Survey Methodology, December 2005
Vol. 31, No. 2, pp. 183-195
Statistics Canada, Catalogue No. 12-001-XPB

Spatio-Temporal Models in Small Area Estimation

Bharat Bhushan Singh, Girja Kant Shukla and Debasis Kundu [1]

Abstract

A spatial regression model in a general mixed effects model framework has been proposed for the small area estimation problem. A common autocorrelation parameter across the small areas has resulted in the improvement of the small area estimates. It has been found to be very useful in the cases where there is little improvement in the small area estimates due to the exogenous variables. A second order approximation to the mean squared error (MSE) of the empirical best linear unbiased predictor (EBLUP) has also been worked out. Using the Kalman filtering approach, a spatial temporal model has been proposed. In this case also, a second order approximation to the MSE of the EBLUP has been obtained. As a case study, the time series monthly per capita consumption expenditure (MPCE) data from the National Sample Survey Organisation (NSSO) of the Ministry of Statistics and Programme Implementation, Government of India, have been used for the validation of the models.

Key Words: Mixed effects linear model; Spatial autocorrelation; Weight matrix; Best linear unbiased predictor; Empirical best linear unbiased predictor; Kalman filtering; NSSO rounds.

1. Introduction

Local level planning requires reliable data at the appropriate level. The complete enumeration or large sample surveys with adequate sample size is expensive and time consuming. The censuses are usually carried out once in a decade, while the sample surveys are often planned to provide estimates at much higher level. One such large sample survey is socio-economic survey of National Sample Survey Organisation (NSSO). Here the direct survey estimates are available at small area (district) level as most of the districts are stratum in the sampling procedure adopted by the NSSO. However, the estimates are exceedingly unreliable due to unacceptably large standard errors. This requires strengthening of such estimates with the use of information from similar small areas or with the help of some relatable exogenous variables, easily available and related to the variable under study.

Various model based approaches have been suggested to improve the direct estimators. The model-based approach facilitates its validation through the sample data. The simple area specific model suggested is two stage model of Fay and Herriot (1979).

$$y_i = \theta_i + \varepsilon_i,\ \ E(\varepsilon_i \mid \theta_i) = 0,\ \ \mathrm{Var}(\varepsilon_i \mid \theta_i) = \sigma_i^2, \qquad (1.1)$$

$$\theta_i = X_i^T \beta + v_i z_i,\ E(v_i) = 0,\ \mathrm{Var}(v_i) = \sigma_v^2,\ i = 1, 2, \ldots, m. \quad (1.2)$$

Here y_i's are direct survey estimators of θ_i's, the characteristic under study. θ_i's may be population small area means. $X_i = (X_{i1}, \ldots, X_{ip})^T$'s are exogenous variables which are available and assumed to be closely related to θ_i's and z_i's are known positive constants. $\beta(p \quad 1)$ is the vector of regression parameters.

The first equation (1.1) is the design model while the second (1.2) is the linking model. The ε_i's are sampling errors. Estimators y_i's are design unbiased and the sampling variances σ_i^2's are known. Further the ε_i's and v_i's are identically and independently distributed random variables. Normality of the random errors and random effects are often assumed. For this model, best linear unbiased predictor (BLUP) on the line of the best linear unbiased estimator (BLUE) has been suggested. The estimate is design consistent and model unbiased (Ghosh and Rao 1994). It is typically the weighted average of the direct survey estimator y_i and the regression synthetic estimator $X_i^T\beta$. The BLUP estimator depends on variance component σ_v^2 which is unknown in pratical applications. Various methods of estimating variance components in general mixed effects linear model are available (Cressie 1992). By replacing σ_v^2 with an asymptotically consistent estimator $\hat{\sigma}_v^2$, an empirical best linear unbiased predictor (EBLUP) has also been obtained.

The main problem associated with the data in the Indian context is the non-availability of administrative or civic registration data at small area level. Often, it is difficult to find out the exogenous variables closely related (multiple correlation coefficient $R^2 > 0.5$) to the variable under study.

In the present paper, the exploitation of spatial autocorrelation amongst the small area units in the form of spatial model, has been considered for improving the small area estimators. Besides this, for the time series data, a spatial temporal model on the line of Kalman filtering has been utilised to further improve the estimators. Time series data on monthly per capital consumption expenditure

1. Bharat Bhushan Singh, Girja Kant Shukla and Debasis Kundu, Department of Mathematics, I.I.T. Kanpur-208016. E-mail: drbbsingh@hotmail.com.

(MPCE) as estimated from a large sample survey carried out by the National Sample Survey Organisation (NSSO) has been studied. In the present paper, we propose suitable models in the framework of mixed effects linear model to provide better estimators of the MPCE at small area level.

Rest of the paper has been organized as follows. In Section 2, we consider a Spatial Model on the line of general mixed effects linear model with the introduction of spatial autocorrelation among the small area units. The BLUP and EBLUP of the mixed effects have been presented. A second order approximation to the MSE of the EBLUP and to the estimator of the MSE has also been obtained. Section 3 deals with the time series extension of Spatial Model in form of Spatial Temporal Model, using the Kalman filtering approach. The BLUP and the EBLUP of the mixed effects along with a second order approximation to the MSE of the EBLUP and to the estimator of the MSE have been discussed. Section 4 presents and analyses estimates of the MPCE from a large sample survey carried out periodically in India. The conclusions of the data analysis are reported in Section 5. All the proofs have been provided in the Appendix.

2. Spatial Model

The small area characteristics usually have the spatial dependence in terms of neighbourhood similarities. Cressie (1990) used conditional spatial dependence among random effects, in the context of adjustment for census undercounts. Here, we use simultaneous spatial dependence (Cliff and Ord 1981) among the random effects which has certain advantage over conditional dependence (Ripley 1981). We have thus tried to explain a portion of the random error unaccounted for and left over by explanatory variables which makes it possible to improve the direct survey estimators. The proposed model is a three stage area specific model (Ghosh and Rao 1994).

$$y = \theta + \varepsilon, \qquad \varepsilon \sim N_m(0, R), \tag{2.1}$$

$$\theta = X\beta + u, \tag{2.2}$$

$$u = \rho W u + v, \qquad v \sim N_m(0, \sigma_v^2 I), \tag{2.3}$$

where θ is a m–component vector (corresponding to number of small areas) for the characteristic under study and y is its direct surevy estimator obtained through small sample data. In the above model, the first equation (2.1) shows the design (sampling) model, the second equation (2.2) shows regression model and the third one (2.3) shows spatial model on the residuals, the later two are linked in the first equation. The above model can be expressed as

$$y = X\beta + Zv + \varepsilon, \qquad Z = (I - \rho W)^{-1}, \tag{2.4}$$

where $X(m \quad p)$ is the design matrix of full column rank p, $\beta(p \quad 1)$ is a column vector of regression parameters and $Z(m \quad m)$ represents the coefficients of random effects *v*. $W(m \quad m)$ is a known spatial weight matrix which shows the amount of interaction between any pair of small areas. The elements of $W \equiv [W_{ij}]$ with $W_{ii} = 0 \quad \forall i$ may depend on the distance between the centers of small areas or on the length of common boundary between them. As a simple alternative, it may have binary values $W_{ij} = 1$ (unscaled) if j^{th} area is physically contiguous to i^{th} area and $W_{ij} = 0$, otherwise. The matrix has been standardised so as to satisfy $\sum_{j=1}^{m} W_{ij} = 1$ for $i = 1, 2, \ldots, m$. The constant ρ is a measure of the overall level of spatial autocorrelation and its magnitude reflects the suitability of *W* for given *y* and *X*. Further v and ε are assumed to be independent of each other. *R* is a diagonal matrix of order *m* which may be expressed as $R = \operatorname{diag}(\sigma_1^2, \sigma_2^2, \ldots, \sigma_m^2)$ where σ_i^2's are known sampling variances corresponding to the i^{th} area. The parameter vector $\psi = [\rho, \sigma_v^2]^T$ has two elements.

In this model the strength is borrowed from the similar small areas through two common parameters *viz.* regression parameter β and autocorrelation parameter ρ. Note that the present model is a more general model and the model of Fay and Herriot (1979) can be obtained from this by taking $\rho = 0$.

By adopting the mixed effects linear model approach (Henderson 1975), the best linear unbiased predictor (BLUP) of $\theta = X\beta + Zv$ and the mean squared error (MSE) of the BLUP may be obtained as

$$\begin{aligned}\hat{\theta}(\psi) &= X\hat{\beta}(\psi) + \Lambda(\psi)[y - X\hat{\beta}(\psi)] \\ &= \sigma_v^2 A^{-1}(\psi)\Sigma^{-1}(\psi)y + R\Sigma^{-1}(\psi)X\hat{\beta}(\psi),\end{aligned} \tag{2.5}$$

$$\begin{aligned}&\mathrm{MSE}[\hat{\theta}(\psi)] = \\ &E[(\hat{\theta}(\psi) - \theta)(\hat{\theta}(\psi) - \theta)^T] = g_1(\psi) + g_2(\psi),\end{aligned} \tag{2.6}$$

$$g_1(\psi) = \Lambda(\psi)R = R - R\Sigma^{-1}(\psi)R, \tag{2.7}$$

$$g_2(\psi) = R\Sigma^{-1}(\psi)X(X^T\Sigma^{-1}(\psi)X)^{-1}X^T\Sigma^{-1}(\psi)R, \tag{2.8}$$

$$\hat{\beta}(\psi) = [X^T\Sigma^{-1}(\psi)X]^{-1}X^T\Sigma^{-1}(\psi)y,$$

$$\Sigma(\psi) = \sigma_v^2 A^{-1}(\psi) + R,$$

$$\Lambda(\psi) = \sigma_v^2 A^{-1}(\psi)\Sigma^{-1}(\psi), \; A(\psi) = (I - \rho W)^T(I - \rho W).$$

Here $\hat{\beta}, \Sigma$ and *A*, all are the functions of ψ and usually have been expressed as $\hat{\beta}(\psi), \Sigma(\psi)$ and $A(\psi)$ respectively. However, sometimes due to brevity, the suffix ψ has been omitted. The first term, $g_1(\psi)$ in the expression for the MSE, shows the variability of $\hat{\theta}$ when all the parameters are known and is of order $O(1)$. The second term, $g_2(\psi)$, due to estimating the fixed effects β, is of order $O(m^{-1})$ for large *m*. Further, with $\rho = 0$, the above model reduces to the standard mixed effects linear regression model while for $X\beta =$, we obtain a purely spatial scheme with only intercept term.

