Estimating small area proportions with kernel logistic regressions models[†]

Jooyong Shim[1] · Changha Hwang[2]

[1]Department of Data Science, Inje University
[2]Department of Applied Statistics, Dankook University

Abstract

Unit level logistic regression model with mixed effects has been used for estimating small area proportions, which treats the spatial effects as random effects and assumes linearity between the logistic link and the covariates. However, when the functional form of the relationship between the logistic link and the covariates is not linear, it may lead to biased estimators of the small area proportions. In this paper, we relax the linearity assumption and propose two types of kernel-based logistic regression models for estimating small area proportions. We also demonstrate the efficiency of our proposed models using simulated data and real data.

Keywords: Logistic regression, mixed effect, proportion, small area estimation, spatial effect, support vector machine.

1. Introduction

Small area estimation (SAE) is a methodology for producing the estimates of parameters for small areas for which there is not enough sample to construct reliable estimates directly based on the survey sample. From a statistical point of view the small area is a small sub-population constituted by specific demographic and socioeconomic group of people, within a larger geographical areas. SAE has received considerable attention in recent years due to a growing demand for reliable small area statistics for policy analysis and planning purposes. There are two typical types of models for SAE, which are unit level and area level small area models. The most popular method to tackle SAE employs linear mixed effects model (Jiang and Lahiri, 2006). In fact, this model incorporates area-specific random effects to account for between-area variation which cannot be explained by the model covariates (Rao, 2003; Pfeffermann, 2013). However parametric formulation may not always be desirable because in many situations the dependence of the response or the link on the covariates exhibits more

complicated manners. The linear mixed effects model is useful for estimating the small area means efficiently under normality assumptions. Nonlinear SAE can be found in Opsomer *et al.* (2008), Salvati *et al.* (2011), Shim and Hwang (2012) and Shim *et al.* (2013). In this paper, we consider the situation when the estimate to be produced is proportion at small area level.

To estimate small area proportions, unit level mixed effects logistic regression (MELR) model with small area specific effects is commonly used. This model treats the spatial effects as random effects. In this paper we propose a mixed effects kernel logistic regression (MEKLR) that combines small area random effects with a smooth, nonparametrically specified trend, which is based on kernel technique of support vector machine (SVM). We also propose a kernel logistic regression (KLR) with mixed input variables, which utilizes a categorical covariate instead of using random effects to treat spatial effects. We call this KLRMIV. Similar idea was used in Shim (2012). The SVM, first developed by Vapnik (1995) and his group at AT&T Bell Laboratories, has been successfully applied to a number of real world problems related to classification and regression problems. For applications of SVM see Cho *et al.* (2010), Hwang and Shim (2012) and Shim and Hwang (2013).

The rest of this paper is organized as follows. Section 2 describes the MELR for estimating small area proportions. Section 3 illustrates MEKLR and KLRMIV models for estimating small area proportions. Section 4 and section 5 present numerical studies and conclusion, respectively.

2. MELR model for estimating small area proportions

To estimate small area proportions, unit level MELR model with small area specific effects is commonly used. Let a finite population U of size N be partitioned into m small areas of interest such that $\cup_{i=1}^{m} U_i = U$ and $\sum_{i=1}^{m} N_i = N$. Suppose y_{ij} is the value of a study variable y on unit j in area i for $i = 1, 2, \cdots, m, \; j = 1, 2, \cdots, N_i$. Here we consider the case y_{ij} takes on a value of zero or one, depending upon whether or not the jth individual within the ith small area possesses the characteristic of interest. Note that m is the number of small areas and N_i is the number of population units in area i. We are interested in estimating the small area proportion given by $p_i = N_i^{-1} \sum_{j \in U_i} y_{ij}$. Then, the following unit level MELR model is commonly used:

$$\log \frac{p_{ij}}{1 - p_{ij}} = b_0 + \boldsymbol{\beta}^t \boldsymbol{x}_{ij} + b_i, \; j = 1, \cdots, N_i, \; i = 1, \cdots, m, \tag{2.1}$$

