Scandinavian Journal of Statistics Theory and Applications

Scandinavian Journal of Statistics
doi: 10.1111/sjos.12218

Functional Mixed Effects Model for Small Area Estimation

TAPABRATA MAITI
Department of Statistics and Probability, Michigan State University
SAMIRAN SINHA
Department of Statistics, Texas A&M University
PING-SHOU ZHONG
Department of Statistics and Probability, Michigan State University

ABSTRACT. Functional data analysis has become an important area of research because of its ability of handling high-dimensional and complex data structures. However, the development is limited in the context of linear mixed effect models and, in particular, for small area estimation. The linear mixed effect models are the backbone of small area estimation. In this article, we consider area-level data and fit a varying coefficient linear mixed effect model where the varying coefficients are semiparametrically modelled via B-splines. We propose a method of estimating the fixed effect parameters and consider prediction of random effects that can be implemented using a standard software. For measuring prediction uncertainties, we derive an analytical expression for the mean squared errors and propose a method of estimating the mean squared errors. The procedure is illustrated via a real data example, and operating characteristics of the method are judged using finite sample simulation studies.

Key words: best linear unbiased predictor, B-splines, functional data analysis, linear mixed models, mean squared error, restricted maximum likelihood, small area estimation, variance components

1. Introduction

The word 'small area' or 'small domain' generally defines a sub-population in a small geographical region or by age–sex–race categorized within a large population. The parameter of interest could be the total or mean of a variable of interest for the sub-populations. The need for reliable small area estimates is felt by many agencies, both public and private, for making useful policy decisions, for example, the small area statistics being used to monitor socio-economic and health condition for different age–sex–race groups or the pattern over small geographical areas. Thus, the small area estimation and the statistical techniques therein have become a topic of growing importance in recent years. The aim of this paper is to propose a method of estimating the sub-population mean and estimating the corresponding mean squared error (MSE) using a varying coefficient model based on aggregated data.

It is now widely recognized that the direct survey estimates for small areas are usually unreliable, being accompanied with large standard errors and large coefficients of variations. Therefore, it is necessary to use models, either explicit or implicit, to connect the small areas, and to obtain estimates of improved precision by 'borrowing strength' from related areas. This estimation is usually performed by using shrinkage estimators where the direct survey estimates are shrunken towards the regression mean. The technique can also be viewed as a smoothing technique where regression is being used as the smoother. The survey-based direct small area estimates and their variance estimates are the main ingredients to build aggregate-level small area models. Typical modelling strategy assumes that the sampling variances are known while a suitable linear regression model is assumed for the means. The fundamental of small area

estimation can be found in the works of Ghosh & Rao (1994), Pfeffermann (2002), Rao (2003) and Datta & Ghosh (2012).

While most research in small area estimation is focused on cross-sectional data, there are many surveys that are repeated over time. Therefore, intuitively, more precise small area estimates can be obtained by borrowing information not only across small areas but also across time. Previous research on small area estimation by combining cross-sectional and time series data includes Pfeffermann & Burck (1990), Rao & Yu (1994), Ghosh *et al.* (1996), Datta *et al.* (1999), Datta *et al.* (2002) and Bell (2012). These research extended the linear parametric models developed for cross-sectional data to accommodate time series nature of data by modifying the area-specific random effect assumption. The modifications are generally adding an area adjusted time-specific random effects with a simple time series model, such as an AR(1) or a random walk model. However, none of these articles considered varying coefficient models in the context of small area estimation.

The purpose of this article is to develop functional regression models for small area estimation that combines time series and cross-sectional data. Functional data is an active area of research for analysing high-dimensional data. Typically, a functional data consists of several functions or curves sampled on a fine grid. Functional models are well suited when the regression surfaces are not easy to express in a simple parametric form. This technique has been increasingly popular for growth curve models and longitudinal data analysis. One of the aspects of functional data analysis is to combine multiple observations across the curves. This is in spirit of small area estimation methodology where we combine information across areas, and an area, in our context, is represented by a curve observed at multiple time points. Thus, this article is connecting functional data analysis to small area estimation research.

In real small area applications, the data are recorded in some prespecific discrete time points despite being outcome of a (approximated) continuous time process. For example, consider the unemployment data that are usually reported quarterly or monthly. However, the actual process changes in much finer timescale. The top left part of Fig. 1 shows the quarterly unemployment rate for all the US states over a period from 2001 to 2011. For policymaking, an interest could be estimating the state profiles of quarterly unemployment rates. The figure indicates that (i) the statewise quarterly unemployment rate profiles are not linear and (ii) there is a variation among the curves and also among the state averages over time, and these are common characteristics of functional data. For a general description and the methodology of functional data analysis, we refer to Ramsay & Silverman (2005), Ferraty & Vieu (2006) and Guo (2002). Additionally, Morris & Carroll (2006) provided an excellent review of functional regression in the context of linear mixed models.

Our approach is to consider a varying coefficient linear regression model for each area. Within an area, the regression coefficients are assumed to be a smooth function of time. Hoover *et al.* (1998) introduced the idea of varying coefficient linear regression model, and it is a special case of functional linear model as discussed in the work of Ramsay & Silverman (2005). In our context, the curves over different small areas are combined through random effects to borrow strength. Our random effects distribution is more general and flexible compared with that used in existing small area estimation literature. In particular, our model can handle a general stationary covariance function, which includes AR(1), heteroskedastic AR(1) or random walk models.

In the functional data context, we first reduce the dimension of the curves (varying coefficients) by approximating them using a semiparametric method. There are several approaches to approximate the curves semiparametrically. For example, Ramsay & Silverman (2005) and Guo (2002) used smoothing splines, and Morris & Carroll (2006) used wavelet basis. However, we model the curves as a linear combination of B-spline basis functions for flexible smoothing,

and these basis functions are computationally more stable (Eilers & Marx, 1996). The semiparametric modelling is not common in small area estimation context, although the approach is quite popular in other areas of statistics. Opsomer *et al.* (2008) used a penalized regression splines for small area estimation in the context of cross-sectional data. Their model is a semiparametric extension of the Battese, Harter & Fuller (1988) model. Briefly, they modelled the mean of the response conditional on the covariate using a penalized regression model. Unlike ours, they considered neither the time-course data nor a combination of time-series and cross-sectional data. Finally, but most importantly, we derive an analytical expression for the MSE of prediction and derive a consistent estimator of this MSE. In this context, we deduce several theoretical results that are new in the context of functional data analysis and small area estimation. Note that Das *et al.* (2004) derived theoretically valid mean squared prediction error under a general linear mixed model. However, our theoretical results cannot be directly derived from Das *et al.* (2004), as the parameter characteristics in our model are different compared with standard linear mixed models and consequently the model estimation requires separate theoretical developments. Some of the other recent non/semiparametric small area estimation research includes Pratesi *et al.* (2008), Salvati *et al.* (2010, 2011), Giusti *et al.* (2012), Ugarte *et al.* (2009) and Militano *et al.* (2012). However, our approach is different from these works in terms of implementation and theoretical validation.

The remainder of the paper is organized as follows. Section 2 contains the functional linear small area models with flexible covariance structure. Section 3 contains the proposed method of estimation, while prediction and MSE calculations are collected in Section 4. The finite sample performance of the proposed method is judged via simulation studies, and the details are given in Section 5. Section 6 contains a real data example followed by a discussion given in Section 7. All technical details are collected to an online supporting information.

2. Models and assumptions

Let $\{Y_i(t_j), X_{i1}(t_j), \ldots, X_{ip}(t_j)\}$ be the observed data at time points $t_1, \ldots, t_m \in [0, \tau]$ and for $i = 1, \cdots, n$. Here, Y denotes the response, and X's denote the covariates. In the small area context, i stands for area, and t_j's are time points when the response variable and covariates are observed. Our proposed model is

$$Y_i(t) = \beta_{i0}(t) + \sum_{k=1}^{p} X_{ik}(t)\beta_{ik}(t) + \epsilon_i(t), \tag{1}$$

where $\beta_{i0}(t)$ is the time-varying intercept and $\beta_{i1}(t), \ldots, \beta_{ip}(t)$ are the time-varying regression coefficients for the p time-varying covariates. Most models in small area estimation assume a constant slope parameter with respect to time and space except Pfeffermann & Burck (1990) who considered a general state-space model for the regression coefficients. However, their model is fully parametric and for discrete time points, and thus, it is not applicable in our functional regression context. Following the mixed model literature, we break the regression coefficients as

$$\beta_{i0}(t) = \beta_0(t) + U_{i0}(t), \ \beta_{ik}(t) = \beta_k(t) + U_{ik}(t), \ k = 1, \ldots, p,$$

where $\beta_k(t)$ is the deterministic component of the area-specific function $\beta_{ik}(t)$, capturing the common mean, and $U_{ik}(t)$ are the area-specific time-varying random effects, measuring the deviation from the common mean, for $k = 0, 1, \ldots, p$. Here, $\epsilon_i(t)$ denotes the sampling error associated with area $i, i = 1, \ldots, n$. This type of break-up (one deterministic and one random) is commonly used in shrinkage estimation (Datta & Ghosh, 2012) and in related applications, like in geostatistics (Cressie & Wikle 2011). Following the standard customs (Rao, 2003),

we assume that $(\epsilon_i(t_1), \ldots, \epsilon_i(t_m))$ follows an m-variate normal distribution with mean zero and known covariance Ω_i. The random effect $U_{ik}(t)$ is assumed to be a Gaussian process with mean zero and stationary covariance function $\text{cov}\{U_{ik}(t), U_{ik}(s)\} = \Sigma_{u_ik}(|t-s|)$. In this paper, we assume $\Sigma_{u_ik}(|t-s|)$ is independent of i and known up to a $(q+1)$-dimensional parameter vector θ_k. Hence, we write the covariance function as $\Sigma_{uk} := \Sigma_k(|t-s|;\theta_k) = \psi_k A_m(\rho_k)$, for $k = 0, 1, \ldots, p$ with $\theta_k = \left(\psi_k, \rho_k^T\right)^T$, where ψ_k is a scalar and ρ_k is a q-dimensional vector, and A_m is a correlation matrix. This structure of the covariance is quite general. For example, a commonly used isotropic covariance function is the exponential function with $\Sigma_k(t;\theta_k) = \psi \exp(-\rho t)$, where $\theta_k = (\psi, \rho)^T$, and this covariance function can be written as the product of variance function and correlation functions. Recently, Bell (2012) mentioned several parametric models for the covariance matrix $\Sigma_k(t;\theta_k)$ in the context of multivariate random intercept Fay–Herriot model that includes the stationary AR(1), heteroskedastic AR(1) and random walk AR(1) processes, and these covariances are special cases of our general covariance function.

