|) + o_p(|\widehat{\boldsymbol{\theta}}_v - \boldsymbol{\theta}_{v,0}|) \\
&\stackrel{\text{def}}{=} F_{mv,0}(\boldsymbol{\delta}_{mv}) + T_{1,k}(\varpi_m, \varpi_v)' (\widehat{\boldsymbol{\theta}}_m - \boldsymbol{\theta}_{m,0}) \\
&+ T_{2,K}(\varpi_m, \varpi_v)' (\widehat{\boldsymbol{\theta}}_v - \boldsymbol{\theta}_{v,0}) \\
&+ o_p(|\widehat{\boldsymbol{\theta}}_m - \boldsymbol{\theta}_{m,0}|) + o_p(|\widehat{\boldsymbol{\theta}}_v - \boldsymbol{\theta}_{v,0}|).
\end{aligned}$$

It follows from Theorem 2.37 of Pollard (1984, p. 34) that

$$\begin{aligned}
& \sup_{\varpi_{mv} \in \boldsymbol{\Omega}_{mv}} |T_{1,K}(\varpi_{mv}) - E\{T_{1,K}(\varpi_{mv})\}| \\
&= o \left\{ \log K / (K h_m^3 h_v)^{1/2} \right\}.
\end{aligned}$$

By Lemma A1, we obtain $E\{T_{1,K}(\varpi_{mv})\} - \Gamma_m^{(100)}(\delta_m, \delta_v, \varpi_m) = O_p(h_m^{r_m} + h_v^{r_v})$, where r_t is the order of kernel \mathcal{K}_t , $t = m, v$. Hence, under Condition B3,

$$(16) \quad T_{1,K}(\varpi_{mv}) - \Gamma_m^{(100)}(\delta_m, \delta_v, \varpi_m) = o_p(1).$$

Similar arguments result in

$$(17) \quad T_{2,K}(\varpi_{mv}) - \Gamma_v^{(010)}(\delta_m, \delta_v, \varpi_v) = o_p(1).$$

Then, Lemma A2 follows from (15), (16) and (17). \square

Lemma A3. Denote for $r = 0, 1$,

$$\begin{aligned}
\widehat{G}_{m,r}(\delta_m) &= \frac{1}{N_m h_m} \sum_{i=1}^K \sum_{j=p+1}^{n_i} y_{ij}^r \mathcal{K}_m \left(\frac{\widehat{d}_{m,ij} - \delta_m}{h_m} \right), \\
G_{m,r}(\delta_m) &= \frac{1}{N_m h_m} \sum_{i=1}^K \sum_{j=p+1}^{n_i} y_{ij}^r \mathcal{K}_m \left(\frac{d_{m,ij} - \delta_m}{h_m} \right).
\end{aligned}$$

For large K , with $\widehat{\delta}_t = \varpi_t' \widehat{\boldsymbol{\theta}}_t$, $t = m, v$,

$$\begin{aligned}
\widehat{G}_{m,0}(\widehat{\delta}_m) &= G_{m,0}(\delta_m) + (\widehat{\boldsymbol{\theta}}_m - \boldsymbol{\theta}_{m,0})' \Upsilon_m^{(10)}(\delta_m, \varpi_m) \\
&+ o_p(|\widehat{\boldsymbol{\theta}}_m - \boldsymbol{\theta}_{m,0}|) + o_p(|\widehat{\boldsymbol{\theta}}_v - \boldsymbol{\theta}_{v,0}|), \\
\widehat{G}_{m,1}(\widehat{\delta}_m) &= G_{m,1}(\delta_m) + (\widehat{\boldsymbol{\theta}}_m - \boldsymbol{\theta}_{m,0})' \Lambda_m^{(10)}(\delta_m, \varpi_m) \\
&+ o_p(|\widehat{\boldsymbol{\theta}}_m - \boldsymbol{\theta}_{m,0}|) + o_p(|\widehat{\boldsymbol{\theta}}_v - \boldsymbol{\theta}_{v,0}|),
\end{aligned}$$

where $\Upsilon_m(\delta_m, \varpi_m) = -\{E(\mathbf{w}_m | d_m = \delta_m) - \varpi_m\} p_m(\delta_m)$, and $\Lambda_m(\delta_m, \varpi_m) = -\{E(y \mathbf{w}_m | d_m = \delta_m) - m(\delta_m) \varpi_m\} p_m(\delta_m)$.