where b_0 is constant term, b_i can be interpreted as the small area effect, $\boldsymbol{x}_{ij} \in R^d$ is covariate vector, and p_{ij} denotes the probability of a response for the the jth observation in the ith small area such that

$$p_{ij} = \frac{\exp\left(b_0 + \boldsymbol{\beta}^t \boldsymbol{x}_{ij} + b_i\right)}{1 + \exp\left(b_0 + \boldsymbol{\beta}^t \boldsymbol{x}_{ij} + b_i\right)}. \tag{2.2}$$

Then, the small area estimator of the proportion p_i is

$$\hat{p}_i = \frac{1}{N_i} \left(\sum_{j \in S_i} y_{ij} + \sum_{j \in R_i} \hat{y}_{ij} \right), \tag{2.3}$$

where S_i and R_i denote the sampled and non-sampled sets of units in area i, respectively, with $U_i = S_i \cup R_i$. Here, $\hat{y}_{ij}$ is the predicted value of response for population unit $j \in R_i$. In fact, we use the biased-adjusted estimator given as

$$\hat{p}_i = \frac{1}{N_i} \left\{ \sum_{j \in S_i} y_{ij} + \sum_{j \in R_i} \hat{y}_{ij} + \frac{N_i - n_i}{n_i} \sum_{j \in S_i} (y_{ij} - \hat{y}_{ij}) \right\}, \tag{2.4}$$

where n_i denotes the number of the units in S_i. The predictor (2.4) will have higher variability and so should only be used when there are large outlying data points. Refer to Salvati *et al.* (2011) for details. Here, $\hat{y}_{ij}$ is the estimated or predicted value of response for population unit $j \in S_i$ or $j \in R_i$.

In fact, $\hat{y}_{ij}$'s in (2.3) and (2.4) are obtained by utilizing $\hat{p}_{ij}$ given as

$$\hat{p}_{ij} = \frac{\exp\left(\hat{b}_0 + \hat{\boldsymbol{\beta}}^t \boldsymbol{x}_{ij} + \hat{b}_i\right)}{1 + \exp\left(\hat{b}_0 + \hat{\boldsymbol{\beta}}^t \boldsymbol{x}_{ij} + \hat{b}_i\right)}, \tag{2.5}$$

where $\boldsymbol{x}_{ij}$ is covariate vector for population unit $j \in S_i$ or $j \in R_i$. That is, if $\hat{p}_{ij} \geq 0.5$, then $\hat{y}_{ij} = 1$. Otherwise, $\hat{y}_{ij} = 0$.

3. KLR models for estimating small area proportions

In this section we illustrate MEKLR and KLRMIV models for estimating small area proportions.

3.1. MEKLR model

For the sampled units we rewrite (2.1) as follows:

$$\eta_{ij} \equiv \log \frac{p_{ij}}{1 - p_{ij}} = b_0 + \boldsymbol{\beta}^t \boldsymbol{x}_{ij} + \boldsymbol{b}^t \boldsymbol{u}_{ij}, j = 1, \cdots, n_i, \ i = 1, \cdots, m, \tag{3.1}$$

where $\boldsymbol{b} = (b_1, \cdots, b_m)^t$, $\boldsymbol{u}_{ij}$ is an $m \times 1$ indicator vector for the individual effect b_i. Then, the negative log-likelihood can be written as

$$\ell(b_0, \boldsymbol{\beta}, \boldsymbol{b}) = -\sum_{i=1}^{m} \sum_{j=1}^{n_i} y_{ij} \eta_{ij} + \sum_{i=1}^{m} \sum_{j=1}^{n_i} \log\left(1 + \exp\left(\eta_{ij}\right)\right). \tag{3.2}$$

