Spatial Modeling and Prediction of County-Level Employment Growth Data

N. Ganesh*

Abstract
For correlated sample survey estimates, a linear model with covariance matrix in which small areas are grouped into clusters by a similarity measure based on spatial locations is proposed. In the context of correlated data, a novel asymptotic framework, a hybrid of infill asymptotics and increasing domain asymptotics is introduced. The hybrid asymptotic framework assumes that the number of clusters and the number of small areas in each cluster grows with sample size. Under the previously mentioned asymptotic framework, the proposed parameter estimators are $k^{\frac{1}{2}}$-consistent, where k is the number of clusters. The proposed model is implemented for county-level civilian employment growth data.

Key Words: empirical best linear unbiased predictor, Fay-Herriot model, increasing domain asymptotics, infill asymptotics, Small Area Estimation, Spatial Statistics.

1. Introduction

The Fay-Herriot model (Fay and Herriot, 1979), a popular area-level model in small area estimation consists of two-levels:

- Level 1 (sampling model): $y_i|\theta_i \overset{\text{ind}}{\sim} N(\theta_i, \psi_i),\ i = 1, \cdots, m;$
- Level 2 (linking model): $\theta_i \overset{\text{ind}}{\sim} N(x_i'\beta, \tau^2),\ i = 1, \cdots, m.$

In the above model, Level 1 is used to account for the sampling variability of the direct survey estimates y_i of the true small area means θ_i. Level 2 links the true small area means θ_i to a vector of q known auxiliary variables x_i, often obtained from various administrative and census records. The parameters β and τ^2 of the linking model are unknown and are estimated from the available data. In order to estimate the sampling variability ψ_i, Fay and Herriot (1979) employed the generalized variance function method [see Wolter (1985)] that uses some external information from the survey. For a comprehensive review of the theory and applications of the above model, see Rao (2003).

The Fay-Herriot model can also be viewed as an area-level mixed regression model:

$$y_i = \theta_i + e_i = x_i'\beta + v_i + e_i,\ i = 1, \cdots, m,$$

where v_i's and e_i's are independent with $v_i \overset{iid}{\sim} N(0, \tau^2)$ and $e_i \overset{ind}{\sim} N(0, \psi_i)$.

In this paper, the Fay-Herriot model is generalized by a model in which the random effects v_i are spatially correlated. Similar spatial models in small area estimation can be found in Cressie (1991), Rao (2003) and Singh et. al (2005). However, a major difference is the asymptotic framework that is considered in this paper. In Section 2, two well-known asymptotic frameworks for spatial data are summarized. In Section 3, the proposed asymptotic framework is introduced. In Section 4, a spatial Fay-Herriot model is introduced, and in Section 5, parameter estimators of the model, large sample properties of these parameter estimators, and empirical best linear unbiased predictors of the small area means θ_i are summarized. Finally, in Section 6, a county-level civilian employment growth data set is analyzed.

2. Asymptotics for Spatial Data

For spatial data, two distinct asymptotic frameworks have been studied. Increasing domain asymptotics refers to more and more observations being sampled over an increasing domain $\mathcal{D} \subset \mathbb{R}^2$ such that the Lebesgue measure of $\mathcal{D}$, $|\mathcal{D}| \to \infty$. When referring to increasing domain asymptotics, it is assumed that the spatial locations of the observations do not become dense.

Infill asymptotics refers to observations being increasingly sampled over a bounded domain. There are very few asymptotic results under infill asymptotics. For example, it is known that some covariance parameters of a zero mean Gaussian process can not be consistently estimated, and for the remaining covariance parameters, the maximum likelihood estimator is consistent and asymptotically normal. For such results, see Abt and Welch (1998), Chen et al. (2000), Ying (1993), Zhang (2004) and Zhang and Zimmerman (2005).

*National Opinion Research Center, 4350 East-West Highway, Suite 800, Bethesda, MD 20814

One of the most popular covariance models for spatial data is given by

$$C(h_i, h_j) = \begin{cases} \sigma^2 + \delta & \text{if} \quad i = j, \\ \delta \exp(-\lambda \|h_i - h_j\|) & \text{if} \quad i \neq j, \end{cases} \tag{1}$$

