Robust Small Area Estimation Using Penalized Spline Mixed Models

J. N. K. Rao, S. K. Sinha and M. Roknossadati
School of Mathematics and Statistics, Carleton University, Ottawa, Canada K1S 5B6

Abstract

Small area estimation has been extensively studied under linear mixed models. In particular, empirical best linear unbiased prediction (EBLUP) estimators of small area means and associated estimators of mean squared prediction error (MSPE) that are nearly unbiased have been developed. However, EBLUP estimators can be sensitive to outliers. Sinha and Rao (2009) developed a robust EBLUP method and demonstrated its advantages over the EBLUP under a unit level linear mixed model in the presence of outliers in the random small area effects and/or unit level errors. A bootstrap method of estimating MSPE of the robust EBLUP estimator was also proposed. In this paper, we relax the assumption of linear regression for the fixed part of the model and replace it by a weaker assumption of a penalized spline regression and develop robust EBLUP estimators. Bootstrap estimators of MSPE are also developed. Results of a limited simulation study are summarized.

Key words: Bootstrap, mean squared prediction error, outliers, random effects, small area mean, unit level model

1. Introduction

Area level and unit level linear mixed models have been extensively used in small area estimation. In this paper, we focus on unit level mixed models based on a single, continuous auxiliary variable, x, related to the variable of interest, y. A basic unit level linear mixed model, called nested error linear regression model (Battese et al 1988), is given by

$$y_{ij} = \beta_0 + \beta_1 x_{ij} + v_i + e_{ij}, i = 1,...,m; j = 1,..., N_i \qquad (1)$$

where m is the number of small areas, N_i is the number of population units in area i, $v_i \sim_{iid} N(0,\sigma_v^2)$ denote random small area effects that account for variation not explained by the auxiliary variables x_{ij}, and v_i is independent of the unit errors $e_{ij} \sim_{iid} N(0,\sigma_e^2)$. A sample of $n_i (\geq 1)$ units is drawn from each area i, and sampling is assumed to be ignorable in the sense that the population model (1) also holds for the sample. Here our interest is in estimating the small area mean $\overline{Y}_i$ of the N_i population values y_{ij}. If the sampling fraction n_i / N_i is negligible, then $\bar{Y}_i \approx \mu_i = \beta_0 + \beta_1 \bar{X}_i$ where $\bar{X}_i$ is the known population mean of the x_{ij} for area i. An empirical best (or empirical Bayes) estimator, abbreviated EB, of μ_i is obtained as a weighted sum of a "sample regression" estimator $\bar{y}_i + \hat{\beta}_1(\bar{X}_i - \bar{x}_i)$ and a synthetic estimator $\hat{\beta}_0 + \hat{\beta}_1 \bar{X}_i$, where $(\bar{y}_i, \bar{x}_i)$ are the sample means for area i and $\beta = (\hat{\beta}_0, \hat{\beta}_1)'$ are consistent estimators of the regression parameters in (1). The optimal weights depend on the variance components $\theta = (\sigma_v^2, \sigma_e^2)'$ which are replaced by consistent estimators, for example maximum likelihood (ML) or restricted ML (REML) estimators. The EB estimator is also an empirical best linear unbiased prediction (EBLUP) estimator without assuming normality. Rao (2003, chapter 7) gives a detailed account of EB and EBLUP estimation, associated mean squared prediction error (MSPE) approximation and a nearly unbiased estimator of MSPE.

The EBLUP estimator can be sensitive to outliers in v_i and e_{ij}. Sinha and Rao (2009) studied robust EBLUP (REBLUP) estimation of μ_i, using some general results of Fellner(1986) who studied robust estimation of random effects in linear mixed models. Fellner obtained robust "mixed model" equations to estimate β and $v = (v_1, ..., v_m)'$ for given θ, using Huber's $\psi -$function and proposed a two-step iterative procedure for getting a robust estimator of θ and in turn robust estimators of β and v. Sinha and Rao (2009) used an alternative method of estimating β and θ by solving robust score equations for β and θ and the resulting robust ML (RML) estimators are then substituted in the mixed model equation for v to get REBLUP of v. Their simulation results suggested that the proposed method of estimating θ can be significantly more efficient than Fellner's method, but the efficiency gains were small when estimating the small area mean μ_i.