A nonlinear form of MELR, known as MEKLR, can be obtained via the so-called "kernel trick", whereby a conventional MELR model is constructed in a high dimensional feature space induced by a Mercer (1909)'s kernel. More formally, given training data, $\cup_{i=1}^{m} S_i$, with $S_i = \{(\boldsymbol{x}_{ij}, y_{ij})\}_{j=1}^{n_i}$, $\boldsymbol{x}_{ij} \in \mathcal{X} \subset R^d$, a feature space $\mathcal{F}$ $(\boldsymbol{\phi} : \mathcal{X} \to \mathcal{F})$, is defined by a kernel function, $K : \mathcal{X} \times \mathcal{X} \to R$, that evaluates the inner product between the images of input vectors in the feature space, i.e., $K(\boldsymbol{x}_{ik}, \boldsymbol{x}_{il}) = \boldsymbol{\phi}(\boldsymbol{x}_{ik})^t \boldsymbol{\phi}(\boldsymbol{x}_{il})$. The kernel function used here is the Gaussian kernel,

$$K(\boldsymbol{x}_{ik}, \boldsymbol{x}_{il}) = \exp\left(-\frac{1}{\sigma^2} \|\boldsymbol{x}_{ik} - \boldsymbol{x}_{il}\|^2\right), \tag{3.3}$$

where σ^2 is the kernel parameter.

The penalized negative log-likelihood function of the MELR model constructed in the feature space is given as follows:

$$\ell(b_0, \boldsymbol{w}, \boldsymbol{b}) = -\sum_{i=1}^{m}\sum_{j=1}^{n_i} y_{ij}\eta_{ij} + \sum_{i=1}^{m}\sum_{j=1}^{n_i} \log\left(1 + \exp(\eta_{ij})\right) + \frac{\lambda}{2}\|\boldsymbol{w}\|^2, \tag{3.4}$$

where $\eta_{ij} = b_0 + \boldsymbol{w}^t\boldsymbol{\phi}(\boldsymbol{x}_{ij}) + \boldsymbol{b}^t\boldsymbol{u}_{ij}$ and λ is the penalty parameter. For convenience, we rearrange y_{ij}'s using single index and then denote each response by $y_i, i = 1, \cdots, n$. That is, y_{ij}'s are denoted as follows:

$$y_1 = y_{11}, y_2 = y_{12}, \cdots, y_{n_1} = y_{1,n_1}, y_{n_1+1} = y_{2,1}, \cdots, y_n = y_{m,n_m}.$$

We also rearrange $(\boldsymbol{x}_{ij}, \boldsymbol{u}_{ij})$'s and then denote these pairs using single index in accordance with y_{ij}'s. Then, the representer theorem (Kimeldorf and Wahba, 1971) guarantees that the minimizer of the penalized negative log-likelihood (3.4) to be $\eta_i = b_0 + \boldsymbol{k}_i\boldsymbol{\alpha} + \boldsymbol{b}^t\boldsymbol{u}_i$, where $\boldsymbol{k}_i$ is the ith row of the $n \times n$ kernel matrix $\boldsymbol{K}$ with elements $K(\boldsymbol{x}_i, \boldsymbol{x}_j), i, j = 1, \cdots, n$.

Thus, the penalized negative log-likelihood (3.4) can be rewritten as

$$\ell(\tilde{\boldsymbol{\alpha}}) = -\boldsymbol{y}\boldsymbol{W}\tilde{\boldsymbol{\alpha}} + \boldsymbol{1}_n^t \log\left(\boldsymbol{1}_n + \exp(\boldsymbol{W}\tilde{\boldsymbol{\alpha}})\right) + \frac{\lambda}{2}\tilde{\boldsymbol{\alpha}}^t\boldsymbol{K}_0\tilde{\boldsymbol{\alpha}}, \tag{3.5}$$