3. Estimation methodology

Note that in the small area estimation, the main interest is in prediction of small area-specific parameters, which in turn requires prediction of random effects and estimation of the deterministic functional regression coefficients $\beta_0(\cdot), \beta_1(\cdot), \ldots, \beta_p(\cdot)$. Our objective is to develop a method that can exploit the well-established linear mixed model machinery and hence will be simple from practitioners' perspective. We assume that the regression functions belong to a normed space of continuous functions with finite $(d-1)$th derivatives. This assumption allows us to write $\beta_0, \beta_1, \ldots, \beta_p$ as a linear combination of several dth-degree B-spline basis functions, that is,

$$\beta_k(t) = \sum_{\ell=1}^{L} B_{\ell,d}(t) b_{k\ell}, \; k = 0, 1, \ldots, p,$$

where $B_{1,d}(\cdot), \ldots, B_{L,d}(\cdot)$ are L B-spline bases based on L knot points $(\tau_1 \leq \tau_2 \ldots \leq \tau_L)$. The dth-degree spline basis is defined as $B_{\ell,d}(t) = (t-\tau_\ell)B_{\ell,d-1}(t)/(\tau_{\ell+d-1} - \tau_{\ell-1}) + (\tau_{\ell+d} - t)B_{\ell+1,d-1}(t)/(\tau_{\ell+d} - \tau_\ell)$, and $B_{\ell,0}(t) = I(t \in (\tau_{\ell-1}, \tau_l])$. Then writing $b_k = (b_{k1}, \ldots, b_{kL})^T$, for $k = 0, 1, \ldots, p$, $Z_i^{(0)}(t) = (B_{1,d}(t), \ldots, B_{L,d}(t))^T$, and $Z_i^{(k)}(t) = (B_{1,d}(t)X_{ik}(t), \ldots, B_{L,d}(t)X_{ik}(t))^T$, $k = 1, \ldots, p$, model (1) becomes

$$Y_i(t) = \sum_{k=0}^{p} b_k^T Z_i^{(k)}(t) + U_{i0}(t) + \sum_{k=1}^{p} X_{ik}(t)U_{ik}(t) + \epsilon_i(t). \tag{2}$$

In order to avoid subjective choice of the knot points, usually moderately large number of knot points are chosen within $[0, \tau]$. Then the parameters $b_{k\ell}, \ell = 1, \ldots, L, k = 0, 1, \ldots, p$ are estimated by maximizing a penalized likelihood to avoid overfitting. The main idea is that the regression functions should be smooth, meaning one should associate a penalty for large curvature that is measured via $(\Delta_2 b_k)^T(\Delta_2 b_k)$, for the kth coefficient, $\beta_k(t)$, where Δ_2 represents the matrix of the second-order finite differences (Eilers & Marx, 1996). In particular, the lth row of $\Delta_2 b_k$ is $b_{k(l+2)} - 2b_{k(l+1)} + b_{kl}$, for $l = 1, \ldots, (L-2)$. We shall denote the penalty corresponding to $\beta_k(t)$ by σ_{bk}^2. Now, we write the entire model in the linear mixed effect model framework.

For this purpose, we introduce some notations. Let $\Delta_2^T \Delta_2 = VSV^T$, where V is a matrix of the orthonormal eigenvectors of $\Delta_2^T \Delta_2$ and S is a diagonal matrix containing the eigenvalues

of $\Delta_2^T \Delta_2$ along the diagonal. Particularly, write $S = \text{Diag}(\lambda_{L_1}, \lambda_{L_2})$, where λ_{L_1} represents the L_1 large (non-zero) eigenvalues and $V = (V_{L_1}, V_{L_2})$ is the partition corresponding to $(\lambda_{L_1}, \lambda_{L_2})$ with $L_1 + L_2 = L$. Define $Y = \left(Y_1^T, \ldots, Y_n^T\right)^T$, where $Y_i = (Y_i(t_1), \ldots, Y_i(t_m))^T$; $Z_F = \left(Z_{F1}^T, \ldots, Z_{Fn}^T\right)^T$, where $Z_{Fi} = (Z_{Fi}(t_1), \ldots, Z_{Fi}(t_m))^T$, $Z_{F_i}(t_j) = \left(Z_i^{(0)^T}(t_j) V_{L_2}, Z_i^{(1)^T}(t_j) V_{L_2}, \ldots, Z_i^{(p)^T}(t_j) V_{L_2}\right)^T$; $b_F = \left(b_0^T V_{L_2}, b_1^T V_{L_2}, \ldots, b_p^T V_{L_2}\right)^T$; $W = (Z_R, M_0, M_1, \ldots, M_p)$, where $Z_R^T = \left(Z_{R_1}^T, \ldots, Z_{R_n}^T\right)$, $Z_{R_i} = (Z_{R_i}(t_1), \ldots, Z_{R_i}(t_m))^T$, $Z_{R_i}(t_j) = \left(Z_i^{(0)^T}(t_j) V_{L_1}, Z_i^{(1)^T}(t_j) V_{L_1}, \ldots, Z_i^{(p)^T}(t_j) V_{L_1}\right)^T$, $M_0 = I_{mn}$, $M_k = \text{Diag}\left(X_{1k}^T, \ldots, X_{nk}^T\right)$, $k = 1, \ldots, p$; $\nu = \left(b_R^T, U_0^T, U_1^T, \cdots, U_p^T\right)^T$, where $b_R = \left(b_0^T V_{L_1}, b_1^T V_{L_1}, \ldots, b_p^T V_{L_1}\right)^T$, $U_k = \left(U_{1k}^T, \ldots, U_{nk}^T\right)^T$, $U_{ik}^T = (U_{ik}(t_1), \ldots, U_{ik}(t_m))$; and $\Upsilon = \left(\epsilon_1^T, \ldots, \epsilon_n^T\right)^T$, where $\epsilon_i = (\epsilon_i(t_1), \ldots, \epsilon_i(t_m))^T$, for $i = 1, \ldots, n$. Then we can rewrite the model as

$$Y = Z_F b_F + W\nu + \Upsilon, \tag{3}$$

where $E(\Upsilon) = 0$, $\text{cov}(\Upsilon) = \text{Diag}(\Omega_1, \ldots, \Omega_n)$, and $E(\nu) = 0, G = \text{cov}(\nu) = \text{Diag}\,(\text{cov}(b_R), I_n \otimes \Sigma_{u0}, \ldots, I_n \otimes \Sigma_{up})$, $\text{cov}(b_R) = \text{Diag}\left(\sigma_{b0}^2 \text{Diag}(\lambda_{L_1}^{-1}), \ldots, \sigma_{bp}^2 \text{Diag}\left(\lambda_{L_1}^{-1}\right)\right)$. Thus, $\text{cov}(Y) \equiv \Sigma = \Sigma(\delta) = \text{Diag}(\Sigma_1(\delta), \ldots, \Sigma_n(\delta))$, where $\Sigma_i = \Sigma_i(\delta) = Z_{Fi}\,\text{cov}(b_R) Z_{Fi}^T + \text{cov}(U_{i0}) + \sum_{k=1}^p \text{Diag}(X_{ik})\,\text{cov}(U_{ik})\text{Diag}(X_{ik}) + \Omega_i$ for $i = 1, \ldots, n$. For identifiability of $\text{cov}(U_{i0})$, the covariance Ω_i has to be known. Here, $\delta = \left(\sigma_{b_0}^2, \sigma_{b_1}^2, \ldots, \sigma_{b_p}^2, \psi_0, \ldots, \psi_p, \rho_0^T, \ldots, \rho_p^T\right)^T$ includes all the variance component parameters of Σ. The variance components $\psi = (\psi_0, \ldots, \psi_p)^T$ and $\rho = \left(\rho_0^T, \ldots, \rho_p^T\right)^T$ are model-induced variance components due to the covariance of $(U_{i0}, \ldots, U_{ip})^T$, whereas the variance components $\left(\sigma_{b0}^2, \sigma_{b1}^2, \ldots, \sigma_{bp}^2\right)$ are the penalties for B-splines.

Because (3) represents a linear mixed model, the generalized least square estimator for the fixed effects b_F and the best linear unbiased predictors (BLUP) of the random effects are

$$\left.\begin{aligned} \widehat{b}_F(\delta) &= \left\{Z_F^T \Sigma^{-1}(\delta) Z_F\right\}^{-1} Z_F^T \Sigma^{-1}(\delta) Y, \\ \widehat{b}_R(\delta) &= \text{cov}(b_R) Z_R \Sigma^{-1}(\delta)\{Y - Z_F \widehat{b}_F(\delta)\}, \\ \widehat{U}_k(\delta) &= \text{cov}(U_k) M_k \Sigma^{-1}(\delta)\{Y - Z_F \widehat{b}_F(\delta)\},\ k = 0, 1, \ldots, p, \\ \widehat{\nu}(\delta) &= G W^T \Sigma^{-1}(\delta)(Y - Z_F \widehat{b}_F). \end{aligned}\right\} \tag{4}$$

To simplify notations, we will suppress the dependence on δ in the previous expressions. Note that the aforementioned estimators and the predictors involve Σ, a function of unknown δ. The variance components δ are estimated by restricted maximum likelihood method, that is, by maximizing $L(\delta) = |\sum_{i=1}^n Z_{Fi}^T \Sigma_i^{-1} Z_{Fi}|^{-1/2} \prod_{i=1}^n |\Sigma_i|^{-1/2} \exp\{-(1/2) \sum_{i=1}^n (Y_i - Z_F \widehat{b}_F)^T \Sigma_i^{-1} (Y_i - Z_F \widehat{b}_F)\}$ with respect to δ. The estimated variance components $\widehat{\delta}$ are then plugged into (4) to obtain $\widehat{b}_F(\widehat{\delta}), \widehat{b}_R(\widehat{\delta}), \widehat{U}_k(\widehat{\delta})$ and $\widehat{\nu}(\delta)$.

Remark 1. Knot selection is an important practical issue in this procedure. Many heuristic procedures are suggested in literature that often work well in many applications. Ruppert *et al.* (2003) noted that automatic knot selection could be complicated and computationally intensive. One of their recommendations for a non-parametric regression of Y on X is to choose the knots on quantile points of X's with the approximate number of knots being $L = \min(0.25 \times \text{ number of unique } X, 35)$.

4. Prediction and prediction error

4.1. Point prediction

Our objective is to predict $\overline{Y}_i(t_m)$, the ith area mean at time point t_m conditional on the random effects, and

$$\overline{Y}_i(t_m) = Z_{Fi}^T(t_m)b_F + Z_{Ri}^T(t_m)b_R + U_{i0}(t_m) + \sum_{k=1}^{p} X_{ik}(t_m)U_{ik}(t_m) := \tilde{l}^T b_F + \tilde{m}^T v,$$

where $\tilde{l}^T = Z_{Fi}^T(t_m)$ and $\tilde{m}^T = \left(Z_{R_i}^T(t_m), e_{im,0}^T, e_{im,x_1}^T, \ldots, e_{im,x_p}^T\right)$, with $e_{im,0}$ being an $mn \times 1$ column vector of zeros but the (im)th component is 1, and e_{im,x_j} being an $mn \times 1$ column vector of zeros but the (im)th component is $X_{ij}(t_m)$, for $j = 1, \ldots, p$. Then the BLUP for $\overline{Y}_i(t_m)$ is $\widehat{Y}_i(t_m;\delta) = \tilde{l}^T\widehat{b}_F(\delta) + \tilde{m}^T\widehat{v}(\delta)$, and the estimated BLUP (EBLUP) is $\widehat{Y}_i(t_m;\widehat{\delta})$. The estimators $\widehat{b}_F(\delta), \widehat{v}(\delta)$ and $\widehat{\delta}$ are defined in the previous section.