The assumption of linear regression in (1) may be restrictive in practice. To get around this difficulty, $\beta_0 + \beta_1 x_{ij}$ in (1) is replaced by an unknown smooth function $f(x_{ij})$ which is assumed to be approximated sufficiently well by a penalized spline (P-spline) function (Rupert et al 2003, Opsomer et al 2008 and Ugarte et al 2009). Using a mixed model representation of the P-spline in (1), an EBLUP estimator of small area mean μ_i may be obtained. The estimation of MSPE of the EBLUP estimator presents some difficulties because the mixed model does not have a block diagonal covariance structure, unlike (1). Opsomer et al (2008) studied this problem and also proposed a bootstrap estimator of MSPE. Rubin-Bleuer et al (2009) studied P-spline area level models and obtained an EBLUP estimator of small area mean and associated bootstrap estimator of MSPE.

The EBLUP estimator under the P-spline mixed model can be sensitive to outliers in v_i and e_{ij}, as in the case of EBLUP under the nested error linear regression model (1). In this paper, we propose to obtain REBLUP estimator of small area mean μ_i under the P-spline version of (1) using Fellner's (1986) general results on robust mixed model equations and his two step iterative method; the approach of Sinha and Rao (2009) for robust estimation of β and θ runs into difficulty in the context of P-spline mixed models. We also propose a bootstrap estimator of MSPE of the REBLUP estimator. Results of a simulation study are summarized.

2. P-spline Mixed Model

We assume that the true mean specification $f(x_{ij})$ is well approximated by a P-spline approximation $sp(x_{ij}, K)$ based on the mixed model formulation. The resulting P-spline mixed model for the sample is given by

$$y_{ij} = sp(x_{ij}, K) + v_i + e_{ij}, i = 1,..., m; j = 1,..., n_i \qquad (2)$$

where

$$sp(x_{ij}, K) = \beta_0 + \beta_1 x_{ij} + \sum_{k=1}^{K} u_k (x_{ij} - q_k)_+ \qquad (3)$$

with the q_k denoting the K knots $q_1,..., q_k$, $(x-q)_+ = \max(0, x-q)$ and $u_k \sim_{iid} N(0, \sigma_u^2)$. Regarding the choice of K and q_k, we have followed Rupert et al (2003) in our simulation study: (1) One needs “enough” knots to ensure sufficient flexibility to fit the data, but after that additional knots do not change the fit much. (2) Place the knots at the sample quantiles of the unique $x-$values which gives equal or nearly equal number of $x-$values between knots. Note that $\mu_i = N_i^{-1} \sum_j f(x_{ij}) + v_i$ which is approximated by $\mu_{iP} = N_i^{-1} \sum_j sp(x_{ij}, K) + v_i$. In the simulation study, we generated $x_{ij} \sim_{iid} N(1,1) \equiv X$ in which case we have

$$\mu_i = E\{f(X)\} + v_i \qquad (4)$$

and

$$\mu_{iP} = \beta_0 + \beta_1 E(X) + \sum_{k=1}^{K} u_k E\{(X - q_k)_+\} + v_i \qquad (5)$$

We express the P-spline mixed model (2) in matrix form as

$$y = X\beta + Wu + Zv + e \qquad (6)$$

where $u \sim N(0, \sigma_u^2 I_K), v \sim N(0, \sigma_v^2 I_m)$ and $e \sim N(0, \sigma_e^2 I_n)$ with $n = \sum_i n_i$. Here u is the $K-$vector of the spline effects u_k, v is the $m-$vector of the small area random effects

v_i and e is the $n-$vector of the unit errors e_{ij}. The matrix form of (1) is given by (6) without the term Wu .Note that (6) does not have a block diagonal structure, unlike the matrix version of (1), because of the additional term Wu .