where $\tilde{\boldsymbol{\alpha}} = (\boldsymbol{\alpha}^t, \boldsymbol{b}^t, b_0)^t$, $\boldsymbol{y} = (y_1, \cdots, y_n)^t$, $\boldsymbol{W} = (\boldsymbol{K}, \boldsymbol{U}, \boldsymbol{1}_n)$, $\boldsymbol{U} = (\boldsymbol{u}_1, \cdots, \boldsymbol{u}_n)^t$, $\boldsymbol{1}_n$ is the $n \times 1$ vector of ones, $\log(\cdot), \exp(\cdot)$ are componentwise functions, and

$$\boldsymbol{K}_0 = \begin{pmatrix} \boldsymbol{K} & \boldsymbol{O}_1 \\ \boldsymbol{O}_1^t & \boldsymbol{O}_2 \end{pmatrix}$$

with the $n \times (m+1)$ zero matrix $\boldsymbol{O}_1$ and $(m+1) \times (m+1)$ zero matrix $\boldsymbol{O}_2$.

By minimizing the penalized negative log-likelihood (3.5) we obtain the estimator of parameter vector $\tilde{\boldsymbol{\alpha}}$, but not in a explicit form, which leads to use the IRWLS procedure. At $(l+1)$th iteration, the parameter vector is estimated as follows:

$$\tilde{\boldsymbol{\alpha}}^{(l+1)} = (\boldsymbol{W}^t\boldsymbol{D}\boldsymbol{W} + \lambda\boldsymbol{K}_0)^{-1}\boldsymbol{W}^t\boldsymbol{D}\boldsymbol{y}^*, \tag{3.6}$$

where $\boldsymbol{D}$ is a diagonal matrix of $\boldsymbol{p}^{(l)} \odot (\boldsymbol{1}_n - \boldsymbol{p}^{(l)})$, $\boldsymbol{p}^{(l)}$ is the estimate of $\boldsymbol{p} = (p_1, \cdots, p_n)^t$ obtained at the lth iteration, and $\boldsymbol{y}^* = \boldsymbol{D}^{-1}(\boldsymbol{y} - \boldsymbol{p}^{(l)}) + \boldsymbol{W}\tilde{\boldsymbol{\alpha}}^{(l)}$ is the working response vector. Here, $\odot$ represents the componentwise multiplication.

We now consider computing the estimator of the small area proportion p_i. If we obtain $\tilde{\boldsymbol{\alpha}}$ through the IRWLS procedure, then, given a covariate vector $\boldsymbol{x}_o$ for a population unit in S_i or R_i for $i = 1, \cdots, m$, we can obtain the corresponding estimated probability for $\boldsymbol{x}_o$ as follows:

$$\hat{p}_o = \frac{\exp\left(\hat{b}_0 + \boldsymbol{k}_o\hat{\boldsymbol{\alpha}} + \hat{\boldsymbol{b}}^t\boldsymbol{u}\right)}{1 + \exp\left(\hat{b}_0 + \boldsymbol{k}_o\hat{\boldsymbol{\alpha}} + \hat{\boldsymbol{b}}^t\boldsymbol{u}\right)}, \tag{3.7}$$

where $\boldsymbol{k}_o = (K(\boldsymbol{x}_o, \boldsymbol{x}_1), \cdots, K(\boldsymbol{x}_o, \boldsymbol{x}_n))$. Thus, $\hat{p}_o \geq 0.5$, then $\hat{y}_o = 1$. Otherwise, $\hat{y}_o = 0$. Then we can obtain the estimator of the small area proportion p_i by applying $\hat{y}_o$ to (2.3) or (2.4).

3.2. KLRMIV model

We now illustrate KLRMIV model. As mentioned before, this model treats spatial effects using a categorical covariate instead of using small area random effect. Thus, the penalized negative log-likelihood function for the KLRMIV model constructed in the feature space is given as follows:

$$\ell(b_0, \boldsymbol{w}) = -\sum_{i=1}^{m}\sum_{j=1}^{n_i} y_{ij}\eta_{ij} + \sum_{i=1}^{m}\sum_{j=1}^{n_i} \log\left(1 + \exp\left(\eta_{ij}\right)\right) + \frac{\lambda}{2}\|\boldsymbol{w}\|^2, \tag{3.8}$$

where $\eta_{ij} = b_0 + \boldsymbol{w}^t\boldsymbol{\phi}(\boldsymbol{x}_{ij})$ and λ is the penalty parameter.