where h_i, h_j are the spatial locations of the observations, $\|.\|$ is the Euclidean norm and $\delta \geqslant 0$, $\lambda \geqslant 0$, $\sigma^2 \geqslant 0$, $\eta = (\delta, \lambda, \sigma^2)'$, see Cressie (1993), Stein (1999) and Zimmermann and Harville (1991). The above model is referred to as "exponential covariance model with nugget effect". Under infill asymptotics and assuming that the spatial process is Gaussian, when the covariance model is given by (1) and the spatial locations h_i are situated on a lattice in $[0,1]$, Chen et al. (2000) showed that the maximum likelihood estimator for σ^2 is $m^{\frac{1}{2}}$-consistent. Moreover, δ and λ can not be simultaneously consistently estimated, but the maximum likelihood estimator for $\delta\lambda$ is $m^{\frac{1}{4}}$-consistent. There are no asymptotic results for the maximum likelihood estimator for η under infill asymptotics when either the spatial locations h_i are irregularly spaced on $[0,1]$ or for any spatial pattern $h_i \in [0,1]^2$. On the other hand, under increasing domain asymptotics and assuming the spatial locations do not become dense, from Mardia and Marshall (1984) it follows that the maximum likelihood estimator for η is $m^{\frac{1}{2}}$-consistent.

3. Proposed Asymptotic Framework

It is assumed that the spatial locations h_i of the small areas are in an increasing domain, but are scaled such that the translated and scaled $h_i^\star$ are in a bounded domain. That is, if h_i and h_j are the spatial locations of two small areas, it is assumed there is a scaling factor M^p such that

$$\|h_i - h_j\| = M^p \|h_i^\star - h_j^\star\|, \tag{2}$$

where $h_i^\star$ and $h_j^\star$ are in a bounded domain, $0 < p < \frac{1}{2}$ is a user specified parameter and M is the total number of small areas. Note that when $p = \frac{1}{2}$ and the spatial location h_i, h_j do not become dense, we have the increasing domain asymptotic framework. Moreover, when $p = 0$, that is, when the original spatial locations h_i, h_j are in a bounded domain, we have the infill asymptotic framework.

Furthermore, it is assumed that the *set of small areas U can be partitioned into k ($= k(M)$ increasing to ∞ with M) clusters $C_1, \ldots, C_k$, with cluster sizes $N_1, \ldots, N_k$ such that $\sum_{l=1}^k N_l = M$. From each cluster C_l, n_l of the N_l small areas are sampled such that $\sum_{l=1}^k n_l = m$. The n_l's are assumed to be non-random. The asymptotic framework that is considered is $k \to \infty$ and for each l, $N_l \to \infty$, $n_l \to \infty$ such that $0 < \lim_{n_l, N_l \to \infty} \frac{n_l}{N_l} < \infty$.*

Moreover, for $l = 1, \ldots, k$, and for $i, j \in C_l$,

$$\limsup_{M \to \infty} M^p \sup_{i,j \in C_l} \|h_i^\star - h_j^\star\| < \infty,$$

and for all $l_1 \neq l_2$,

$$\liminf_{M \to \infty} \frac{M^p}{\log M} \inf_{i \in C_{l_1}, j \in C_{l_2}} \|h_i^\star - h_j^\star\| = \infty.$$

Note the slightly unusual definition of what it means for two small areas to be in the same cluster. They are defined to be in the same cluster only if asymptotically their scaled distance from one another is finite. Moreover, the clusters should not shrink to a point, that is, it is assumed that for $l = 1, \ldots, k$, $\exists\ \epsilon_l > 0$ such that

$$\lim_{N_l \to \infty} \frac{1}{N_l^2} \sum_{i,j \in C_l} I_{[M^p \|h_i^\star - h_j^\star\| \geqslant \epsilon_l]} = c_l,$$

where $1 \geqslant c_l > 0$.

In Figure 1, examples of the spatial locations of the small areas under each of the asymptotic frameworks is given. Under infill asymptotics [panels (A1)-(A3)], the domain remains fixed and the spatial locations of the small areas become dense in $[0,16] \times [0,16]$. However, under increasing domain asymptotics [panels (B1)-(B3)], as the total number of small areas M increase, the domain also increases and the spatial locations do not become dense in the increasing domain. Finally, under the proposed asymptotic framework [panels (C1)-(C3)], as the number of clusters k and the number of small areas in each cluster N increase (in this example, for $l = 1, \ldots, k$, $N_l = N$), the domain also increases and the clusters move away from one another. However, within each cluster, the spatial locations of the small areas become dense, and the clusters do not shrink to a point.