4.2. Asymptotic expression for the mean squared error

Assessing the uncertainty of the EBLUP under the linear mixed model is a challenging task. It is well known that there is no closed-form expression for the MSE of the EBLUP. Prasad & Rao (1990), Das *et al.* (2004), Kackar & Harville (1984) and Booth & Hobert (1998) among others derived the asymptotic expressions for the MSE under parametric models. To simplify the notations, we define the BLUP estimator of $\overline{Y}_i(t_m)$ as $t(\hat{\delta}) = \widehat{Y}_i(t_m;\hat{\delta})$. Next, we break up the MSE of $t(\widehat{\delta})$ (Kackar & Harville, 1984) as

$$\text{MSE}\left\{t(\widehat{\delta})\right\} = E\left\{t(\widehat{\delta}) - \overline{Y}_i(t_m)\right\}^2 = E\left\{t(\delta) - \overline{Y}_i(t_m)\right\}^2 + E\left\{t(\widehat{\delta}) - t(\delta)\right\}^2. \tag{5}$$

The first term on the right-hand side of (5) could be obtained by direct calculation using the moments of the multivariate normal distribution. One of the key steps in obtaining the asymptotic expression for MSE in (5) is to establish an approximation of $E\{t(\widehat{\delta}) - t(\delta)\}^2$ using the first-order Taylor expansion of $t(\widehat{\delta})$ around δ. Namely, under some regularity conditions, we wish to establish

$$E\left\{t(\widehat{\delta}) - t(\delta)\right\}^2 = E\left\{\frac{\partial t(\delta)}{\partial \delta}\left(\widehat{\delta} - \delta\right)\right\}^2 + o\left(n^{-1}\right). \tag{6}$$

Theorem 3.1 in the work of Das *et al.* (2004) provides a general approximation of the expression (6) for the MSE of empirical BLUP in a linear mixed effect model. Unfortunately, we are not able to apply their result, because some of the regularity conditions in their theorems do not hold in our model. In particular, we observe that $\widehat{Y}_i(t_m,\delta)$ can be written as $\widehat{Y}_i(t_m,\delta) = \sum_{k=1}^n\sum_{j=1}^m a_{kj}Y_k(t_j) + \sum_{k=1}^n\sum_{j=1}^m \zeta_{kj}Y_k(t_j)$, where $(a_{k1}, \ldots, a_{km}) = \tilde{l}^T\left(\sum_{k=1}^n Z_{F_k}^T\Sigma_k^{-1}Z_{F_k}\right)^{-1} Z_{F_k}^T\Sigma_k^{-1}$, $\zeta^T(\delta) = \tilde{m}^T G W^T \Sigma^{-1} := \left(\widehat{\zeta}_1^T(\delta), \cdots, \widehat{\zeta}_n^T(\delta)\right)$, so that $\widehat{\zeta}_k^T(\delta) = (\widehat{\zeta}_{k1}, \ldots, \widehat{\zeta}_{km}) = Z_{R_i}^T(t_m)\text{cov}(b_R)Z_{R_k}^T\Sigma_k^{-1}$ for $k \neq i$. If $k = i$, $\widehat{\zeta}_i^T(\delta) = \left\{Z_{R_i}^T(t_m)\text{cov}(b_R)Z_{R_i}^T + \Sigma_{u_0}^{(m)} + \sum_{j=1}^p X_{ij}(t_m)[\Sigma_{u_j}\text{diag}(X_{ij})]^{(m)}\right\}\Sigma_i^{-1}$, where $Q^{(m)}$ denotes the mth row of a generic matrix Q. Now, condition (iii) of theorem 3.1 in the work of Das *et al.* (2004) requires that $\sum_{k=1}^n\sum_{j=1}^m |\partial a_{kj}/\partial\delta_l|$ and $\sum_{k=1}^n\sum_{j=1}^m |\partial\zeta_{ij}/\partial\delta_l|$ to be bounded, and they do hold if $\text{cov}(b_R) = 0$. In our case, $\text{cov}(b_R)$ is not zero, and consequently, $\sum_{k=1}^n\sum_{j=1}^m |\partial\zeta_{ij}/\partial\delta_l|$ is no longer bounded. Therefore, one of the key assumptions of Das *et al.* (2004) does not hold here.

For the asymptotic expressions, we introduce some notations. Let $s = (p+1)(q+2)$ be the number of parameters in the variance component δ, and $\mathcal{V}_l = \text{diag}(\mathcal{V}_{l1}, \cdots, \mathcal{V}_{ln})$, where $1 \le l \le s$ is the index of parameters and $\mathcal{V}_l$ is an $mn \times mn$ matrix. The ith ($1 \le i \le n$) block diagonal $m \times m$ matrix component of $\mathcal{V}_l$ is

$$\mathcal{V}_{li} = \begin{cases} Z_{R_i}^{(l)} \text{diag}(\lambda_{L1}^{-1}) Z_{R_i}^{(l)^T} & \text{for } 1 \le l \le p+1; \\ A_m(\rho_0) & \text{for } l = p+2; \\ \text{diag}(X_{il'}) A_m(\rho_k) \text{diag}\,(X_{il'}) & \text{for } p+3 \le l \le 2(p+1) \\ & \text{and } k = l-(p+2); \\ \partial A_m(\rho_0)/\partial \rho_{0,l'} & \text{for } 2(p+1)+1 \le l \le 2(p+1)+q; \\ & \text{and } l' = l-2(p+1); \\ \text{diag}(X_{il'})\,(\partial A_m(\rho_k)/\partial \rho_{k,l'})\,\text{diag}\,(X_{il'}) & \text{for } 2(p+1)+(k-1)q+1 \le l \le 2(p+1)+kq, \\ & 2 \le k \le p+1 \text{ and } l' = l-2(p+1)+(k-1)q, \end{cases}$$

where $Z_{R_i}^{(l)} = \left(Z_{R_i}^{(l)}(t_1), \ldots, Z_{R_i}^{(l)}(t_m)\right)^T$ is an m-dimensional column vector, $\rho_{k,l'}$ is the l'th component of ρ_k and l' is used to index the components in the q-dimensional variance component vector ρ_k. Define $D = I - Z_F \left(Z_F^T \Sigma^{-1} Z_F\right)^{-1} Z_F^T \Sigma^{-1}$ and $P = \Sigma^{-1} D$. Let $\nabla\zeta = [\partial\zeta(\delta)/\partial\delta_1, \ldots, \partial\zeta(\delta)/\partial\delta_s]^T$ and $\eta = A^{-1}h$, where $l(\delta) = \log\{L(\delta)\}$, $A = E\{\partial^2 \ell(\delta)/\partial\delta^2\} = \{-0.5\text{tr}(P\mathcal{V}_{k_1} P\mathcal{V}_{k_2})\}_{1\le k_1,k_2\le s}$ is an $s \times s$ matrix and $h = (h_1^T, \ldots, h_s^T)^T$, $h_i = \{\tilde{l}^T - \zeta^T(\delta) Z_F\} \left(Z_F^T \Sigma^{-1} Z_F\right)^{-1} Z_F^T (\partial\Sigma/\partial\delta_i) D + \{\partial\zeta(\delta)/\partial\delta_i\} D$. Next, to establish (6) for our model, we assume that the following regularity conditions hold.

(a) The elements of Z_R and Z_F are uniformly bounded, and $n^{-1}\sum_{i=1}^n Z_{F_i}^T \Sigma_i^{-1} Z_{F_i} = O_p(1)$.
(b) The smallest and largest eigenvalues of Σ_i, $\partial\Sigma_i/\partial\delta_j$, $\partial^2\Sigma_i/\partial\delta_j\partial\delta_k$, $n^{-1}\sum_{i=1}^n Z_{F_i}^T \Sigma_i^{-1} Z_{F_i}$, $n^{-1}\sum_{i=1}^n Z_{F_i}^T (\partial\Sigma_i/\partial\delta_j) Z_{F_i}$ and $n^{-1}\sum_{i=1}^n Z_{F_i}^T \left(\partial^2\Sigma_i/\partial\delta_j\partial\delta_k\right) Z_{F_i}$ are uniformly (for all $1 \le i \le n$) bounded away from 0 and infinity.
(c) $A_m(\rho_k)$ has bounded second derivatives with respect to $\rho_{k,l'}$ for $1 \le l' \le q$.
(d) The eigenvalues $\lambda_{L_1}^{-1} = o(n^{-1/2})$.

Condition (a) is similar to the condition given in the work of Prasad & Rao (1990). Condition (b) can be easily satisfied because all the matrices are of fixed $m \times m$ dimension. Condition (c) is a mild condition on smoothness of $A_m(\rho_k)$ with respect to the parameters in ρ_k. Condition (d) is used to guarantee that the higher-order terms in the Taylor expansion in (5) are of order $o(n^{-1})$. In lemma 4 in the Supporting Information, we establish (6) under conditions (a)–(d). This result is one of the key steps in proving the following theorem that summarizes the MSE of $t(\widehat{\delta})$.

Theorem 1. *Under conditions (a)–(d), the MSE of the predictor* $t(\widehat{\delta}) = \widehat{Y}_i(t_m, \widehat{\delta})$ *is*

$$MSE\left\{t(\widehat{\delta})\right\} = g_1(\delta) + g_2(\delta) + g_3(\delta) + g_4(\delta) + o\left(n^{-1}\right), \tag{7}$$

where $g_1(\delta) = \tilde{m}^T (G - GW^T \Sigma^{-1} WG)\tilde{m}$, $g_2(\delta) = (\tilde{l} - Z_F^T \Sigma^{-1} WG\tilde{m})^T \left(Z_F^T \Sigma^{-1} Z_F\right)^{-1} \left(\tilde{l} - Z_F^T \Sigma^{-1} WG\tilde{m}\right)$, $g_3(\delta) = -tr\,(\nabla\zeta\Sigma\nabla\zeta^T A^{-1})$ *and* $g_4(\delta) = 4\sum_{j=1}^s \sum_{l=1}^s \eta_j^T \Sigma(P\mathcal{V}_j P\mathcal{V}_l P + P\mathcal{V}_l P\mathcal{V}_j P)\Sigma\eta_l$.

Note that this MSE expression is different from those derived for the longitudinal model (Prasad & Rao, 1990; Datta & Lahiri, 2000). There are two important differences. First, unlike longitudinal models, all areas share the same random effects b_R. Then, W in (3) cannot be

written as a block diagonal matrix (Prasad & Rao, 1990; Datta & Lahiri, 2000), and consequently, the coefficient vector $\zeta(\delta)$ corresponding to the random effects in the prediction is no longer a sparse vector. By sparsity of a vector, we refer to the scenario where many components are 0 and only a few of them are non-zero. Sparsity of $\zeta(\delta)$ is crucial in showing that the $g_4(\delta)$ is a smaller order of n^{-1} (see Prasad & Rao, 1990, and the proof in their Appendix). Second, unlike the assumptions made in previous work (Datta & Lahiri, 2000; Das *et al.*, 2004), the variance components in G are not linear in parameters $\rho_0, \ldots, \rho_p$. Now, we outline the main steps in the proof of Theorem 1.