We now divide covariate vector $\boldsymbol{x}_{ij}$ into two components $\boldsymbol{x}_{ij}^{(n)}$ and $\boldsymbol{x}_{ij}^{(c)}$, where $\boldsymbol{x}_{ij}^{(n)}$ is a vector of numerical covariates and $\boldsymbol{x}_{ij}^{(c)}$ is a vector of categorical covariates. As done before, we rearrange y_{ij}'s and $\boldsymbol{x}_{ij}$'s using single index, and then denote each response by $y_i, i = 1, \cdots, n$ and each covariate vector by $\boldsymbol{x}_i = (\boldsymbol{x}_i^{(n)}, \boldsymbol{x}_i^{(c)}), i = 1, \cdots, n$. Then, the representer theorem (Kimeldorf and Wahba, 1971) guarantees that the minimizer of the penalized negative log-likelihood (3.8) to be $\eta_i = b_0 + \boldsymbol{k}_i\boldsymbol{\alpha}$, where $\boldsymbol{k}_i$ is the ith row of the $n \times n$ kernel matrix $\boldsymbol{K}$ with elements $K(\boldsymbol{x}_i, \boldsymbol{x}_j), i, j = 1, \cdots, n$. Here, we propose the use of the kernel function defined as follows:

$$\begin{aligned} K(\boldsymbol{x}_i, \boldsymbol{x}_j) &= K^{(n)}(\boldsymbol{x}_i^{(n)}, \boldsymbol{x}_j^{(n)})K^{(c)}(\boldsymbol{x}_i^{(c)}, \boldsymbol{x}_j^{(c)}) \\ &= \exp\left(-\frac{1}{\sigma^2}\sum_{k=1}^{d_n}\frac{|x_{ik}^{(n)} - x_{jk}^{(n)}|}{d_n R_k} - \frac{1}{\beta}\sum_{l=1}^{d_c}\frac{I(x_{il}^{(c)} \neq x_{jl}^{(c)})}{d_c}\right), \end{aligned} \tag{3.9}$$

where d_n and d_c are the numbers of numerical and categorical covariates, respectively. Here, R_k is the range of the kth numerical covariate, σ^2 and β are kernel parameters. This kernel function can be expressed as a componentwise product of the weighted exponential kernel function and Hamming distance kernel function (Aradhye and Dorai, 2002), which leads an easy verification of kernelness (Genton, 2001).

Then, the penalized negative log-likelihood (3.8) can be written as

$$\ell(\tilde{\boldsymbol{\alpha}}) = -\boldsymbol{y}\boldsymbol{W}\tilde{\boldsymbol{\alpha}} + \boldsymbol{1}_n^t\log\left(\boldsymbol{1}_n + \exp(\boldsymbol{W}\tilde{\boldsymbol{\alpha}})\right) + \frac{\lambda}{2}\tilde{\boldsymbol{\alpha}}^t\boldsymbol{K}_0\tilde{\boldsymbol{\alpha}}, \tag{3.10}$$

where $\tilde{\boldsymbol{\alpha}} = (\boldsymbol{\alpha}^t, b_0)^t$, $\boldsymbol{y} = (y_1, \cdots, y_n)^t$, $\boldsymbol{W} = (\boldsymbol{K}, \boldsymbol{1}_n)$, $\boldsymbol{1}_n$ is the $n \times 1$ vector of ones, $\log(\cdot), \exp(\cdot)$ are componentwise functions, and

$$\boldsymbol{K}_0 = \begin{pmatrix} \boldsymbol{K} & \boldsymbol{0}_n \\ \boldsymbol{0}_n^t & 0 \end{pmatrix}$$

with the $n \times 1$ zero vector $\boldsymbol{0}_n$.