In practice parameter ψ is unknown and is estimated from the data. The maximum likelihood estimator (MLE) of the parameter, ψ is obtained by maximizing the following log likelihood function of ψ

$$
\begin{aligned}
l = \text{const} - \frac{1}{2}\log[|\Sigma(\psi)|] \\
- \frac{1}{2}[y - X\hat{\beta}(\psi)]^T \Sigma^{-1}(\psi)[y - X\hat{\beta}(\psi)]
\end{aligned} \tag{2.9}
$$

with respect to the parameter ψ. The empirical best linear unbiased predictor (EBLUP), $\hat{\theta}(\psi)$ and the naive estimator of the MSE are obtained from the equations (2.5) and (2.6) respectively, by replacing the parameter vector ψ by its estimator $\hat{\psi}$.

$$\hat{\theta}(\hat{\psi}) = \hat{\sigma}_v^2 A^{-1}(\hat{\psi})\Sigma^{-1}(\hat{\psi})y + R\Sigma^{-1}(\hat{\psi})X\hat{\beta}(\hat{\psi}), \tag{2.10}$$

$$\text{MSE}[\hat{\theta}(\hat{\psi})] = g_1(\hat{\psi}) + g_2(\hat{\psi}), \tag{2.11}$$

where $\Sigma(\hat{\psi}) = \hat{\sigma}_v^2 A^{-1}(\hat{\psi}) + R$

and $A(\hat{\psi}) = (I - \hat{\rho}W)^T(I - \hat{\rho}W)$.

This expression for the MSE of the EBLUP severely underestimates the true MSE as the variability due to the estimation of the parameters through the data has been ignored. We obtain a second order approximation to the $\text{MSE}[\hat{\theta}(\hat{\psi})]$ in case $\hat{\psi}$ is the maximum likelihood estimator (MLE) or the restricted maximum likelihood estimator (REMLE) of ψ, with the assumption of large m and by neglecting all the terms of the order $o(m^{-1})$, under the following regularity conditions. The approximation has been worked out along the lines of Prasad and Rao (1990) and Datta and Lahiri (2000) which are heuristic in nature.

Regularity Conditions 1

(a) The elements of X are uniformly bounded such that $X^T\Sigma^{-1}(\psi)X = [O(m)]_{p\;p}$, where $\Sigma(\psi) = [\sigma_v^2 A^{-1}(\psi) + R]$;

(b) m is finite;

(c) $\Lambda(\psi)X = [O(1)]_{m\;p}$, $(\partial[\Lambda(\psi)X])/(\partial\psi_d) = [O(1)]_{m\;p}$, $(\partial^2[\Lambda(\psi)])/(\partial\psi_d\partial\psi_e) = [O(1)]_{m\;m}$ for $d, e = 1, 2$;

(d) $\hat{\psi}$ is the estimator of ψ which satisfies $\hat{\psi} - \psi = O_p(m^{-1/2}), \hat{\psi}(-y) = \hat{\psi}(y), \hat{\psi}(y + xh) = \hat{\psi}(y)\ \forall h \in R^p$ and $\forall y$.

These regularity conditions are satisfied in this case. The special standardised form of the weight matrix W satisfies the condition (c) for $|\rho| < 1$ as it has only a finite number of nonzero elements and its row sum is equal to 1. It may be mentioned here that the matrix $\sigma_v^2 A^{-1}\Sigma^{-1}$ has finite number of nonzero elements and the order of $W, (I - \rho W), W(I - \rho W), \Sigma, \Sigma^{-1}$ or any sum or product combination of these and their derivatives mentioned in condition (c) do not increase. The MLE and the REMLE, in addition satisfy the condition (d). A second order approximation to the MSE of the EBLUP has been shown in Theorem A.1 of the Appendix as

$$
\begin{aligned}
\text{MSE}[\hat{\theta}(\hat{\psi})] &= E[(\hat{\theta}(\hat{\psi}) - \theta)(\hat{\theta}(\hat{\psi}) - \theta)^T] \\
&= g_1(\psi) + g_2(\psi) + g_3(\psi) + o(m^{-1}).
\end{aligned} \tag{2.12}
$$

Here the third term $g_3(\psi)$ comes from estimating the unknown parameter vector from the sample data and it is of the same order $O(m^{-1})$ as that of $g_2(\psi)$. Further $g_3(\psi)$ may be expressed as

$$g_3(\psi) = L^T(\psi)[I_\psi^{-1}(\psi) \otimes \Sigma(\psi)]L(\psi), \tag{2.13}$$

where

$$L(\psi) = \underset{1 \le d \le 2}{\text{Col}}[L_d(\psi)] = [L_\rho(\psi), L_{\sigma_v^2}(\psi)]^T,$$

$$L_d(\psi) = \frac{\partial\Lambda(\psi)}{\partial\psi_d}, d = 1, 2. \quad I_\psi(\psi) = E[-\frac{\partial^2 l}{\partial\psi\partial\psi^T}]$$

is the information matrix and $\otimes$ represents Kronecker product. Further $g_3(\psi)$ may also be written as

$$g_3(\psi) = \sum_{d=1}^{2}\sum_{e=1}^{2} L_d(\psi)\Sigma(\psi)L_e^T(\psi)I_{de}^{-1}(\psi) \tag{2.14}$$

where $I_\psi^{-1}(\psi) \equiv (I_{de}^{-1}(\psi))$.

It is common practice to estimate the MSE of the EBLUP by replacing the unknown parameters including components of the variance by their respective estimators. This procedure can lead to severe underestimation of the true MSE (Prasad and Rao 1990, Singh, Stukel and Pfeffermann 1998). We obtain the estimator of the MSE of the EBLUP in Theorem A.2 of the Appendix for large m neglecting all terms of order $o(m^{-1})$. As a result we have the expressions

$$E[g_1(\hat{\psi}) + g_3(\hat{\psi}) - g_4(\hat{\psi}) - g_5(\hat{\psi})] = g_1(\psi) + o(m^{-1}), \tag{2.15}$$

$$E[g_2(\hat{\psi})] = g_2(\psi) + o(m^{-1})$$

$$\text{and} \quad E[g_3(\hat{\psi})] = g_3(\psi) + o(m^{-1}), \tag{2.16}$$

and finally the estimator of the MSE of $\hat{\theta}(\hat{\psi})$ as

$$
\begin{aligned}
&\text{mse}[\hat{\theta}(\hat{\psi})] = \\
&[g_1(\hat{\psi}) + g_2(\hat{\psi}) + 2g_3(\hat{\psi}) - g_4(\hat{\psi}) - g_5(\hat{\psi})] + o(m^{-1}),
\end{aligned} \tag{2.17}
$$

where $E[\text{mse}(\hat{\theta}(\hat{\psi}))] = \text{MSE}[\hat{\theta}(\hat{\psi})] + o(m^{-1})$.

Obviously the additional terms, $g_3(\hat{\psi}), g_4(\hat{\psi})$ and $g_5(\hat{\psi})$ are the contributions, due to estimation of unknown parameter vector ψ by $\hat{\psi}$. The expressions for $g_4(\psi)$ and $g_5(\psi)$ up to order $o(m^{-1})$ are given by

$$g_4(\psi) = [b^T_{\hat{\psi}}(\psi) \otimes I_m] \frac{\partial g_1(\psi)}{\partial \psi},$$

$$b_{\hat{\psi}}(\psi) = \frac{1}{2} I^{-1}_{\hat{\psi}}(\psi) \underset{1 \le d \le 2}{\mathrm{Col}} \left[\mathrm{Trace} \left[I^{-1}_{\beta}(\psi) \frac{\partial I_{\beta}(\psi)}{\partial \psi_d} \right] \right], \quad (2.18)$$

$$g_5(\psi) = \frac{1}{2} \underset{m}{\mathrm{Trace}} \begin{bmatrix} [I_2 \otimes (R\Sigma^{-1}(\psi))] \\ \dfrac{\partial^2 \Sigma(\psi)}{\partial \psi \partial \psi^T} [I^{-1}_{\psi}(\psi) \otimes (\Sigma^{-1}(\psi) R)] \end{bmatrix}. \quad (2.19)$$

Here $b_{\hat{\psi}}(\psi)$ is the bias of $\hat{\psi}$ *i.e.*, $E(\hat{\psi}) - \psi$ up to order $o(m^{-1})$ and $(\partial g_1(\psi))/(\partial \psi)$ is a partitioned matrix $[(\partial \mathrm{g}_1(\psi))/(\partial \rho), (\partial g_1(\psi))/(\partial \sigma^2_v)]^T$ of order $(2m \quad m)$ having 2 matrices of order $m \quad m$ in a column. In the same way $(\partial^2 \Sigma(\psi))/(\partial \psi \partial \psi^T)$ is a partitioned matrix of order $(2m \quad 2m)$ having 2 partitions, row and column wise with $(\partial^2 \Sigma(\psi))/(\partial \psi_d \partial \psi_e)$ being a general sub matrix of order $m \quad m$ therein. $\mathrm{Trace}(B) = \sum^2_{d=1} B_{dd}$, where B is a square partitioned matrix with square sub matrices of similar order. In addition $g_4(\psi)$ and $g_5(\psi)$ may also be written as

$$g_4(\psi) = \frac{1}{2} \sum^2_{d=1} \sum^2_{e=1} I^{-1}_{de}(\psi) \mathrm{Trace} \left[I^{-1}_{\beta}(\psi) \frac{\partial I_{\beta}(\psi)}{\partial \psi_d} \right] \frac{\partial g_1(\psi)}{\partial \psi_e}, \quad (2.20)$$

$$g_5(\psi) = \frac{1}{2} \sum^2_{d=1} \sum^2_{e=1} \left[R\Sigma^{-1}(\psi) \frac{\partial^2 \Sigma(\psi)}{\partial \psi_d \partial \psi_e} \Sigma^{-1}(\psi) R I^{-1}_{de}(\psi) \right]. \quad (2.21)$$

The expression (2.17) gives the matrix of the estimator of the MSE of EBLUP, $\hat{\theta}(\hat{\psi})$ and the MSE of the individual small area estimators may be obtained as the respective diagonal element. In case of simple model without the spatial autocorrelation, similar expressions can be obtained. In this case $g_5(\psi)$, however, becomes zero.

