



EXPLOITING SPATIAL DEPENDENCE TO IMPROVE MEASUREMENT OF NEIGHBORHOOD SOCIAL PROCESSES

*Natalya Verbitsky Savitz**
Stephen W. Raudenbush†

A number of recent studies have used surveys of neighborhood informants and direct observation of city streets to assess aspects of community life such as collective efficacy, the density of kin networks, and social disorder. Raudenbush and Sampson (1999a) have coined the term “ecometrics” to denote the study of the reliability and validity of such assessments. Random errors of measurement will attenuate the associations between these assessments and key outcomes. To address this problem, some studies have used empirical Bayes methods to reduce such biases, while assuming that neighborhood random effects are statistically independent. In this paper we show that the precision and validity of ecometric measures can be considerably improved by exploiting the spatial dependence of neighborhood social processes within the framework of empirical Bayes shrinkage. We compare three estimators of a neighborhood social process: the ordinary least

This work was supported by National Science Foundation Grant 218966. We thank Ben Hansen, Ed Ionides, Jeff Morenoff, Susan Murphy, and the anonymous reviewers for their thoughtful comments. Direct correspondence to Natalya Verbitsky Savitz, Mathematica Policy Research, Inc., 600 Maryland Ave., SW, Suite 550, Washington, DC 20024; email: NVSavitz@mathematica-mpr.com.

*Mathematica Policy Research, Inc.

†University of Chicago

squares estimator (OLS), an empirical Bayes estimator based on the independence assumption (EBE), and an empirical Bayes estimator that exploits spatial dependence (EBS). Under our model assumptions, EBS performs better than EBE and OLS in terms of expected mean squared error loss. The benefits of EBS relative to EBE and OLS depend on the magnitude of spatial dependence, the degree of neighborhood heterogeneity, as well as neighborhood's sample size. A cross-validation study using the original 1995 data from the Project on Human Development in Chicago Neighborhoods and a replication of that survey in 2002 show that the empirical benefits of EBS approximate those expected under our model assumptions; EBS is more internally consistent and temporally stable and demonstrates higher concurrent and predictive validity. A fully Bayes approach has the same properties as does the empirical Bayes approach, but it is preferable when the number of neighborhoods is small.

1. INTRODUCTION

Social scientists have long known that urban neighborhoods vary substantially in rates of crime (Shaw and McKay 1942), disease rates (Zubrick 2007), mental health problems (Leventhal and Brooks-Gunn 2000), birth weight (Buka et al. 2003; Morenoff 2003), and education, fertility, and earnings (Galster et al. 2007). One major debate concerns the extent to which such associations are causal rather than attributable to the background characteristics of persons and families who migrate into these neighborhoods (see reviews by Duncan and Raudenbush [1999]; Oakes [2004]; Diez-Roux [2004]). Closely related is the theoretical question of how social processes might produce these outcomes and how those processes might be measured in order to test theories about neighborhood influences. However, answers to these questions require reliable and valid measures of the social processes in urban neighborhoods.

In this paper, we propose that one can exploit spatial dependence to improve measures of these social processes. Recent advances in statistical theory enable the study of spatially dependent random effects (Banerjee, Carlin, and Gelfand 2004; Verbitsky 2007). We adopt this approach here. In particular, we formulate a first-order Markov model for spatial dependence (Anselin 1988) within a hierarchical linear model with normal-theory random effects. We then provide a series of empirical tests of internal consistency, temporal stability,

concurrent validity, and predictive validity using data collected in 1995 and 2002 by the Project on Human Development in Chicago Neighborhoods (PHDCN).

2. THE LOGIC OF NEIGHBORHOOD MEASUREMENT

Sampson, Raudenbush, and Earls (1997) found that the “collective efficacy” of urban neighborhoods—defined as the fusion of social cohesion and informal social control—significantly predicted low rates of perceived violence, violent victimization, and homicide in Chicago neighborhoods after controlling for demographic characteristics of neighborhoods obtained from the U.S. Census and found to be predictive of crime in past studies. In a similar vein, Browning (2002) found that collective efficacy predicted partner violence, and Browning, Leventhal, and Brooks-Gunn (2005) found a significant association between collective efficacy and low rates of early sexual initiation. Sampson, Morenoff, and Earls (1999) considered the association between neighborhood composition and various aspects of social capital, including collective efficacy as well as the density of kinship and friendship networks and the intensity of reciprocal exchanges among neighbors.

In these studies, survey researchers measured neighborhood social processes by sampling adults within spatially defined units (“neighborhood clusters”), regarding each respondent as an informant about relations among neighbors. Responses to conceptually related questions were combined to form scales intended to measure each informant’s assessment of a latent construct such as collective efficacy. Next, analysts combined these scales across informants within neighborhoods to generate indicators of neighborhood-level latent variables. These neighborhood-level indicators then became the indicators of social process used in the studies just cited. Variation between items within informants and between informants within neighborhoods generates errors of measurement of the neighborhood-level latent variable.

Raudenbush and Sampson (1999a) coined the term “ecometrics” to describe the study of the reliability and validity of assessments of ecological units such as neighborhoods. This work parallels earlier work on the assessment of school climate using teachers as informants (Raudenbush, Rowan, and Kang 1991). Just as psychometrics

identifies sources of error in assessments of cognitive skill and personality, econometrics identifies sources of error in studies of social settings. Using this logic, we can readily see that the reliability of a neighborhood social process measured by interviewing informants will depend on item consistency (the association between item responses within a scale), the number of items in the scale, the degree to which neighborhood informants agree on social relations in a local area, and the number of informants sampled per local area.

The logic of econometrics applies similarly when researchers measure neighborhood characteristics through direct observation. Raudenbush and Sampson (1999a) studied physical and social disorder of Chicago neighborhoods using such observational data. For example, to assess physical disorder, observers coded each city "face block" (one side of a street) on the presence or absence of garbage, broken bottles, abandoned cars, graffiti, cigarette butts, needles or syringes, and condoms. Using a three-level hierarchical logistic regression model, the authors combined responses to these items within a face block and across face blocks within a neighborhood cluster to produce a measure of physical disorder within that neighborhood cluster and to estimate the variance of errors of measurement. The reliability of such a measure will depend on the internal consistency of the items, the number of items, the similarity of face blocks within a neighborhood, the number of face blocks sampled per neighborhood, and the time of day.

Researchers have used such observational measures to study the association between neighborhood disorder and children's physical activity (Molnar et al. 2004), and violent crime (Sampson and Raudenbush 1999). McCrea and colleagues (2005) used a survey-based measure of perceived neighborhood disorder to predict fear of crime, while Sampson and Raudenbush (2003) studied the association between observed and perceived disorder, showing that perceptions of disorder are influenced not only by observable disorder but also by the demographic composition of the local area.

Regardless of whether interviews of key informants or direct observations are used to measure a neighborhood social process, budget constraints will impose some limits on reliability of measurement. In the case of interviews, the sample size of informants per neighborhood cluster constrains reliability. Using almost 8000 informants to measure 343 Chicago neighborhoods, Raudenbush and Sampson (1999a) showed that the reliability of measurement of key social processes ranged

between 0.70 and 0.85. In the case of direct observation, they showed that the number of face blocks observed per neighborhood imposed the key constraint on reliability. Reliability ranged between 0.70 for social disorder and 0.98 for physical disorder. However, their study, part of the Project on Human Development in Chicago Neighborhoods (PHDCN), sampled about 200 face blocks per neighborhood, requiring a budget that often is unavailable.