Proof of theorem 1: Using a standard computation (Das *et al.*, 2004), we have $\text{MSE}\{t(\delta)\} = g_1(\delta) + g_2(\delta)$. Combining with (5), we obtain

$$\text{MSE}\left\{t(\widehat{\delta})\right\} = g_1(\delta) + g_2(\delta) + E\left\{t(\widehat{\delta}) - t(\delta)\right\}^2. \tag{8}$$

Next, using lemmas 3 and 4 given in the Supporting Information, $E\{t(\widehat{\delta}) - t(\delta)\}^2 = E\left[\{\partial t(\delta)/\partial\delta\}\left(\widehat{\delta} - \delta\right)\right]^2 + o\left(n^{-1}\right) = E\left\{h^T(\widehat{\delta} - \delta)\right\}^2 + o\left(n^{-1}\right) = E\left(h^T A^{-1} a\right)^2 + o\left(n^{-1}\right)$, where h and A are defined before, $a = \partial\ell(\delta)/\partial\delta := (a_1, \cdots, a_s)^T$ with $a_i = \partial\ell(\delta)/\partial\delta_i = (1/2)\left\{\epsilon^T P\mathcal{V}_i P\epsilon - \text{tr}(P\mathcal{V}_i)\right\}$ and $\epsilon = Y - Z_F b_F$. Therefore, by lemma A.2 in Prasad & Rao (1990), we know

$$\begin{aligned} E\left(h^T A^{-1} a\right)^2 &= -\text{tr}\left(h\Sigma h^T A^{-1}\right) + 4\sum_{j=1}^{s}\sum_{l=1}^{s}\eta_j^T\Sigma\left(P\mathcal{V}_j P\Sigma P\mathcal{V}_l P + P\mathcal{V}_l P\Sigma P\mathcal{V}_j P\right)\Sigma\eta_l \\ &= -\text{tr}\left(h\Sigma h^T A^{-1}\right) + g_4(\delta). \end{aligned} \tag{9}$$

Notice that $h_j = h_j^{(0)} + \partial\zeta(\delta)/\partial\delta_j$, where $h_j^{(0)} = \left\{\tilde{l}^T - \zeta^T(\delta)Z_F\right\}\left(Z_F^T\Sigma^{-1}Z_F\right)^{-1} Z_F^T(\partial\Sigma/\partial\delta_j)D - \{\partial\zeta(\delta)/\partial\delta_j\}Z_F\left(Z_F^T\Sigma^{-1}Z_F\right)^{-1}Z_F^T\Sigma^{-1}$. By condition (b), it can be seen that each element in $Z_F\left(Z_F^T\Sigma^{-1}Z_F\right)^{-1}Z_F^T$ is of order n^{-1}. Because Σ is a block diagonal matrix with finite-dimensional blocks, all the elements in $Z_F\left(Z_F^T\Sigma^{-1}Z_F\right)^{-1}Z_F^T\Sigma^{-1}$ are of order n^{-1}, and D is a matrix with all the off-diagonal elements at the order of n^{-1}. By condition (d), we know each element in $\zeta(\delta)$ is of order $o(n^{-1/2})$. It then can be shown that each element in $h_j^{(0)}(\delta)$ are $o(n^{-1/2})$. Then $\text{tr}\ (h\Sigma h^T A^{-1}) = \text{tr}\ \left(\nabla\zeta\Sigma\nabla\zeta^T A^{-1}\right) + o\left(n^{-1}\right) = g_3(\delta) + o(n^{-1})$, and combining (8) and (9), we obtain (7). □

4.3. Estimation of the mean squared error

Because the approximated MSE (7) is a function of unknown parameters δ, it is not computable. The seminal work of Prasad & Rao (1990) showed that simply plugging in a consistent estimator $\widehat{\delta}$ in place of δ may produce an underestimated MSE. Therefore, in the following theorem, we state how to consistently estimate the MSE.

Theorem 2. *Under conditions (a)–(d), the MSE* $\text{MSE}\{t(\widehat{\delta})\}$ *can be estimated by*

$$eMSE\left\{t(\widehat{\delta})\right\} = g_1\left(\widehat{\delta}\right) + g_2\left(\widehat{\delta}\right) + 2g_3\left(\widehat{\delta}\right) + g_4\left(\widehat{\delta}\right),$$

and the bias of $eMSE\{t(\widehat{\delta})\}$ *in estimating* $MSE\left\{t(\widehat{\delta})\right\}$ *is a smaller order of* n^{-1}*, which means* $E\left[eMSE\{t(\widehat{\delta})\}\right] = MSE\left\{t\left(\widehat{\delta}\right)\right\} + o\left(n^{-1}\right)$.

Note that this approximation is based on the number of curves rather than the number of times points where the data were observed. This second-order correction often has a significant impact in real applications as shown in the work of Rao (2003).

Proof of theorem 2: Let $\widehat{\delta}$ be the REML estimate. From lemma 3 given in the Supporting Information, we know that $\widehat{\delta}-\delta = -A^{-1}a + o_p(n^{-1})$. The Taylor series expansion of $g_1(\widehat{\delta})$ around δ gives

$$g_1\left(\widehat{\delta}\right) = g_1(\delta) + \left(\widehat{\delta}-\delta\right)^T \nabla g_1(\delta) + \frac{1}{2}\,\mathrm{tr}\,\left\{\left(\widehat{\delta}-\delta\right)\left(\widehat{\delta}-\delta\right)^T H_{g_1}(\delta)\right\} + R(\delta^*), \quad (10)$$

where $R(\delta^*) = 0.5\mathrm{tr}\left\{\left(\widehat{\delta}-\delta\right)\left(\widehat{\delta}-\delta\right)^T \Delta H_{g_1}\right\}$, $\Delta H_{g_1} = H_{g_1}(\delta^*) - H_{g_1}(\delta)$, $\|\delta^*-\delta\| \le \|\widehat{\delta}-\delta\|$, and H_{g_1} is the second derivative of g_1 with respect to δ. We want to show that $R(\delta^*)$ is of order $o_p\left(n^{-1}\right)$. Let $\gamma^T(\delta) = \tilde{m}^T G W$. Then the (i,j)th component of the second derivative of $g_1(\delta)$ is

$$\begin{aligned} H_{g_1}^{ij} = \frac{\partial^2 g_1(\delta)}{\partial\delta_i\partial\delta_j} &= \tilde{m}^T \frac{\partial^2 G}{\partial\delta_i\partial\delta_j}\tilde{m} - 2\frac{\partial^2\gamma^T(\delta)}{\partial\delta_i\partial\delta_j}\Sigma^{-1}\gamma(\delta) - 2\frac{\partial\gamma^T(\delta)}{\partial\delta_i}\Sigma^{-1}\frac{\partial\gamma(\delta)}{\partial\delta_j} \\ &+ 2\frac{\partial\gamma^T(\delta)}{\partial\delta_i}\Sigma^{-1}\frac{\partial\Sigma}{\partial\delta_j}\Sigma^{-1}\gamma(\delta) + 2\frac{\partial\gamma^T(\delta)}{\partial\delta_j}\Sigma^{-1}\frac{\partial\Sigma}{\partial\delta_i}\Sigma^{-1}\gamma(\delta) \\ &- 2\gamma^T(\delta)\Sigma^{-1}\frac{\partial\Sigma}{\partial\delta_j}\Sigma^{-1}\frac{\partial\Sigma}{\partial\delta_i}\Sigma^{-1}\gamma(\delta) + \gamma^T(\delta)\Sigma^{-1}\frac{\partial^2\Sigma}{\partial\delta_j\partial\delta_i}\Sigma^{-1}\gamma(\delta) := K_1 + \ldots + K_7. \end{aligned}$$

Because $\widehat{\delta}-\delta = O_p(n^{-1/2})$ and $\mathrm{tr}\{(\widehat{\delta}-\delta)(\widehat{\delta}-\delta)^T\Delta H_{g_1}\} = \sum_{i=1}^s\sum_{j=1}^s(\widehat{\delta}_i-\delta_i)(\widehat{\delta}_j-\delta_j)\Delta H_{g_1}^{ij}$, where $\Delta H_{g_1}^{ij}$ is the (i,j)th element of ΔH_{g_1}, we want to show that $|\Delta H_{g_1}^{ij}| \le C\|\widehat{\delta}-\delta\|$. Let K_l^* denote the corresponding K_l evaluated at δ^*. Then we know that $|\Delta H_{g_1}^{ij}| \le \sum_{l=1}^7 |K_l^* - K_l|$. In the Supporting Information, we show that $|K_l^* - K_l| \le C\|\delta^*-\delta\|$ for $l = 1,\ldots,7$. Therefore,

$$\begin{aligned} E\left[\mathrm{tr}\left\{\left(\widehat{\delta}-\delta\right)\left(\widehat{\delta}-\delta\right)^T\Delta H_{g_1}\right\}\right] &\le \sum_{i=1}^s\sum_{j=1}^s E\left|\left(\widehat{\delta}_i-\delta_i\right)\left(\widehat{\delta}_j-\delta_j\right)\right||\Delta H_{g_1}^{ij}| \\ &\le C\sum_{i=1}^s\sum_{j=1}^s E\left|\left(\widehat{\delta}_i-\delta_i\right)\left(\widehat{\delta}_j-\delta_j\right)\right|\|\widehat{\delta}-\delta\| = o\left(n^{-1}\right). \end{aligned}$$

From (10), we have $E\left\{g_1(\widehat{\delta})\right\} = g_1(\delta) + (1/2)\mathrm{tr}\left\{\mathrm{var}\left(\widehat{\delta}\right)H_{g_1}(\delta)\right\} + o\left(n^{-1}\right) = g_1(\delta) - g_3(\delta) + o(n^{-1})$. Next, we would like to show that

$$E\left\{g_2(\widehat{\delta})\right\} = g_2(\delta) + o\left(n^{-1}\right). \quad (11)$$

Again, by Taylor's expansion of $g_2(\widehat{\delta})$ around δ, we have $g_2(\widehat{\delta}) = g_2(\delta) + \{\partial g_2(\delta^*)/\partial\delta\}(\widehat{\delta}-\delta)$, where $\|\delta^*-\delta\| \le \|\widehat{\delta}-\delta\|$. The first derivative of $g_2(\delta)$ is

$$\begin{aligned} \frac{\partial g_2(\delta)}{\partial\delta_i} &= 2\left\{\tilde{l} - Z_F^T\frac{\partial\zeta(\delta)}{\partial\delta}\right\}^T\left(Z_F\Sigma^{-1}Z_F\right)^{-1}\left\{\tilde{l} - Z_F^T\zeta(\delta)\right\} \\ &+ \left\{\tilde{l} - Z_F^T\zeta(\delta)\right\}^T\left(Z_F'\Sigma^{-1}Z_F\right)^{-1}Z_F^T\Sigma^{-1}\frac{\partial\Sigma}{\partial\delta_i}\Sigma^{-1}Z_F\left(Z_F^T\Sigma^{-1}Z_F\right)^{-1}\left\{\tilde{l} - Z_F^T\zeta(\delta)\right\}. \end{aligned}$$

Because of $|\partial g_2(\delta)/\partial\delta_i| = o(n^{-1/2})$ and $|\partial g_2(\delta^*)/\partial\delta_i - \partial g_2(\widehat{\delta})/\partial\delta_i| \le C\|\widehat{\delta}-\delta\|$, we have

$$\left|E\left\{\frac{\partial g_2(\delta^*)}{\partial\delta}(\widehat{\delta}-\delta)\right\}\right| \le \sum_{i=1}^{s}\left|\frac{\partial g_2(\delta)}{\partial\delta_i}\right| E|\widehat{\delta}_i-\delta_i|\{1+o(1)\} = o\left(n^{-1}\right).$$

