Local Polynomial Regression for Small Area Estimation

Pushpal K Mukhopadhyay[1]
SAS Institute Inc.

Tapabrata Maiti[2]
Iowa State University

Abstract

Estimation of small area means in the presence of area level auxiliary information is considered. A class of estimators based on local polynomial regression is proposed. The assumptions on the area level regression are considerably weaker than standard small area models. Both the small area mean functions and the between area variance function are modeled as smooth functions of area level covariates. A composite estimator that is a convex combination of the design weighted mean and the prediction from the non-parametric model is developed. The estimator is shown to be asymptotically consistent under mild regularity conditions. An approximation of the mean squared error (MSE) based on Taylor linearization is proposed.

Keywords: Nonparametric smoothing, Multi-level modeling.

1 Introduction

The term "Small Area" refers to a subpopulation or domain where the domain sample size is not large enough to produce direct estimates with adequate precision. Composite estimators are often used to provide reliable small area statistics. A composite estimator is a convex combination of a direct estimator and a synthetic estimator. Several model-based composite estimators under realistic small area models are proposed in the literature. Rao (2003) gives an extensive review of the most commonly used estimators including empirical best linear predictors (EBLUP), empirical Bayes (HB), and hierarchical Bayes (HB) approaches. However all the model-based approaches in use for small area estimation have relied on parametric, most often linear, modeling techniques. In this article we propose a small area estimator that relies on a nonparametric model formulation.

In section 2, we describe a nonparametric area level small area model. Both the small area mean functions and the between area variance function are assumed to be smooth functions of some area level covariates. Design variance of direct area means are estimated from unit level information and then assumed to be known. The assumption of known design variance of area means are common for area level small area models (Rao, 2003). Local polynomial regression estimators are used as synthetic estimators of the small area mean functions. The observed residuals are adjusted for the known within area variance. The between area variance function is estimated using local polynomial regression by smoothing the adjusted residuals. The small area predictions are obtained by a convex combination of the direct mean and the synthetic mean.

Section 3 describes the theoretical properties of the proposed estimators. The exact and the asymptotic properties of the estimated mean functions and the between area variance function are derived under the nonparametric mixed effects model. It is shown than both estimators are consistent under mild regularity assumptions. The properties of the small area predictions are derived from the properties of the estimated mean functions and the estimated variance function.

In section 4 we propose an approximation of the mean squared error of the proposed small area predictions. The order of the approximation depends on the number of small areas. The mean squared error is derived under the assumed small area model. A practical approach for bandwidth selection by minimizing the mean squared error of small area predictions is also discussed. Finally the conclusions are given in section 5.

2 Framework for Local Polynomial Regression

For each area $i = 1, 2, ..., n$, assume that y_i is the Horvitz-Thompson estimator (Särndal et al., 1991) of the true mean θ_i with design variance D_i. Let $\mathbf{x}_i$ be a vector of known area level covariates. The basic area level small area model can be written as a special case of a linear mixed effect model,

$$y_i = \mathbf{x}_i^T \boldsymbol{\beta} + u_i + e_i, \quad i = 1, 2, ..., n, \tag{1}$$

where $\boldsymbol{\beta}$ is a vector of regression parameters, u_i's are random effects and e_i's are sampling errors. Note that the design-induced error e_i accounts for within

[1] 600 Research Drive, SAS Institute Inc., Cary, NC 27513; E-mail: pushpal.mukhopadhyay@sas.com.
[2] 221 Snedecor Hall, Ames, IA 50011; E-mail: taps@iastate.edu

area variation and the model-induced error, u_i, accounts for between area variation. We also assume $e_i \overset{ind}{\sim} (0, D_i)$, $u_i \overset{iid}{\sim} (0, \sigma_u^2)$ and that they are independent. For estimation purposes, D_i is usually assumed to be known; see Rao (2003).