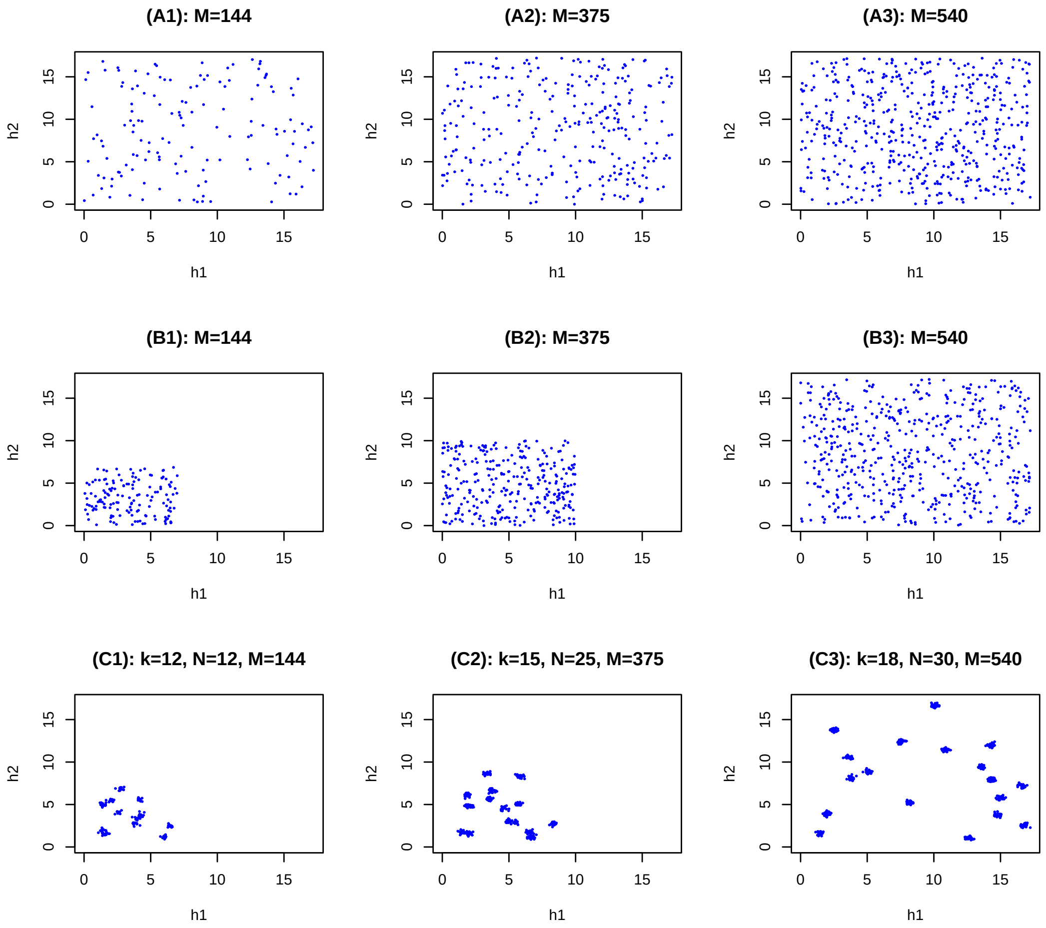

Figure 1: Spatial locations of the small areas under (i) infill asymptotics [panels (A1)-(A3)], (ii) increasing domain asymptotics [panels (B1)-(B3)], (iii) proposed asymptotic framework [panels (C1)-(C3)].

4. Spatial Fay-Herriot Model

In the Fay-Herriot model, the small area effects are assumed independent, though in many data problems neighboring areas ought to be correlated, and by modeling the correlation, better predictors of the small area means could be achieved. A spatial generalization of the Fay-Herriot model is given by

$$y_i = \theta_i + e_i,\ \ i \in S,$$
$$\theta_i = x_i'\beta + v_i,\ \ i \in U,$$

where U is the set of all small areas, with $|U| = M$, and $S \subset U$ is the set of sampled small areas, with $|S| = m$. As in the Fay-Herriot model, the survey estimates $\{y_i : i \in S\}$ are observed, $\{x_i : i \in U\}$ are vector valued covariates, the sampling errors $e_i \overset{\text{ind}}{\sim} N(0, \psi_i)$, and the e_i's and the v_i's are independent. Moreover, $v_{\text{U}} = (v_1, \ldots, v_{\text{M}})'$ is a mean zero normal random vector with covariance matrix Σ_{U} given by

$$\Sigma_{\text{U}} = \sigma^2 I_M + \delta A_{\text{U}},$$

where the $(i,j)^{th}$ entry of A_{U} is given by

$$A_{ij} = \exp(-\lambda M^p \|h_i^\star - h_j^\star\|),$$

where $\delta \geqslant 0$, $\lambda \geqslant 0$, $\sigma^2 \geqslant 0$, $\eta = (\delta, \lambda, \sigma^2)'$ and $M^p\|h_i^\star - h_j^\star\|$ is as defined in (2).