Proof of Lemma A3. Lemma A3 can be shown on a similar line of the proof for Lemma A2. \square

Proof of Theorem 2. First, it follows from (4) that $\widehat{E}(y | \varpi_m) \equiv \widehat{E}(y | \widehat{\delta}_m) = \widehat{G}_{m,1}(\widehat{\delta}_m) / \widehat{G}_{m,0}(\widehat{\delta}_m)$. Applying the results of Lemma A3 concerning $\widehat{G}_{m,1}$ and $\widehat{G}_{m,0}$, we obtain

(18)

$$\begin{aligned}
\widehat{E}(y | \varpi_m) - m(\delta_m) &= \frac{G_{m,1}(\delta_m) - m(\delta_m) G_{m,0}(\delta_m)}{G_{m,0}(\delta_m)} \\
&+ \frac{(\widehat{\boldsymbol{\theta}}_m - \boldsymbol{\theta}_{m,0})' \left\{ \Lambda_m^{(10)}(\delta_m, \varpi_m) - m(\delta_m) \Upsilon_m^{(10)}(\delta_m, \varpi_m) \right\}}{G_{m,0}(\delta_m)} \\
&+ o_p(|\widehat{\boldsymbol{\theta}}_m - \boldsymbol{\theta}_{m,0}|) = \frac{G_{m,1}(\delta_m) - m(\delta_m) G_{m,0}(\delta_m)}{p_m(\delta_m)} \\
&+ \frac{(\widehat{\boldsymbol{\theta}}_m - \boldsymbol{\theta}_{m,0})' \left\{ \Lambda_m^{(10)}(\delta_m, \varpi_m) - m(\delta_m) \Upsilon_m^{(10)}(\delta_m, \varpi_m) \right\}}{p_m(\delta_m)} \\
&+ o_p(|\widehat{\boldsymbol{\theta}}_m - \boldsymbol{\theta}_{m,0}|) + O_p \left(\frac{1}{N_m h_m h_v} \right).
\end{aligned}$$

Similarly, from (5) we have $\widehat{E}(y^2 | \varpi_{mv}) \equiv \widehat{E}(y^2 | \widehat{\boldsymbol{\delta}}_{mv}) = \widehat{F}_{mv,1}(\widehat{\boldsymbol{\delta}}_{mv}) / \widehat{F}_{mv,0}(\widehat{\boldsymbol{\delta}}_{mv})$. It follows from Lemma A2 that

(19)

$$\begin{aligned}
\widehat{E}(y^2 | \varpi_{mv}) - M(\boldsymbol{\delta}_{mv}) &= \frac{F_{mv,1}(\boldsymbol{\delta}_{mv}) - M(\boldsymbol{\delta}_{mv}) F_{mv,0}(\boldsymbol{\delta}_{mv})}{F_{mv,0}(\boldsymbol{\delta}_{mv})} \\
&+ \frac{(\widehat{\boldsymbol{\theta}}_m - \boldsymbol{\theta}_{m,0})' \left\{ \Delta_m^{(100)}(\boldsymbol{\delta}_{mv}, \varpi_m) - M(\boldsymbol{\delta}_{mv}) \Gamma_m^{(100)}(\boldsymbol{\delta}_{mv}, \varpi_m) \right\}}{F_{mv,0}(\boldsymbol{\delta}_{mv})} \\
&+ \frac{(\widehat{\boldsymbol{\theta}}_v - \boldsymbol{\theta}_{v,0})' \left\{ \Delta_v^{(010)}(\boldsymbol{\delta}_{mv}, \varpi_v) - M(\boldsymbol{\delta}_{mv}) \Gamma_v^{(010)}(\boldsymbol{\delta}_{mv}, \varpi_v) \right\}}{F_{mv,0}(\boldsymbol{\delta}_{mv})} \\
&+ o_p(|\widehat{\boldsymbol{\theta}}_m - \boldsymbol{\theta}_{m,0}|) + o_p(|\widehat{\boldsymbol{\theta}}_v - \boldsymbol{\theta}_{v,0}|) + O_p \left(\frac{1}{N_m h_m h_v} \right),
\end{aligned}$$

uniformly over $\varpi_{mv} \in \boldsymbol{\Omega}_{mv}$.