The linearity assumption and the assumption of homoscedastic between area variance are restrictive in many applications. A limited simulation study indicates that a violation of the linear relationship may reduce the efficiency of the current method. We consider an extension of model (1). We assume that y_i and $\mathbf{x}_i$ are related through a smooth function $m(\cdot)$ and the between area variance component is also a smooth function of $\mathbf{x}_i$. Let $\mathbf{X}$ be the random vector of predictors. Thus

$$y_i = m(\mathbf{x}_i) + u_i + e_i, \quad i = 1, 2, ..., n, \tag{2}$$

where $u_i|\mathbf{X} \overset{ind}{\sim} (0, v(\mathbf{x}_i))$, $e_i|\mathbf{X} \overset{ind}{\sim} (0, D_i)$, and u_i and e_i are conditionally independent. We call m the mean function and v the between area variance function. The small area mean functions

$$\theta_i(\mathbf{x}_i) = m(\mathbf{x}_i) + u_i \tag{3}$$

are linear combinations of the mean $m(x_i)$ and the random effects u_i. We propose an estimator of the mean function using a linear smoother. By this, we mean $\hat{\mathbf{m}} = P_1\mathbf{y}$ for some $n \times n$ matrix P_1, often referred to as the smoother matrix, and $\mathbf{y}$ and $\mathbf{m}$ denote the column vectors with elements of y_i and $m(x_i)$ respectively. Examples of linear smoothers include smoothing splines, regression splines, and local polynomial regression (Hastie and Tibshirani, 1990).

We concentrate on local polynomial regression estimators of $\mathbf{m}$ and $\mathbf{v}$; see Fan and Gijbels (1996), or Wand and Jones (1995) for an introduction. With one dimensional covariate x, we estimate $m(x)$ by fitting a p_1th-degree polynomial to the data using weighted least squares. As commonly used in the literature, we will use the weight

$$K_{h_1}(X_i - x) = h_1^{-1} K\left(h_1^{-1}(X_i - x)\right), \tag{4}$$

where K is a probability density function known as the kernel function and h_1 is a bandwidth parameter. The weighted least squares estimators of m(x) is

$$\begin{aligned}\hat{m}(x) &= \mathbf{e}_1^T [X_{p_1}(x)^T W_{p_1}(x) X_{p_1}(x)]^{-1} X_{p_1}(x)^T \\ &\quad W_{p_1}(x)\mathbf{y},\end{aligned} \tag{5}$$

where

$$X_{p_1}(x) = \begin{bmatrix} 1 & X_1 - x & \cdots & (X_1 - x)^{p_1} \\ 1 & X_2 - x & \cdots & (X_2 - x)^{p_1} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & X_n - x & ... & (X_n - x)^{p_1} \end{bmatrix},$$

$W_{p_1}(x) = \text{diag}_{1 \le i \le n}\{K_1(X_i - x)\}$, $\mathbf{e}_i$ denotes the unit vector of appropriate order with 1 in the ith-position, and $\text{diag}_{1 \le i \le n}\{a_i\}$ denotes the diagonal matrix with $a_1, a_2, ..., a_n$ on the diagonal. The (i, j) entry of the pth-degree local polynomial smoother for the mean function is

$$\begin{aligned}[P_1]_{ij} &= \mathbf{e}_1^T [X_{p_1}(x_i)^T W_{p_1}(x_i) X_{p_1}(x_i)]^{-1} X_{p_1}(x_i)^T \\ &\quad W_{p_1}(x_i)\mathbf{e}_j.\end{aligned} \tag{6}$$

To the best of our knowledge we are the first to consider a nonparametric variance components model of the form (2) where one part of the variance is known from the survey design and the other part of the variance is assumed to be a smooth function of covariate x. We estimate the between area variance function by smoothing the adjusted observed residuals using a p_2th-degree polynomial

$$\hat{\mathbf{v}} = \frac{P_2(\mathbf{r}^2 - \mathbf{\Delta}_2)}{\mathbf{1} + P_2\mathbf{\Delta_1}}, \tag{7}$$

where P_2 is a smoother matrix similar to P_1 except it uses a p_2th-degree polynomial and a different bandwidth h_2, $\mathbf{y} = (y_1, y_2, \ldots, y_n)^T$, $\mathbf{r} = \mathbf{y} - P_1\mathbf{y}$ is the observed residual,

$$\mathbf{\Delta}_1 = \text{diag}\{P_1 P_1^T - 2P_1\}, \tag{8}$$

$$\mathbf{\Delta}_2 = \text{diag}\{D + P_1 D P_1^T - 2P_1 D\}, \tag{9}$$

$D = \text{diag}_{1 \le i \le n} D_i$, $\mathbf{1}$s is a column vector of ones, $\text{diag}\{A\}$ is the column vector containing the diagonal elements of any square matrix A, and vector multiplications and divisions are elementwise.