3. Spatial Temporal Model

In this section, State Space Models via Kalman filtering have been used to take the advantage of the time series data along with the common regression parameter and common autocorrelation parameter to strengthen the direct survey estimators at any point of time. This is especially advantageous in the case where the past survey estimates are more reliable. The models used in this category are the following

$$y_t = X_t \beta + Z v_t + \varepsilon_t, \ \varepsilon_t \overset{\text{ind}}{\sim} N_m(0, R_t), \ Z = (I - \rho W)^{-1}, \quad (3.1)$$

$$v_t = k v_{t-1} + \eta_t, \ \eta_t \overset{\text{ind}}{\sim} N_m(0, \sigma^2_v I) \ \ t = 1, 2, \ldots, T \ \text{ and}$$
ε_t and η_t are independent of each other. (3.2)

Here the parameters have usual meaning as explained in the previous section. Weight matrix $W(m \quad m)$ and design matrices $X_t(m \quad p)$ are known, $Z(m \quad m)$ is a matrix of coefficients of random effects and ρ is an unknown autocorrelation coefficient. R_t is a diagonal matrix of order m which may be expressed as $R_t = \mathrm{diag}(\sigma^2_{1t}, \sigma^2_{2t}, \ldots, \sigma^2_{mt})$ where σ^2_{it} 's are known sampling variances corresponding to the i^{th} small area and t^{th} time point. β is unknown vector of fixed effects and $\psi = [\rho, \sigma^2_v, k]^T$ is a vector of three unknown parameters. These parameters are independent of time *t*. It may be noted that the random effects v_t have been allowed to change in accordance with (3.2) and k is temporal autoregressive parameter. For stationarity $|k| < 1$.

The estimators of fixed and random effects and the MSE of these estimators are obtained in stages, starting with assumption of mixed effects linear model approach at time $t = 1$, and by taking $v_1 \sim N_m(0, \sigma^2_v I)$ (Sallas and Harville 1994). In the standard form we write the model as

$$y_t = U_t \alpha_t + \varepsilon_t, \ \alpha_t = T\alpha_{t-1} + \zeta_t, \ T = \mathrm{diag}[I_p, kI_m], \quad (3.3)$$

$$\zeta_t \sim N_{p+m}(0, Q), \quad Q = \mathrm{diag}[0_p, \sigma^2_v I_m]$$
$$U_t = [X_t, Z], \ \alpha_t = [\beta_t, v_t]^T. \quad (3.4)$$

Here I_m and 0_m are the unit and zero matrices of order m and by $\mathrm{diag}[I_p, kI_m]$ we mean the matrix

$$\begin{bmatrix} I_{p \quad p} & 0_{p \quad m} \\ 0_{m \quad p} & kI_{m \quad m} \end{bmatrix}.$$

In case β is assumed fixed but dependent on time, there is no change in the model except that $T = \mathrm{diag}[0_p, kI_m]$.

The initial estimates of the effects α_t and their variances (based on $t = 1$) are obtained as

$$\hat{\beta}_1 = (X^T_1 H^{-1}_1 X_1)^{-1} X^T_1 H^{-1}_1 y_1, \ \hat{v}_1 = \sigma^2_v Z^T H^{-1}_1 (y_1 - X_1 \hat{\beta}_1),$$

$$H_1 = R_1 \sigma^2_v A^{-1}, \quad \Sigma_1 = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix},$$

$$\Sigma_{11}(p \quad p) = (X^T_1 H^{-1}_1 X_1)^{-1},$$
$$\Sigma_{12}(p \quad m) = \Sigma^T_{21} = -\sigma^2_v (X^T_1 H^{-1}_1 X_1)^{-1} X^T_1 H^{-1}_1 Z$$
$$\text{and } \ \Sigma_{22}(m \quad m) = \sigma^2_v I_m - \sigma^4_v Z^T H^{-1}_1 Z$$
$$+ \sigma^4_v Z^T H^{-1}_1 X_1 (X^T_1 H^{-1}_1 X_1)^{-1} X^T_1 H^{-1}_1 Z.$$

The recurring Kalman filtering equations for updation of the estimators at subsequent stages are

$$\Sigma_{t|t-1} = T\Sigma_{t-1}T^T + Q, \hat{\alpha}_{t|t-1} = T\hat{\alpha}_{t-1}, \ H_t = R_t + U_t\Sigma_{t|t-1}U_t^T,$$
$$\hat{\alpha}_t = \hat{\alpha}_{t|t-1} + \Sigma_{t|t-1}U_t^T H_t^{-1}(y_t - U_t\hat{\alpha}_{t|t-1}),$$
$$\Sigma_t = \Sigma_{t|t-1} - \Sigma_{t|t-1}U_t^T H_t^{-1}U_t\Sigma_{t|t-1}$$

where $\hat{\alpha}_{t|t-1}$ are the estimators of the effects α_t given the observations $[y_1, y_2, \ldots, y_{t-1}]$ and the $\Sigma_{t|t-1}$ are the mean squared errors of $\hat{\alpha}_{t|t-1}$. H_t are the conditional variance covariance matrix of y_t given $[y_1, y_2, \ldots, y_{t-1}]$. With the help of the above recurring filtering equations, the best linear unbiased predictor (BLUP) of $\theta_t = X_t\beta + Zv_t$, and the mean squared error (MSE) of the BLUP may be obtained as

$$\begin{aligned}\hat{\theta}_t(\psi) &= U_t(\psi)\hat{\alpha}_t(\psi)\\ &= y_t - R_t H_t^{-1}(\psi)[y_t - U_t(\psi)\hat{\alpha}_{t|t-1}(\psi)]\\ &= U_t(\psi)\hat{\alpha}_{t|t-1}(\psi) + \Lambda_t(\psi)e_t(\psi),\end{aligned} \tag{3.5}$$

$$\text{MSE}[\hat{\theta}_t(\psi)] = g_{12t}(\psi) = U_t(\psi)\Sigma_t(\psi)U_t^T(\psi), \tag{3.6}$$

$$\begin{aligned}\text{where}\quad \Lambda_t(\psi) &= U_t(\psi)\Sigma_{t|t-1}(\psi)U_t^T(\psi)H_t^{-1}(\psi)\\ &= I_m - R_t H_t^{-1}(\psi)\\ \text{and}\quad e_t(\psi) &= y_t - U_t(\psi)\hat{\alpha}_{t|t-1}(\psi).\end{aligned}$$

It may be noted that $g_{12t}(\psi)$ is the spatial counterpart of $g_1(\psi) + g_2(\psi)$. As usual in practice, the parameter vector ψ is unknown and its restricted maximum likelihood estimators (REMLE) can be obtained by maximizing the following log likelihood function, based on the sample data covering all time points

$$\begin{aligned} l = \text{const.} &- \frac{1}{2}\log[|X_1^T H_1^{-1}X_1|] - \frac{1}{2}\sum_{t=1}^{T}\log[|H_t|]\\ &- \frac{1}{2}(y_1 - X_1\hat{\beta}_1)^T H_1^{-1}(y_1 - X_1\hat{\beta}_1)\\ &- \frac{1}{2}\sum_{t=2}^{T}(y_t - U_t\hat{\alpha}_{t|t-1})^T H_t^{-1}(y_t - U_t\hat{\alpha}_{t|t-1})\end{aligned} \tag{3.7}$$

with respect to the parameter ψ. With the help of the above, the estimator, $\hat{\psi}$ is obtained and the EBLUP of θ_t and the naive estimator of the MSE of the EBLUP are given by

$$\hat{\theta}_t(\hat{\psi}) = U_t(\hat{\psi})\hat{\alpha}_t(\hat{\psi}) = U_t(\hat{\psi})\hat{\alpha}_{t|t-1}(\hat{\psi}) + \Lambda_t(\hat{\psi})e_t(\hat{\psi}), \tag{3.8}$$

$$\text{MSE}[\hat{\theta}_t(\hat{\psi})] = g_{12t}(\hat{\psi}) = U_t(\hat{\psi})\Sigma_t(\hat{\psi})U_t^T(\hat{\psi}). \tag{3.9}$$

As explained earlier in section 2, the MSE of the EBLUP underestimates the true MSE as it does not take care of the variability due to replacing parameters by their estimates. A second order approximation to the $\text{MSE}[\hat{\theta}_t(\hat{\psi})]$ for large m and neglecting all the terms of order $o(m^{-1})$, has been obtained in Theorem A.3 of the Appendix, under the following regularity conditions satisfied by our model. These conditions are analogous to the regularity conditions 1.

Regularity Conditions 2

(a) The elements of $X_t, t = 1, 2, \ldots, T$ are uniformly bounded such that $X_t^T\Sigma_t^{-1}(\psi)X_t = [O(m)]_{p \ p}$, where $\Sigma_t(\psi) = [\sigma_v^2 A^{-1}(\psi) + R_t]$;

(b) m and T are finite;

(c) $\Lambda_t(\psi)U_t(\psi) = [O(1)]_{m \ p}$, $(\partial[\Lambda_t(\psi)U_t(\psi)])/(\partial\psi_d) = [O(1)]_{m \ p}$, $([\partial^2\Lambda_t(\psi)])/(\partial\psi_d\partial\psi_e) = [O(1)]_{m \ m}$, $t = 1, 2, \ldots, T$ and $d, e = 1, 2, 3$;

(d) $\hat{\psi}$ is the estimator of ψ which satisfies $\hat{\psi} - \psi = O_p(m^{-1/2})$, $\hat{\psi}(-y) = \hat{\psi}(y)$, $\hat{\psi}(y + xh) = \hat{\psi}(y)\ \forall h \in R^p$ and $\forall$ y.