Second, noting that $N_t = O(K)$, $t = m, v$, the estimates $\widehat{\boldsymbol{\theta}}_m$ and $\widehat{\boldsymbol{\theta}}_v$ are the roots of the following equations,

$$\begin{aligned}
\frac{1}{K} \sum_{i=1}^K \sum_{j=p+1}^{n_i} \{y_{ij} - \widehat{E}(y | \mathbf{w}_{m,ij})\} \mathbf{w}_{m,ij} &= 0, \\
\frac{1}{K} \sum_{i=1}^K \sum_{j=p \vee q+1}^{n_i} \{y_{ij}^2 - \widehat{E}(y^2 | \mathbf{w}_{mv,ij})\} \mathbf{w}_{v,ij} &= 0,
\end{aligned}$$

which can be rewritten as follows,

$$(20) \quad \frac{1}{K} \sum_{i=1}^K \sum_{j=p+1}^{n_i} \{y_{ij} - m(d_{m,ij})\} \mathbf{w}_{m,ij} - \frac{1}{K} \sum_{i=1}^K \sum_{j=p+1}^{n_i} \{\widehat{\mathbb{E}}(y_{ij} | \mathbf{w}_{m,ij}) - m(d_{m,ij})\} \mathbf{w}_{m,ij} = 0,$$

$$(21) \quad \frac{1}{K} \sum_{i=1}^K \sum_{j=p \vee q+1}^{n_i} \{y_{ij}^2 - M(d_{mv,ij})\} \mathbf{w}_{v,ij} - \frac{1}{K} \sum_{i=1}^K \sum_{j=p \vee q+1}^{n_i} \{\widehat{\mathbb{E}}(y^2 | \mathbf{w}_{mv,ij}) - M(d_{mv,ij})\} \mathbf{w}_{v,ij} = 0.$$

Plugging (18) into (21) and (19) into (21), exchanging the order of two summations, and applying Condition B1, we obtain

$$\begin{aligned} & \frac{1}{K} \sum_{i=1}^K \sum_{j=p+1}^{n_i} \{y_{ij} - m(d_{m,ij})\} \mathbf{w}_{m,ij} \\ & - \frac{1}{K} \sum_{i=1}^K \sum_{j=p+1}^{n_i} \tau_m(d_{m,ij}) \{y_{ij} - m(d_{m,ij})\} \\ & - \mathbf{A}_1(\widehat{\boldsymbol{\theta}}_m - \boldsymbol{\theta}_{m,0}) = o_p(K^{-1/2}), \end{aligned}$$

and

$$\begin{aligned} & \frac{1}{K} \sum_{i=1}^K \sum_{j=p \vee q+1}^{n_i} \{y_{ij}^2 - M(d_{mv,ij})\} \mathbf{w}_{v,ij} \\ & - \frac{1}{K} \sum_{i=1}^K \sum_{j=p \vee q+1}^{n_i} \tau_v(d_{mv,ij}) \{y_{ij}^2 - M(d_{mv,ij})\} \\ & - \mathbf{B}_1(\widehat{\boldsymbol{\theta}}_m - \boldsymbol{\theta}_{m,0}) - \mathbf{B}_2(\widehat{\boldsymbol{\theta}}_v - \boldsymbol{\theta}_{v,0}) = o_p(K^{-1/2}). \end{aligned}$$