We define a composite estimator of small area means by taking a convex combination of the survey weighted mean and the mean function $m(\cdot)$ from model (2),

$$\theta_i^* = \gamma_i y_i + (1 - \gamma_i) m_i, \tag{10}$$

where $m_i = m(x_i)$ and $\theta_i = \theta_i(x_i) = m_i + u_i$. The ratio γ_i is obtained by minimizing the mean squared error of θ_i^*. The mean squared error for θ_i^* can be written as

$$\begin{aligned}E[\theta_i^* - \theta_i]^2 &= E[\gamma_i y_i + (1 - \gamma_i) m_i + m_i - u_i]^2 \\ &= E[\gamma_i (y_i - m_i) - u_i]^2 && (11)\\ &= E[\gamma_i (e_i + u_i) - u_i]^2 && (12)\\ &= E[\gamma_i e_i + (1 - \gamma_i) u_i]^2 && (13)\\ &= \gamma_i^2 D_i + (1 - \gamma_i)^2 v_i && (14)\\ &= \gamma_i^2 (v_i + D_i) - 2\gamma_i v_i + v_i, && (15)\end{aligned}$$

since $E[e_i u_i|\mathbf{X}] = 0$. Therefore $\gamma_i = (v_i + D_i)^{-1} v_i$ minimizes the mean squared error of θ_i^*. Assume D_i's are known. Estimate γ_i by

$$\hat{\gamma}_i = (\hat{v}_i + D_i)^{-1} \hat{v}_i. \tag{16}$$

Thus, a two stage estimator for θ_i is given by,

$$\hat{\theta}_i = \hat{\gamma}_i y_i + (1-\hat{\gamma}_i)\hat{m}_i \quad (17)$$
$$= \hat{m}_i + \hat{\gamma}_i(y_i - \hat{m}_i), \quad (18)$$

where

$$\hat{m}_i = \mathbf{e}_i^T P_1 \mathbf{y}, \quad (19)$$

$$\hat{v}_i = \mathbf{e}_i^T \frac{P_2(\mathbf{r}^2 - \mathbf{\Delta}_2)}{\mathbf{1} + P_2\mathbf{\Delta_1}} \quad (20)$$

and $\hat{\gamma}_i$ is given in (16). For the linear mixed effects model (1), the plug-in estimator of γ_i gives an empirical best linear unbiased predictor for θ_i (Rao, 2003).

Remark 1. Ruppert et al. (1997) proposed similar estimators of $\hat{\mathbf{m}}$ and $\hat{\mathbf{v}}$ for the model $y_i = m(x_i)+e_i$, where $e_i \overset{ind}{\sim} (0, v(x_i))$ and $v(\cdot)$ is a smooth function. The nonparametric model we considered is different in the sense that it accounts for separate within area variabilities and between area variability. The effect of estimating the within area sampling variance D_i is generally ignored in small area estimation (Rao, 2003).

Remark 2. If the between area variance function is assumed to be the same for all x then one should simply replace the smoother matrix P_2 by $n^{-1}\mathbf{1}\mathbf{1}^T$.

Remark 3. For theoretical convenience we consider only one covariate X; however the results can be extended to a vector of covariates.

Remark 4. We assume the random matrix $X_{p_1}(x_0)^T W_{p_1}(x_0) X_{p_1}(x_0)$ is invertible. In other words, the P_X probability that $X_{p_1}(x_0)^T W_{p_1}(x_0) X_{p_1}(x_0)$ is singular is zero. This assumption is not new in sample survey (Breidt and Opsomer, 2000) and is meaningful in small area estimation.

Remark 5. The nature of the marginal mean function $m(\cdot)$ and the variance function $v(\cdot)$ are quite different. We should be able to model high spikes on $m(\cdot)$ but usually in practice the variance function $v(\cdot)$ is more smooth. Hence, a lower degree polynomial fit is often used for $v(\cdot)$. Ruppert et al. (1997) recommended the use of $p_1 = 2$ and $p_2 = 1$ in most situations.