The second order approximation to the MSE of the EBLUP is

$$\begin{aligned}\text{MSE}[\hat{\theta}_t(\hat{\psi})] &= E[(\hat{\theta}_t(\hat{\psi}) - \theta_t)(\hat{\theta}_t(\hat{\psi}) - \theta_t)^T]\\ &= g_{12t}(\psi) + g_{3t}(\psi) + o(m^{-1}).\end{aligned} \tag{3.10}$$

Here $g_{3t}(\psi)$ is the bias due to the estimation of the parameters from the sample data and is of the order $O(m^{-1})$ and it is given by

$$g_{3t}(\psi) = L_t^T(\psi)I_\psi^{-1}(\psi)K_\psi(\psi)H_t I_\psi^{-1}(\psi)L_t(\psi) \tag{3.11}$$

where $K_\psi(\psi) \equiv (K_{de}(\psi))$

$$\text{and}\quad K_{de}(\psi) = \frac{1}{2}\sum_{i=1}^{T}\text{Trace}\left[H_i^{-1}\frac{\partial H_i}{\partial\psi_d}H_i^{-1}\frac{\partial H_i}{\partial\psi_e}\right]. \tag{3.12}$$

Further

$L_t(\psi) = \underset{1 \le d \le 3}{\text{Col}}[L_{td}(\psi)]$ and $L_{td}(\psi) = (\partial\Lambda_t(\psi))/(\partial\psi_d)$ for $d = 1, 2, 3$.

In a proper form, we may write $g_{3t}(\psi)$ as

$$g_{3t}(\psi) =$$
$$\sum_{d=1}^{3}\sum_{e=1}^{3}L_{td}(\psi)\left[\begin{array}{l}\displaystyle\sum_{f=1}^{3}\sum_{g=1}^{3}I_{df}^{-1}(\psi)\\ \qquad\displaystyle\sum_{i=1}^{T}\text{Trace}\left(H_i^{-1}\frac{\partial H_i}{\partial\psi_f}H_i^{-1}\frac{\partial H_i}{\partial\psi_g}\right)\\ \qquad H_t I_{ge}^{-1}(\psi)\end{array}\right]L_{te}^T(\psi).$$

The expression for the information matrix involved here, may be given as

$$I_{de}(\psi) = E\left[-\frac{\partial^2 l}{\partial\psi_d \partial\psi_e}\right]$$

$$= \frac{1}{2}\sum_{t=1}^{T} \text{Trace}\left[H_t^{-1}\frac{\partial H_t^{-1}}{\partial\psi_d} H_t^{-1}\frac{\partial H_t}{\partial\psi_e}\right] + \sum_{t=1}^{T}\left[\frac{\partial e_t^T}{\partial\psi_d} H_t^{-1}\frac{\partial e_t}{\partial\psi_e}\right]$$

$$-\frac{1}{2}\text{Trace}\left[\begin{array}{l}(X_1^T H_1^{-1} X_1)^{-1} X_1^T H_1^{-1} \\ \quad\left(\frac{\partial^2 H_1}{\partial\psi_d\partial\psi_e} - 2\frac{\partial H_1}{\partial\psi_d} H_1^{-1}\frac{\partial H_1}{\partial\psi_e}\right) H_1^{-1} X_1\end{array}\right]$$

$$-\frac{1}{2}\text{Trace}\left[\begin{array}{l}(X_1^T H_1^{-1} X_1)^{-1} X_1^T H_1^{-1}\frac{\partial H_1}{\partial\psi_d} H_1^{-1} X_1 \\ \quad (X_1^T H_1^{-1} X_1)^{-1} X_1^T H_1^{-1}\frac{\partial H_1}{\partial\psi_e} H_1^{-1} X_1)\end{array}\right].$$

Estimator of the MSE of the EBLUP has also been obtained with the assumption of large m and neglecting all terms of order $o(m^{-1})$ in Theorem A.4 of the Appendix as

$$\text{mse}[\hat{\theta}_t(\hat{\psi})] = [g_{12t}(\hat{\psi}) + g_{3t}(\hat{\psi}) + g_{31t}(\hat{\psi}) - g_{4t}(\hat{\psi}) - g_{5t}(\hat{\psi})] + o(m^{-1}), \tag{3.13}$$

where $g_{31t}(\psi)$, $g_{4t}(\psi)$ and $g_{5t}(\psi)$ are given as

$$g_{31t}(\psi) = L_t^T(\psi)[I_\psi^{-1}(\hat{\psi}) \otimes H_t(\psi)] L_t(\psi), \tag{3.14}$$

$$g_{4t}(\psi) = [b_{\hat{\psi}}^T(\psi) \otimes I_m]\frac{\partial g_{12t}(\psi)}{\partial(\psi)},$$

$$b_{\hat{\psi}} = \frac{1}{2} I_\psi^{-1}(\psi) \underset{1\le d\le 3}{\text{Col}}\left[\text{Trace}\left[I_\beta^{-1}(\psi)\frac{\partial I_\beta(\psi)}{\partial\psi_d}\right]\right], \tag{3.15}$$

$$g_{5t}(\psi) = \frac{1}{2}\underset{m}{\text{Trace}}\left[\begin{array}{l}[I_3 \otimes (R_t H_t^{-1})] \\ \frac{\partial^2 H_t}{\partial\psi\partial\psi^T}[I_\psi^{-1}(\psi) \otimes (H_t^{-1} R_t)]\end{array}\right]. \tag{3.16}$$

4. Analysis of the NSSO Data

National Sample Survey Organisation (NSSO) of the Ministry of Statistics and Programme Implementation (Government of India) conducts quinquennial large sample surveys (QS) on household consumption expenditure and employment, almost every five years in India. The surveys cover more than hundred thousand households spread over a number of villages and urban blocks. In order to fill the gaps in data between the successive QSs, the NSSO conducts annual consumer expenditure survey (CES) in almost every round (equivalent to six months or one year duration). The annual series covers only 10–30 thousand households depending on the number of villages and urban blocks surveyed all over the country. Each round of NSS normally has more than one subject of enquiry. The annual series has a different principal subject of enquiry. However schedule 1.0 of the annual surveys is designed to collect data on household consumption expenditure among other characteristics on employment.

The NSSO adopts two stage stratified sampling design, the first stage units being census villages in the rural sector selected through circular systematic sampling with probability proportional to size (PPS) and the ultimate-stage units being the households selected circular systematically with independent random starts. India has been divided into States and the Districts are the second level administrative units in the States. There is not much difference between the annual and quinquennial surveys excepting that normally in annual series, a small sample of four households per first stage units are surveyed while in the case of quinquennial survey, ten to twelve households per first stage units are surveyed. Besides this, in NSSO surveys, we have two samples *viz*, the first one as central sample surveyed by the investigators of the NSSO, and the second one as state sample surveyed by the State authorities. Regarding the estimation procedure, the first stage units are selected in the form of two independent sub-samples. The estimate of the population mean and its variance based on the two sub-samples are separately obtained. The pooled mean $y_i = (\hat{y}_{1i} + \hat{y}_{2i})/2$ and $R_i = (\hat{y}_{1i} - \hat{y}_{2i})^2/4$ for $i = 1, 2, \ldots, m$, where $\hat{y}_{1i}, \hat{y}_{2i}$ are the sub-sample means, estimate respectively the population mean and its variance for a particular district (small area). In case of round 55, first stage units are selected in the form of eight independent sub-samples and the estimate of the population mean and its variance are based on these sub-samples. In view of the problems related to the estimates of R_i's with 1 d.f., the R_i for each small area were analysed and compared over time. In case of any abnormal R_i, it was smoothed out by taking the average of R_i's over neighboring time points and in some cases, over neighboring small areas also. The survey estimates y_i's are the direct estimates, and the smoothed R_i's are the diagonal elements of the sampling variance covariance matrix R, in our model equations (2.1), (2.4) and (3.1), referred in this paper.

In this paper, we have used data from central sample only. The estimates of monthly per capita consumption expenditure (MPCE) and of the standard errors(SE) of the estimators have been obtained under various mixed effects models for the rural 63 districts (small areas) of a large state in India, namely, Uttar Pradesh. We have used data from the six rounds of the NSSO *viz* round 50 (July 1993–June 1994), round 51 (July 1994–June 1995), round 52 (July 1995–June 1996), round 53 (January–December 1997), round 54 (January–June 1998) and round 55 (July 1999–June 2000). Out of these rounds 50 and 55 are based on quinquennial surveys. The selected exogenous variables used in the models are i) number of households, ii) gross area sown and iii) per capita net area sown in the districts. The agricultural data are available on annual basis while the estimates of the households and the population were obtained through the interpolation techniques based on the 1971, 1981 and 1991 decennial census data. These exogenous variables have been selected from a host of variables ranging from 1991 census to annual agricultural data through the covariate analysis. Different weight matrices such as length of common boundary between a pair of districts, distance between centres of two districts and the binary weights were considered. Binary weights give larger estimate of spatial autocorrelation coefficient, therefore they (standardised by making row sum of the weight matrix as one) have been used for further analysis in this paper. In the whole exercise, maximization of log likelihood function and the estimation of the parameters have been carried out by using the Nelder and Mead simplex method on the software MATLAB.

Various mixed effects models, used for finding out improved estimates of MPCE are given in Table 1. The parameters in the models have usual meaning as shown in sections 2 and 3. Further, in case of each model, sampling variance R or R_t (in case of temporal model) are assumed to be known.

Table 1
Mixed Effects Models

Model−1	Direct Estimates	
Model−2	Regression Model	$y = X\beta + v + \varepsilon$
Model−3	Spatial Model	$y = X\beta + Zv + \varepsilon$
Model−3A	Spatial Model (intercept)	$y = \quad + Zv + \varepsilon$
Model−4	Regression Temporal	$y_t = X_t\beta + v_t + \varepsilon_t, v_t = kv_{t-1} + \eta_t$
Model−5	Spatial Temporal	$y_t = X_t\beta + Zv_t + \varepsilon_t, v_t = kv_{t-1} + \eta_t$

Table 2 presents the round wise estimates of the parameters for the simple mixed effects regression and spatial models. The value of the multiple correlation coefficients R^2 between MPCE estimates and the auxiliary variables, in case of each round has also been shown here. The figures in bracket show the Standard Errors (SE) of the parameter estimates. Note that $\lambda(=\lambda_1, \lambda_2)$ is the likelihood ratio test (LRT) statistics defined as $-2\log L \sim \chi^2_k$, where L is the ratio of nested likelihoods at the hypothesised parameter values for two competing models under different hypotheses and k is the difference between the number of parameters under two models. Here λ_1 compares regression model and spatial model, under $H_0: \rho = 0$ against $H_1: \rho \neq 0$ and is distributed as χ^2_1 under H_0, and λ_2 compares spatial model and spatial (intercept) model, under $H_0: \beta = 0$ against $H_1: \beta \neq 0$ [β does not include intercept term β_0] and is distributed as χ^2_3 under H_0.