For notational convenience, the set U is re-indexed so that the first m elements of U consist of the sampled small areas. Given the set of sampled small areas, the vector of survey estimates $y = (y_1, \ldots, y_m)'$ can be modeled as

$$y = \theta + e = X\beta + v + e,$$

where $X = (x_1, \ldots, x_m)'$, $v = (v_1, \ldots, v_m)'$ and $e = (e_1, \ldots, e_m)'$ are independent with $v \sim \mathcal{N}(0_m, \Sigma)$ and $e \sim \mathcal{N}(0_m, \Psi)$. Here, Σ is the sub-matrix of Σ_U that corresponds to the sampled small areas and $\Psi = \text{diag}(\psi_1, \ldots, \psi_m)$. Also,

$$\text{var}(y) \equiv V(\eta) = V = \Sigma + \Psi. \tag{3}$$

5. Parameter Estimation and Prediction

As mentioned in Section 2, for spatial models under infill asymptotics, there are very few asymptotic results for the maximum likelihood estimator. Moreover, under infill asymptotics, all known asymptotic results of the maximum likelihood estimator assume restrictive conditions on the spatial locations h_i to be able to write the inverse of the variance-covariance matrix of y in manageable form (Loh and Lam, 2000 and Ying, 1991). Similar technical difficulties are encountered in trying to show the maximum likelihood estimator of (β, η) is consistent and asymptotically normal under the proposed asymptotic framework. Hence, alternate methods of estimation are considered.

5.1 Estimator of (β, τ^2)

The parameter τ^2 is defined as $\tau^2 = \delta + \sigma^2$. An estimator $(\hat{\beta}, \hat{\tau}^2)$ for (β, τ^2) is given by

$$(\hat{\beta}, \hat{\tau}^2) = \underset{\beta \in \mathbb{R}^q, \tau^2 > 0}{\text{argmax}} \; g(\beta, \tau^2; y), \tag{4}$$

where

$$g(\beta, \tau^2; y) = -\frac{m}{2} \log 2\pi - \frac{1}{2} \sum_{i=1}^{m} \log(\tau^2 + \psi_i) - \frac{1}{2} \sum_{i=1}^{m} \frac{(y_i - x_i'\beta)^2}{\tau^2 + \psi_i}. \tag{5}$$

Note that (5) is the log likelihood when the direct survey estimates y_i are assumed to follow the Fay-Herriot model. That is, (β, τ^2) is estimated by maximizing a misspecified log likelihood.

Theorem 1. Under the asymptotic framework described in Section 3 and under certain regularity conditions [for details and proof, see Ganesh (2007)], $(\hat{\beta}, \hat{\tau}^2)$ is consistent for (β, τ^2). Moreover,

$$\begin{pmatrix} (X'D^{-1}VD^{-1}X)^{-\frac{1}{2}} X'D^{-1}X & 0_q \\ 0_q' & \sum_{i=1}^{m} (\tau^2 + \psi_i)^{-2} / [2\text{tr}(D^{-2}VD^{-2}V)]^{\frac{1}{2}} \end{pmatrix} \begin{pmatrix} \hat{\beta} - \beta \\ \hat{\tau}^2 - \tau^2 \end{pmatrix} \xrightarrow{d} \text{N}(0_{q+1}, I_{q+1}),$$

where $D = \text{diag}(\tau^2 + \psi_1, \ldots, \tau^2 + \psi_m)$ and V is given by (3).

5.2 Estimator of (δ, λ)

An estimator $(\hat{\delta}, \hat{\lambda})$ for (δ, λ) is given by

$$(\hat{\delta}, \hat{\lambda}) = \underset{\delta \geqslant 0, \lambda \geqslant 0}{\text{argmax}} \; h(\delta, \lambda; y), \tag{6}$$

where

$$h(\delta, \lambda; y) = -\sum_{l=1}^{k} \sum_{\substack{i,j \in C_l \\ i \neq j}} \left(\hat{\epsilon}_i \hat{\epsilon}_j - \delta \exp(-\lambda M^p \| h_i^\star - h_j^\star \|) \right)^2,$$

where $\hat{\epsilon}_i = y_i - x_i'\hat{\beta}$.