In this paper, we consider the problem of bias that arises in estimating the association between a neighborhood social process measured with error and an outcome of interest. Random errors of measurement at the level of the informant, combined with small neighborhood sample sizes, will attenuate the estimated associations. To address this problem, some studies have used empirical Bayes methods to reduce such bias (Sampson, Raudenbush, and Earls 1997; Morenoff, Sampson, and Raudenbush 2001). Using this approach, latent variables of interest are regarded as independently distributed across neighborhoods. The posterior mean of the random effect, given the estimated variance components, is a weighted average of the neighborhood sample mean and the overall mean. Under the model assumptions, this “shrinkage” toward the overall mean eliminates the bias that arises from measurement error of the neighborhood social process.

Given the spatial contiguity of neighborhoods, however, the independence assumption regarding the neighborhood random effects is implausible. In this paper, we show that the precision and predictive validity of econometric measures can be considerably improved by exploiting the spatial dependence of neighborhood social processes within the framework of empirical Bayes shrinkage. We compare three estimators of a neighborhood social process: the ordinary least squares estimator (OLS), an empirical Bayes estimator based on the independence assumption where random effects are regarded as exchangeable (EBE), and an empirical Bayes estimator that exploits spatial dependence (EBS). Under our model assumptions, EBS performs better than EBE and OLS in terms of expected mean squared error loss. The benefits of EBS relative to EBE and OLS depend on the magnitude of spatial dependence and the degree of neighborhood heterogeneity, as well as a neighborhood’s sample size.

Of course, the superiority of EBS under our model assumptions does not prove that EBS will be useful in practice, because our model assumptions may not hold up in practice. Thus, the failure of these

assumptions may negate the expected benefit of EBS. To investigate this possibility, we conduct a cross-validation study using the original 1995 PHDCN data (Sampson, Raudenbush, and Earls 1997) and a replication of that survey that was done in 2002. The results show that the empirical benefits of EBS approximate those expected under our model assumptions.

3. STATISTICAL BACKGROUND

3.1. *Shrinkage Estimation When Means Are Independent: A Brief History*

The empirical Bayes approach adopted in past studies of neighborhood social processes draws upon a long tradition of research on the problem of simultaneous estimation of J means. This problem arises when data are collected on a number of independent groups, but the amount of data on each group is insufficient to estimate the group mean precisely. Previous research has shown that a “shrinkage” estimator, such as an empirical Bayes estimator, performs better than the sample mean in predicting the population means of independent groups (e.g., Stein 1956; James and Stein 1961; Lindley 1971; Efron and Morris 1972a, 1973, 1975, 1977) despite the fact that the sample mean is the maximum likelihood as well as the uniform minimum variance unbiased (UMVU) estimator of the population mean in each group when the data are normally distributed. This apparent contradiction is known as Stein’s paradox.

Statisticians later extended James and Stein’s work from a Bayesian perspective. Lindley (1971) noted that this problem could be seen as estimating the means in the analysis of variance (ANOVA) case. If the ANOVA is approached from Bayesian perspective, then the mean of the posterior distribution has a form similar to that proposed by James and Stein and is admissible. Jackson, Novick, and Thayer (1971) extended Lindley’s theoretical results and demonstrated their utility on several real data sets. In a later work, Novick et al. (1972) performed a cross-validation study to demonstrate some of the advantages of using the empirical Bayes estimator over the least squares in predicting future grade point averages of students. Efron and Morris (1972a, 1972b, 1973) extended the statistical theory and provided simulation results

demonstrating the practical utility of James-Stein estimator. They also noted that an empirical Bayes estimator is a James-Stein estimator. Efron and Morris (1977) discussed and illustrated Stein's paradox using various real data sets and provided some cross-validation study results. The basic idea of Stein's paradox is that one can borrow strength when estimating a group mean by using the data from other groups when these groups are independent.

In geographical or spatial settings, however, the groups may be dependent. For example, cities are not generally homogeneous; city neighborhoods may be clustered by ethnicity, socioeconomic status, or age. Furthermore, since these neighborhoods generally do not have impenetrable bounds, spillover effects occur. For example, a crime committed in a proximate neighborhood may affect residents' perceptions of their neighborhood's safety as well as their well-being. This may result in the correlation of area means, since surrounding neighborhoods may contain information about the focal neighborhood.

Empirical Bayes methodology has been used in geographical or spatial research in the past, where estimating county, city, or neighborhood parameters was of interest. Tsutakawa, Shoop, and Marienfeld (1985) derived estimates of rates of mortality in Missouri cities from bladder cancers using an empirical Bayes approach and demonstrated that these estimates are more stable than the standard estimates of mortality rates. Later researchers used the empirical Bayes approach to estimate neighborhood characteristics in Chicago's neighborhoods, such as collective efficacy, and social and physical disorder (Sampson, Raudenbush, and Earls 1997; Sampson and Raudenbush 1999; Morenoff, Sampson, and Raudenbush 2001). These measures were later used as predictors to test various substantive hypotheses. For a review of the use of empirical Bayes methodology in public health, see Bingenheimer and Raudenbush (2004:66–68).

3.2. *Incorporating Spatial Dependence*

However, with some exceptions, the majority of previous research on estimation of site means in a spatial setting did not take into account the dependence of these sites. Carter and Rolph (1974) considered estimation of the probability of a false fire alarm for alarm boxes of the borough of the Bronx in New York City in the late 1960s. They created

larger neighborhoods consisting of several locations and proposed an estimator that shrinks the observed probability at a particular location toward the average probability of the larger neighborhood. However, in this approach, all of the location measures within a particular neighborhood are shrunk toward the same neighborhood measure without taking into account how far a certain location is from the center of the neighborhood. Marshall (1991) improved on this estimator by defining a different neighborhood for each location. Assunção et al. (2005) extended Marshall's (1991) work using Longford's (1999) vector approach and proposed a multivariate EB shrinkage estimator. Clayton and Kaldor (1987) derived formulas for an empirical Bayes estimator for mapping relative risks of disease using a multivariate log normal distribution framework with log relative risks conditionally autocorrelated (Besag 1974; Banerjee et al. 2004) and compared it to with other types of estimators using the lip cancer in Scotland counties example. This approach was later used by Britt et al. (2005) in analyzing the relationship between alcohol outlet density and criminal violence in neighborhoods of the city of Minneapolis, Minnesota.

Recent advances in statistical theory now suggest an approach to the measurement of neighborhood social processes that we believe should be superior to available methods. The approach requires assessment data from all neighborhoods within a well-defined region such as a city. Like many past approaches, the approach specifies a prior distribution for the true or latent neighborhood characteristics and bases inference about these latent variables on their posterior distribution. However, we postulate that these latent variables are *a priori* spatially dependent, following a first-order Markov contiguity process. In principle, incorporating prior information about spatial dependence should strengthen inference by reducing posterior uncertainty about the latent variables that are the objects of measurement. The statistical theory upon which our measurement approach is built was developed generally by Banerjee et al. (2004), who developed a fully Bayesian approach to spatial dependence with many applications. Closely related, Verbitsky (2007) explicated the connections between hierarchical linear models for multilevel data (Raudenbush and Bryk 2002; Goldstein 2002) and spatial dependence models (c.f., Ord 1975; Anselin 1988; Clayton and Kaldor 1987) using maximum likelihood estimation of model parameters and empirical Bayes (EB) estimation of spatially random effects.