3. P-spline EBLUP Estimators

We first consider EBLUP estimation of the random effects in the P-spline mixed model (6), following Fellner (1986), and then modify the equations to get robust EBLUP estimators in the presence of outliers. For fixed $\theta = (\sigma_u^2, \sigma_v^2, \sigma_e^2)'$, the BLUP estimators $\tilde{\beta} = \tilde{\beta}(\theta), \tilde{u} = \tilde{u}(\theta)$ and $\tilde{v} = \tilde{v}(\theta)$ of β, u and v are obtained by solving the following "mixed model" equations of Henderson (1963):

$$
\begin{aligned}
&\sigma_e^{-2} X'(y - X\beta - Wu - Zv) = 0 \\
&\sigma_e^{-2} W'(y - X\beta - Wu - Zv) - \sigma_u^{-2} u = 0 \qquad (7) \\
&\sigma_e^{-2} Z'(y - X\beta - Wu - Zv) - \sigma_v^{-2} v = 0
\end{aligned}
$$

Now replacing θ by the REML estimator $\hat{\theta}$ we get the EBLUP estimators $\hat{\beta} = \tilde{\beta}(\hat{\theta}), \hat{u} = \tilde{u}(\hat{\theta})$ and $\hat{v} = \tilde{v}(\hat{\theta})$. Fellner (1986), following Harville (1977), obtained REML equations which are solved iteratively in conjunction with (7) to get the REML estimator of θ. Fellner's REML equations for the P-spline mixed model (6) may be written as

$$
\begin{aligned}
&\sigma_u^2 = \sum_{k=1}^{K} \tilde{u}_k^2 / (K - t_1) \\
&\sigma_v^2 = \sum_{i=1}^{m} \tilde{v}_i^2 / (m - t_2) \qquad (8) \\
&\sigma_e^2 = \sum_i \sum_j \tilde{e}_{ij}^2 / \{(n-2) - (K - t_1) - (m - t_2)\}
\end{aligned}
$$

where $\tilde{e}_{ij} = y_{ij} - \tilde{\beta}_0 - \tilde{\beta}_1 x_{ij} - \sum_k \tilde{u}_k (x_{ij} - q_k)_+ - \tilde{v}_i$, $t_1 = tr(T_{11}) / \sigma_u^2$ and $t_2 = tr(T_{22}) / \sigma_v^2$ with T_{11} and T_{22} denoting the diagonal blocks of a partitioned matrix T which is the inverse of the partitioned matrix with diagonal blocks given by $\sigma_e^{-2} W'W + \sigma_u^{-2} I_K$ and $\sigma_e^{-2} Z'Z + \sigma_v^{-2} I_m$ and off-diagonals given by $\sigma_e^{-2} W'Z$ and its transpose.

An EBLUP estimator, $\hat{\mu}_{iP}$, of the P-spline approximation to area i mean μ_i is then obtained from (5) by replacing β_0, β_1, u_k and v_i by the estimators $\hat{\beta}_0, \hat{\beta}_1, \hat{u}_k$ and $\hat{v}_i$:

$$
\hat{\mu}_{iP} = \hat{\beta}_0 + \hat{\beta}_1 E(X) + \sum_{k=1}^{K} \hat{u}_k E(X - q_k)_+ + \hat{v}_k \qquad (9)
$$

The mean squared prediction error (MSPE) of the P-spline EBLUP estimator $\hat{\mu}_{iP}$ is given by $MSPE(\hat{\mu}_{iP}) = E(\hat{\mu}_{iP} - \mu_i)^2$ where the expectation is with respect to the true underlying model.

4. Robust P-spline EBLUP Estimators

We now obtain robust P-spline EBLUP estimators, using Huber's (1973) robust M-estimation approach with ψ −function given by $\psi_b(u) = u\min(1, b/|u|)$, where $b > 0$ is a tuning constant, commonly chosen as $b = 1.345$. Robust estimators of β, u and v for fixed θ, denoted $\tilde{\beta}_M = \tilde{\beta}_M(\theta), \tilde{u}_M = \tilde{u}_M(\theta)$ and $\tilde{v} = \tilde{v}_M(\theta)$, are obtained by solving robust mixed model equations. The latter equations are obtained from (7) by replacing $y - X\beta - Wu - Zv$ by $\sigma_e\psi\{\sigma_e^{-1}(y - X\beta - Wu - Zv)\}$, u by $\sigma_u\psi(\sigma_u^{-1}u)$ and v by $\sigma_v\psi(\sigma_v^{-1}v)$.