On comparison of the simple regression model (Model 2) and spatial model (Model 3) through LRT, we find that under $H_0(\rho = 0)$, the spatial autocorrelation ρ for Model 3 has been found highly significant for the two rounds 52 and 55, obviously for these rounds, use of spatial model results in much improvement in the estimates of MPCE. On the other hand, in case of rounds 50 and 53, and for these only, the regression coefficients β have been found nearly significant for the Model 3 in comparison to Model 3A which shows that the spatial model with intercept term may improve the estimates for these rounds without any help of the exogenous variables.

Table 3 presents the parameter estimates and their SE in case of regression temporal model and spatial temporal model.

For Model 4, unconstrained iterative maximisation process converged the value of k greater than 1, which is inadmissible under the assumption of stationarity. For this

Table 2
Estimates of Parameters for Small Area Estimates of MPCE Under Regression and Spatial Models

Round	R^2	Model 2	Model 3		LRT	Model 3A		LRT
		σ^2_v	ρ	σ^2_v	λ_1	ρ	σ^2_v	λ_2
Rd. 50	0.27	1,724.48	0.30	1,635.70	1.80	0.59	1,724.68	6.64
		(356.19)	(0.18)	(346.45)		(0.13)	(378.66)	
Rd. 51	0.27	3,424.21	0.48	3,156.90	0.66	0.67	3,022.32	4.54
		(820.89)	(0.19)	(815.24)		(0.13)	(824.54)	
Rd. 52	0.17	2,150.54	0.87	714.96	13.46	0.86	768.11	0.90
		(540.23)	(0.07)	(257.15)		(0.07)	(272.27)	
Rd. 53	0.13	6,312.99	−0.39	5,822.99	1.56	0.09	7,141.60	7.66
		(1,397.92)	(0.27)	(1,374.70)		(0.23)	(1,561.72)	
Rd. 54	0.22	3,437.67	0.61	2,793.24	1.30	0.66	2,888.66	3.00
		(806.87)	(0.14)	(742.35)		(0.13)	(768.84)	
Rd. 55	0.31	2,989.73	0.87	1,060.21	20.30	0.86	1,186.58	1.56
		(712.28)	(0.06)	(362.40)		(0.07)	(394.27)	

λ_1 and λ_2 compare models 2,3 and models 3,3A respectively. $\chi^2_{1,.05}$=3.841 for λ_1 and $\chi^2_{3,.05} = 7.815$ for λ_2.

case, estimates were obtained by taking $k = 1$ and Model 4 was accordingly modified. Table 3 reports the results for $k = 1$ in case of regression temporal model. The spatial temporal model shows higher value of common auto-correlation coefficient and far lower value of the estimate of σ_v^2. A summary of the round wise average estimates of MPCE (based on all the 63 districts), their estimated standard errors (SE) and the coefficient of variation (CV) under each model has been presented in Table 4.

The results of Table 4 have been summarized below.

The Direct survey estimates are less precise and all the models involving mixed effects improve it. The estimates for the rounds 50 and 55 (based on large samples) are more precise than the estimates based on other rounds. Spatial model, depending on the value of ρ improves the estimates considerably. In case of rounds 52 and 55, where the autocorrelation have been found significant, the reduction in the average SE of the estimates in comparison to the model without spatial autocorrelation, is considerable. Model 3A with spatial effect and without auxiliary variables is equally good. The spatial temporal model further improves the estimates taking into advantage of the state space considerations. It may be noted that for the round 52 (very high spatial autocorrelation), the estimates based on temporal models are worse than the estimates based on models without temporal considerations. Perhaps due to fixed regression and autocorrelation parameters, the estimates tend towards the average of the five rounds.

In order to judge the performances of the estimators under various models vis-a-vis under the most general model (spatial temporal model), data have been simulated under the spatial temporal model and true MSEs of the replicated estimates under each of the assumed models have been obtained. For this, we have conducted the simulation by taking the estimated parameters from the spatial temporal model, given in Table 2 and obtained the true replicated small area mean $\theta(b)$ for b^{th} replication $(b = 1, 2, \ldots, B)$ along with simulated observations $y(b)$ for a large number of replications. On this simulated dataset, for each replication, different models including spatial temporal model

Table 3
Estimates of Parameters for Small Area Estimation of MPCE Under Regression Temporal and Spatial Temporal Models

	ρ		σ_v^2		k	
Models	Estimate	S.E.	Estimate	S.E.	Estimate	S.E.
Model 4	–	–	4,715.64	431.00	–	–
Model 5	0.79	0.04	2,163.50	245.50	0.53	0.07

Table 4
Average EBLUP for MPCE (Rs.), their Estimated SE and CV Under Regression, Spatial, Regression Temporal and Spatial Temporal Models

Models	NSSO Rounds 50	51	52	53	54	55
	Average Small Area Estimates					
Model 1	276.10	321.26	373.07	408.52	411.25	482.00
Model 2	272.87	312.53	354.45	397.52	400.87	471.99
Model 3	272.98	313.14	351.51	398.21	400.78	471.09
Model 3A	273.56	314.19	352.01	396.40	399.91	471.91
Model 4	274.13	305.62	345.54	383.53	399.56	463.32
Model 5	273.75	312.21	351.79	391.61	399.50	473.57
	Average Standard Errors (SE)					
Model 1	25.09	66.06	64.18	74.19	53.87	45.45
Model 2	17.10	33.65	29.09	39.85	32.68	30.59
Model 3	16.88	32.84	21.51	39.98	30.87	24.84
Model 3A	16.56	31.29	20.79	40.03	30.23	24.37
Model 4	19.51	34.91	35.19	37.79	35.14	33.15
Model 5	17.18	28.99	28.33	30.02	28.76	28.10
	Average Coefficient of Variation (CV) (%)					
Model 1	9.09	20.56	17.20	18.16	13.10	9.43
Model 2	6.27	10.79	8.21	10.01	8.15	6.48
Model 3	6.18	10.49	6.12	10.04	7.70	5.27
Model 3A	6.05	9.96	5.91	10.10	7.56	5.17
Model 4	7.12	11.42	10.18	9.85	8.79	7.15
Model 5	6.28	9.29	8.05	7.67	7.20	5.93

Table 5
Percentage Relative Efficiency [RMSE] of the Temporal Models in Comparison to other Models for MPCE

	NSSO Rounds					
	50	51	52	53	54	55
	Spatial Temporal Model [Model 5]					
Model 2	123.63	170.54	193.68	203.55	204.72	169.76
Model 3	100.24	133.82	149.70	165.46	165.85	154.23
Model 4	125.81	141.50	141.93	137.55	139.11	129.88
	Regression Temporal Model [Model 4]					
Model 2	100.71	134.50	156.35	165.30	163.13	152.56

have been applied and the small area mean estimators under each of them are obtained. While fitting the regression and spatial temporal models on the simulated datasets, the iterative maximisation process have the constrained value of $k <= 1$. Here we have taken $B = 5{,}000$ replications. The true MSEs of the estimators for i^{th} small area under a particular model $(k = 2 - 4)$ may be defined as

$$\text{MSE}(\theta_i^k) = \frac{1}{B}\sum_{k=1}^{B}[\hat{\theta}_i^k(b) - \theta_i(b)]^2, \quad i = 1, 2, \ldots, m.$$

The relative efficiency of the estimators under spatial temporal model (Model 5) against the estimators under models 2–4 have been judged by the ratio of their mean squared errors (RMSE) as

$$\text{RMSE}(k, \text{Temp}) = 100\frac{\sum_{i=1}^{m}\text{MSE}(\hat{\theta}_i^k)}{\sum_{i=1}^{m}\text{MSE}(\hat{\theta}_i^{\text{Temp}})}$$

where 'Temp' denotes the spatial temporal model and k denotes models 2, 3 and 4. Likewise the relative efficiency of the regression temporal model (Model 4) against the simple regression model (Model 2) has been found by simulating data with the estimated parameters given in Table 3, under the regression temporal model. The results have been shown in Table 5.

The results confirm the superiority of the spatial temporal model in comparison to other models for these parameters. The regression temporal model has also been found better than the simple regression model.

5. Conclusions

The Direct survey estimates based on the small sample can be considerably improved by using the area specific small area models. The spatial autocorrelation amongst the neighboring areas may be exploited for improving the direct survey estimates. However, the model must be applied after studying the significant correlation amongst the small areas by virtue of their neighborhood effects. In case of poor relation between the dependent and exogenous variables, the simple spatial model with intercept only, may equally improve the estimates. This model uses only the spatial autocorrelation to strengthen the small area estimates and do not require the use of exogenous variables. The spatial models, by using the appropriate weight matrix *W*, or a combination of *W* matrices, can considerably improve the estimates. Weight matrix should be based on logical considerations and it may be used effectively for the cases, where due to some reasons, reliable exogenous variables are not available. This aspect can be further exploited to find out the small area estimates for the areas which have been recently created/demarcated.

One has to be careful about the increase in the MSE due to the variability caused by replacing the parameters by their estimates. This gets reflected through the second order approximation to the MSE dealt in the paper. That is why many times the simple spatial model (with intercept) performs better than the spatial model involving more parameters. Use of time series data with fixed regression parameters across the time, further improves the small area estimates especially for the time points where the direct survey estimates have larger MSE. Spatial temporal models have advantage over temporal models without spatial consideration due to the inclusion of fixed spatial auto-correlation across the small areas. However, for some time points for which ρ may be very different than the rest, this may not hold due to estimates tending towards the average of five rounds. Here the temporal consideration can be started from a suitable initial time point. Finally the exogenous variables *X* and the weight matrix *W* supplement each other through the regression parameter β and the autocorrelation parameter ρ and a judicious use of them may result in considerable improvement in the small area estimates.