By minimizing the penalized negative log-likelihood (3.10) we obtain the estimator of parameter vector $\tilde{\boldsymbol{\alpha}}$, but not in a explicit form, which leads to use the IRWLS procedure. At $(l+1)$th iteration, the parameter vector is estimated as follows:

$$\tilde{\boldsymbol{\alpha}}^{(l+1)} = (\boldsymbol{W}^t\boldsymbol{D}\boldsymbol{W} + \lambda\boldsymbol{K}_0)^{-1}\boldsymbol{W}^t\boldsymbol{D}\boldsymbol{y}^*, \tag{3.11}$$

where $\boldsymbol{D}$ is a diagonal matrix of $\boldsymbol{p}^{(l)} \odot (\mathbf{1}_n - \boldsymbol{p}^{(l)})$, $\boldsymbol{p}^{(l)}$ is the estimate of $\boldsymbol{p} = (p_1, \cdots, p_n)^t$ obtained at the lth iteration, and $\boldsymbol{y}^* = \boldsymbol{D}^{-1}(\boldsymbol{y} - \boldsymbol{p}^{(l)}) + \boldsymbol{W}\tilde{\boldsymbol{\alpha}}^{(l)}$ is the working response vector. Here, $\odot$ represents the componentwise multiplication.

We now consider computing the estimator of the small area proportion p_i. If we obtain $\tilde{\boldsymbol{\alpha}}$ through the IRWLS procedure, then, given a covariate vector $\boldsymbol{x}_o$ for a population unit in S_i or R_i for $i = 1, \cdots, m$, we can obtain the corresponding estimated probability for $\boldsymbol{x}_o$ as follows:

$$\hat{p}_o = \frac{\exp\left(\hat{b}_0 + \boldsymbol{k}_o\hat{\boldsymbol{\alpha}}\right)}{1 + \exp\left(\hat{b}_0 + \boldsymbol{k}_o\hat{\boldsymbol{\alpha}}\right)}, \tag{3.12}$$

where $\boldsymbol{k}_o = (K(\boldsymbol{x}_o, \boldsymbol{x}_1), \cdots, K(\boldsymbol{x}_o, \boldsymbol{x}_n))$. Thus, $\hat{p}_o \geq 0.5$, then $\hat{y}_o = 1$. Otherwise, $\hat{y}_o = 0$. Then we can obtain the estimator of the small area proportion p_i by applying $\hat{y}_o$ to (2.3) or (2.4).

3.3. Model selection

The functional structures of the proposed two types of KLR models are characterized by the hyperparameters, that is, kernel and penalty parameters. To choose optimal values of hyperparameters of the model we define a leave-one-out cross validation (LOOCV) function as follows:

$$CV(\boldsymbol{\theta}) = \frac{1}{n} \sum_{i=1}^{n} \left(y_i - \hat{p}_i^{(-i)}\right)^2, \tag{3.13}$$

where $\boldsymbol{\theta}$ is a set of hyperparameters, and $\hat{p}_i^{(-i)} = \frac{\exp\left(\hat{\eta}_i^{(-i)}\right)}{1+\exp\left(\hat{\eta}_i^{(-i)}\right)}$ is the estimate of p_i without the ith observation. Among the candidate values of hyperparameters, we choose the values of hyperparameters which minimize the LOOCV function (3.13).

4. Numerical studies

In this section, we illustrate the performance of two types of KLR models using a simulated example and a real example. We note that we use (2.4) for the estimator of p_i in numerical studies.

Example 4.1 We generate the artificial data set consisting of 10 small areas by following two steps. The results are described in Table 4.1.

1. The size N_i of each $U_i = S_i \cup R_i$, $i = 1, \cdots, 10$, is given as $N_i = [100u_i]$, where u_i is generated from uniform distribution $U(0.5, 1.5)$ and $[\cdot]$ is round-off function. The size n_i of each s_i is given as $n_i = [0.1N_i]$, $i = 1, \cdots, 10$.