Theorem 2. Under the asymptotic framework described in Section 3 and under certain regularity conditions [for details and proof, see Ganesh (2007)], $(\hat{\delta}, \hat{\lambda})$ is consistent for (δ, λ). Moreover,

$$\frac{\sum_{l=1}^{k} n_l^2}{(\sum_{l=1}^{k} n_l^4)^{\frac{1}{2}}} K^{-\frac{1}{2}} L \begin{pmatrix} \hat{\delta} - \delta \\ \hat{\lambda} - \lambda \end{pmatrix} \xrightarrow{d} \text{N}(0_2, I_2),$$

where

$$K = \frac{8}{\sum_{l=1}^{k} n_l^4} \begin{pmatrix} \mathrm{tr}[GVGV] & -\mathrm{tr}[GVHV] \\ -\mathrm{tr}[GVHV] & \mathrm{tr}[HVHV] \end{pmatrix}, \quad L = \frac{2}{\sum_{l=1}^{k} n_l^2} \begin{pmatrix} \mathrm{tr}[G^2] & -\mathrm{tr}[GH] \\ -\mathrm{tr}[GH] & \mathrm{tr}[H^2] \end{pmatrix}$$

$$G_{ij} = \begin{cases} \exp(-\lambda M^p \|h_i^\star - h_j^\star\|) & \text{if } i \neq j, \ i,j \in C_l \text{ for some } l, \\ 0 & \text{otherwise} \end{cases}$$

$$H_{ij} = \begin{cases} \delta M^p \|h_i^\star - h_j^\star\| \exp(-\lambda M^p \|h_i^\star - h_j^\star\|) & \text{if } i \neq j, \ i,j, \in C_l \text{ for some } l, \\ 0 & \text{otherwise} \end{cases}$$

where G_{ij}, H_{ij} are respectively the $(i,j)^{th}$ entries of G, H.

From Theorem 2, it follows that the asymptotic variances of $\hat{\delta}$ and $\hat{\lambda}$ are of the order $O\big(\sum_{l=1}^{k} n_l^4/(\sum_{l=1}^{k} n_l^2)^2\big)$. If all the n_l's grow at the same rate, that is, there exists n such that for $l = 1, \ldots, k$, $0 < \lim_{n_l, n \to \infty} \frac{n_l}{n} < \infty$, then $\hat{\delta}$ and $\hat{\lambda}$ are $k^{\frac{1}{2}}$-consistent.

Finally, since $\tau^2 = \delta + \sigma^2$, σ^2 can be estimated by

$$\hat{\sigma}^2 = \max(\hat{\tau}^2 - \hat{\delta}, 0), \tag{7}$$

which is a consistent estimator of σ^2.

5.3 Predictor of θ_i

As mentioned previously, one of the objectives of spatially modeling the random effects v_i is to obtain better predictors of the small area means $\theta_i = x_i'\beta + v_i$. For the proposed model, the best linear unbiased predictor (BLUP) of θ_i can be derived along the same lines as the BLUP for a general linear model, see Rao (2003). For the model given in Section 4, the BLUP of θ_i is given by

$$\tilde{\theta}_i = x_i'\tilde{\beta}(\eta) + f_i'\Delta(\eta)[V(\eta)]^{-1}(y - X\tilde{\beta}(\eta)), \tag{8}$$

where f_i is the i^{th} standard basis vector in $\mathbb{R}^M$, $\Delta(\eta) = \mathrm{cov}(v_{\mathrm{U}}, v)$, $\tilde{\beta}(\eta)$ is the best linear unbiased estimator of β, that is, $\tilde{\beta}(\eta) = \big(X'[V(\eta)]^{-1}X\big)^{-1}X'[V(\eta)]^{-1}y$ and $V(\eta)$ is given by (3). Since (8) depends on unknown variance components $\eta = (\delta, \lambda, \sigma^2)'$, an empirical best linear unbiased predictor of θ_i is given by

$$\hat{\theta}_i = x_i'\tilde{\beta}(\hat{\eta}) + f_i'\Delta(\hat{\eta})[V(\hat{\eta})]^{-1}(y - X\tilde{\beta}(\hat{\eta})), \tag{9}$$

where $\hat{\eta} = (\hat{\delta}, \hat{\lambda}, \hat{\sigma}^2)'$ is given by (6) and (7).

Moreover, for the Fay-Herriot model, the empirical best linear unbiased predictor of θ_i is given by

$$\breve{\theta}_i = \begin{cases} [\psi_i/(\hat{\tau}^2 + \psi_i)]\, x_i'\hat{\beta} + [\hat{\tau}^2/(\hat{\tau}^2 + \psi_i)]\, y_i & \text{if } i \in S, \\ x_i'\hat{\beta} & \text{if } i \in S^c, \end{cases} \tag{10}$$

where $(\hat{\beta}, \hat{\tau}^2)$ is given by (4) and S^c is the set of non-sampled small areas.

6. Data Analysis

In this section, we analyze a U.S. county-level data set that was previously analyzed by Wheeler (2003a), (2003b) for a different purpose. The data set consists of civilian employment growth rates for all U.S. counties between 1980 and 1990 and includes 14 county-level covariates. The set of covariates included in the data set were (the year is given in parenthesis): log employment (1980), log population (1980), employment density (1980), population density (1980), log land area (1980), fraction of adult population with bachelor's degree (1980), fraction of employment in manufacturing (1980), unemployment rate (1980), per capita income (1979), urban/rural indicator (1990), share of local government spending on education (1982), share of local government spending on police (1982), share of local government spending on highways (1982) and fraction of population that is not white (1980).