Acknowledgements

The unit level data for the research have been made available by the National Sample Survey Organisation (NSSO), Ministry of Statistics and Programme Implementation under a research collaboration between IIT Kanpur and the NSSO. The weight matrix containing the length of the boundary between different small areas (districts) have been provided by the National Informatics Centre (NIC) of the Ministry of Information Technology, Government of India. We would like to thank the referees for their helpful comments which has considerably improved the paper.

Appendix

Theorem A.1: Under Regularity Conditions 1

$$\mathrm{MSE}[\hat{\theta}(\hat{\psi})] = g_1(\psi) + g_2(\psi) + g_3(\psi) + o(m^{-1}). \quad (5.1)$$

For proof of the Theorem, we use the following well known results (Srivastawa and Tiwari 1976). Let $U \sim N(0, \Sigma)$ then for the symmetric matrices A, B and C

$$E[U(U^T AU)U^T] = \mathrm{Trace}(A\Sigma)\Sigma + 2\Sigma A\Sigma$$

$$\begin{aligned}E[U(U^T AU)(U^T BU)U^T] &= \mathrm{Trace}(A\Sigma)\mathrm{Trace}(B\Sigma)\Sigma\\ &+ 2[\mathrm{Trace}(A\Sigma)\Sigma B\Sigma + \mathrm{Trace}(B\Sigma)\Sigma A\Sigma + \mathrm{Trace}(A\Sigma B\Sigma)\Sigma]\\ &+ 4[\Sigma A\Sigma B\Sigma + \Sigma B\Sigma A\Sigma].\end{aligned}$$

Proof of Theorem A.1

Kackar and Harville (1984) showed that $\mathrm{MSE}[\hat{\theta}(\hat{\psi})] = \mathrm{MSE}[\hat{\theta}(\psi)] + E[(\hat{\theta}(\hat{\psi}) - \hat{\theta}(\psi))(\hat{\theta}(\hat{\psi}) - \hat{\theta}(\psi))^T]$. It is straight forward to show that $\mathrm{MSE}[\hat{\theta}(\psi)] = g_1(\psi) + g_2(\psi)$. We need to prove that $g_3(\psi) = E[(\hat{\theta}(\hat{\psi}) - \hat{\theta}(\psi))(\hat{\theta}(\hat{\psi}) - \hat{\theta}(\psi))^T] + o(m^{-1})$. Taylor Series expansion of $\hat{\theta}(\hat{\psi})$ around ψ and using $(\hat{\psi} - \psi) = O_p(m^{-1/2})$ and $(\partial^2\hat{\theta}(\psi))/(\partial\psi_d\partial\psi_e)|_{\psi=\hat{\psi}^*} = O_p(1)$ when $\|\hat{\psi}^* - \hat{\psi}\| \leqq \|\hat{\psi} - \psi\|$ we get

$$[\hat{\theta}(\hat{\psi}) - \hat{\theta}(\psi)] = [(\hat{\psi} - \psi) \otimes I_m]^T \nabla\hat{\theta}(\psi) + O_p(m^{-1}). \quad (5.2)$$

Here $\nabla\hat{\theta}(\psi) = (\partial\hat{\theta}(\psi))/(\partial\psi) = [(\partial\hat{\theta}(\psi))/(\partial\rho), (\partial\hat{\theta}(\psi))/(\partial\sigma_v^2)]^T$. Using

$$\frac{\partial\hat{\theta}(\psi)}{\partial\psi_d} = \sum_{\alpha=1}^{p} \frac{\partial\hat{\theta}^*(\beta,\psi)}{\partial\beta_\alpha}\Big|_{\beta=\hat{\beta}(\psi)} \frac{\partial\hat{\beta}(\psi)}{\partial\psi_d} + \frac{\partial\hat{\theta}^*(\beta,\psi)}{\partial\psi_d}\Big|_{\beta=\hat{\beta}(\psi)}$$

$d = 1, 2$

where $\hat{\theta}^*(\beta,\psi) = X\beta(\psi) + \Lambda(\psi)[y - X\beta(\psi)]$, and the fact that $(\partial\hat{\beta}_\alpha(\psi))/(\partial\psi_d) = O_p(m^{-1/2})$ (Cox and Reid (1987)), we get from the above

$$[\hat{\theta}(\hat{\psi}) - \hat{\theta}(\psi)] = [(\hat{\psi} - \psi)^T \otimes I_m]\nabla\hat{\theta}^*(\psi) + O_p(m^{-1}) \quad (5.3)$$

where $$\nabla\hat{\theta}^*(\psi) = \left[\frac{\partial\hat{\theta}^*(\beta,\psi)}{\partial\rho}, \frac{\partial\hat{\theta}^*(\beta,\psi)}{\partial\sigma_v^2}\right]^T\Big|_{\beta=\hat{\beta}(\psi)} = \mathrm{L}(\psi)[y - X\hat{\beta}(\psi)].$$

Using the Regularity Conditions 1 and the fact that $\hat{\beta}(\psi) - \beta = O_p(m^{-1/2})$ we have

$$\begin{aligned}[\hat{\theta}(\hat{\psi}) - \hat{\theta}(\psi)] &= [(\hat{\psi} - \psi) \otimes I_m]^T L(\psi)[y - X\hat{\beta}(\psi)] + O_p(m^{-1})\\ &= \sum_{d=1}^{2} (\hat{\psi}_d - \psi_d)L_d(\psi)[y - X\hat{\beta}(\psi)] + O_p(m^{-1}).\end{aligned}$$

Further using the Taylor Series expansion of the Likelihood $S(\hat{\eta}) = 0$ around ψ where

$$S(\eta) = [S_\beta^T(\eta), S_\psi^T(\eta)]^T, S_\beta^T(\eta) = \underset{1\le\alpha\le p}{\mathrm{Col}}\left[\frac{\partial\ell}{\partial\beta_\alpha}\right]$$

and the orthogonality of β and ψ, it follow that

$$(\hat{\psi} - \psi) = I_\psi^{-1}(\psi)S_\psi(\eta) + O_p(m^{-1}).$$

Writing

$$S_\psi(\psi) = \underset{1\le d\le 2}{\mathrm{Col}}[S_d(\psi)] = [S_\rho(\psi), S_{\sigma_v^2}(\psi)]^T,$$

$$S_d(\psi) = \frac{\partial\ell}{\partial\psi_d} = -\frac{1}{2}\mathrm{Trace}\left[\Sigma^{-1}\frac{\partial\Sigma}{\partial\psi_d}\right] + \frac{1}{2}[u^T B_d(\psi)u],$$

$$B_d(\psi) = \Sigma^{-1}\frac{\partial\Sigma}{\partial\psi_d}\Sigma^{-1}, u = y - X\beta(\psi) \text{ and}$$

$$I_{de}(\psi) = \frac{1}{2}\mathrm{Trace}\left[\Sigma^{-1}\frac{\partial\Sigma}{\partial\psi_d}\Sigma^{-1}\frac{\partial\Sigma}{\partial\psi_e}\right]$$

we get

$$[\hat{\theta}(\hat{\psi}) - \hat{\theta}(\psi)] = L^T(\psi)[I_\psi^{-1}(\psi) \otimes I_m][\mathrm{S}_\psi(\psi) \otimes u]$$

and thus the expression

$[\hat{\theta}(\hat{\psi}) - \hat{\theta}(\psi)][\hat{\theta}(\hat{\psi}) - \hat{\theta}(\psi)]^T$ upto order $o(m^{-1})$

$$\begin{aligned}&= L^T(\psi)[I_\psi^{-1}(\psi) \otimes I_m]\underset{1\le d\le 2}{\mathrm{Col}}[uS_d(\psi)]\underset{1\le e\le 2}{\mathrm{Concat}}[S_e(\psi)u^T]\\ &\qquad [I_\psi^{-1}(\psi) \otimes I_m]L(\psi)\\ &= L^T(\psi)[I_\psi^{-1}(\psi)\otimes I_m]\underset{1\le d\le 2}{\mathrm{Col}}\underset{1\le e\le 2}{\mathrm{Concat}}[uS_d(\psi)S_e(\psi)u^T]\\ &\qquad [I_\psi^{-1}(\psi) \otimes I_m]L(\psi).\end{aligned} \quad (5.4)$$

Now we can write the likelihood and its derivative as

$$\ell = \log L = \mathrm{const.} - \frac{1}{2}\log[|\Sigma|] - \frac{1}{2}u^T\Sigma^{-1}u$$

$$\frac{\partial\ell}{\partial\psi_d} = -\frac{1}{2}\mathrm{Trace}\left[\Sigma^{-1}\frac{\partial\Sigma}{\partial\psi_d}\right] + \frac{1}{2}u^T B_d(\psi)u,$$

$$B_d(\psi) = \Sigma^{-1}\frac{\partial\Sigma}{\partial\psi_d}\Sigma^{-1}$$

$$E\left[-\frac{\partial^2\ell}{\partial\psi_d\partial\psi_e}\right] = \frac{1}{2}\mathrm{Trace}\left[\Sigma^{-1}\frac{\partial\Sigma}{\partial\psi_d}\Sigma^{-1}\frac{\partial\Sigma}{\partial\psi_e}\right] = I_{de}(\psi)$$

where information matrix $I_\psi(\psi) \equiv I_{de}(\psi)$.