Robust REML equations for estimating the variance components in the P-spline mixed model are obtained from Fellner's REML equations (8) by making the following changes: Replace $\tilde{u}_k$ by $\sigma_u\psi(\sigma_u^{-1}\tilde{u}_{kM})$, $\tilde{v}_i$ by $\sigma_v\psi(\sigma_v^{-1}\tilde{v}_{iM})$ and $\tilde{e}_{ij}$ by $\sigma_e\psi(\sigma_e^{-1}\tilde{e}_{ijM})$, where

$$\tilde{e}_{ijM} = \tilde{e}_{ijM}(\theta) = y_{ij} - \tilde{\beta}_{0M} - \tilde{\beta}_{1M}x_{ij} - \sum_k \tilde{u}_{kM}(x_{ij} - q_k)_+ - \tilde{v}_{iM}$$

Now solving the modified equations corresponding to (7) and (8) iteratively as desribed in Section 3, we get a robust estimator of the variance component vector θ denoted by $\hat{\theta}_M = (\hat{\sigma}_{uM}^2, \hat{\sigma}_{vM}^2, \hat{\sigma}_{eM}^2)'$ and robust EBLUP estimators of β, u and v as $\hat{\beta}_M = \tilde{\beta}_M(\hat{\theta}_M), \hat{u}_M = \tilde{u}_M(\hat{\theta}_M)$ and $\hat{v}_M = \tilde{v}_M(\hat{\theta}_M)$. Now substituting the robust EBLUP estimators for β, u and v in the P-spline approximation μ_{iP} given by (5), we get the robust EBLUP (REBLUP) estimator of μ_{iP} as

$$\hat{\mu}_{iPM} = \hat{\beta}_{0M} + \hat{\beta}_{1M}E(X) + \sum_{k=1}^{K}\hat{u}_{kM}E\{(X - q_k)_+\} + \hat{v}_{iM} \qquad (10)$$

5. Bootstrap Estimation of MSPE

It seems difficult to get an analytical formula for an MSPE estimator that is nearly unbiased. Therefore, in this paper we focus on bootstrap estimation of MSPE of the robust P-spline EBLUP estimator $\hat{\mu}_{iPM}$. The basic idea is to mimic the MSPE by using simulated samples generated from an estimated model, following Sinha and Rao (2009) who considered REBLUP estimation and associated bootstrap MSPE estimation for the nested error linear regression model.

The steps in implementing bootstrap estimation are as follows:

1. Generate independently $u_k^* \sim_{iid} N(0, \hat{\sigma}_{uM}^2), v_i^* \sim_{iid} N(0, \hat{\sigma}_{vM}^2)$ and $e_{ij}^* \sim_{iid} N(0, \hat{\sigma}_{eM}^2)$. Let $y_{ij}^* = \hat{\beta}_{0M} + \hat{\beta}_{1M} E(X) + \sum_{k=1}^{K} u_k^* E(X - q_k)_+ + v_i^* + e_{ij}^*$, $j = 1,...,n_i; i = 1,...,m$ and $\mu_i^* = \hat{\beta}_{0M} + \hat{\beta}_{1M} E(X) + \sum_{k=1}^{K} u_k^* E(X - q_k)_+ + v_i^*$, $i = 1,...,m$.

2. Calculate robust P-spline EBLUP $\hat{\mu}_{iMP}^*$ from the bootstrap data $\{(y_{ij}^*, x_{ij}) : j = 1,...,n_i; i = 1,...,m\}$.

3. Theoretical bootstrap estimator of MSPE of $\hat{\mu}_{iPM}^*$ is then given by

$$mspe_B(\hat{\mu}_{iPM}) = E_*(\hat{\mu}_{iPM}^* - \mu_i^*)^2 \qquad (11)$$

where E_* denotes bootstrap expectation.

4. In practice, we generate a large number, B, of bootstrap samples $b(=1,...,B)$ and calculate $\hat{\mu}_{iPM}^{*(b)}$ and $\mu_i^{*(b)}$ from each sample b. We then approximate (11) by

$$mspe_{B(a)}(\hat{\mu}_{iPM}) = B^{-1} \sum_{b=1}^{B} (\hat{\mu}_{iPM}^{*(b)} - \mu_i^{*(b)})^2 . \qquad (12)$$