3 Theory for Local Polynomial Regression

Estimators of the following quantities are proposed: (i) the mean function $m(\cdot)$, (ii) the between area variance function $v(\cdot)$, and (iii) the small area mean function $\theta_i(\cdot)$. In this section, the exact matrix algebraic expressions for the bias and the variance of the proposed estimators are obtained. Asymptotic approximations for the bias and the variance of the proposed estimators are also obtained under certain regularity conditions. Asymptotic approximations are useful for choosing bandwidths or evaluating the performances of the proposed estimators. Proofs can be obtained from the author.

In practice, X_i can either be fixed or random. For theoretical convenience, we assume X_is are random and $X_i \overset{iid}{\sim} f_X(\cdot)$. Let $\mathbf{X} = (X_1, X_2, \ldots, X_n)^T$ be the random vector of predictors. Results for exact bias and covariance are conditional on $\mathbf{X}$ and therefore do not depend on a particular form of the distribution of $\mathbf{X}$. Results are derived for an interior point x_0 and for odd integers p_1 and p_2. Similar to local polynomial estimators for a fixed effects model (Fan and Gijbels, 1996), our results can easily be derived for boundary points and for even integers.

3.1 Exact Bias and Variance

Theorem 3.1 *The expectation and variance of $\hat{m}(x_0)$ are given by*

$$E[\hat{m}(x_0)|\mathbf{X}] = m(x_0) + \mathbf{e}_1[[X_{p_1}(x_0)^T W_{p_1}(x_0) X_{p_1}(x_0)]]^{-1} X_{p_1}(x_0)^T W_{p_1}(x_0)\mathbf{t}_{x_0}, \quad (21)$$

and

$$V[\hat{m}(x_0)|\mathbf{X}] = \mathbf{e}_1[X_{p_1}(x_0)^T W_{p_1}(x_0) X_{p_1}(x_0)]^{-1} X_{p_1}(x_0)^T \{\Sigma^{(1)}_{1,x_0} + \Sigma^{(2)}_{1,x_0}\} X_{p_1}(x_0) [X_{p_1}(x_0)^T W_{p_1}(x_0) X_{p_1}(x_0)]^{-1}\mathbf{e}_1^T \quad (22)$$

where $\mathbf{t}_{x_0} = \mathbf{m} - X_{p_1}(x_0)\boldsymbol{\beta}(x_0)$ is the remainder from the Taylor series expansion of $\mathbf{m}$ around $m(x_0)$, $\Sigma^{(1)}_{1,x_0} = diag\{K^2_{h_1}(X_i - x_0)v(X_i)\}$ and $\Sigma^{(2)}_{1,x_0} = diag\{K^2_{h_1}(X_i - x_0)D_i\}$ are the weighted variance components.

Proposition 3.2 *For a homoscedastic model, the expected value of the squared residuals after fitting the conditional mean function is given by*

$$E[\mathbf{r}^2|\mathbf{X}] = \{E[P_1\mathbf{y} - \mathbf{m}]|\mathbf{X}\}^2 + \sigma^2(\mathbf{1} + \mathbf{\Delta}_1) + \mathbf{\Delta_2}, \quad (23)$$

where $v(x_i) = \sigma^2$, $D_i = \psi$ for all i, and $\mathbf{\Delta}_1$ and $\mathbf{\Delta}_2$ are defined in (8) and (9).

Theorem 3.3 *Let $G_u = diag\{Eu_i^3\}$, $T_u = diag\{Eu_i^4\}$, $G_e = diag\{Ee_i^3\}$, $T_e = diag\{Ee_i^4\}$, $P_2\mathbf{1} = P_2$, $D = diag_{1\le i\le n}\{D_i\}$, $\boldsymbol{\psi} = (\psi_1, ..., \psi_n)^T$ and $\Sigma = diag_{1\le i\le n}\{v(x_i)\}$. The expectation and* *variance of* $\hat{\mathbf{v}}$ *are given by*

$$
\begin{aligned}
&E[(\hat{\mathbf{v}}-\mathbf{v})|\mathbf{X}] \\
&= \big[(P_2-I)\mathbf{v}+P_2\{\mathbf{b}^2+diag\{P_1\Sigma P_1^T-2P_1\Sigma\}\} \\
&\quad -P_2\mathbf{\Delta}_1\mathbf{v}\big]/\{\mathbf{1}+P_2\mathbf{\Delta}_1\}, \qquad (24)
\end{aligned}
$$