The expectation of a typical element of the inner most terms in the expression (5.4) becomes

$$E[uS_d(\psi)S_e(\psi)u^T] = E\begin{bmatrix} u[u^T B_d(\psi)u][u^T B_e(\psi)u]u^T \\ -u\,\text{Trace}\left[\Sigma^{-1}\dfrac{\partial\Sigma}{\partial\psi_d}\right][u^T B_e(\psi)u]u^T \\ -u[u^T B_d(\psi)u]\text{Trace}\left[\Sigma^{-1}\dfrac{\partial\Sigma}{\partial\psi_e}\right]u^T \\ +u\,\text{Trace}\left[\Sigma^{-1}\dfrac{\partial\Sigma}{\partial\psi_d}\right]\text{Trace}\left[\Sigma^{-1}\dfrac{\partial\Sigma}{\partial\psi_e}\right]u^T \end{bmatrix}$$

and by applying the results of Srivastawa and Tiwari (1976), it becomes

$$E[uS_d(\psi)S_e(\psi)u^T] = \frac{1}{2}\text{Trace}\left[\Sigma^{-1}\frac{\partial\Sigma}{\partial\psi_d}\Sigma^{-1}\frac{\partial\Sigma}{\partial\psi_e}\right]\Sigma + 2\left[\frac{\partial\Sigma}{\partial\psi_d}\Sigma^{-1}\frac{\partial\Sigma}{\partial\psi_e}\right].$$

Substituting these in the expression (5.4) and also the second expression being of order $O(m^{-1}),$ we can get the following upto order $o(m^{-1})$

$$[\hat\theta(\hat\psi)-\hat\theta(\psi)][\hat\theta(\hat\psi)-\hat\theta(\psi)]^T$$
$$= L^T(\psi)[I_\psi^{-1}(\psi)\otimes I_m]\underset{1\le d\le 2}{\text{Col}}\ \underset{1\le e\le 2}{\text{Concat}}[I_{de}(\psi)\Sigma]$$
$$[I_\psi^{-1}(\psi)\otimes I_m]L(\psi)$$
$$= L^T(\psi)[I_\psi^{-1}(\psi)\otimes I_m][I_\psi(\psi)\otimes\Sigma][I_\psi^{-1}(\psi)\otimes I_m]L(\psi)$$
$$= L^T(\psi)[I_\psi^{-1}(\psi)\otimes\Sigma]L(\psi).$$

Theorem A.2: Under Regularity Conditions 1

$$E[g_1(\hat\psi)+g_3(\hat\psi)-g_4(\hat\psi)-g_5(\hat\psi)] = g_1(\psi)+o(m^{-1}), \quad (5.5)$$

$$E[g_2(\hat\psi)] = g_2(\psi)+o(m^{-1}),$$
$$E[g_3(\hat\psi)] = g_3(\psi)+o(m^{-1}) \quad (5.6)$$

$$\text{and } E[g_5(\hat\psi)] = g_5(\psi)+o(m^{-1}). \quad (5.7)$$

Proof of Theorem A.2

Taylor Series expansion of $g_1(\hat\psi)$ around ψ and using $\hat\psi-\psi = O_p(m^{-1/2})$ when $\|\hat\psi^*-\psi\| <= \|\hat\psi-\psi\|,$ we get

$$g_1(\hat\psi) = g_1(\psi)+[(\hat\psi)-(\psi)^T\otimes I_m]\nabla g_1(\psi)$$
$$+\frac{1}{2}[(\hat\psi-\psi)^T\otimes I_m]\nabla^2 g_1(\psi)[(\hat\psi-\psi)\otimes I_m]$$
$$+o_p(m^{-1})$$

$$\nabla g_1(\psi) = \left[\frac{\partial g_1(\psi)}{\partial\rho}\ \frac{\partial g_1(\psi)}{\partial\sigma_v^2}\right]^T,$$

$$\nabla^2 g_1(\psi) = \underset{1\le d\le 2}{\text{Col}}\left[\underset{1\le e\le 2}{\text{Concat}}\frac{\partial^2 g_1(\psi)}{\partial\psi_d\partial\psi_e}\right]$$

$$\frac{\partial g_1(\psi)}{\partial\psi_d} = R\Sigma^{-1}\frac{\partial\Sigma}{\partial\psi_d}\Sigma^{-1}R$$

$$\frac{\partial^2 g_1(\psi)}{\partial\psi_d\partial\psi_e} = -2R\Sigma^{-1}\frac{\partial\Sigma}{\partial\psi_d}\Sigma^{-1}\frac{\partial\Sigma}{\partial\psi_e}\Sigma^{-1}R$$
$$+R\Sigma^{-1}\frac{\partial^2\Sigma}{\partial\psi_d\partial\psi_e}\Sigma^{-1}R.$$

Using the fact that $\Sigma(\psi)$ and its derivatives are symmetric, we have the second term of the expression as

$$[(\hat\psi-\psi)^T\otimes I_m]\nabla^2 g_1(\psi)[(\hat\psi-\psi)\otimes I_m]$$
$$= -L^T(\psi)[I_\psi^{-1}(\psi)\otimes\Sigma]L(\psi)$$
$$+\frac{1}{2}\underset{m}{\text{Trace}}\left[[I_2\otimes(R\Sigma^{-1})]\frac{\partial^2\Sigma}{\partial\psi\partial\psi^T}[I_\psi^{-1}(\psi)\otimes(\Sigma^{-1}R)]\right]$$
$$= -g_3(\psi)+g_5(\psi)$$

where $I_\psi^{-1}(\psi) = \text{Var}(\psi)$ is information matrix, the asymptotic variance of $\psi.$ The first term in the expression $[(\hat\psi-\psi)^T\otimes I_m]\nabla g_1(\psi)$ reduces to $g_4(\psi)$ because of $E(\hat\psi-\psi) = b_{\hat\psi}(\psi)$ up to order $o(m^{-1})$ (Peers and Iqbal 1985).

The second part of the Theorem follows from the Taylor series expansion of $g_2(\hat\psi), g_3(\hat\psi)$ and $g_5(\hat\psi),$ each around ψ and using $\hat\psi-\psi = O_p(m^{-1/2})$ and $(\partial^2 g_2(\psi))/(\partial\psi_d\,\partial\psi_e)|_{\psi=\hat\psi} = O_p(m^{-1}),$ $(\partial^2 g_3(\psi))/(\partial\psi_d\partial\psi_e)|_{\psi=\hat\psi} = O_p(m^{-1})$ and $(\partial^2 g_5(\psi))/(\partial\psi_d\partial\psi_e)|_{\psi=\hat\psi} = O_p(m^{-1}),$ respectively where $\|\hat\psi^*-\psi\|\le\|\hat\psi-\psi\|.$

Theorem A.3: Under Regularity Conditions 2

$$\text{MSE}([\hat\theta_t(\hat\psi)] = g_{12t}(\psi)+g_{3t}(\psi)+o(m^{-1}). \quad (5.8)$$

Proof of Theorem A.3

The proof is basically on the line of Theorem A.1 and with the use of the results of (Srivastawa and Tiwari (1976)) mentioned therein.

$$\text{MSE}([\hat\theta_t(\hat\psi)]$$
$$= \text{MSE}([\hat\theta_t(\psi)]+E([\theta_t(\psi)-\theta_t)(\theta_t(\psi)-\theta_t)^T]$$
$$= g_{12t}(\psi)+E[(\theta_t(\psi)-\theta_t)(\theta_t(\psi)-\theta_t)^T]. \quad (5.9)$$

Taylor series expansion of $\theta_t(\psi)$ around ψ and using $(\hat{\psi}-\psi)=O_p(m^{-1/2})$ and $(\partial^2\hat{\theta}(\psi))/(\partial\psi_d\partial\psi_e)|_{\psi=\hat{\psi}^*}=O_p(1)$ when $\|\hat{\psi}^*-\psi\|\leqq\|\hat{\psi}-\psi\|$ we have

$$[\hat{\theta}_t(\hat{\psi})-\hat{\theta}_t(\psi)]$$
$$=[(\hat{\psi}-\psi)\otimes I_m]^T\nabla\hat{\theta}_t(\psi)+O_p(m^{-1})$$
$$=\sum_{d=1}^{3}[(\hat{\psi}_d-\psi_d)L_{td}(\psi)e_t(\psi)]+O_p(m^{-1}). \qquad (5.10)$$

Further using the Taylor series expansion of the Likelihood equation $S(\hat{\eta})=0$ and the orthogonality of β and ψ, it follows

$$(\hat{\psi}-\psi)=I_\psi^{-1}(\psi)S(\psi)+O_p(m^{-1}). \qquad (5.11)$$

Substituting the expression for $(\hat{\psi}-\psi)$ in equation (5.10), we have up to order $o(m^{-1})$

$$[\hat{\theta}_t(\hat{\psi})-\hat{\theta}_t(\psi)]=L_t^T(\psi)[I_\psi^{-1}(\psi)\otimes I_m][S_\psi(\psi)\otimes e_t] \qquad (5.12)$$

and

$$[(\hat{\theta}_t(\hat{\psi})-\hat{\theta}_t(\psi))(\hat{\theta}_t(\hat{\psi})-\hat{\theta}_t(\psi))^T]$$
$$=L^T(\psi)[I_\psi^{-1}(\psi)\otimes I_m]\underset{1\le d\le 3}{\text{Col}}\ \underset{1\le e\le 3}{\text{Concat}}$$
$$[e_tS_d(\psi)S_e(\psi)e_t^T][I_\psi^{-1}(\psi)\otimes I_m]L(\psi) \qquad (5.13)$$

where

$$S_\psi(\psi)=\underset{1\le d\le 3}{\text{Col}}[S_d(\psi)],\quad S_d(\psi)=\frac{\partial\ell}{\partial\psi_d}.$$

Using the expression for derivatives of likelihood, we have

$$S_d(\psi)=\frac{1}{2}\left[\begin{array}{l}\text{Trace}[C_{1d}(\psi)]-\sum_{t=1}^{T}\text{Trace}\left[H_t^{-1}\frac{\partial H_t}{\partial\psi_d}\right]\\+\sum_{t=1}^{T}[e_t^TB_{td}(\psi)e_t]\end{array}\right]$$

$$-\left[e_t^TH_t^{-1}\frac{\partial e_t}{\partial\psi_d}\right]C_{1d}(\psi)=\left[(X_1^TH_1^{-1}X_1)^{-1}X_1^TH_1^{-1}\frac{\partial H_1}{\partial\psi_d}H_1^{-1}X_1\right],$$

$$B_{td}(\psi)=H_t^{-1}\frac{\partial H_t}{\partial\psi_d}H_t^{-1}.$$

By applying the considerations $e_t\sim N(0,H_t)$, $\text{Corr}(e_i,e_j)=0$ for $i\neq j$, $\text{Corr}(e_i,(\partial e_i)/(\partial\psi_d))=0$ and $\text{Corr}(e_i,(\partial^2e_i)/(\partial\psi_d\partial\psi_e))=0$ due to the fact that $(\partial e_i)/(\partial\psi_d)=(\partial(y_i-U_t\hat{\alpha}_{t|t-1}))/(\partial\psi_d)$ being linear function of $(y_1,y_2,\ldots,y_{t-1})$ is uncorrelated with e_t, we get the expectation of the inner most terms of the expression (5.13) as

$$E[e_tS_d(\psi)S_e(\psi)E_t^T]=K_{de}(\psi)H_t+2\left[\frac{\partial H_t}{\partial\psi_d}H_t^{-1}\frac{\partial H_t}{\partial\psi_d}\right]$$
$$+\frac{1}{2}\left[\text{Trace}[B_{td}(\psi)]\frac{\partial H_t}{\partial\psi_e}+\text{Trace}[B_{te}(\psi)]\frac{\partial H_t}{\partial\psi_d}\right]$$
$$+\frac{1}{4}\text{Trace}[B_{td}(\psi)]\text{Trace}[B_{te}(\psi)]H_t$$

where

$$K_{de}(\psi)=\frac{1}{2}\sum_{t=1}^{T}\text{Trace}\left[H_t^{-1}\frac{\partial H_t}{\partial\psi_d}H_t^{-1}\frac{\partial H_t}{\partial\psi_e}\right].$$