The EB approach is simpler than the fully Bayes approach in that, while both approaches specify a prior distribution for the random effects, the EB approach does not specify a prior for the model parameters. Rather, using the EB approach, posterior inference about the random effects is conditioned on maximum likelihood estimates of the model parameters. The EB approach may be regarded as a first-order approximation to the fully Bayes approach, as inferences using the two methods converge as the number of neighborhoods increases. The fully Bayes approach is especially valuable when the number of neighborhoods is modest. In our sample, with 343 neighborhoods, we adopt the simpler EB approach, recognizing the utility of the fully Bayes approach more generally. Our focus is on the extent to which exploiting the spatial dependence of the random effects (deviations of the latent variables from the overall mean) in the context of two-level data can be expected to increase the accuracy of measurement of neighborhood social processes.

4. THEORETICAL COMPARISON OF ALTERNATIVE ESTIMATORS

4.1. *The Model*

Consider the case where data are collected on N individuals living in J sites, with n_j individuals in site j , $j = 1, 2, \dots, J$. These sites are mutually exclusive and comprise the entire geographic area under investigation. A researcher would like to describe the geographic area by aggregating interview responses or observer reports within that area. Following the hierarchical linear modeling terminology, individuals (level-1 units) are nested in sites (level-2 units). For a positive integer n , let $\mathbf{1}_n$ be an $n \times 1$ vector with each element equal to unity. The model we consider is

$$\mathbf{Y} = \gamma \mathbf{1}_N + \mathbf{V}\mathbf{b} + \epsilon \quad (1)$$

$$\mathbf{b} = \rho \mathbf{W}\mathbf{b} + \mathbf{u}, \quad (2)$$

where \mathbf{Y} is an $N \times 1$ outcome vector, γ is a scalar fixed effect, $\mathbf{V} = \bigoplus_{j=1}^J \mathbf{1}_{n_j}$ is an $N \times J$ block diagonal design matrix with the j th element

$\mathbf{1}_{n_j}$, \mathbf{b} is a $J \times 1$ vector of level-2 random spatially autoregressive effects, ϵ is an $N \times 1$ vector of level-1 errors, ρ is a scalar spatial parameter, \mathbf{W} is a $J \times J$ spatial weight matrix, and \mathbf{u} is a $J \times 1$ vector of level-2 errors. Assume $\epsilon \sim N(0, \sigma^2 \mathbf{I}_N)$, $\mathbf{u} \sim N(0, \tau \mathbf{I}_J)$, where \mathbf{u} is independent of ϵ , and σ^2 and τ are scalar level-1 (within-site) and level-2 (between-site) variances, respectively. Thus, \mathbf{V} assigns to each level-2 unit j the appropriate element b_j of the random effects vector \mathbf{b} . The spatial weight matrix \mathbf{W} is defined such that the element $w_{jj'}$ is negatively related to the distance between neighborhoods j and j' . However, the rows of \mathbf{W} may be standardized so that the total distance between neighborhood j and all other neighborhoods is unity. This helps ensure that the spatial dependence parameter ρ will have an absolute value less than unity.

When $\rho \neq 0$, the spatial dependence of the sites is introduced. For example, $\rho > 0$ indicates that a site is typically surrounded by other sites with similar values on the outcome. Thus, a site with a high value of the outcome tends to be surrounded by other similarly high outcome-valued sites, and a site with a low value is typically surrounded by other low outcome-valued sites. On the other hand, $\rho < 0$ indicates that high-value sites are typically surrounded by low-value sites, and vice versa. Finally, $\rho = 0$ indicates no spatial dependence.

Consistent with many other spatial models, the spatial process is specified *a priori* through the spatial weight matrix, \mathbf{W} (Ord 1975; Anselin 1988). In the simplest case, \mathbf{W} is a binary contiguity matrix (or a row-standardized binary contiguity matrix) indicating that sites are contiguous to each other. In the more complex case, the specification of \mathbf{W} can incorporate the distance between the sites, their relative area, the proportion of the common boundary, and other characteristics that are substantively relevant. In the empirical example presented later in this paper, we use the row-standardized binary contiguity matrix where nonzero entries are used for neighborhoods that share a common boundary (i.e., “rook” criterion). Specifically, set $w_{jj'} = 1$ if neighborhood j' is contiguous to neighborhood j while $w_{jj'} = 0$ if not. (By convention, no neighborhood is contiguous to itself.) Set $w_{jj'}^* = w_{jj'} / \sum_{h=1}^J w_{jh}$ so that $\sum_{j'=1}^J w_{jj'}^* = 1$. This induces a first-order Markov process in which the association between first-order neighbors is proportional to ρ , the association between second-order neighbors is proportional to ρ^2 , and so on (Anselin 1988).

As in other spatial models, this model can be written in a “reduced form” (Anselin 2003) by solving equation (2) for \mathbf{b} and

substituting that expression into equation (1):

$$\mathbf{Y} = \gamma \mathbf{1}_N + \mathbf{V}(\mathbf{I}_J - \rho \mathbf{W})^{-1} \mathbf{u} + \epsilon. \quad (3)$$

In contrast to the single-level spatial models, this two-level spatial hierarchical linear model (SHLM) takes into account the hierarchical structure of the data and thus improves the estimation of parameters and provides appropriate standard errors. Note that if $\rho = 0$, then this model reduces to the standard one-way random-effects ANOVA model, $\mathbf{Y} = \gamma \mathbf{1}_N + \mathbf{V}\mathbf{u} + \epsilon$. Also, when $\sigma^2 = 0$ then this model reduces to a linear regression model with a spatially autoregressive disturbance (Anselin 1988), $\mathbf{Y} = \gamma \mathbf{1}_N + \mathbf{V}(\mathbf{I}_J - \rho \mathbf{W})^{-1} \mathbf{u}$.

4.2. Estimation

To estimate the model parameters, we used the maximum likelihood (ML) approach via the expectation-maximization (EM) algorithm (Dempster, Laird, and Rubin 1977). In the maximization step (M-step), parameters were estimated given the complete data (\mathbf{Y} and \mathbf{b}). Since it is not possible to find a closed-form solution for the M-step parameter estimators, a Fisher scoring algorithm was used in the M-step. In the expectation step (E-step), the complete-data sufficient statistics were estimated by their conditional expectations given prior parameter estimates from the M-step and observed data (\mathbf{Y}). The algorithm iterates between the M-Step and the E-Step until the difference in the observed-data log likelihood between two consecutive iterations falls below some specified tolerance level. There are two ways to implement the EM algorithm in this case: Use \mathbf{u} as part of the complete data or use \mathbf{b} as part of the complete data. We found that using \mathbf{b} as part of the complete data results in a faster convergence of the algorithm both in terms of the number of iterations used as well as the time per iteration required. The EM algorithm for this case is presented in Appendix A.