and

$$
\begin{aligned}
&Cov[\hat{\mathbf{v}}|\mathbf{X}] \\
&= P_2[\{(P_1-I)\odot(P_1-I)\}(T-3V^2) \\
&\quad \{(P_1-I)\odot(P_1-I)\}^T \\
&\quad +2diag\{\mathbf{b}\}(P_1-I)G\{(P_1-I)\odot(P_1-I)\}^T \\
&\quad +2\{(P_1-I)\odot(P_1-I)\}G(P_1-I)diag\{\mathbf{b}\} \\
&\quad +2\{(P_1-I)V(P_1-I)^T\} \\
&\quad \odot\{(P_1-I)V(P_1-I)^T\} \\
&\quad +4\{(P_1-I)V(P_1-I)^T\} \\
&\quad \odot(\mathbf{b}\mathbf{b}^T)]P_2^T/\{(\mathbf{1}+P_2\mathbf{\Delta}_1)(\mathbf{1}+P_2\mathbf{\Delta}_1)^T\}, (25)
\end{aligned}
$$

where $\mathbf{b}=E[\hat{\mathbf{m}}-\mathbf{m}|\mathbf{X}]$ *is the bias due to the estimation of the mean,* $G=G_u+G_e$, $T=T_u+T_e$ *and* $\odot$ *denotes element wise matrix multiplication.*

3.2 Asymptotics for Local Polynomial Estimators

The exact bias and variance expressions involve unknown quantities. Approximations of the bias and the variance are required for most applications. Asymptotic approximations are derived under certain regularity assumptions about the nature of $f_X(\cdot)$, $m(\cdot)$, $v(\cdot)$ and D_i. Most of these assumptions are standard for local polynomial regression (Fan and Gijbels, 1996).

Theorem 3.4 *Assume the following:*

(A1) x_0 *is an interior point in the* $\mathbf{X}$ *space.*

(A2) $f_X(x_0)\geq 0$.

(A3) There exists a $\delta_1 \geq 0$ *such that* $f_X(\cdot)$, $m^{(p_1+1)}(\cdot)$, *and* $v(\cdot)$ *are continuous and* $m^{(p_1+2)}(\cdot)$ *is bounded on* $N_{\delta_1}(x_0)$.

(A4) $n^{-1}\sum_i D_i$ *and* $n^{-1}\sum_i D_i^2$ *are bounded.*

(A5) p_1 *is an odd integer.*

(A6) $h_1\to 0$ *and* $nh_1\to\infty$ *as* $n\to\infty$.

The asymptotic bias and variance of $\hat{m}(x_0)$ *are given by*

$$
\begin{aligned}
&Bias[\hat{m}(x_0)|\mathbf{X}] \\
&= \mathbf{e}_1^T S_1^{-1}\mathbf{c}_{p_1}\frac{1}{(p_1+1)!}m^{(p_1+1)}(x_0)h_1^{p_1+1} \\
&\quad +o_P(h_1^{p_1+1}), \qquad (26)
\end{aligned}
$$

and

$$
\begin{aligned}
&Var[\hat{m}(x_0)|\mathbf{X}] \\
&= \mathbf{e}_1^T S_1^{-1}S_1^* S_1^{-1}\mathbf{e}_1\frac{v(x_0)+n^{-1}\sum_i D_i}{f_X(x_0)nh_1} \\
&\quad +o_P(n^{-1}h_1^{-1}), \qquad (27)
\end{aligned}
$$

where

$$\mathbf{c}_{p_1}=(\mu_{p_1},...,\mu_{2p_1+1})^T, \qquad (28)$$

$$S_1=[\mu_{j+l}]_{0\leq j+l\leq p_1}, \qquad (29)$$

$$S_1^*=[\nu_{j+l}]_{0\leq j+l\leq p_1}, \qquad (30)$$

$$\mu_j=\int u^j k(u)du, \qquad (31)$$

and

$$\nu_j=\int u^j K^2(u)du. \qquad (32)$$

Theorem 3.5 *Assume that (A1) - (A6) hold. In addition, assume the following:*

(A7) There exists a $\delta_2>0$ *such that* $f_X(\cdot)$ *and* $v^{(p_2+1)}(\cdot)$ *are continuous and* $v^{(p_2+2)}(\cdot)$ *is bounded on* $N_{\delta_2}(x_0)$.