Among the 3106 U.S. counties, 4 counties with missing covariates were deleted. Unfortunately, the deleted counties were all large counties; they were Bronx, New York, Queens and Richmond with employment growth rates of 0.0926, 0.0972, 0.0992 and 0.1976. Among the deleted counties, the first 3 counties have approximately the median employment growth rate among all U.S. counties while the last county has an employment growth rate in the 75^{th} percentile.

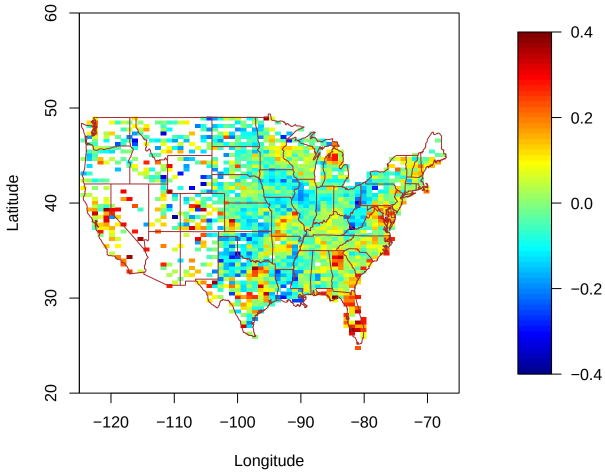

Figure 2: Plot of the residuals from the model given by (11).

Initially, for the employment growth rates, a linear model with independent errors was considered:

$$y_i = x_i'\beta + \varepsilon_i, \tag{11}$$

where $\varepsilon_i \overset{iid}{\sim} N(0, \sigma_\varepsilon^2)$. In order to choose the best set of covariates, a stepwise AIC criterion was used. The covariates selected by the AIC criterion were: log employment, urban/rural indicator, fraction of adult population with bachelor's degree, fraction of employment in manufacturing, log population, share of local government spending on police, fraction of population that is not white, and interaction between the following pairs of covariates: urban/rural indicator and fraction of employment in manufacturing, urban/rural indicator and log population, share of local government spending on police and fraction of employment in manufacturing, share of local government spending on police and fraction of adult population with bachelor's degree. Among all coefficients for the fixed effects, the largest p-value was 2.2×10^{-11}.

A histogram of the employment growth rates shows a slightly fatter tail to the right, but is otherwise symmetric and unimodal. Having fitted the model with the above mentioned covariates, a plot of the residuals against covariates and against the fitted values showed no pattern. Moreover, for the residuals, a normal Q-Q plot indicated that for the middle and the left tail, the plot looked fine, however beyond the second standard deviation the plot deviated significantly from the normal quantiles. This is associated with the slightly fatter tail to the right of the histogram of the observations.

In Figure 2 (the pixels are county centroids), a plot of the residuals from fitting the model given by (11) (with x_i chosen to be the above selected covariates) indicates strong spatial correlation among the residuals, and also, large residuals for counties in the North East, South East and West/South West regions of the U.S. Hence, in addition to the previously selected covariates, indicator variables were included for each of these regions in the model given by (11). All three indicator variables were significant with p-values $< 2 \times 10^{-16}$. Moreover, all the previously selected covariates were still significant with largest p-value 2.56×10^{-11}.

A sample of 1240 from the 3102 U.S. counties were drawn, that is, $m = 1240$, $M = 3102$, $m/M \approx 0.4$. Counties with a population of at least 500000 were self-represented, there were 81 such counties, and the remaining 1159 counties were chosen by simple random sample from the 3021 counties with a population of less than 500000. We pretend that only the sampled counties were observed. Moreover, since the sampling variances ψ_i were not included in the data set, noise was added to the data in order to mimic a typical setting in which the Fay-Herriot model is used. Also, we pretend that the employment growth rates given in the data set were the true employment growth rates.

Four different data sets were generated by adding noise to the true employment growth rates of the sampled counties with differing sampling variances. The sampling variances ψ_i were chosen so that $\psi_i \propto 1/P_i$, where P_i

Table 1: Parameter estimates for each of the four data sets.

c	$\hat{\delta}$	$\hat{\lambda}$	$\hat{\sigma}^2$	$\hat{\tau}^2$
90.95	0.01256	0.00282	0.00250	0.01506
155.92	0.01231	0.00277	0.00377	0.01607
242.54	0.01170	0.00274	0.00319	0.01489
363.81	0.01145	0.00282	0.00388	0.01533

Table 2: Average squared error for the sampled and non-sampled counties for each of the four data sets.