4.3. Shrinkage With and Without Spatial Dependence

The primary interest of this paper, however, is not in estimating the fixed-effects or the variance components of the model, but in the best way of estimating the neighborhood intercepts—that is, $\beta = \gamma \mathbf{1}_J + \mathbf{b}$,

which are also the neighborhood means, with each mean representing the latent social process of interest within that neighborhood. The standard ANOVA model assumes exchangeability of sites. To differentiate the empirical Bayes estimator under the ANOVA model from that under the proposed spatial model, we shall introduce new terminology. We refer to the first as an empirical Bayes exchangeable (EBE) estimator and denote it by $\hat{\beta}_{EBE}$, while the latter one is referred to as an empirical Bayes spatial (EBS) estimator and is denoted by $\hat{\beta}_{EBS}$. The ordinary least squares (OLS) estimator is denoted by $\hat{\beta}_{OLS}$.

The OLS estimator is the observed average of that site:

$$\begin{aligned}\hat{\beta}_{OLSj} &= E[\beta_j | Y_j, \tau^{-1} = 0, \rho = 0] \\ &= \frac{1}{n_j} \sum_{i=1}^{n_j} Y_{ij} = \bar{Y}_{.j}.\end{aligned}\tag{4}$$

Denote the estimated reliability of site j 's OLS estimator under the ANOVA model by $\hat{\lambda}_{Ej} = \hat{\tau}_E / (\hat{\tau}_E + \hat{\sigma}_E^2 / n_j)$, where $\hat{\tau}_E$ and $\hat{\sigma}_E^2$ are the ML estimates of between- and within-site variances, respectively, under the ANOVA model. Then the EBE estimator of the mean for site j is the posterior mean of β_j given observed data, Y_j , and ML estimates of parameters under the ANOVA model, $\hat{\gamma}_E$, $\hat{\tau}_E$, and $\hat{\sigma}_E^2$:

$$\begin{aligned}\hat{\beta}_{EBEj} &= E[\beta_j | Y_j, \gamma = \hat{\gamma}_E, \tau = \hat{\tau}_E, \sigma^2 = \hat{\sigma}_E^2, \rho = 0] \\ &= \hat{\lambda}_{Ej} \bar{Y}_{.j} + (1 - \hat{\lambda}_{Ej}) \hat{\gamma}_E\end{aligned}\tag{5}$$

$$= \hat{\gamma}_E + \hat{\lambda}_{Ej} (\bar{Y}_{.j} - \hat{\gamma}_E).\tag{6}$$

To compare the EBE with the EBS estimator, it is useful to express the former in a matrix notation. Denote $\mathbf{N} = \bigoplus_{j=1}^J n_j$. Then

$$\hat{\mathbf{\Lambda}}_E = \bigoplus_{j=1}^J \hat{\lambda}_{Ej} = \hat{\tau}_E (\hat{\tau}_E \mathbf{I}_J + \hat{\sigma}_E^2 \mathbf{N}^{-1})^{-1}\tag{7}$$

$$\hat{\beta}_{EBE} = \hat{\gamma}_E \mathbf{1}_J + \hat{\mathbf{\Lambda}}_E \begin{bmatrix} \bar{Y}_{.1} - \hat{\gamma}_E \\ \vdots \\ \bar{Y}_{.J} - \hat{\gamma}_E \end{bmatrix}.\tag{8}$$

Denote the reliability of the OLS estimator under the spatial hierarchical linear model Λ_S . Then

$$\hat{\Lambda}_S = \hat{\tau} [\hat{\tau} \mathbf{I}_J + \hat{\sigma}^2 \mathbf{N}^{-1} (\mathbf{I}_J - \hat{\rho} \mathbf{W})^T (\mathbf{I}_J - \hat{\rho} \mathbf{W})]^{-1}, \quad (9)$$

and the empirical Bayes spatial estimator, $\hat{\beta}_{EBS}$, is the posterior mean of β given the observed data and maximum likelihood estimates of the parameters under the spatial hierarchical linear model:

$$\begin{aligned} \hat{\beta}_{EBS} &= E[\beta | \mathbf{Y}, \gamma = \hat{\gamma}, \tau = \hat{\tau}, \sigma^2 = \hat{\sigma}^2, \rho = \hat{\rho}] \\ &= \hat{\gamma} \mathbf{1}_J + \hat{\Lambda}_S \begin{bmatrix} \bar{Y}_{.1} - \hat{\gamma} \\ \vdots \\ \bar{Y}_{.J} - \hat{\gamma} \end{bmatrix}. \end{aligned} \quad (10)$$

Since under the standard ANOVA model an exchangeability of sites is assumed, the OLS estimate of the mean for site j' by itself provides no information about the mean for site j . This is reflected in the reliability of the OLS estimators under the ANOVA model, Λ_E , which is a diagonal matrix. However, under the SHLM, the mean for site j' contains information about the mean for site j due to the spatial dependence in the data. The amount of the information is determined by the structure of the spatial dependence (\mathbf{W}), the strength of the spatial relationship (ρ), as well as the between- and within-site variabilities (τ and σ^2 , respectively).

4.4. Comparing Expected Mean Squared Errors

Similar to the EBE estimator, whose properties were previously reviewed by Efron and Morris (1975), Morris (1983), and Raudenbush (1988), the EBS estimator proposed here is also biased. Therefore, we use the expected mean squared error (EMSE) to compare the performance of the three estimators. The EMSE formulas under the SHLM and large J assumptions are presented below, while their derivation is shown in Appendix B:

$$\begin{aligned} \text{trEMSE}(\hat{\beta}_{EBS}) &= \tau \text{tr}(\mathbf{I}_J - \Lambda_S)^T (\mathbf{I}_J - \Lambda_S) (\mathbf{I}_J - \rho \mathbf{W})^{-1} (\mathbf{I}_J - \rho \mathbf{W})^{-1T} \\ &\quad + \sigma^2 \text{tr} \mathbf{N}^{-1} \Lambda_S^T \Lambda_S, \end{aligned} \quad (11)$$

$$\begin{aligned} \text{trEMSE}(\hat{\beta}_{EBE}) &= \tau \text{tr}(\mathbf{I}_J - \mathbf{\Lambda}_E)(\mathbf{I}_J - \mathbf{\Lambda}_E)(\mathbf{I}_J - \rho \mathbf{W})^{-1}(\mathbf{I}_J - \rho \mathbf{W})^{-1T} \\ &\quad + \sigma^2 \sum_{j=1}^J \lambda_{Ej}^2 / n_j, \end{aligned} \quad (12)$$

$$\text{trEMSE}(\hat{\beta}_{OLS}) = \sigma^2 \text{tr} \mathbf{N}^{-1} = \sigma^2 \sum_{j=1}^J n_j^{-1}. \quad (13)$$

Note that $E[\hat{\sigma}_E^2] = \sigma^2$, while $E[\hat{\tau}_E] = \bar{V} - \bar{C}$, where $\bar{V} = \frac{\tau}{J} \text{tr} \mathbf{M}$, $\bar{C} = \frac{\tau}{J(J-1)} \sum_{j=1}^J \sum_{k \neq j} M_{jk}$, $\mathbf{M} = (\mathbf{I}_J - \rho \mathbf{W})^{-1}(\mathbf{I}_J - \rho \mathbf{W})^{-1T}$, and M_{jk} is a j -th row and k -th column element of $J \times J$ matrix \mathbf{M} .