(A8) $h_2\to 0$, $nh_2\to\infty$, *as* $n\to\infty$.

(A9) p_2 *is an odd integer.*

(A10) $h_1^{2(p_1+1)}+(nh_1)^{-1}=o(h_2^{p_2+1})$.

(A11) $Ee_i^4=\kappa_i$ *where* $n^{-1}\sum\kappa_i=O(1)$, $n^{-1}\sum\kappa_i^2=O(1)$ *and* k_i *are known from the sampling design.*

(A12) Skewness in both error components can be ignored. i.e., $Ee_i^3=0$ *and* $Eu_i^3=0$.

The asymptotic bias and variance of $\hat{v}(x_0)$ *are given by*

$$
\begin{aligned}
&Bias[\hat{v}(x_0)|\mathbf{X}] \\
&= \mathbf{e}_1^T S_2^{-1}c_{p_2}\frac{1}{(p_2+1)!}v^{(p_2+1)}(x_0)h_2^{p_2+1} \\
&\quad +o_p(h_2^{p_2+1}), \qquad (33)
\end{aligned}
$$

and

$$
\begin{aligned}
&Var[\hat{v}(x_0)|\mathbf{X}] \\
&= \mathbf{e}_1^T S_2^{-1}S_2^* S_2^{-1}\mathbf{e}_1[\{f_X(x_0)\}^{-1} \\
&\quad \{\eta(x_0)+\bar{\eta}_\pi-2v(x_0)\bar{\psi}\}](nh_2)^{-1} \\
&\quad +o_p(nh_2)^{-1} \qquad (34)
\end{aligned}
$$

where

$$\bar{\eta}_\pi=\bar{\kappa}-(\bar{\psi})^2, \qquad (35)$$

$$\mathbf{c}_{p_2}=(\mu_{p_2},...,\mu_{2p_2+1})^T, \qquad (36)$$

$$S_2=[\mu_{j+l}]_{0\leq j+l\leq p_2}, \qquad (37)$$

$$S_2^*=[\nu_{j+l}]_{0\leq j+l\leq p_2}, \qquad (38)$$

$$\mu_j=\int u^j k(u)du \qquad (39)$$

$$\nu_j = \int u^j K^2(u)du, \quad (40)$$

and

$$\eta(x_i) = Eu_i^4 - (Eu_i^2)^2. \quad (41)$$

Remark 6. Assumption (A10) is satisfied if $p_1 = p_2$ and for any optimal selection of bandwidth. Ruppert et al. (1997) used the same assumption for variance function estimators of nonparametric fixed effect models.

Remark 7. D_i's are variances of county means. For many survey designs, the assumptions about D_i and e_i are valid (Fuller, 2006).

Remark 8. We are smoothing the observed residuals, not the true residuals. Asymptotically $\hat{v}(x_0)$ behaves like a local polynomial smooth of the true residuals (Theorem 3.5). There is no loss in asymptotic efficiency of $\hat{\mathbf{v}}$ due to the estimation of $\hat{\mathbf{m}}$.

4 Approximation of the MSE

We provide an approximation for the MSE of $\hat{\theta}_i$ using the formulas derived in Section 3. Let

$$\tilde{\theta}_i = \gamma_i y_i + (1-\gamma_i)\hat{m}_i \quad (42)$$

and

$$\theta_i^* = \gamma_i y_i + (1-\gamma_i) m_i. \quad (43)$$

We write

$$\begin{aligned} &E(\hat{\theta}_i - \theta_i)^2 \\ &= E(\theta_i^* - \theta_i)^2 + E(\tilde{\theta}_i - \theta_i^*)^2 + E(\hat{\theta}_i - \tilde{\theta}_i)^2 \\ &\quad + E(\hat{\theta}_i - \tilde{\theta}_i)(\tilde{\theta}_i - \theta_i^*), \quad (44) \end{aligned}$$

since the expected values for the other product terms vanish. The first two terms on the right side of (44) have similar expressions to those in Prasad and Rao (1990). However, the last two terms are not tractable in general. We approximate the last two terms by the Taylor series expansions. More formally, Theorem 4.1 can be shown.