In the simulation study (Section 6) we generated a large number of samples $r = 1,...,R$ from the assumed true model with specified parameters and from each sample we generated B bootstrap samples and calculated (12). The values of $\hat{\mu}_{iPM}$ and μ_i for sample r are denoted by $\hat{\mu}_{iPM}^{(r)}$ and $\mu_i^{(r)} = \beta_0 + \beta_1 E(x) + \sum_{k=1}^{K} u_k^{(r)} E\{(X - q_k)_+\} + v_i^{(r)}$ where $u_k^{(r)}$ and $v_i^{(r)}$ are the values of u_k and v_i for the simulation run r generated from specified distributions, for example contaminated normal distributions, see Section 6. The MSPE of $\hat{\mu}_{iPM}$ is approximated by

$$MSPE(\hat{\mu}_{iPM}) \approx R^{-1} \sum_{r=1}^{R} (\hat{\mu}_{iPM}^{(r)} - \mu_i^{(r)})^2 \qquad (13)$$

Similarly, the bootstrap MSPE estimator (12) is calculated for each simulation run r and then averaged over r to get an approximation to the expectation of the bootstrap MSPE estimator. The simulated relative bias (RB) of the MSPE estimator is then calculated from the latter quantity and the approximation (13) using the formula

$$RB = \{E(mspe_{B(a)}) - MSPE\}/(MSPE). \qquad (14)$$

6. Simulation Study

In this section we report some simulation results on the performance of the proposed robust P-spline EBLUP estimator and the associated bootstrap estimator of MSPE. The following functions $f(x)$ were used in generating the samples:

Model 1: $f(x) = 1 + x$ (linear),
Model 2: $f(x) = 1 + x + x^2$ (quadratic) and
Model 3: $f(x) = 1 + 2(x-1) + \exp\{-4(x-1)^2\}$ (bump function, Breidt et al 2005).

We assumed that $X \sim N(1,1)$ so that we have $E\{f(X)\}$ equal to 2 for model 1, 4 for model 2 and $1+\frac{1}{3}$ for model 3 respectively. We generated $\{x_{ij} : j = 1,...,4; i = 1,...,40\}$ from $N(1,1)$ and held them fixed for generating the sample responses y_{ij} from the assumed true model $y_{ij} = f(x_{ij}) + v_i + e_{ij}$, where the random effects v_i and the unit errors e_{ij} are drawn either from contaminated normal distributions or from t distributions with 3 degrees of freedom to reflect outliers either in v_i or in e_{ij} or in both. For the contaminated distributions we assumed that

$$v_i \sim_{iid} (1-\gamma_1)N(0,\sigma_v^2) + \gamma_1 N(0,\sigma_{v1}^2),\ e_{ij} \sim_{ind} (1-\gamma_2)N(0,\sigma_e^2) + \gamma_2 N(0,\sigma_{e1}^2)$$

where $\sigma_v^2 = \sigma_e^2 = 1$ and $\sigma_{v1}^2 = \sigma_{e1}^2 = 25$. Four combinations of distributions for v and e, denoted $(0,0), (v,0), (0,e)$ and (v,e), were studied, where $(0,0)$ indicates no contamination $(\gamma_1 = \gamma_2 = 0)$, $(v,0)$ indicates contamination in v only $(\gamma_1 = 0.1, \gamma_2 = 0)$, $(0,e)$ indicates contamination in e only $(\gamma_1 = 0, \gamma_2 = 0.1)$ and (v,e) indicates contamination in both v and e $(\gamma_1 = 0.1, \gamma_2 = 0.1)$.

For specified distributions of v_i and e_{ij} we generated $R = 500$ samples $\{v_i^{(r)}, e_{ij}^{(r)} : j = 1,...,4; i = 1,...,40\}$ and then the associated responses $\{y_{ij}^{(r)}; j = 1,...,4; i = 1,...,40\}$ from the true model $(r = 1,..., R)$. Then the P-spline EBLUP and robust EBLUP estimators $\hat{\mu}_{iP}$ and $\hat{\mu}_{iPM}$ were computed from each simulated sample $\{(y_{ij}^{(r)}, x_{ij}) : j = 1,..,4; i = 1,...,40\}$ for specified number of knots ($q = 0,20,30$). The simulated $MSPE$ of the estimators were then computed using the generated estimates $\hat{\mu}_{iP}^{(r)}$ and $\hat{\mu}_{iPM}^{(r)}$, and the small area means $\mu_i^{(r)}$ $(r = 1,..., R = 500)$, using (13) for $\hat{\mu}_{iPM}$ and a similar expression for $\hat{\mu}_{iP}$. It may be noted that $q = 0$ corresponds to the standard nested error linear regression model (1).