2. Generate the continuous covariate x_{ij}'s from uniform distribution $U(0, 1)$. The covariate x_{ij} is nonlinearly related to the canonical parameter as $\eta_{ij} = \sin(\pi x_{ij}) + b_i$, where b_i is generated from a normal distribution $N(0, 1)$. The proportion p_{ij}'s are obtained as $p_{ij} = \exp(\eta_{ij})/(1 + \exp(\eta_{ij}))$. We generate each response y_{ij} from Bernoulli distribution with $p_i = \frac{1}{N_i} \sum_{j=1}^{N_i} p_{ij}$.

We note that for KLRMIV $x_{ij}^{(n)}$'s are x_{ij}'s generated as above and $x_{ij}^{(c)} = i$, $i = 1, \cdots, 10$. After N_i's, n_i's and p_i's are determined and y_{ij}'s for all small areas are generated, we repeat 100 times the procedure of making S_i of size n_i from U_i, $i = 1, \cdots, 10$. In this sense we carry out a total of 100 simulations. For each small area, we compute the Monte Carlo estimate of the mean absolute error (MAE) regarding small area proportion given as

$$MAE_i = \frac{1}{100} \sum_{t=1}^{100} |\hat{p}_{it} - p_i|, \ i = 1, \cdots, m. \tag{4.1}$$

The results of MEKLR, KLRMIV and MELR are shown in Tables 4.1. The boldfaced figure in each column signifies the smallest averages of MAEs. From this table we can see that overall two types of KLR models work better than the MELR.

Table 4.1 Average of MAEs for the estimates of p_i's for artificial data

State	N_i	n_i	p_i	MEKLR	KLRMIV	MELR
1	145	15	0.8245	0.0906	**0.0859**	0.0924
2	73	7	0.9226	0.0956	**0.0890**	0.0958
3	111	11	0.9039	**0.0769**	0.0772	0.0776
4	99	10	0.6793	**0.1196**	0.1231	0.1331
5	139	14	0.1009	0.0784	**0.0765**	0.0789
6	126	13	0.6224	**0.1154**	0.1252	0.1261
7	96	10	0.3054	**0.1155**	0.1208	0.1227
8	52	5	0.1480	0.1378	**0.1363**	0.1386
9	132	13	0.2732	**0.0949**	0.0951	0.0956
10	94	9	0.1270	**0.0679**	0.0682	0.0683

Example 4.2 We now consider the numerical study on the 1996 presidential election data set which can be obtained in the spatial-econometrics library at http://www.spatial- economet- rics.com/ html/jplv7.zip. The data set contains approval indices of Clinton for presidential election in 853 counties of 13 states. The variables, N_i's, n_i's and p_i's are described in Table 4.3. Here, p_i denotes the approving rating of Clinton for each state. We note that for KL-RMIV $x_{ij}^{(n)}$'s are composed of variables described in Table 4.2 and $x_{ij}^{(c)} = i$, $i = 1, \cdots, 13$. As in Example 4.1, we carry out a total of 100 simulations and then compute MAEs using (4.1). The results of MEKLR, KLRMIV and MELR are reported in Tables 4.3. The bold-faced figure in each column signifies the smallest averages of MAEs. From this table we can see that overall two types of KLR models work better than the MELR.

Table 4.2 Description of input variables used in election data

Variable	Description
x_1	Logarithm of urban population
x_2	Logarithm of rural population
x_3	Population with high school or GED graduates as a proportion of educated
x_4	Population with some college as a proportion of educated
x_5	Population with associate degrees as a proportion of educated
x_6	Population with college degrees as a proportion of educated
x_7	Population with grad/professional degrees as a proportion of educated