Hence, we obtain (11). Next, consider $g_3(\delta)$. By the Taylor expansion, we have

$$g_3(\widehat{\delta}) = g_3(\delta) + \frac{\partial g_3(\delta^*)}{\partial\delta}\left(\widehat{\delta}-\delta\right),$$

where $\|\delta^*-\delta\| \le \|\widehat{\delta}-\delta\|$ and

$$\begin{aligned}\frac{\partial g_3(\delta)}{\partial\delta_i} &= -\mathrm{tr}\left\{\frac{\nabla\zeta(\delta)}{\delta_i}\Sigma\nabla\zeta(\delta)^T A^{-1}\right\} - \mathrm{tr}\left\{\nabla\zeta(\delta)\Sigma\frac{\nabla\zeta(\delta)}{\delta_i}^T A^{-1}\right\}\\ &\quad - \mathrm{tr}\left\{\nabla\zeta(\delta)\frac{\partial\Sigma}{\partial\delta_i}\nabla\zeta(\delta)^T A^{-1}\right\} + \mathrm{tr}\left\{\nabla\zeta(\delta)\Sigma\nabla\zeta(\delta)^T A^{-1}\frac{\partial A}{\partial\delta_i}A^{-1}\right\}.\end{aligned}$$

Because the derivation for each terms in $\partial g_3(\delta)/\partial\delta_i$ are similar, we only show the third term. Let A^{lk} be the (l,k)th component of A^{-1}. Then

$$\begin{aligned}\mathrm{tr}\left\{\nabla\zeta(\delta)\frac{\partial\Sigma}{\partial\delta_i}\nabla\zeta(\delta)^T A^{-1}\right\} &= \sum_{l=1}^{s}\sum_{k=1}^{s}\frac{\partial\zeta^T(\delta)}{\partial\delta_k}\frac{\partial\Sigma}{\partial\delta_i}\frac{\partial\zeta(\delta)}{\partial\delta_l}A^{lk}\\ &= \sum_{l=1}^{s}\sum_{k=1}^{s}\sum_{j=1}^{n}\frac{\partial\zeta_j^T(\delta)}{\partial\delta_k}\frac{\partial\Sigma_j}{\partial\delta_i}\frac{\partial\zeta_j(\delta)}{\partial\delta_l}A^{lk}.\end{aligned}$$

It can be shown that $|\left\{\partial\zeta_j^T(\delta^*)/\partial\delta_k\right\}\left\{\partial\Sigma_j^*/\partial\delta_i\right\}\{\partial\zeta_j(\delta^*)/\partial\delta_l\} - \left\{\partial\zeta_j^T(\delta)/\partial\delta_k\right\}\{\partial\Sigma_j/\partial\delta_i\}\{\partial\zeta_j(\delta)/\partial\delta_l\}| \le C\|\widehat{\delta}-\delta\|$ and $|A^{*lk}-A^{lk}| \le C\|\widehat{\delta}-\delta\|$. Then $|\mathrm{tr}\left\{\nabla\zeta(\delta^*)(\partial\Sigma^*/\partial\delta_i^*)\nabla\zeta(\delta)^T A^{*-1}\right\}| \le |\mathrm{tr}(\nabla\zeta(\delta)(\partial\Sigma/\partial\delta_i)\nabla\zeta(\delta)^T A^{-1})|(1+C\|\widehat{\delta}-\delta\|)$. Similarly, the other terms in $\partial g_3(\delta^*)/\partial\delta_i$ can be shown to be bounded. Notice that $\partial g_3(\delta)/\partial\delta_i = o(n^{-1/2})$. Hence, $|E[\{\partial g_3(\delta^*)/\partial\delta\}(\widehat{\delta}-\delta)]| \le E[|\{\partial g_3(\delta)/\partial\delta_i\}(\widehat{\delta}-\delta)|\{1+o(1)\}] = o(n^{-1})$. Therefore, $E\{g_3(\widehat{\delta})\} = g_3(\delta)+o(n^{-1})$. The proof for $g_4(\widehat{\delta})$ is similar to that for g_2 and g_3 by noting that $\partial g_4(\delta)/\partial\delta_i = o(n^{-1/2})$, and the details are given in the Supporting Information. Therefore, it can be shown that $E\{g_4(\widehat{\delta})\} = g_4(\delta)+o(n^{-1})$, and hence, the proof follows. □

5. Simulation study

In this section, we conduct a simulation study to illustrate the performance of our new method of estimation and compare it with some existing approaches.

Simulation design: We considered $n = 32$ small areas, and on each area, we observed $m = 16$ timescaled observations. The data were generated from three designs, and for all designs, the structure of the small area-specific errors remained the same. First, we simulated iid X from Normal$(0,1)$ distribution. Then, we generated responses based on the following three designs: design 1: $Y_i(t) = 2+2X_i(t)+U_{i0}(t)+U_{i1}(t)X_i(t)+\epsilon_i(t)$; design 2: $Y_i(t) = \beta_0(t)+\beta_1(t)X_i(t)+U_{i0}(t)+U_{i1}(t)X_i(t)+\epsilon_i(t)$, with $\beta_0(t) = 2\cos(2\pi t/m)$, $\beta_1(t) = 0.4$ for all t; and design 3: $Y_i(t) = \beta_0(t)+\beta_1(t)X_i(t)+U_{i0}(t)+U_{i1}(t)X_i(t)+\epsilon_i(t)$ with $\beta_0(t) = \beta_1(t) = 2\sin(1.5\pi t/m)$. Here, the time variable t takes on values $1,\ldots,m$. For all

three designs, we took $(U_{i0}(1),\ldots,U_{i0}(m))^T \overset{iid}{\sim} \text{Normal}_m(0,\Sigma_0)$, $(U_{i1}(1),\ldots,U_{i1}(m))^T \overset{iid}{\sim} \text{Normal}_m(0,\Sigma_1)$, with $\Sigma_0 = \Sigma_1 = \psi_0 A_m(0.1)$, and the (j,k)th element of $A_m(\rho)$ is defined as $\rho^{|j-k|}$, $j,k = 1,\ldots,m$. Here, Normal_m denotes an m-variate normal distribution. We set $\psi_0 = 0.1$ and took $\epsilon_i(t) \sim \text{Normal}\left(0, \sqrt{0.025}^2\right)$, $i = 1,\ldots,8$, $\epsilon_i(t) \sim \text{Normal}\left(0, \sqrt{0.05}^2\right)$, $i = 9,\ldots,16$, $\epsilon_i(t) \sim \text{Normal}\left(0, \sqrt{0.1}^2\right)$, $i = 17,\ldots,24$ and $\epsilon_i(t) \sim \text{Normal}\left(0, \sqrt{0.2}^2\right)$, $i = 25,\ldots,32$. Observe that we took four different error variances for four subsets of the small areas; in particular, the ratio $\text{var}\{\epsilon_i(t)\}/\psi_0$ is a non-decreasing function of i. We want to point out the difference between the three designs. In design 1, the coefficients are not varying; in design 2, the intercept is a time-varying function, while the slope is a constant; and in design 3, both the intercept and the slope are time varying. This simulation study will be referred to as scenario 1.

Method of analyses: We analysed the simulated data using the proposed method referred to as method I, the following model referred to as method II (Rao & Yu, 1994)

$$\begin{cases} Y_i(t) = \gamma_0 + \gamma_1 X_i(t) + a_i + \alpha_i(t) + \epsilon_i(t), \\ a_i \sim \text{Normal}\left(0,\sigma_a^2\right),\ \alpha_i(t) = \rho\alpha_i(t-1) + \xi_i(t),\ \xi_i(t) \overset{iid}{\sim} \text{Normal}\left(0,\sigma_\xi^2\right), |\rho| < 1,\ t = 1,\ldots,m, \end{cases}$$

and using the following linear mixed model referred to as method III (Datta *et al.*, 2002)

$$\begin{cases} Y_i(t) = \gamma_0 + \gamma_1 X_i(t) + a_i + \alpha_i(t) + \epsilon_i(t), \\ a_i \sim \text{Normal}\left(0,\sigma_a^2\right),\ \alpha_i(t) = \alpha_i(t-1) + \zeta_i(t),\ \zeta_i(t) \overset{iid}{\sim} \text{Normal}\left(0,\sigma_\zeta^2\right),\ t = 1,\ldots,m. \end{cases}$$

Additionally, we also considered direct survey estimates, and that is referred to as method IV. In method I, and for all designs, we have used $(2,5,8,11,14)$ as the inner knot points for the B-splines. Also, in the simulation and in the real data analysis, we have used cubic $(d = 3)$ B-splines as they are the lowest degree splines with added flexibility at the end points. The performance of these methods are judged via prediction and its accuracy. We considered prediction of Y for four area indices $i = 8,16,24,32$, selected from four different variance groups, and at time point $m = 16$. Our results are based on $R = 500$ replications. For the ith area, let $\widehat{Y}_i^r(16)$ be predicted value for $\overline{Y}_i^r(16)$ at the rth replication and the corresponding eMSE be denoted by $\text{eMSE}_i^r(16)$, for $r = 1,\ldots,500$. We compare the performance of the methods based on the following statistics: absolute bias (AB), absolute relative bias (ARB), coefficient of variation (CV), MSE, estimated MSE (eMSE), relative bias of eMSE (RBeM) and 95 per cent coverage probability (CP), where $\text{AB} = (1/500)\sum_{r=1}^{500} |\widehat{Y}_i^r(16) - \overline{Y}_i^r(16)|$, $\text{ARB} = (1/500)\sum_{r=1}^{500} |\left\{\widehat{Y}_i^r(16) - \overline{Y}_i^r(16)\right\}/\overline{Y}_i^r(16)|$, $\text{CV} = (1/500)\sum_{r=1}^{500} \sqrt{\text{eMSE}_i^r(16)}/\widehat{Y}_i^r(16)$, $\text{MSE} = (1/500)\sum_{r=1}^{500}(\widehat{Y}_i^r(16) - \overline{Y}_i^r(16))^2$, $\text{RBeM} = \left\{(1/500)\sum_{r=1}^{500}\text{eMSE}_i^r(16) - \text{MSE}\right\}/\text{MSE}$, $\text{CP} = (1/500)\sum_{r=1}^{500} I\left\{\overline{Y}_i^r(16) \in \widehat{Y}_i^r(16) \pm 1.96\sqrt{\text{eMSE}_i^r(16)}\right\}$, where I denotes the indicator function. Additionally, each table contains $\text{MSE2} = \{g_1(\delta) + g_2(\delta)\}_{\delta=\widehat{\delta}}$, to show the underestimation of uncertainty when δ is unknown and is estimated from the data. Additionally, for assessing efficiency, in the tables, we provide ratio of the MSE to the MSE of method I. Although Rao & Yu (1994) proposed a separate approach for estimating their variance components, in our simulations, we used the same Restricted maximum likelihood (REML) approach for estimating variance components for all three methods for a fair comparison. For method IV, $Y_i^r(16)$ was the direct survey estimator for $\overline{Y}_i^r(16)$, for $r = 1,\ldots,500$, and $\text{var}\{\epsilon_i(16)\}$ was used as the corresponding eMSE. Here, $Y_i^r(16)$ denotes the observed data $Y_i(16)$ in the rth replication.