The middle three terms in the expression being of order $O(1)$ which along with $I_\psi^{-1}(\psi)$ in the expression given below makes them of order $o(m^{-1})$,

$$E[(\hat{\theta}_t(\hat{\psi})-\hat{\theta}_t(\psi))(\hat{\theta}_t(\hat{\psi})-\hat{\theta}_t(\psi))^T]=g_{3t}(\psi)$$
$$=L^T(\psi)[I_\psi^{-1}(\psi)\otimes I_m][K_\psi(\psi)\otimes H_t]$$
$$[I_\psi^{-1}(\psi)\otimes I_m]L(\psi)+o(m^{-1})$$
$$=L^T(\psi)[I_\psi^{-1}(\psi)K_\psi(\psi)I_\psi^{-1}(\psi)\otimes H_t]L(\psi)+o(m^{-1}).$$

Theorem A.4: Under Regularity Conditions 2

$$E[g_{12t}(\hat{\psi})+g_{3t}(\hat{\psi})+g_{31t}(\hat{\psi})-g_{4t}(\hat{\psi})-g_{5t}(\hat{\psi})]$$
$$=g_{12t}(\psi)+o(m^{-1})$$
$$E[g_{3t}(\hat{\psi})]=g_{3t}(\psi)+o(m^{-1})$$

and

$$E[g_{5t}(\hat{\psi})]=g_{5t}(\psi)+o(m^{-1}).$$

Proof of Theorem A.4

The proof is essentially based on the line suggested in proving Theorem A.2. Using Taylor series expansion of $g_{12t}(\hat{\psi})$ around ψ, we get

$$g_{12t}(\hat{\psi})=g_{12t}(\psi)+[(\hat{\psi}-\psi)\otimes I_m]^T\nabla g_{12t}(\psi)$$
$$+\frac{1}{2}[(\hat{\psi}-\psi)^T\otimes I_m]\nabla^2g_{12t}(\psi)[(\hat{\psi}-\psi)\otimes I_m]$$
$$+o_p(m^{-1})$$

$$\nabla g_{12t}(\psi)=\underset{1\le d\le 3}{\text{Col}}[\nabla g_{12td}(\psi)],\ \nabla g_{12td}(\psi)=\frac{\partial g_{12t}(\psi)}{\partial\psi_d}$$

$$\nabla^2g_{12t}(\psi)=\underset{1\le d\le 3}{\text{Col}}\left[\underset{1\le e\le 3}{\text{Concat}}\frac{\partial^2g_{12t}(\psi)}{\partial\psi_d\partial\psi_e}\right]$$

$$\frac{\partial g_{12t}(\psi)}{\partial\psi_d}=R\Sigma^{-1}\frac{\partial R}{\partial\psi_d}\Sigma^{-1}R$$

$$\frac{\partial^2g_{12t}(\psi)}{\partial\psi_d\partial\psi_e}=-2R\Sigma^{-1}\frac{\partial\Sigma}{\partial\psi_d}\Sigma^{-1}\frac{\partial\Sigma}{\partial\psi_e}\Sigma^{-1}R$$
$$+R\Sigma^{-1}\frac{\partial^2\Sigma}{\partial\psi_d\partial\psi_e}\Sigma^{-1}R.$$

Using the fact that $\Sigma(\psi)$ and its derivatives are symmetric, we have the second term of the expression as

$$\begin{aligned}
&[(\hat{\psi}-\psi)^T \otimes I_m]\nabla^2 g_{12t}(\psi)[(\hat{\psi}-\psi)\otimes I_m] \\
&= \; -L^T(\psi)[I_\psi^{-1}(\psi)\otimes\Sigma]L(\psi) \\
&\quad +\frac{1}{2}\underset{m}{\text{Trace}}\left[[I_3\otimes(R\Sigma^{-1})]\frac{\partial^2\Sigma}{\partial\psi\partial\psi^T}[I_\psi^{-1}(\psi)\otimes(\Sigma^{-1}R)]\right] \\
&= \; -g_{3t}(\psi) = g_{5t}(\psi)
\end{aligned}$$

where $I_\psi^{-1}(\psi) = \text{Var}(\psi)$ is the asymptotic variance of ψ. The first term in the expression $[(\hat{\psi}-\psi)^T \otimes I_m]\nabla g_{12t}(\psi)$ reduces to $g_{4t}(\psi)$ because of $E(\hat{\psi}-\psi) = b_{\hat{\psi}}(\psi)$ up to order $o(m^{-1})$ (Peers and Iqbal 1985).

The second part of the Theorem follows from the Taylor series expansion of $g_{3t}(\hat{\psi})$ and $g_{5t}(\hat{\psi})$, each around ψ and using $\hat{\psi}-\psi = O_p(m^{-1/2})$ and $(\partial^2 g_{3t}(\psi))/(\partial\psi_d\,\partial\psi_e)|_{\psi=\hat{\psi}} = O_p(m^{-1})$ and $(\partial^2 g_{5t}(\psi))/(\partial\psi_d\partial\psi_e)|_{\psi=\hat{\psi}} = O_p(m^{-1})$, respectively where $\|\hat{\psi}^* - \hat{\psi}\| \le \|\hat{\psi}-\psi\|$.

References

Cliff, A.D., and Ord, J.K. (1981). *Spatial Processes, Models and Applications*. Pion Limited, London.

Cox, D.R., and Reid, N. (1987). Parameter orthogonality and approximate conditional inference. *Journal of the Royal Statistical Society,* Series B, 49, 1-39 (with discussion).

Cressie, N. (1990). Small-Area prediction of undercount using the general linear model. *Proceedings of Statistics Canada Symposium 90, Measurement and Improvement of Data Quality*, October 1990.

Cressie, N. (1992). REML estimation in empirical bayes smoothing of census undercount. *Survey Methodology*, 18, 75-94.

Datta, G.S., and Lahiri, P. (2000). A unified measure of uncertainty of estimates for best linear unbiased predictors in small area estimation problem. *Statistica Sinica*, 10, 613-627.

Fay, R.E., and Herriot, R.A. (1979). Estimates of income for small places: An application of James-Stein procedure to census data. *Journal of the American Statistics Association*, 74, 267-277.

Ghosh, M., and Rao, J.N.K. (1994). Small area estimation: An appraisal. *Statistical Science*, 9, 55-93.

Henderson, C.R. (1975). Best linear unbiased estimation and prediction under a selection model. *Biometrics*, 31, 423-447.

Kackar, R.N., and Harville, D.A. (1984). Approximations for standard errors of estimators of fixed and random effects in mixed linear models. *Journal of the American Statistics Association*, 79, 853-862.

National Sample Survey Organisation, Department of Statistics, Govt. of India (1997). Report No. 404(50/1.0/1), Consumption of some important commodities in India. *NSS* 50th *Round, July 1993-June 1994.*

National Sample Survey Organisation, Department of Statistics, Govt. of India (1998). Report No. 436(51/1.0/1), Household consumption expenditure and employment situation in India. *NSS* 51st *Round, July 1994-June 1995*.

National Sample Survey Organisation, Department of Statistics, Govt. of India (1998). Report No. 440(52/1.0/1), Household consumption expenditure and employment situation in India, *NSS* 52nd *Round, July 1995-June 1996*.

National Sample Survey Organisation, Department of Statistics, Govt. of India (1998). Report No. 442(53/1.0/1), Household consumption expenditure and employment situation in India. *NSS* 53rd *Round, January-December 1997.*

National Sample Survey Organisation, Department of Statistics, Govt. of India (1999). Report No. 448(54/1.0/1), Household consumption expenditure and employment situation in India. *NSS* 54th *Round, January-June 1998.*

National Sample Survey Organisation, Ministry of Statistics and Programme Implementaion, Govt. of India (2001). Report No. 472(55/1.0/1), Differences in level of consumption among socio-economic groups. *NSS* 55th *Round, July 1999-June 2000.*

Peers, H.W., and Iqbal, M. (1985). Asymptotic expansions for confidence limits in the presence of nuisance parameters, with applications. *Journal of the Royal Statistical Society*, Series B, 47, 547-554.

Prasad, N.G.N., and Rao, J.N.K. (1990). The estimation of mean square error of small area estimates. *Journal of the American Statistics Association*, 85, 163-171.

Rao, C.R., and Toutenbourg, Heldge (1999). *Linear Models*: *Least Squares and Alternatives*. Second Edition, New York: Springer.

Rao, J.N.K. (1999). Some recent advances in model based small area estimation. *Survey Methodology*, 25, 175-186.

Ripley, B.D. (1981). *Spatial Statistics*. New York: John Wiley & Sons, Inc.

Sallas W.M., and Harville D.A. (1994). Noninformative priors and restricted likelihood estimation in the Kalman filter. *Bayesian Analysis of Time Series and Dynamic Models*, (Ed. James C. Spall). New York: Marcel Dekker Inc., 477-508.

Singh, A.C., Mantel, H.J. and Thomas B.W. (1994). Time series EBLUPs for small area using survey data. *Survey Methodology*, 20, 33-43.

Singh, A.C., Stukel, D. and Pfeffermann, D. (1998). Bayesian versus frequentist measures of error in small area estimation. *Journal of the Royal Statistical Society*, Series B, 377-396.

Srivastava, V.K., and Tiwari, R. (1976). Evaluation of expectations of products of stochastic matrices. *Scandinavian Journal of Statistics*, 3, 135-138.