Consider three cases to see how the EMSE of the three estimators compare. First, if $\rho = 0$ then EBS and EBE estimators are equivalent, as are their respective EMSEs, which are given by

$$\begin{aligned} \text{trEMSE}(\hat{\beta}_{EBS}) &= \text{trEMSE}(\hat{\beta}_{EBE}) \\ &= \tau \sum_{j=1}^J (1 - \lambda_{Ej})^2 + \sigma^2 \sum_{j=1}^J \lambda_{Ej}^2 / n_j \\ &= \sigma^2 \sum_{j=1}^J \lambda_{Ej} / n_j, \end{aligned} \quad (14)$$

where $\lambda_{Ej} = \tau_E / (\tau_E + \sigma_E^2 / n_j)$. Moreover, since $0 \leq \lambda_{Ej} \leq 1$, their EMSE is smaller than that of the OLS estimator and the asymptotic relative efficiency of $\hat{\beta}_{EBS}$ to $\hat{\beta}_{OLS}$ is

$$\bar{\lambda} = \left(\sum_{j=1}^J n_j^{-1} \lambda_{Ej} \right) / \left(\sum_{j=1}^J n_j^{-1} \right). \quad (15)$$

Furthermore, if every site has the same sample size—that is, if $n_j = n$ for all j —then $\lambda_{Ej} = \lambda_E = \tau_E / (\tau_E + \sigma_E^2 / n)$ for all j , and the relative efficiency of the EBS to the OLS estimator is λ_E . A similar result for the EBE estimator in a nonspatial setting was shown in Raudenbush (1988). Second, if $\rho = 0$ and $\tau \gg \sigma^2$, then the three estimators are equivalent as are their respective EMSEs. Finally, in the general case, when $\rho \neq 0$ and $\tau \neq 0$, the asymptotic relative efficiency is given by the ratio of the corresponding EMSEs.

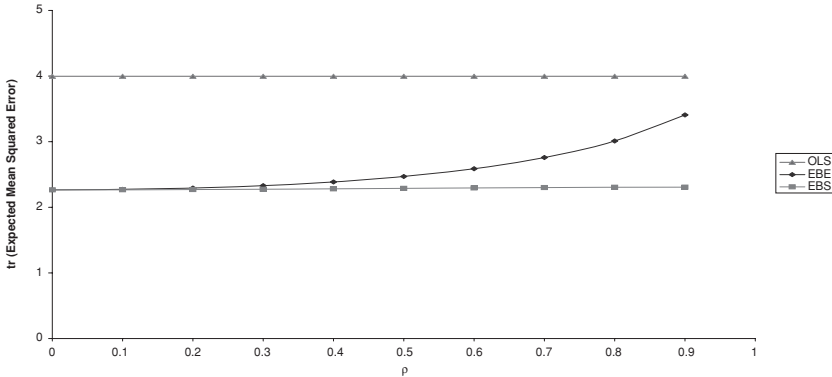


FIGURE 1. Expected mean squared error for the three estimators.

Using the \mathbf{W} and \mathbf{N} from the 1995 PHDCN data, $\tau = \hat{\tau} = 0.01702$ and $\sigma^2 = \hat{\sigma}^2 = 0.21122$ estimated using the model introduced above, and ρ ranging from 0.0 to 0.9 we calculated the three sets of EMSEs. As seen in Figure 1, even though the EMSE for the EBS estimator increases slightly as ρ increases, the EBS strictly dominates the EBE and the OLS estimators.¹ Furthermore, while the EMSE for EBE starts at the same value as that for EBS (at $\rho = 0$), it approaches that of the OLS as ρ increases. Finally, the EMSE for the OLS estimator remains constant (for all values of ρ) and consistently has the largest value out of the three EMSEs. These results held for all other values of τ and σ^2 that were examined as well.

Note that as n_j increases, $\mathbf{\Lambda}_E$ and $\mathbf{\Lambda}_S$ converge to an identity matrix and the three estimators converge. Therefore, it is most beneficial to use the EBS estimator when the site sample size is relatively small. In that case, the EBS estimator borrows information from the surrounding sites in estimating the focal site's mean and improves estimation. It may also prove cost efficient, as a smaller sample size per site may be needed to obtain the same precision.

¹ In statistical decision theory, estimator A strictly dominates estimator B with respect to a loss function if the expected loss of A is always lower than the expected loss of B. Thus, to say that the EBS estimator strictly dominates the EBE estimator with respect to squared error loss is to say that the expected sum of squared errors associated with using EBS is always lower than the expected sum of squared error associated with using EBE.

5. CROSS-VALIDATION STUDY: COLLECTIVE EFFICACY IN CHICAGO NEIGHBORHOODS

In the previous section, we presented evidence suggesting that the EBS estimator outperforms the EBE and OLS estimators with respect to squared error loss under our assumptions. However, failure of the model assumptions may negate these advantages of EBS in real life. In this section, we conduct a cross-validation study using data from two waves of the PHDCN community survey, with wave 1 in 1995 and wave 2 in 2002.

We use four strategies for testing this approach. First, we ask how well the three estimators perform when within-neighborhood sample sizes are small. To do this, we first assess the neighborhoods using the entire sample of 7672 interviews. Given the large sample sizes per neighborhood, we find that the three methods produce very similar estimates. Regarding any one of these as the “gold standard,” we then draw a small random sample from each neighborhood and reassess the neighborhoods based on these small samples. We hypothesize that EBS will more accurately recover the large-sample estimates than will EBE or OLS. In measurement theory, this test is akin to an “internal consistency” reliability analysis.

Second, we assess the temporal stability of the three estimators. A substantial literature suggests that the social processes of interest should be reasonably stable over relatively short periods of time. Indeed, Shaw and McKay (1942) found considerable stability in neighborhood social processes during the industrial expansion of Chicago early in the twentieth century even in the face of neighborhood demographic change because of the ecological niches of these neighborhoods within the geography of the city. Moreover, neighborhood demography is quite stable. Even during 1970–1990, a time of comparatively rapid neighborhood change, Morenoff and Sampson (1997) found substantial stability of neighborhood socioeconomic status, ethnic composition, age, residential mobility, and crime. For these reasons, we expect reasonable continuity in neighborhood collective efficacy between the two waves of the PHDCN neighborhood survey, conducted in 1995 and 2002.

Third, we consider the question of concurrent validity. Past research suggests that collective efficacy is related to neighborhood concentrated disadvantage, residential instability, and concentration of immigration (Sampson et al. 1997). However, if a measure of collective

efficacy is internally inconsistent because of small samples, we would expect an attenuation of this correlation. As Raudenbush and Sampson (1999b) have shown, such attenuation would have unfortunate consequences in studies that attempt to assess the association between collective efficacy and outcomes, controlling for neighborhood demographic structure. So we ask whether EBS as compared to EBE or OLS relieves the attenuation of correlations between collective efficacy and theoretically linked neighborhood demography.

Fourth, we examine the benefits of EBS relative to EBE and OLS in predicting future crime. We rely on past theory and evidence suggesting that neighborhood collective efficacy is strongly related to crime (Sampson et al. 1997) and that crime and neighborhood social processes should be comparatively stable over time, as discussed above. If a measure of collective efficacy is inconsistent or temporally instable, its utility in predicting later crime would be diminished. If EBS produces higher internal consistency and temporal stability than do the other methods, it should also demonstrate higher predictive validity as a result.

None of these four methods provides decisive evidence in favor of one approach over the other. However, if EBS consistently outperforms the others across the four tests and the magnitude of the improvement is consistent with what we might expect from statistical theory under model assumptions, we reason that the resulting web of theory and evidence provides support for the superiority of EBS.