Hence, Theorem 2 is proved. \square

Received 16 May 2009

REFERENCES

- [1] CARROLL, R. J. (2003). Variances are not always nuisance parameters. *Biometrics*, **59**, 211–220. [MR1987387](#)
- [2] CARROLL, R. J., FAN, J., GIJBELS, I. and WAND, M. P. (1997). Generalized partially linear single-index models. *Journal of the American Statistical Association*, **92**, 477–489. [MR1467842](#)
- [3] CARROLL, R. J. and RUPPERT, D. (1988). Transformation and Weighting in Regression. *Chapman and Hall*. [MR1014890](#)
- [4] CHIOU, J. M. and MÜLLER, H. (1998). Quasi-likelihood regression with unknown link and variance functions. *Journal of the American Statistical Association*, **93**, 1376–1387. [MR1666634](#)
- [5] CROWDER, M. (1995). On the use of a working correlation matrix in using generalized linear models for repeated measurements. *Biometrika*, **82**, 407–410.
- [6] CROWDER, M. (2001). On repeated measures analysis with misspecified covariance structure. *Journal of Royal Statistical Society, B*, **63**, 55–62. [MR1811990](#)
- [7] DIGGLE, P. J., HEAGERTY, P., LIANG, K. Y. and ZEGER, S. L. (2002). *Analysis of Longitudinal Data (second edition)*. Oxford: Oxford University Press. [MR2049007](#)
- [8] DIGGLE, P. J. and VERBYLA, A. P. (1998). Nonparametric estimation of covariance structure in longitudinal data, *Biometrics*, **54**, 401–415.
- [9] FAN, J., HUANG, T. and LI, R. Z. (2007). Analysis of longitudinal data with semiparametric estimation of covariance function. *Journal of the American Statistical Association*, **35**, 632–641. [MR2370857](#)
- [10] FAN, J. and YAO, Q. (1998). Efficient estimation of conditional variance functions in stochastic regression. *Biometrika*, **85**, 645–660. [MR1665822](#)
- [11] HALL, D. B. and SEVERINI, T. A. (1998). Extended generalized estimating equations for clustered data. *Journal of the American Statistical Association*, **93**, 1365–1375. [MR1666633](#)
- [12] HALL, P. and CARROLL, R. J. (1989). Variance function estimation in regression: the effect of estimating the mean. *Journal of the Royal Statistical Society, Series B*, **51**, 3–14. [MR0984989](#)
- [13] HALL, P. and MARRON, J. (1990). On variance estimation in nonparametric regression. *Biometrika*, **77**, 415–419. [MR1064818](#)
- [14] HARLOW, S. D. and MATANOSKI, G. M. (1991). The association between weight, physical activity, and stress and variation in the length of the menstrual cycle. *American Journal of Epidemiology*, **133**, 38–49.
- [15] HARLOW, S. D. and ZEGER, S. (1991). An application of longitudinal methods to the analysis of menstrual diary data. *Journal of Clinical Epidemiology*, **44**, 1015–1025.
- [16] HASTIE, T. J., and TIBSHIRANI, R. (1990). *Generalized Additive Models*. London: Chapman and Hall. [MR1082147](#)
- [17] HOROWITZ, J. L. (1996). Semiparametric estimation of a regression model with an unknown transformation of the dependent variable. *Econometrica*, **64**, 103–137. [MR1366143](#)
- [18] JØRGENSEN, B. and KNUDSEN, S. J. (2004). Parameter orthogonality and bias adjustment for estimating functions. *Scandinavian Journal of Statistics*, **31**, 93–114. [MR2042601](#)
- [19] LIN, X., RAZ, J. and HARLOW, S. (1997). Linear mixed models with heterogeneous within-cluster variances. *Biometrics*, **53**, 910–923. [MR1475054](#)
- [20] MÜLLER, H. G. (1984). Smooth optimum kernel estimators of densities, regression curves and modes, *Annals of Statistics*, **12**, 766–774. [MR0740929](#)
- [21] MÜLLER, H. G. and STADTMULLER, U. (1987). Estimation of heteroscedasticity in regression analysis. *Annals of Statistics*, **15**, 610–625. [MR0888429](#)
- [22] NEUMANN, M. (1994). Fully data-driven nonparametric variance estimation. *Statistics*, **25**, 189–212. [MR1366825](#)
- [23] PAIK, M. C. (1992). Parametric variance function estimation for nonnormal repeated measurement data. *Biometrics*, **48**, 19–30.
- [24] PAN, W. (2001). Model selection in estimating equations, *Biometrics*, **57**, 529–534. [MR1855687](#)
- [25] POLLARD, D. (1984). Convergence of stochastic processes. New York: Springer-Verlag. [MR0762984](#)
- [26] RUPPERT, D. (1997). Empirical-bias bandwidths for local polynomial nonparametric regression and density estimation. *Journal of the American Statistical Association*, **92**, 1049–1062. [MR1482136](#)
- [27] RUPPERT, D., WAND, M., HOLST, U. and HOSSJER, O. (1997). Local polynomial variance-function estimation. *Technometrics*, **39**, 262–272. [MR1462587](#)
- [28] WANG, Y. G. and CAREY, V. J. (2003). Working correlation structures misspecification, estimation and covariate design: implications for GEE performance. *Biometrika*, **90**, 29–41. [MR1966548](#)

- [29] YE, H. and PAN, J. (2006). Modelling of covariance structures in generalised estimating equations for longitudinal data. *Biometrika*, **93**, 927–941. [MR2285080](#)
- [30] YING, X. and COOK, R. D. (2005). Direction estimation in single-index regressions. *Biometrika*, **92**, 371–384. [MR2201365](#)

Huazhen Lin
School of Statistics
Southwestern University of Finance and Economics
Chengdu, Sichuan
China 611130
E-mail address: huazhenlin@hotmail.com

Peter X.-K. Song
Department of Biostatistics
University of Michigan
Ann Arbor, Michigan 48109-2029
USA
E-mail address: pxsong@umich.edu