Table 4.3 Average of MAEs for the estimates of p_i's for election data

State	N_i	n_i	p_i	MEKLR	KLRMIV	MELR
1	67	10	0.4627	**0.1074**	0.1278	0.1320
2	75	11	0.8800	0.1403	**0.1265**	0.1431
3	15	8	0.6000	0.4022	**0.3611**	0.4116
4	58	9	0.4310	**0.1019**	0.1205	0.1292
5	63	9	0.3492	**0.1111**	0.1208	0.1211
6	9	5	1.000	0.5556	0.5556	0.5556
7	3	2	1.000	0.6667	0.6667	0.6667
8	67	10	0.4776	0.1104	**0.1011**	0.1437
9	159	25	0.4906	**0.0687**	0.0825	0.0848
10	99	15	0.2020	0.0761	**0.0750**	0.0808
11	44	7	0.9091	0.1632	**0.1521**	0.1667
12	102	15	0.3725	**0.0802**	0.0989	0.1072
13	92	14	0.7174	0.1404	**0.1029**	0.1451

5. Conclusions

In this paper, we have studied how kernel-based logistic regression models perform in estimating small area proportions. By the way, it is difficult to compare the proposed MEKLR and KLRMIV models in terms of accuracy. The proposed models take over all advantages of SVM that capture nonlinearities in the data, that have good generalization ability, and that are useful tools when the data are characterized by complex patterns of spatial dependence. In particular, the proposed models can be easily used without heavy computations under high-dimensional covariate settings.

References

Aradhye, H. and Dorai, C. (2002). New kernels for analyzing multimodal data in multimedia using kernel machines. *Proceedings of the IEEE International Conference on Multimedia and Expo*, **2**, 37-40.

Cho, D. H., Shim, J. and Seok, K. H. (2010). Doubly penalized kernel method for heteroscedastic autoregressive data. *Journal of the Korean Data & Information Science Society*, **21**, 155-162.

Genton, M. G. (2001). Classes of kernels for machine learning: a statistics perspective. *Journal of Machine Learning Research*, **2**, 299-312.

Hwang, C. and Shim, J. (2012). Smoothing Kaplan-Meier estimate using monotone suppot vector regression. *Journal of the Korean Data & Information Science Society*, **23**, 1045-1054.

Jiang, J. and Lahiri, P. (2006). Mixed model prediction and small area estimation. *Test*, **15**, 1-96.

Kimeldorf, G. and Wahba, G. (1971). Some results on Tchebycheffian spline functions. *Journal of Mathematical Analysis and Applications*, **33**, 82-95.

Mercer, J. (1909). Functions of positive and negative type and their connection with theory of integral equations. *Philosophical Transactions of Royal Society A*, 415-446.

Opsomer, J. D., Claeskens, G., Ranalli, M. G., Kauemann, G. and Breidt, F. J. (2008). Non-parametric small area estimation using penalized spline regression. *Journal of the Royal Statistical Society B*, **70**, 265-286.

Pfeffermann, D. (2013). New important developments in small area estimation. *Statistical Science*, **28**, 40-68.

Rao, J. N. K. (2003). *Small area estimation*, Wiley, New York.

Salvati, N., Ranalli, M. G. and Pratesi, M. (2011). Small area estimation of the mean using non-parametric M-quantile regression: A comparison when a linear mixed model does not hold. *Journal of Statistical Computation and Simulation*, **81**, 945-964.

Shim, J. (2012). Kernel Poisson regression for mixed input variables. *Journal of the Korean Data & Information Science Society*, **23**, 1231-1239.

Shim, J. and Hwang, C. (2012). M-quantile kernel regression for small area estimation. *Journal of the Korean Data & Information Science Society*, **23**, 794-756.

Shim, J. and Hwang, C. (2013). Expected shortfall estimation using kernel machines. *Journal of the Korean Data & Information Science Society*, **24**, 625-636.

Shim, J., Kim, Y. and Hwang, C. (2013). Generalized kernel estimating equation for panel estimation of small area unemployment rates. *Journal of the Korean Data & Information Science Society*, **24**, 1199-1210.

Vapnik, V. N. (1995). *The nature of statistical learning theory*, Springer, New York.