We begin by describing the data used for this study. In PHDCN data, Chicago's 865 census tracts were combined into 343 neighborhood clusters (NCs), such that each NC was as ecologically meaningful as possible, composed of geographically contiguous census tracts, and internally homogeneous on key census indicators. Geographic boundaries (for example, railroad tracks, parks, and freeways) and knowledge of Chicago's neighborhoods guided this process (Sampson et al. 1997; Sampson et al. 1999; Morenoff et al. 2001). Chicago residents representing all 343 NCs were interviewed in their homes as part of the community survey regarding their perception of their neighborhood's characteristics. One such characteristic is collective efficacy (Sampson et al. 1997; Sampson et al. 1999; Morenoff et al. 2001), defined as social cohesion among neighbors combined with their willingness to intervene on behalf of common good. The collective efficacy scale consists of ten items indicating whether people in this neighborhood know each

TABLE 1
Descriptive Statistics

Variable Name	N	Mean	Std. Dev.	Minimum	Maximum
Neighborhood sample size, n_j , 1995	343	22.37	12.14	7.00	60.00
Neighborhood sample size, n_j , 2002	343	9.01	3.87	1.00	21.00
Neighborhood size, n_j , 1995 subsample	343	7.93	4.86	1.00	23.00
Collective efficacy 1995	7672	3.14	0.50	1.30	4.47
Collective efficacy 2002	3090	3.20	0.55	1.31	4.42

other, trust each other, share common values, and can be relied on in various ways to maintain public order. (See Sampson et al. [1997] for details.) Based on the responses to these items, individual scores were estimated. These individual scores are used as the outcome in this paper.

PHDCN data were collected in two waves, the first in 1995 and the second in 2002. In 1995, measures of collective efficacy were available for 7672 respondents residing in 343 NCs. Due to the budget constraints, in 2002, data on collective efficacy perceptions of only 3090 individuals residing in the same 343 NCs were collected. Descriptive statistics for the NC sample size (n_j) as well as individual measures of collective efficacy in 1995 and 2002 are shown in Table 1.

5.1. Goal 1: Estimating Neighborhood Social Process with Small Neighborhood Sample Sizes

In the previous section, we used analytic methods to demonstrate that the EBS estimator outperforms EBE and OLS with respect to expected mean squared error. However, these theoretical results hold under the assumptions of the spatial hierarchical linear model. Here we examine how well this result holds in real life by comparing the performance of the three estimators in estimating the true parameter. When using real data, however, the true parameters are unknown. The 1995 PHDCN data set is quite large, with 7672 individual measures of collective efficacy available, and an average of 22.37 individuals per NC. How do our estimators compare when neighborhood sample sizes are small? We use the 1995 OLS estimates based on the complete sample as the

“gold” standard and estimate these “true” parameters using data from a subsample.

The subsample was drawn with probability of 0.35 of an individual being included. This probability was chosen to ensure a substantial decrease in the data set as well as to ensure that each site had at least one individual in the subsample. The resulting subsample had 2719 individual measures of collective efficacy. The NC sample size ranged from 1 to 23, with a mean of 7.93 (see again Table 1). Using this subsample, we estimated the mean neighborhood collective efficacy via the three methods (EBS, EBE, and OLS) and then computed the sum of squares of errors for each estimator. Consistent with the theoretical results, the EBS estimator had the smallest sum of squares of errors ($SSE = 5.91$), followed by the EBE estimator ($SSE = 7.14$). As expected, the OLS estimator performed worse than the other two ($SSE = 8.89$).

5.2. Goal 2: Examining Temporal Stability

Research cited above suggests that neighborhood social processes such as collective efficacy should be relatively stable over short time intervals. In this section, we compare the three methods in terms of their temporal stability. Given that EBS is less vulnerable than the other methods to inconsistency arising from small samples, we expect it also to display higher temporal stability. We therefore ask: How do the estimators compare when we correlate 1995 and 2002 estimates?

EBS, EBE, and OLS estimates of collective efficacy for each NC were computed in each year (1995 and 2002). Table 2 shows the correlations between them. Note that the correlations between the three 1995 measures are very high (>0.96). For predictive validity, the actual numbers, of course, depend on which 2002 measure is used as the standard. For example, if OLS 02 estimates are used, then the correlations

TABLE 2
Correlations of Estimated Measures of Neighborhood Collective Efficacy

	EBE 95	OLS 95	EBS 02	EBE 02	OLS 02
EBS 95	0.965	0.963	0.763	0.609	0.584
EBE 95		0.996	0.666	0.575	0.554
OLS 95			0.669	0.575	0.558
EBS 02				0.852	0.811
EBE 02					0.966

with EBS 95, EBE 95, and OLS 95 are 0.584, 0.554, and 0.558, respectively; EBS 95 has the highest correlation here. The advantage of EBS 95 measure increases as the correlations with EBE 02 or EBS 02 estimates are examined. Note that the highest correlation (0.763) is between EBS estimates for 1995 and 2002. There does not appear to be any meaningful difference in correlations of EBE or OLS 1995 estimates with the three 2002 estimates, for example $\text{corr}(\text{EBE95}, \text{EBS02}) = 0.666$ and $\text{corr}(\text{OLS95}, \text{EBS02}) = 0.669$. The most plausible explanation for the higher correlations using EBS is that the EBS measures are less vulnerable to measurement error than are the other methods.

5.3. Goal 3: Examining Concurrent Validity

We have found that EBS is less vulnerable to instability as a function of small sample size. We would therefore expect a pay off in terms of estimating correlations between collective efficacy and theoretically linked constructs that were measured concurrently. We concentrate on three demographic measures derived from the U.S. Census that we expect to be correlated with collective efficacy based on theory and research cited above: concentrated disadvantage, immigrant concentration, and residential stability.

The first demographic measure, concentrated disadvantage, is a composite of six factors: percentage below poverty line, percentage on public assistance, percentage in female-headed families, percentage unemployed, percentage less than 18 years of age, and percentage Black. Higher values on concentrated disadvantage indicate poor and more disadvantaged neighborhoods, which may not have the resources to intervene and to improve public services and safety. Therefore, such neighborhoods are expected to have lower collective efficacy. The second demographic measure, immigrant concentration, is a composite of two factors: percentage Latino and percentage foreign-born. A high value of immigrant concentration is indicative of a larger percentage of immigrant residents. In general, recent immigrants come from diverse cultural backgrounds, may have limited English communication skills, and have a lack of knowledge of the resources available to help troubled neighbors. We might thus expect a neighborhood with higher immigrant concentration to have lower collective efficacy (Shaw and McKay 1942; Sampson et al. 1997). The final demographic measure, residential stability, is a composite of two factors: percentage of residents occupying

TABLE 3
Correlations of Estimated Measures of Neighborhood Collective Efficacy with
Relevant Demographic Covariates

Variable Name	Collective Efficacy 1995			Collective Efficacy 2002		
	EBS	EBE	OLS	EBS	EBE	OLS
Concentrated disadvantage	-0.640	-0.613	-0.617	-0.548	-0.456	-0.457
Immigrant concentration	-0.329	-0.318	-0.311	-0.165	-0.145	-0.131
Residential stability	0.414	0.372	0.365	0.431	0.367	0.350

Note: These correlations are based on 342 neighborhoods as the measures for concentrated disadvantage, immigrant concentration, and residential stability were unavailable for one of the NCs.

same house as in 1985 and percentage in owner-occupied house. Higher values on this measure indicate lower mobility from and to the neighborhood. Individuals residing in a neighborhood with high residential stability would thus tend to know each other. Furthermore, owning a house one resides in ensures higher emotional and financial investment in the neighborhood. Therefore, we expect residential stability to be positively associated with collective efficacy.