Theorem 4.1 *Assume that (A1) - (A12) hold. Assume* $u_i \overset{ind}{\sim} N(0, v_i)$ *and* $e_i \overset{ind}{\sim} N(0, D_i)$. *Then*

$$\begin{aligned} &E[\hat{\theta}_i - \theta_i|\mathbf{X}]^2 \\ &= g_{1i}(v_i) + g_{2i}(v_i, m_i) + g_{3i}(v_i) + g_{4i}(v_i) \\ &\quad + O_P(a_{nh}), \quad (45) \end{aligned}$$

where

$$g_{1i}(v_i) = (v_i + D_i)^{-1} v_i D_i, \quad (46)$$

$$g_{2i}(v_i, m_i) = (1-\gamma_i)^2 MSE(\hat{m}_i), \quad (47)$$

$$\begin{aligned} &g_{3i}(v_i) \\ &= \{b_i^2 + v_i(1 + \mathbf{\Delta}_{1i}) + \mathbf{\Delta}_{2i}\} \\ &\quad (v_i + D_i)^{-4} D_i^2 MSE(\hat{v}_i), \quad (48) \end{aligned}$$

$$\begin{aligned} &g_{4i}(v_i) \\ &= (D_i + v_i)^{-3} D_i^2 \{b_i^2 + v_i(1 + \mathbf{\Delta}_{1i}) + \mathbf{\Delta}_{2i}\} \\ &\quad Bias(\hat{v}_i), \quad (49) \end{aligned}$$

and $a_{nh} = max\{(nh_2)^{-3/2}, h_2^{2p_2+2}\}$. *Asymptotic expressions for* $g_{2i}(v_i)$, $g_{3i}(v_i)$, *and* $g_{4i}(v_i)$ *are given by*

$$\begin{aligned} &g_{2i}(v_i, m_i) \\ &= (1-\gamma_i)^2 \Big[\mathbf{e}_1^T S_1^{-1} \mathbf{c}_{p_1} \mathbf{c}_{p_1} S_1^{-1} \mathbf{e}_1 \{(p_1+1)!\}^{-2} \\ &\quad \{m^{(p_1+1)}\}^2(x_i) h_1^{2p_1+2} + \mathbf{e}_1^T S_1^{-1} S_1^* S_1^{-1} \mathbf{e}_1 \\ &\quad \{v(x_0) + n^{-1}\sum_i D_i\}\{f_X(x_0) n h_1\}^{-1}\Big] \\ &\quad + o_p(b_{nh}), \quad (50) \end{aligned}$$

$$\begin{aligned} &g_{3i}(v_i) \\ &= (v_i + D_i)^{-3} D_i^2 \Big[\mathbf{e}_1^T S_2^{-1} c_{p_2} c_{p_2}^T S_2^{-1} \mathbf{e}_1^T \\ &\quad \{(p_2+1)!\}^{-2} \{v^{(p_2+1)}(x_0)\}^2 h_2^{2(p_2+1)} \\ &\quad + \mathbf{e}_1^T S_2^{-1} S_2^* S_2^{-1} \mathbf{e}_1 [\{f_X(x_0)\}^{-1} \\ &\quad \{\eta(x_0) + \bar{\eta}_\pi - 2v(x_0)\bar{\psi}\}](nh_2)^{-1}\Big] \\ &\quad + o_p(c_{nh}), \quad (51) \end{aligned}$$

$$\begin{aligned} &g_{4i}(v_i) \\ &= (v_i + D_i)^{-2} D_i^2 \mathbf{e}_1^T S_2^{-1} c_{p_2} \{(p_2+1)!\}^{-1} \\ &\quad v^{(p_2+1)}(x_0) h_2^{(p_2+1)} + o_P(h_2^{p_2+1}), \quad (52) \end{aligned}$$

where $b_{nh} = max\{h_1^{2p_1+2}, (nh_1)^{-1}\}$, *and* $c_{nh} = max\{h_2^{2p_2+2}, (nh_2)^{-1}\}$.