Table 1. *Results of the simulation scenario 1 where* $cov\{U_{i0}(t_1,\ldots,t_m)\} = cov\{U_{i1}(t_1,\ldots,t_m)\} = 0.1A_m(0.1)$

		AB	ARB	CV	MS	eMS	RBe	CP	MS2	Eff	AB	ARB	CV	MS	eMS	RBe	CP	MS2	Eff	AB	ARB	CV	MS	eMS	RBe	CP	MS2	Eff
		Design 1 $\beta_0(t) = 2$ and $\beta_1(t) = 2$									Design 2 $\beta_0(t) = 2\cos\{2\pi(t/m)\}$ and $\beta_1(t) = 0.4$									Design 3 $\beta_0(t) = \beta_1(t) = 2\sin(1.5\pi t/m)$								
σ^2/ψ_0																												
0.25	I	1.0	0.5	0.7	0.2	0.2	**−0.3**	9.5	0.2	10	1.0	0.5	0.9	0.2	0.2	**−0.2**	9.5	0.2	10	1.0	0.5	0.1	0.2	0.2	0.8	9.5	0.2	10
0.5		1.5	0.7	0.1	0.4	0.4	**−0.8**	9.4	0.4	10	1.5	0.7	1.2	0.4	0.4	**−0.5**	9.4	0.4	10	1.4	0.6	**−1.2**	0.4	0.5	1.0	9.6	0.4	10
1		1.6	0.8	**−0.1**	0.6	0.6	**−0.1**	9.5	0.6	10	1.7	0.8	1.5	0.7	0.7	0.2	9.5	0.6	10	1.7	0.8	**−1.7**	0.7	0.8	2.1	9.6	0.7	10
2		2.0	1.0	1.9	0.9	0.9	**−0.0**	9.4	0.9	10	2.1	1.0	1.8	1.0	1.0	0.9	9.6	1.0	10	2.1	1.0	0.6	1.0	1.2	1.9	9.7	1.0	10
0.25	II	0.9	0.5	0.7	0.2	0.2	2.2	9.7	0.2	8.2	1.0	0.6	0.8	0.2	0.2	1.4	9.7	0.2	9.6	1.0	0.5	**−0.7**	0.2	0.3	1.5	9.6	0.3	9.1
0.5		1.3	0.4	0.6	0.4	0.4	1.0	9.7	0.4	9.5	1.5	0.8	1.0	0.5	0.5	**−0.2**	9.6	0.5	11.4	1.5	0.5	**−0.6**	0.5	0.5	0.1	9.6	0.5	11.2
1		1.8	3.4	5.6	0.6	0.7	0.4	9.6	0.7	10.2	2.2	1.5	2.0	1.0	0.9	**−0.5**	9.3	0.9	15.1	2.1	4.0	**−4.3**	1.0	1.0	**−0.2**	9.5	1.0	15.0
2		2.3	0.6	0.7	1.3	1.0	**−2.3**	9.2	1.0	14.2	3.2	1.5	1.9	2.2	1.7	**−2.0**	9.3	1.7	22.7	2.9	0.8	**−0.9**	2.1	1.9	**−1.2**	9.3	1.9	21.6
0.25	III	1.0	0.4	0.7	0.2	0.2	0.7	9.6	0.2	10.5	1.0	0.5	0.8	0.2	0.2	1.5	9.6	0.2	9.6	1.0	0.5	**−0.7**	0.2	0.2	1.4	9.6	0.2	9.6
0.5		1.4	0.4	0.6	0.4	0.4	0.5	9.6	0.4	10.8	1.5	0.7	1.0	0.5	0.5	0.5	9.6	0.5	10.7	1.6	0.5	**−0.6**	0.5	0.5	**−0.2**	9.5	0.5	11.2
1		2.1	4.4	6.7	0.9	0.8	**−1.0**	9.4	0.8	13.5	2.0	1.3	2.0	0.9	0.9	**−0.1**	9.5	0.9	13.7	2.1	4.6	**−7.5**	1.1	1.0	**−0.9**	9.3	1.0	16.2
2		2.6	0.7	0.8	1.6	1.3	**−1.8**	9.2	1.3	17.1	2.8	1.1	1.7	1.8	1.6	**−0.8**	9.3	1.6	18.6	3.0	0.8	**−0.9**	2.1	1.9	**−1.0**	9.5	1.9	21.4
0.25	IV	1.2	0.5	0.7	0.2	0.3	1.4	9.6		10.0	1.1	0.6	0.8	0.2	0.3	1.4	9.6		10.0	1.2	0.5	**−0.7**	0.2	0.3	1.4	9.6		9.6
0.5		1.8	0.5	0.6	0.5	0.5	0.2	9.6		12.9	1.8	0.8	1.0	0.5	0.5	0.2	9.6		11.7	1.8	0.5	**−0.6**	0.5	0.5	0.2	9.6		11.4
1		2.5	4.5	4.6	1.0	1.0	**−0.3**	9.5		16.4	2.5	1.3	2.0	1.0	1.0	**−0.3**	9.5		15.9	2.5	4.5	**−4.6**	1.0	1.0	**−0.3**	9.5		15.6
2		3.7	0.8	1.0	2.3	2.0	**−1.2**	9.3		24.4	3.7	1.6	1.9	2.3	2.0	**−1.2**	9.3		23.9	3.7	0.6	**−0.9**	2.3	2.0	**−1.2**	9.3		22.9

Here, methods I, II, III and IV refer to the proposed method, the approach proposed by Rao & Yu (1994), the approach proposed by Datta, Lahiri & Maiti (2002) and the direct survey estimates, respectively. Here, AB = absolute bias, ARB = absolute relative bias, CV = coefficient of variation, MS = mean squared error, eMS = estimated MS, RBe = relative bias of eMS, CP = Wald-type 95 per cent coverage probability, $MS2 = [MSE\{t(\delta)\}]_{\delta=\hat{\delta}} = [MS\{t(\delta)\}]_{\delta=\hat{\delta}}$ and Eff denotes the ratio of the MS of a method to the corresponding MS of method I. All entries of the table are multiplied by 10.

To implement the proposed method both in the simulation and in the real data analysis, we have used R statistical software. In particular, the computation involves two steps, estimation and MSE estimation. The key step of estimation is the estimation of the variance components using an optimization routine. Once they are estimated, the other components are estimated using (4). Next, for the MSE estimation, we write code for calculating $g_1(\widehat{\delta})$, $g_2(\widehat{\delta})$, $g_3(\widehat{\delta})$ and $g_4(\widehat{\delta})$.

Results: Results given in Table 1 show that for designs 2 and 3 where at least one of the two coefficients vary over time, the proposed method (I) generally outperforms the other two approaches in terms of bias and MSE especially when $\text{var}\{\epsilon_i(t)\}/\psi_0 \geq 1$. The maximum ARB of point estimates for the proposed method is 13 per cent, whereas this is 33 per cent for method II and 41 per cent for method III. While the maximum CV is 27 per cent for our method, this is 43 per cent and 74 per cent for methods II and III. The maximum MSE for our method is 6 per cent, while this is 21 per cent for both the methods II and III. The gain is more prominent when $\text{var}\{\epsilon_i(t)\}/\psi_0$ is large, meaning that the small area models are more effective. On the other hand, for design 1, method I is at least as good as the other two methods in terms of all the statistics. Compared with methods I–III, usually method IV has larger bias and MSE. This result is expected and is confirming the effectiveness of functional approach over existing linear model-based approaches, and the direct survey estimators. Also, when $\text{var}\{\epsilon_i(t)\}/\psi_0 = 0.25$, there is a very high correlation between the model-based estimator and the direct estimator, and that correlation decreases with $\text{var}\{\epsilon_i(t)\}/\psi_0$. Furthermore, the results indicate the performance of eMSE is very satisfactory in terms of relative bias of this estimator and the coverage probability. Also, a close comparison between MSE2 and eMSE reveals that failure to consider $g_3(\delta)$ and $g_4(\delta)$ in the MSE estimation results in underestimation of the uncertainty in some cases.

To study the robustness of the proposed method towards a different variance structure of U_{i0} and U_{i1}, we simulated data using $\Sigma_0 = \Sigma_1 = 0.2J_m + 0.1A_m(0.1)$, while the rest of the design remain unchanged, and we shall refer it to as scenario 2. Here, J_m denotes an $m \times m$ matrix with all entries are one. This was the variance structure of Rao & Yu (1994), and this variance–covariance structure was different from the assumed structure in the proposed method of analysis. The results presented in Table 2 show the supremacy of method I over the other approaches for larger values of $\text{var}\{\epsilon_i(t)\}/\psi_0$ even when the model assumption regarding the variances of U_{i0} and U_{i1} is violated.

The simulation results clearly establish the effectiveness of the proposed method for small area estimation when both cross-sectional and times series data are available.

6. Analysis of unemployment data

For the purpose of illustration of the proposed method, we present a secondary data analysis of the unemployment data. Unemployment rate analysis is currently a topic of interest for policymaking. Monthly or quarterly unemployment data are often used for various socioeconomic benefits. From the Bureau of Labor Statistics website http://www.bls.gov/lau/, we obtained monthly unemployment rate for 51 states including Hawaii, and Washington DC, and excluding Puerto Rico from January 2001 to December 2011. We considered logarithm of unemployment rate as the response variable. Then we transformed the data to quarterly data by taking average of 3-month log-transformed unemployment rates and consequently calculated the standard errors based on 3-month data. Because the proposed method was developed under the assumption that the area-level error variance Ω_i is known, the quarterly average will be taken as the small area-specific estimate, and the corresponding standard error will be considered as variance of $\epsilon_i(t)$. By $Y_i(t_j)$ we denote the transformed

Table 2. *Results of the simulation scenario 2 where* $cov\{U_{i0}(t_1,\ldots,t_m)\} = cov\{U_{i1}(t_1,\ldots,t_m)\} = 0.2J + 0.1A_m(0.1)$