Model	c	Sampled	Non-sampled
Proposed	90.95	0.0038	0.0130
Fay-Herriot	90.95	0.0042	0.0174
Proposed	155.92	0.0052	0.0133
Fay-Herriot	155.92	0.0059	0.0175
Proposed	242.54	0.0064	0.0141
Fay-Herriot	242.54	0.0071	0.0175
Proposed	363.81	0.0071	0.0139
Fay-Herriot	363.81	0.0083	0.0176

is the population size for the i^{th} county. Note that for the sampled counties, the empirical best linear unbiased predictor of θ_i given in (10) is a weighted linear combination of the direct estimate and the regression estimate. Hence, the constant of proportionality for ψ_i was chosen so that among the sampled counties, for the county with median population size, the empirical best linear unbiased predictor for the true employment growth rate would have weight approximately equal to 0.8, 0.7, 0.6, 0.5 for the direct estimate. These four cases correspond to the constant proportionality $c = 90.95$, 155.92, 242.54, 363.81.

For each value of c, noise was added to the true employment growth rates of the sampled counties, and the Fay-Herriot model and the proposed model were fitted. The user specified parameter p in (2) was chosen to be 0.25. Table 1 gives the parameter estimates of $(\delta, \lambda, \sigma^2, \tau^2)$. As can be seen, the random effects show strong spatial correlation. For example, for the first data set with $c = 90.95$, the estimated correlation among the random effects when two counties are 0, 20, 40 and 100 miles apart were respectively 0.83, 0.57, 0.40, 0.13.

For each of the four data sets, Table 2 gives the averaged squared error of the empirical best linear unbiased predictor for the sampled and non-sampled counties under the proposed model and the Fay-Herriot model. For example, for the sampled counties S, the average squared error is defined as

$$\frac{1}{m}\sum_{i\in S}(\check{\theta}_i - \theta_i)^2,$$

where $\check{\theta}_i$ is either (9) or (10) and θ_i is the true employment growth rate. Note that the average squared error is larger for the non-sampled counties than the sampled counties. Moreover, the relative efficiency of the empirical best linear unbiased predictor under the proposed model and the Fay-Herriot model differs significantly for the sampled and non-sampled counties. The relative efficiency is computed by taking the ratio of the average squared error of the predictors under the Fay-Herriot model and the proposed model. For example, when $c = 90.95$, the relative efficiency of the sampled counties is $0.0042/0.0038 = 1.105$ and the relative efficiency of the non-sampled counties is $0.0174/0.0130 = 1.338$. Moreover, Table 2 suggests that the relative efficiency of the non-sampled counties decreases as c increases. That is, when noise with increasing sampling variance is added to the true employment growth rates, the relative efficiency of the predictors for the non-sampled counties decreases. There is also evidence to suggest that the relative efficiency of the sampled counties increases as c increases.

Figure 3 gives the true employment growth rates [panel (A)], and for the first data set with $c = 90.95$, the predicted employment growth rates using the proposed model [panel (B)]. In some States such as CO, FL, KS, NE and SD, the empirical best linear unbiased predictor of the true employment growth rates have been over-smoothed, but otherwise both maps look similar.