The results in Table 3 confirm these expectations. Concentrated disadvantage and immigrant concentration are both negatively related to collective efficacy, while residential stability is positively related to collective efficacy. Furthermore, in every case, the EBS measure of collective efficacy correlates more negatively to concentrated disadvantage and immigrant concentration and more positively to residential stability than do the corresponding EBE or OLS measures. For example, the correlation between concentrated disadvantage in 1990 and the EBS measure of collective efficacy in 1995 is -0.640 , while the corresponding correlations with the EBE measure for 1995 and the OLS measure for 1995 are -0.613 and -0.617 , respectively. Note that the EBE measure did not consistently perform better than the OLS measure in 1995, which we believe is due to the large NC sample sizes in 1995 on which these measures were based. For 2002 measures, the EBE measure performs as well as the OLS measure in regards to the relationship with concentrated disadvantage and better in the relationship with the immigrant concentration and residential stability. However, the EBS measure clearly performs better than the other two measures, with 1995 correlation gains ranging from 0.01 to 0.05, while gains in 2002 range from 0.02 to 0.09.

5.4. Goal 4: Predicting Future Crime

If measures based on EBS are less vulnerable to inconsistency arising from small sample sizes and temporal instability, they ought to be more useful in predicting future outcomes. We now examine the relationship between estimates of collective efficacy and crime. Since homicide is generally considered to be the most validly measured crime, we examined correlations of different measures of collective efficacy with homicide rates per 100,000 for 1993, 1995–1998, and 2000–2003. The results for 1993, 1997, 2000, and 2003 are shown in Table 4 as there is an obvious temporal lag between those and the estimates of collective efficacy in 1995 and 2002. However, the results were consistent across all of the years analyzed.

According to substantive theory, collective efficacy and crime are negatively related. Low collective efficacy, due to the unwillingness of the neighbors to intervene on behalf of others, is expected to predict high future crime rates. Moreover, high crime rate tends to undermine social cohesion among neighbors and their willingness to intervene, and thus predicts low future collective efficacy.

As expected, the results showed a negative correlation between measures of collective efficacy and various homicide rates. Moreover, correlations of homicide rates with the EBS measures of collective efficacy were consistently higher (in absolute value) than those with EBE or OLS measures of collective efficacy. For example, the EBS measure of the 1995 collective efficacy was much better at predicting future (2003) homicide rates than the 1995 EBE or OLS measures. The gains for the EBS measures are seen more clearly when examining correlations with 2002 measures of collective efficacy, as the NC sample size for 2002

TABLE 4
Correlations of Estimated Measures of Neighborhood Collective Efficacy with Homicide Rates per 100,000

Variable Name	Collective Efficacy 1995			Collective Efficacy 2002		
	EBS	EBE	OLS	EBS	EBE	OLS
1993 homicide rate	-0.450	-0.413	-0.417	-0.426	-0.371	-0.382
1997 homicide rate	-0.429	-0.396	-0.402	-0.403	-0.338	-0.350
2000 homicide rate	-0.481	-0.459	-0.455	-0.441	-0.403	-0.365
2003 homicide rate	-0.412	-0.380	-0.389	-0.348	-0.265	-0.275

measures is smaller and therefore the EBS measures benefit greater from borrowing strength due to the spatial dependence in the data.

6. DISCUSSION

Valid and reliable measurement of neighborhood social and physical environments is an important challenge in sociology and public health. The standard Bayes or empirical Bayes approaches to this problem borrow strength from the fact that the measurement process in each neighborhood is replicated in many neighborhoods. The mean and variance of the latent variable across these neighborhoods carries useful information about the latent variable in any single neighborhood. However, these approaches have conventionally assumed that the neighborhoods are independent or exchangeable. In this paper, we have investigated whether and to what extent an estimator that exploits the spatial dependence between neighborhoods can add additional information to the measurement of each neighborhood. If so, we reasoned that such an approach may improve the reliability and validity of neighborhood measurement.

Theoretical evidence in favor of this proposition is based on derivation of the expected mean squared error of measurement. Under model assumptions, including a first-order Markov model for spatial dependence and normal theory random effects, the EBS estimator, which exploits spatial dependence in borrowing strength, outperforms the EBE estimator, which assumes neighborhoods to be exchangeable in borrowing strength, and the OLS estimator, which relies solely on the information from each neighborhood in estimating that neighborhood's latent variable. The superiority of EBS is large when spatial dependence is large and when within-neighborhood sample sizes are small.

An empirical cross-validation study gave evidence that the logic of EBS holds up with real data from Chicago, which, of course, may not follow model assumptions to any close approximation. First, we found that EBS was less vulnerable to inconsistency associated with small neighborhood-specific sample sizes and therefore better reproduces large-sample estimates of the latent variable. Second, EBS displayed higher temporal stability. Third, EBS displayed higher construct validity in that it correlated more strongly than did the other methods with theoretically linked variables observed at the same time

using the U.S. Census. Fourth, the more consistent and temporally stable EBS measures also produced a benefit in terms of predicting future crime, demonstrating a potentially important practical advantage.

This work could profitably be extended in two ways. First, it may be worthwhile to conduct simulation studies to check the extent to which known departures from model assumptions degrade the advantages of the approach. Second, it would be useful to investigate the benefits of a fully Bayesian approach to exploiting spatial dependence when the number of neighborhoods is modest to small. As mentioned, we have adopted the empirical Bayes approach, which uses the posterior mean of the latent variable given maximum likelihood estimates (MLE) of model parameters to estimate the true latent variable. This reliance on MLE point estimates is reasonable when the number of neighborhoods is large, as in the case of the PHDCN data. However, researchers will often be interested in data sets that provide fewer neighborhoods. The fully Bayes approach effectively averages the point estimate of the latent variable over all possible values of the model parameters, where each estimate is weighted by the posterior probability of the value of the model parameters. However, care must be taken to assess the sensitivity of inferences to the choice of prior distribution for the variance components when the number of neighborhoods is small (Seltzer, Wong, and Bryk 1996). This may give better point estimates as well as standard errors for the latent variables of interest. The methods for such an approach are provided in Banarjee et al. (2004).

APPENDIX A: THE EM ALGORITHM

This appendix presents computational formulas for the EM algorithm used to estimate parameters for the spatial hierarchical linear model (SHLM) with a scalar spatial parameter, ρ , and a univariate spatial random effect, \mathbf{b} . Note that \mathbf{b} is used as part of the complete data in the M-step of the EM algorithm.