σ^2/ψ_0		AB	ARB	CV	MS	eMS	RBe	CP	MS2	Eff	AB	ARB	CV	MS	eMS	RBe	CP	MS2	Eff	AB	ARB	CV	MS	eMS	RBe	CP	MS2	Eff
		Design 1 $\beta_0(t) = 2$, $\beta_1(t) = 2$									Design 2 $\beta_0(t) = 2\cos\{2\pi(t/m)\}$ and $\beta_1(t) = 0.4$									Design 3 $\beta_0(t) = \beta_1(t) = 2\sin(1.5\pi t/m)$								
0.25	I	1.1	0.5	**−3.2**	0.3	0.2	**−0.8**	9.3	0.2	10	1.1	0.6	1.1	0.2	0.2	**−1.4**	9.3	0.2	10	1.1	0.5	**−0.6**	0.3	0.3	0.1	9.4	0.2	10
0.5		1.5	0.7	0.9	0.5	0.4	**−1.2**	9.5	0.4	10	1.6	0.8	1.2	0.5	0.4	**−1.6**	9.3	0.4	10	1.5	0.7	**−1.3**	0.5	0.5	**−0.0**	9.4	0.4	10
1		1.9	0.9	2.8	0.8	0.8	**−0.9**	9.4	0.7	10	2.0	1.0	2.5	0.9	0.8	**−1.3**	9.3	0.7	10	2.0	0.9	**−1.6**	0.9	1.0	1.4	9.4	0.7	10
2		2.4	1.3	**−7.1**	1.4	1.2	**−1.2**	9.2	1.1	10	2.5	1.2	2.5	1.5	1.3	**−1.3**	9.3	1.1	10	2.7	1.3	0.1	1.6	1.4	**−1.3**	9.3	1.1	10
0.25	II	1.0	0.5	0.7	0.2	0.2	2.0	9.6	0.2	7.6	1.0	0.6	0.8	0.2	0.2	1.4	9.7	0.2	9.6	1.0	0.5	**−0.7**	0.2	0.2	1.5	9.6	0.2	8.5
0.5		1.4	0.5	0.6	0.5	0.4	**−0.2**	9.6	0.4	9.4	1.5	0.8	1.1	0.5	0.5	**−0.2**	9.5	0.5	9.4	1.6	0.5	**−0.6**	0.5	0.5	0.1	9.6	0.5	9.2
1		1.9	3.4	3.8	0.8	0.8	0.4	9.5	0.8	9.4	2.0	2.2	2.5	1.0	0.9	**−0.4**	9.4	0.9	11.1	2.1	3.3	1.3	1.0	1.0	**−0.2**	9.5	1.0	11.1
2		2.7	0.7	0.8	1.7	1.4	**−2.1**	9.2	1.4	12.5	3.2	2.0	2.3	2.2	1.8	**−1.8**	9.3	1.8	14.7	2.9	0.8	**−1.0**	2.2	1.9	**−1.2**	9.3	1.9	13.5
0.25	III	1.0	0.4	0.7	0.2	0.2	1.2	9.6	0.2	8.8	1.0	0.5	0.8	0.2	0.2	1.4	9.6	0.2	9.1	1.0	0.5	**−0.7**	0.2	0.3	1.4	9.6	0.2	8.5
0.5		1.5	0.4	0.6	0.5	0.5	0.3	9.6	0.5	9.4	1.5	0.7	1.1	0.5	0.5	0.3	9.6	0.5	8.9	1.6	0.5	**−0.6**	0.5	0.5	0.3	9.5	0.5	9.3
1		2.1	4.1	13.3	1.0	0.9	**−1.2**	9.4	0.9	12.1	2.0	1.4	5.2	1.0	0.9	**−0.5**	9.4	0.9	10.9	2.2	3.5	1.1	1.1	1.0	**−0.8**	9.4	1.0	11.9
2		2.7	0.6	0.9	1.7	1.6	**−0.6**	9.4	1.6	12.2	2.8	1.2	2.6	1.8	1.7	**−0.6**	9.3	1.7	12.3	3.0	0.8	**−0.9**	2.1	1.9	**−1.0**	9.4	1.9	13.4
0.25	IV	1.2	0.5	0.7	0.2	0.3	1.9	9.6		8.4	1.2	0.6	0.8	0.2	0.3	1.4	9.6		9.6	1.2	0.5	**−0.7**	0.2	0.3	1.4	9.6		8.5
0.5		1.8	0.5	0.6	0.5	0.5	0.2	9.6		10.2	1.8	0.8	1.1	0.5	0.5	0.2	9.6		9.4	1.8	0.5	**−0.6**	0.5	0.5	0.2	9.6		9.6
1		2.5	3.7	3.4	1.0	1.0	**−0.3**	9.5		12.7	2.5	1.4	1.9	1.0	1.0	**−0.3**	9.5		11.8	2.5	3.4	**−3.2**	1.0	1.0	**−0.3**	9.5		11.7
2		3.7	0.8	1.1	2.3	2.0	**−1.2**	9.3		16.5	3.7	2.2	**−8.2**	2.3	2.0	**−1.2**	9.3		15.3	3.7	0.8	**−1.0**	2.3	2.0	**−1.2**	9.3		14.3

Here, methods I, II, III and IV refer to the proposed method, the approach proposed by Rao & Yu (1994), the approach proposed by Datta, Lahiri & Maiti (2002) and the direct survey estimates, respectively. Here, AB = absolute bias, ARB = absolute relative bias, CV = coefficient of variation, MS = mean squared error, eMS = estimated MS, RBe = relative bias of eMS, CP = Wald-type 95 per cent coverage probability, MS2 $= [\mathrm{MSE}\{t(\delta)\}]_{\delta=\hat{\delta}} = [\mathrm{MS}\{t(\delta)\}]_{\delta=\hat{\delta}}$ and Eff denotes the ratio of the MS of a method to the corresponding MS of method I. All entries of the table are multiplied by 10.

unemployment rate for the ith state and for the jth quarter, $j = 1, \ldots, 44$, and $i = 1, \ldots, 51$. We do not have any covariate in the data. For method I, we fit $Y_i(t_j) = \beta_{i0}(t_j) + \epsilon_i(t_j)$, $\beta_{i0}(t_j) = \beta_0(t_j) + U_{0i}$, and $\beta_0(t_j)$ is semiparametrically modelled via cubic B-splines with equidistant knots $(4, 9, 14, 19, 24, 29, 34, 39)$. This data set is also analysed by method II (Rao & Yu, 1994) and method III (Datta *et al.*, 2002). Note that in the last two methods, the fixed effect component is only the intercept that is assumed to be constant over time. In particular, we fit $Y_i(t_j) = \gamma + a_i + \alpha_{it_j} + \epsilon_i(t_j)$, where $\alpha_{it_j} = \rho\alpha_{it_{j-1}} + \zeta_{it_j}$, $|\rho| < 1$, $a_i \sim \text{Normal}(0, \sigma_a^2)$ and $\zeta_{it_j} \sim \text{Normal}(0, \sigma_\zeta^2)$ for method II, and $Y_i(t_j) = \gamma + a_i + \alpha_{it_j} + \epsilon_i(t_j)$, $\alpha_{it_j} = \alpha_{it_{j-1}} + \zeta_{it_j}$, $a_i \sim \text{Normal}(0, \sigma_a^2)$ and $\zeta_{it_j} \sim \text{Normal}(0, \sigma_\zeta^2)$ for method III.

The results of the proposed method are insensitive to the choice of the knot locations as long as they are large in number and cover the entire spectrum of the timescale. Because of a very small sampling variability (i.e., small $\text{var}\{\epsilon_i(t)\}$), we were not able to see any appreciable

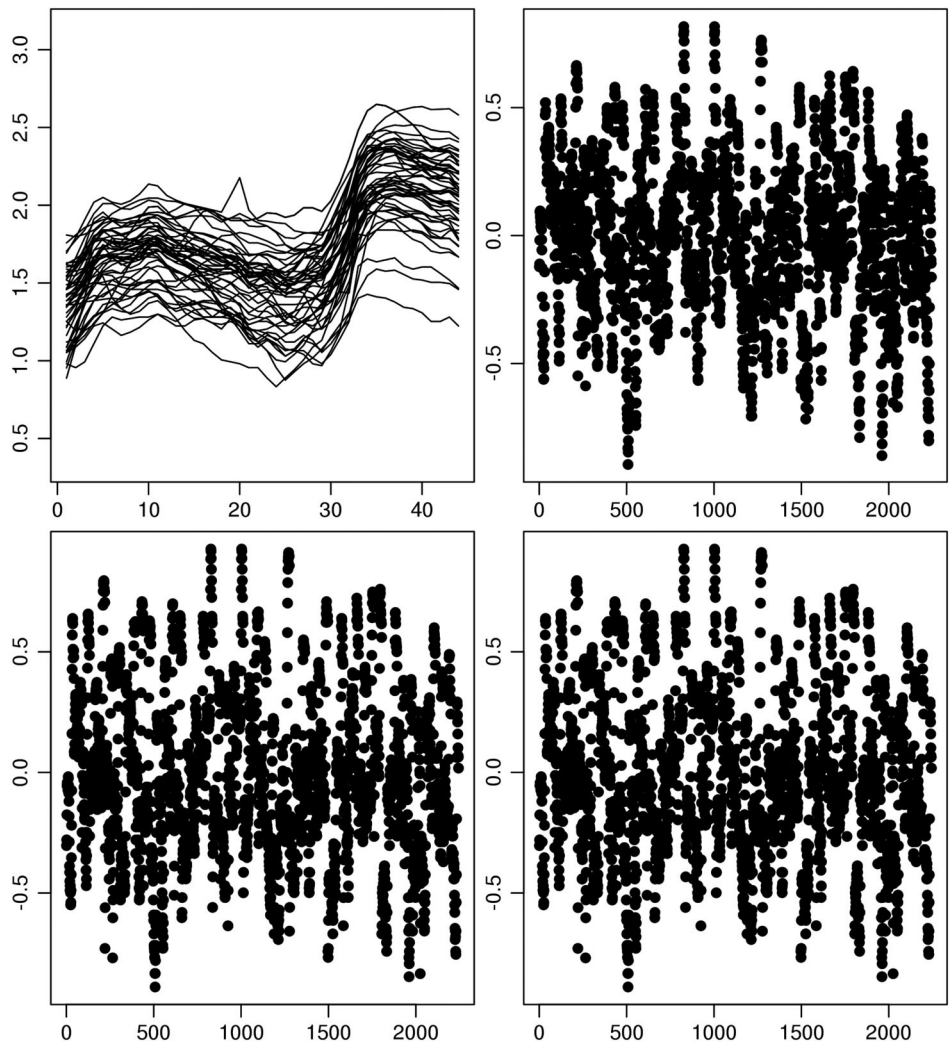

Fig. 1. Top left panel shows the quarterly unemployment rate for the period from 2001 to 2011 for 51 states. The top right, bottom left and bottom right panels show residual plots due to methods I, II and III, respectively.

difference among the methods in terms of the point prediction and prediction interval. Figure 1 shows the residual plots ($Y - Z_F \widehat{b}_F$), and the corresponding quantile plots given in Fig. 2 indicate that method I has a somewhat better fit to the data among the methods. Also, the left panel of Fig. 3 compares the eMSEs for the methods. To illustrate the effectiveness of our approach, we artificially increased the sampling variability by multiplying the actual sampling variability by 3 and then re-analysed the data. The corresponding eMSEs are given in the right panel of Fig. 3. These comparisons are consistent with the simulation results that the superiority of our method is evident when the sampling variability becomes large.