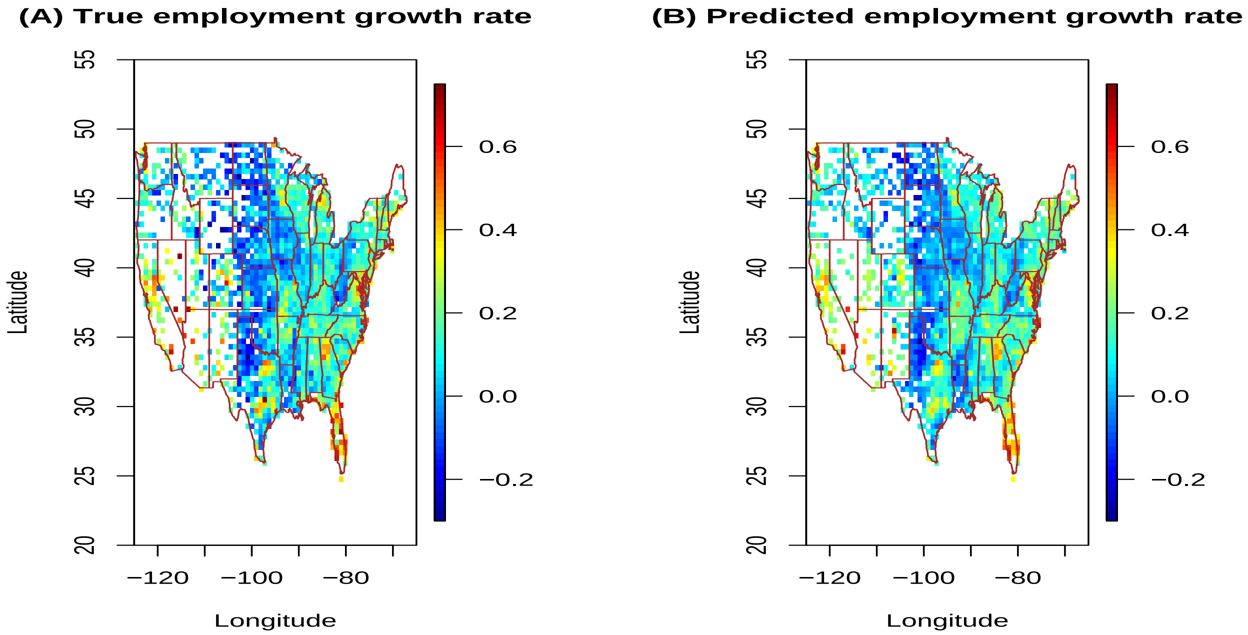

Figure 3: True employment growth rate [panel (A)], and for the data set with $c = 90.95$, predicted employment growth rate using the proposed model [panel (B)].

In summary, when small areas are spatially correlated, significant improvements in prediction can be obtained, especially for the non-sampled small areas, by using the proposed model as opposed to the Fay-Herriot model.

REFERENCES

Abt, M. and Welch, W.J. (1998), "Fisher Information and Maximum Likelihood Estimation of Covariance Parameters in Gaussian Stochastic Processes," *Canadian Journal of Statistics*, 26, 127–137.

Chen, H., Simpson, D.G. and Ying, Z. (2000), "Infill Asymptotics for a Stochastic Process Model with Measurement Error," *Statistical Sinica*, 10, 141–156.

Cressie, N. (1991), "Small Area Prediction of Undercount using the General Linear Model," *Proceedings of Statistics Symposium 90: Measurement and Improvement of Data Quality*, Ottawa: Statistics Canada, 93–105.

Cressie, N. (1993), *Statistics for Spatial Data*, New York: Wiley.

Fay, R.E. and Herriot, R. A. (1979), "Estimates of Income for Small Places: An Application of the James-Stein Procedures to Census Data," *Journal of the American Statistical Association* 74, 269–277.

Ganesh, N. (2007), "Small Area Estimation and Prediction Problems: Spatial Models, Bayesian Multiple Comparisons, and Robust MSE Estimation," *Ph.D. Dissertation*, University of Maryland. http://hdl.handle.net/1903/7241.

Loh, W. and Lam, T. (2000), "Estimating Structured Correlation Matrices in Gaussian Random Field Models," *Annals of Statistics*, 28, 880–904.

Mardia, K.V. and Marshall, R.J. (1984), "Maximum Likelihood Estimation of Models for Residual Covariance in Spatial Regression," *Biometrika*, 71, 135–146.

Rao, J.N.K. (2003), *Small Area Estimation*, New York: Wiley.

Singh, B.B., Shukla, G.K. and Kundu, D. (2005), "Spatial-Temporal Models in Small Area Estimation," *Survey Methodology*, 31 (2), 183–195.

Stein, M.L. (1999), *Interpolation of Spatial Data: Some Theory for Kriging*. New York: Springer-Verlag.

Wheeler, C.H. (2003a), U.S. Counties 1998. http://qed.econ.queensu.ca/jae/ 2003-v18.1/wheeler.

Wheeler, C.H. (2003b), "Evidence on Agglomeration Economies, Diseconomies, and Growth," *Journal of Applied Econometric*, 18, 79–104.

Wolter, K.M. (1985), *Introduction to Variance Estimation*. New York: Springer-Verlag.

Ying, Z. (1993), "Maximum Likelihood Estimation of Parameters under a Spatial Sampling Scheme," *Annals of Statistics*, 21, 1567–1590.

Zhang, H. (2004), "Inconsistent Estimation and Asymptotically Equal Interpolations in Model-Based Geostatistics," *Journal of the American Statistical Association*, 99, 250–261.

Zhang, H. and Zimmerman, D.L. (2005), "Towards Reconciling two Asymptotic Frameworks in Spatial Statistics," *Biometrika*, 92, 921–936.

Zimmerman, D.L. and Harville, D.A. (1991), "A Random Field Approach to the Analysis of Field-Plot Experiments and other Spatial Experiments," *Biometrics*, 47, 223–239.