We summarize the simulation results based on average MSPE over the areas. Detailed results will be reported in a separate paper after implementing further simulations based on a larger number of simulation runs and other scenarios. Some broad results from the simulation study are the following: (1) MSPE for the P-spline estimators is not affected by the choice of q, when we compared the values for $q = 20$ to the corresponding values for $q = 30$. This result suggests that $q = 20$ is a good choice. (2) In the case of contamination in e (model $(0, e)$) or in both v and e (model (v, e)) robust EBLUP leads to significant reduction of MSPE relative to EBLUP. For example, for $q = 20$ and quadratic true model, average MSPE (%) for robust EBLUP is 32.0 compared to 49.3 for EBLUP under $(0, e)$ and 38.2 compared to 68.1 under (v, e). On the other hand, EBLUP is quite robust across the three models under $(v,0)$. Sinha and Rao (2009) observed a similar result for the linear case (model1) and $q = 0$. (3) In the linear case (model1), the increase in MSPE of the P-spline EBLUP over the EBLUP is minimal across the four contamination combinations. On the other hand, EBLUP (with $q = 0$) leads to large increase in average MSPE (%) relative to P-spline EBLUP when the true model is quadratic (model 2). For example, average MSPE (%) of EBLUP is 45.7 compared to 21.2 for the P-spline EBLUP with $q = 20$ in the case of $(0,0)$, and similarly for the robust EBLUP versus robust P-spline EBLUP and the three contamination combinations. In the case of bump function (model 3), however, the increase in MSPE of the EBLUP estimator relative to the P-spline EBLUP estimator is small (similarly for the robust EBLUP versus robust P-spline EBLUP). This is perhaps due to the fact that the bump function is closer to linearity. We need to confirm this result by further study. (4) Results for the t distribution case are similar to the above results for the contaminated distributions.

We also computed the bootstrap estimates of MSPE for each simulation run using the approximation (12) with $B = 200$ and then used (14) to obtain an approximation to the average absolute relative bias of the bootstrap MSPE estimator for the contamination cases. Our results suggest that the bootstrap MSPE estimator performs quite well in terms of average absolute relative bias: less than 10% in most cases. But further study is needed to confirm these results.

All in all, our limited simulation study indicates that the proposed robust P-spline EBLUP estimator with “enough” knots (say $q = 20$) performs well in terms of MSPE when the true mean function is not linear. Also, the proposed bootstrap MSPE estimator seems to track the MSPE quite well.

References

Battese, G. E., Harter, R. M. and Fuller, W. A. (1988). An error component model for prediction of county crop areas using survey and satellite data. *Journal of the American Statistical Association*, 83, 28-36.

Breidt, F. J., Claeskens, G. and Opsomer, J. D. (2005). Model-assisted estimation for complex surveys using penalised splines. *Biometrika*, 92, 831-846.

Fellner, W. H. (1986). Robust estimation of variance components. *Technometrics*, 28, 51-60.

Harville, D. A. (1977). Maximum likelihood approaches to variance component estimation and to related problems. *Journal of the American Statistical Association*, 72, 320-340.

Henderson, C. R. (1963). Selection index and expected genetic advance. In *Statistical Genetics and Plant Breeding*, National Research Council Publication 982, National Academy of Science, pp. 141-163.

Opsomer, J. P., 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, Series B*, 70, 265-286.

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

Rubin-Bleuer, S., Dochitoiu, C. and Rao, J. N. K. (2009). Bootstrap estimation of the mean -squared prediction error in small area models for business data. Paper presented at SAE 2009 Conference, Elche, Spain.

Rupert, D., Wand, M. P. and Carroll, R. J. (2003). *Semiparametric Regression*. Cambridge University Press, New York.

Sinha, S. K. and Rao, J. N. K. (2009). Robust small area estimation. *The Canadian Journal of Statistics*, 37, 381-399.

Ugarte, M. D., Goicoa, T., Militino, A. F. and Durban, M. (2009). Spline smoothing in small area trend estimation and forecasting. *Computational Statistics and Data Analysis*, 53, 3616-3629.