A striking feature of this data is non-linear time trend in unemployment rates. While the proposed method captures this well, the linear models fail to incorporate this. We further investigated the usefulness of functional approach for this application. Thus, we are interested to

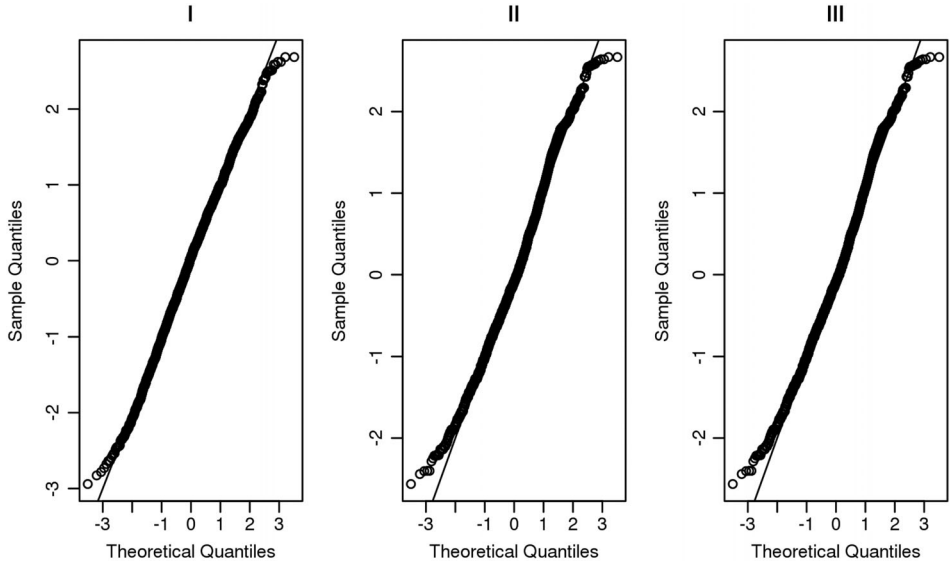

Fig. 2. Plot of quantiles of the re-scaled residuals against the theoretical quantiles of the standard normal distribution.

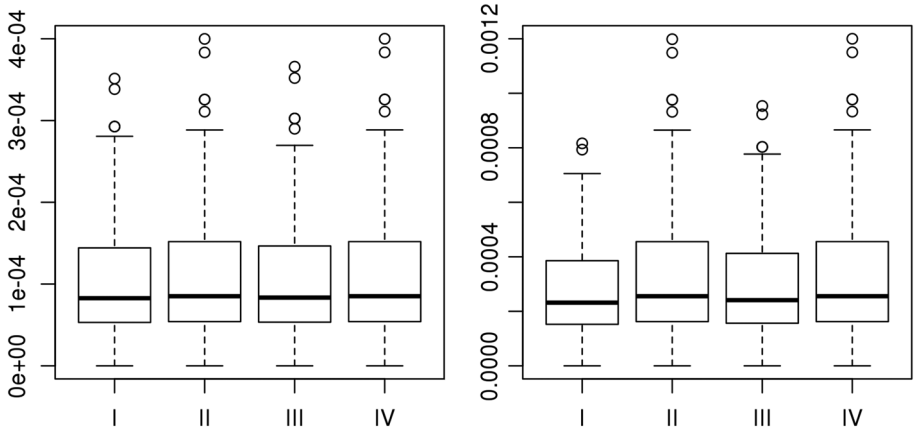

Fig. 3. Boxplot of the estimated mean squared errors for the four methods. The left and right panels correspond to the original sampling variance and the artificially increased sampling variance, respectively.

check if $\beta_0(t)$ is constant across all t. If $\beta_0(t) = \sum_{l=1}^{L} B_{l,d}(t) b_{0l}$ is a constant function, then b_{0l} are all equal as $\sum_{l=1}^{L} B_{l,d}(t) = 1$. Here, for eight inner knot points along with $d = 3$ makes $L = 12$. Now, we can write $\widehat{b}_0 = V(\widehat{b}_{R0}^T, \widehat{b}_{F0}^T)^T = V\Gamma(\widehat{\delta})Y$, where

$$\Gamma(\delta) = \begin{bmatrix} \operatorname{cov}(b_R) Z_R \Sigma^{-1} \left\{ I - Z_F \left(Z_F^T \Sigma^{-1} Z_F \right)^{-1} Z_F^T \Sigma^{-1} \right\} \\ \left(Z_F^T \Sigma^{-1} Z_F \right)^{-1} Z_F^T \Sigma^{-1} \end{bmatrix}.$$

For testing equality of all components of b_0, we test $H_0 : Cb_0 = 0$, where C is a 11×12 matrix and $C_{i,1} = 1$ and $C_{i,(i+1)} = -1$ for all $i = 1, \ldots, 11$ and 0 otherwise. We use $\widehat{b}_0^T C^T \left\{ C V \Gamma(\widehat{\delta}) \Sigma(\widehat{\delta}) \Gamma^T(\widehat{\delta}) V^T C^T \right\}^{-1} C\widehat{b}_0$ as our test statistic. Note that we expect a large value of the test statistic when the null hypothesis is violated. Under the null hypothesis, the model is $Y_i(t_j) = \beta_{i0}(t_j) + \epsilon_i(t_j)$ with $\beta_{i0}(t_j) = \alpha_0 + U_{0i}$, and the model for U_{0i} remains the same as model I. We have applied a parametric bootstrap approach to compute the p-value of our test. The test statistic was 141.31, and the p-value was smaller than 0.001, indicating a strong evidence against the null hypothesis.

7. Concluding remarks

Functional approach has been explored in many areas of statistics in recent years. This article develops functional regression theory and methodology in the context of small area estimation. The dimension reduction in this functional approach has been performed in such a way that the method can be used by the people familiar with small area estimation techniques without having theoretical training of functional analysis. The necessary theory has been developed to support the methodology. This advances the theory of functional mixed models.

The numerical findings support the effectiveness of our functional approach over traditional linear regression-based approach. There is a clear advantage in case of functional data and almost no loss in the case of linear regression. The computational time is minimal. The computer code is based on several in-built routines in R, and it is available from the authors upon request.

Although the estimation procedure was introduced for a balanced designed data set without missing time points, it can be applied to data with missing time points without much of a difficulty. Theoretical and numerical quantification of efficiency loss due to missing data is an interesting problem that can be pursued in a future project. Although the proposed method is developed solely for aggregated data, the techniques can be applied to deal with unit-level data where the sampling variability need not be known.

Acknowledgements

The authors are thankful to the associate editor and two referees for their constructive comments, which substantially improved the presentation. This research was partially supported by NIH grant R03CA176760 and NSF grants DMS 1309156, SES 0961618 and SES 0961649.

Supporting information

Additional information for this article is available online including the detailed proof of Theorems 1 and 2 referenced in Section 4.

References

Battese, G. E., Harter, R. M. & Fuller, W. A. (1988). An error components model for prediction of county crop areas using survey and satellite data. *J. Amer. Statist. Assoc.* **83**, 28–36.

Bell, W. (2012). Notes on a multivariate Fay-Herriot model with AR(1) model errors. Available at: http://www.census.gov/did/www/saipe/publications/files/belljsm2012.pdf.

Booth, J. G. & Hobert, J. P. (1998). Standard error of prediction in generalized linear mixed models. *J. Amer. Statist. Assoc.* **93**, 262–272.
Cressie, C. & Wikle, C. (2011). *Statistics for spatio-temporal data*, Wiley.
Das, K., Jiang, J. & Rao, J. N. K. (2004). Mean squared error of empirical predictor. *Ann. Statist.* **32**, 818–840.
Datta, G. S. & Ghosh, M. (2012). Small area shrinkage estimation. *Statist. Sci.* **27**, 95–114.
Datta, G. S. & Lahiri, P. (2000). A unified measure of uncertainty of estimated best linear unbiased predictor in small area estimation problems. *Statist. Sinica* **10**, 613–627.
Datta, G. S., Lahiri, P. & Maiti, T. (2002). Empirical Bayes estimation of the median income of four-person families by state using time series and cross-sectional data. *J. Statist. Plann.* **102**, 83–97.
Datta, G. S., Lahiri, P., Maiti, T. & Lu, K. L. (1999). Hierarchical Bayes estimation of unemployment rates for the states of the U.S. *J. Amer. Statist. Assoc.* **94**, 1074–1082.
Eilers, P. H. C. & Marx, B. D. (1996). Flexible smoothing with B-splines and penalties. *Statist. Sci.* **11**, 89–121.
Ferraty, F. & Vieu, P. (2006). *Nonparametric functional data analysis: theory and practice*, Springer, New York.
Ghosh, M., Nangia, N. & Kim, D. -H. (1996). Estimation of median income of four-person families: A Bayesian time series approach. *J. Amer. Statist. Assoc.* **91**, 1423–1431.
Ghosh, M. & Rao, J. N. K. (1994). Small area estimation: An appraisal (with discussion). *Statist. Sci.* **9**, 55–93.
Giusti, C. S., Salvati, M. N. & Pratesi, M. (2012). Semiparametric Fay-Herriot model using penalized splines. *Journal of the Indian Society of Agricultural Statistics* **66**, 1–14.
Guo, W. (2002). Functional mixed effects models. *Biometrics* **58**, 121–128.
Hoover, D. R., Rice, J. A., Wu, C. O. & Yang, L. P. (1998). Nonparametric smoothing estimates of time-varying-coefficient models with longitudinal data. *Biometrika* **85**, 809–822.
Kackar, R. N. & Harville, D. A. (1984). Approximations for standard errors of estimators of fixed and random effects in linear mixed models. *J. Amer. Statist. Assoc.* **79**, 853–862.
Militano, A. F., Goicoa, T. & Ugarte, M. D. (2012). Estimating the percentage of food expenditure in small areas using bias-corrected P-spline based estimators. *Comput. Statist. Data Anal.* **56**, 2934–2948.
Morris, J. S. & Carroll, R. J. (2006). Wavelet-based functional mixed models. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **68**, 179–199.
Opsomer, J. D., Claeskens, G., Ranalli, M. G., Kauermann, G. & Breidt, F. J. (2008). Non-parametric small area estimation using penalized spline regression. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **70**, 265–286.
Pfeffermann, D. (2002). Small area estimation – new developments and directions. *Int. Stat. Rev.* **70**, 125–143.
Pfeffermann, D. & Burck, L. (1990). Robust small area estimation combining time series and cross-sectional data. *Survey Methodology* **16**, 217–237.
Prasad, N. G. N. & Rao, J. N. K. (1990). The estimation of mean squared error of small-area estimators. *J. Amer. Statist. Assoc.* **85**, 163–171.
Pratesi, M., Ranalli, M. G. & Salvati, N. (2008). Semiparametric M-quantile regression for estimating the proportion of acid lakes in 8-digit HUCs of northeastern U.S. *Environmetrics* **19**, 687–701.
Rao, J. N. K. (2003). *Small area estimation*, Wiley, New York.
Rao, J. N. K. & Yu, M. (1994). Small area estimation by combining time series and cross-sectional data. *Canad. J. Statist.* **22**, 511–528.
Ramsay, J. O. & Silverman, B. W. (2005). *Functional data analysis*, Springer, New York.
Ruppert, R., Wand, M. & Carroll, R. J. (2003). *Semiparametric regression*, Cambridge University Press, Oxford.
Salvati, N. H., Chandra, H., Ranalli, M. G. & Chambers, R. (2010). Small area estimation using a nonparametric model based direct estimator. *Comput. Statist. Data Anal.* **54**, 2159–2171.
Salvati, N. H., Ranalli, M. G. & Pratesi, M. (2011). Small area estimation of the mean using nonparametric M-quantile regression: A comparison when a linear mixed model does not hold. *J. Stat. Comput. Simul.* **81**, 945–964.
Ugarte, M. D., Goicoa, T., Militino, A. F. & Durbàn, M. (2009). Spline smoothing in small area trend estimation and forecasting. *Comput. Statist. Data Anal.* **53**, 3616–3629.

Received March 2014, in final form December 2015

Tapabrata Maiti, Department of Statistics and Probability, Michigan State University, East Lansing, MI 48824, USA.
E-mail: maiti@stt.msu.edu