4.1 Bandwidth Selection for Local Polynomial Estimators

An important issue is the choice of bandwidth parameters, h_1, and h_2. Local optimal bandwidths and global optimal bandwidths are common in practice (Fan and Gijbels, 1996). We provide a methodology for local optimal bandwidths selection. Ideally, the local optimal bandwidths should minimize the MSE of the small area predicted means. Thus,

$$(h_1^{opt}, h_2^{opt})_i = \text{argmin}_{h_1,h_2} \text{MSE}(\hat{\theta}_i|\mathbf{X}), \quad (53)$$

subject to

$$h_1^{2p_1+2} + (nh_1)^{-1} = o_p(h_2^{p_2+1}), \quad (54)$$

where $MSE[\hat{\theta}_i|\mathbf{X}]$ is given in (45). Finding (h_1^{opt}, h_2^{opt}) by minimizing (53) is difficult to do in practice because the effects of h_1 on the MSE of $\hat{\theta}_i$ are of second order. Using Theorem 3.4 and Theorem 3.5, we propose the following strategy:

1. Select an asymptotically optimal bandwidth, h_1^{opt}, for the estimation of the mean function $m(x_i)$ by minimizing MSE$[\hat{m}(x_i)|\mathbf{X}]$. One can use any bandwidth selection strategy described in Fan and Gijbels (1996).

2. Find the residuals

$$\hat{y}_i - \hat{m}_{h_1^{opt}}(x_i) \tag{55}$$

using the asymptotic optimal bandwidth h_1^{opt}.

3. Assume the observed residuals are true residuals. Apply the same bandwidth selector as in Step 1 to the observed squared residuals from Step 2 and obtain h_2^{opt}. By Remark 8, the proposed bandwidth selector will produce asymptotically optimal bandwidths (Ruppert et al., 1997).

5 Summary

A nonparametric mixed effects model is considered for small area estimation. Local polynomial estimators for both the mean function and the between area variance function are proposed. The between area variance function is estimated by smoothing the observed residuals. Theoretical properties of the proposed estimators are studied. A shrinkage estimator using the direct mean and the local polynomial estimator is proposed for the small area mean function. Asymptotic approximation for the MSE of the proposed estimator of the small area mean is derived. An asymptotic optimal bandwidth selection technique is discussed.

Our main theoretical contributions are results on the bias and the variance of the estimated mean and the estimated variance functions for a nonparametric mixed effects model. We develop predictors of small area mean function. The theoretical properties of the proposed estimators are studied. An optimal bandwidth selection method based on the estimated mean squared error of small area means is discussed. Our framework is expandable to the empirical Bayes estimation under a hierarchical nonparametric model assumption.

ACKNOWLEDGEMENTS

This research was supported in part by Cooperative Agreement No. 63-3A75-4-122 between the USDA Natural Resources Conservation Service and the Center for Survey Statistics and Methodology at Iowa State University. This work was accomplished when Pushpal K Mukhopadhyay was a graduate research assistant at Iowa State University. The authors are grateful to Dr. Jean Opsomer for his constructive suggestions.

References

Breidt, F. J. and Opsomer, J. D. (2000). Local polynomial regression estimators in survey sampling. *The Annals of Statistics*, 28(4):1026–1053.

Fan, J. and Gijbels, I. (1996). *Local Polynomial Modelling and Its Applications*. Chapman & Hall, London.

Fuller, W. A. (2006). Sampling statistics. Unpublished Manuscript, Center for Survey Statistics and Methodology.

Hastie, T. J. and Tibshirani, R. J. (1990). *Generalized Additive Models*. Chapman & Hall, London.

Prasad, N. G. N. and Rao, J. N. K. (1990). The estimation of the mean squared error of the small area estimators. *Journal of the American statistical association*, 85:163–171.

Rao, J. N. K. (2003). *Small Area Estimation*. Wiley Series in Survey Methodology, Hoboken, New Jersey.

Ruppert, D., Wand, M. P., Holst, U., and Hössjer, O. (1997). Local polynomial variance fuction estimation. *Technometrics*, 39(3):262–273.

Särndal, C. E., Swensson, B., and Wretman, J. (1991). *Model assisted survey sampling*. Springer, New York, NY.

Wand, M. P. and Jones, M. C. (1995). *Kernel Smoothing*. Chapman & Hall/CRC, Boca Raton, FL.