Let $i = 1, 2, \dots, n_j$ denote a level-1 unit (e.g., individual) nested within $j = 1, 2, \dots, J$ level-2 clusters (e.g., neighborhoods), such that $\sum_{j=1}^J n_j = N$. Then a “reduced” form of a two-level SHLM is

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\gamma} + \mathbf{V}(\mathbf{I}_J - \rho\mathbf{W})^{-1}\mathbf{u} + \boldsymbol{\epsilon}, \quad (\text{A1})$$

where \mathbf{Y} is an $N \times 1$ outcome matrix, \mathbf{X} is an $N \times p$ covariate matrix, γ is a $p \times 1$ fixed effects matrix, $\mathbf{V} = \bigoplus_{j=1}^J \mathbf{1}_{n_j}$ is an $N \times J$ level-2 random effects design matrix, $\mathbf{b} = (\mathbf{I}_J - \rho \mathbf{W})^{-1} \mathbf{u}$ is a $J \times 1$ matrix of level-2 random spatially correlated effects, ϵ is an $N \times 1$ matrix of level-1 errors, ρ is a scalar spatial parameter, \mathbf{W} is a $J \times J$ spatial weight matrix, and \mathbf{u} is a $J \times 1$ matrix of level-2 errors. Assume $\epsilon \sim N(0, \sigma^2 \mathbf{I}_N)$, $\mathbf{u} \sim N(0, \tau \mathbf{I}_J)$, where σ^2 and τ are scalar level-1 and level-2 variances, respectively, and $\mathbf{u} \perp \epsilon$.

So, the complete data are \mathbf{Y} , \mathbf{X} , \mathbf{V} , \mathbf{W} , and \mathbf{b} . The observed data are \mathbf{Y} , \mathbf{X} , \mathbf{V} , and \mathbf{W} . Parameters are γ , σ^2 , ρ , τ .

A.1. Maximization Step

To maximize the complete-data likelihood for the parameters, $L(\gamma, \sigma^2, \rho, \tau | \mathbf{Y}, \mathbf{b})$, it is enough to maximize the two conditional probability densities (see Equation A2):

$$\begin{aligned} L(\gamma, \sigma^2, \rho, \tau | \mathbf{Y}, \mathbf{b}) &\propto f(\mathbf{Y}, \mathbf{b} | \gamma, \sigma^2, \rho, \tau) \\ &\propto g(\mathbf{Y} | \mathbf{b}, \gamma, \sigma^2, \rho, \tau) * h(\mathbf{b} | \gamma, \sigma^2, \rho, \tau) \\ &\propto g(\mathbf{Y} | \mathbf{b}, \gamma, \sigma^2) * h(\mathbf{b} | \rho, \tau) \end{aligned} \quad (\text{A2})$$

Since \mathbf{b} is assumed to be known in the M-step, we can rewrite equation (A1) as $\mathbf{Y}^* = \mathbf{Y} - \mathbf{Vb} = \mathbf{X}\gamma + \epsilon$. So, $\mathbf{Y}^* \sim N(\mathbf{X}\gamma, \sigma^2 \mathbf{I}_N)$ and the maximum likelihood estimates (MLE) for γ and σ^2 are

$$\hat{\gamma} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{Y}^* = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T (\mathbf{Y} - \mathbf{Vb}) \quad (\text{A3})$$

and

$$\hat{\sigma}^2 = \frac{1}{N} (\mathbf{Y}^* - \mathbf{X}\hat{\gamma})^T (\mathbf{Y}^* - \mathbf{X}\hat{\gamma}) = \frac{1}{N} (\mathbf{Y} - \mathbf{X}\hat{\gamma} - \mathbf{Vb})^T (\mathbf{Y} - \mathbf{X}\hat{\gamma} - \mathbf{Vb}). \quad (\text{A4})$$

Moreover, since $\mathbf{b} = (\mathbf{I}_J - \rho \mathbf{W})^{-1} \mathbf{u}$ and $\mathbf{u} \sim N(0, \tau \mathbf{I}_J)$, then $\mathbf{b} \sim N(0, (\mathbf{I}_J - \rho \mathbf{W})^{-1} \tau (\mathbf{I}_J - \rho \mathbf{W})^{-1T})$. Hence, the log likelihood for ρ and τ is

$$\begin{aligned} l(\rho, \tau) &= -\frac{J}{2} \ln(2\pi) + \ln |\mathbf{I}_J - \rho \mathbf{W}| - \frac{J}{2} \ln(\tau) \\ &\quad - \frac{1}{2} \tau^{-1} \text{tr}[(\mathbf{I}_J - \rho \mathbf{W})^T (\mathbf{I}_J - \rho \mathbf{W}) \mathbf{b} \mathbf{b}^T]. \end{aligned} \quad (\text{A5})$$

So, the score is

$$\frac{\partial l(\rho, \tau)}{\partial \rho} = -\text{tr}[\mathbf{W}(\mathbf{I}_J - \rho\mathbf{W})^{-1}] + \tau^{-1}\text{tr}[(\mathbf{I}_J - \rho\mathbf{W})^T \mathbf{W}\mathbf{b}\mathbf{b}^T] \quad (\text{A6})$$

$$\frac{\partial l(\rho, \tau)}{\partial \tau} = -\frac{J}{2}\tau^{-1} + \frac{1}{2}\tau^{-2}\text{tr}[(\mathbf{I}_J - \rho\mathbf{W})^T (\mathbf{I}_J - \rho\mathbf{W})\mathbf{b}\mathbf{b}^T] \quad (\text{A7})$$

and the hessian is

$$\frac{\partial^2 l(\rho, \tau)}{\partial \rho^2} = -\tau^{-1}\text{tr}[\mathbf{W}^T \mathbf{W}\mathbf{b}\mathbf{b}^T] - \text{tr}[\mathbf{W}(\mathbf{I}_J - \rho\mathbf{W})^{-1} \mathbf{W}(\mathbf{I}_J - \rho\mathbf{W})^{-1}] \quad (\text{A8})$$

$$\frac{\partial^2 l(\rho, \tau)}{\partial \rho \partial \tau} = -\tau^{-2}\text{tr}[(\mathbf{I}_J - \rho\mathbf{W})^T \mathbf{W}\mathbf{b}\mathbf{b}^T] \quad (\text{A9})$$

$$\frac{\partial^2 l(\rho, \tau)}{\partial \tau^2} = \frac{J}{2}\tau^{-2} - \tau^{-3}\text{tr}[(\mathbf{I}_J - \rho\mathbf{W})^T (\mathbf{I}_J - \rho\mathbf{W})\mathbf{b}\mathbf{b}^T] \quad (\text{A10})$$

So, the complete-data sufficient statistics are \mathbf{b} and $\mathbf{b}\mathbf{b}^T$.

A.2. Expectation Step

In the E-step, complete-data sufficient statistics are estimated given observed data and parameters. Since, $\mathbf{Y} | \mathbf{b} \sim N(\mathbf{X}\gamma + \mathbf{V}\mathbf{b}, \sigma^2 \mathbf{I}_N)$, $\mathbf{b} \sim N(0, (\mathbf{I}_J - \rho\mathbf{W})^{-1} \tau (\mathbf{I}_J - \rho\mathbf{W})^{-1T})$, then $\mathbf{b} | \mathbf{Y} \sim N(\mu, \Sigma)$ and the density of $\mathbf{b} | \mathbf{Y}$ is proportional to joint density of \mathbf{Y} and \mathbf{b} .

$$\begin{aligned} f(\mathbf{b} | \mathbf{Y}) &\propto f(\mathbf{Y}, \mathbf{b}) \propto f(\mathbf{Y} | \mathbf{b}) f(\mathbf{b}) \\ &\propto \exp \left[-\frac{1}{2} \left[\mathbf{b}^T \left(\frac{1}{\sigma^2} \mathbf{V}^T \mathbf{V} + (\mathbf{I}_J - \rho\mathbf{W})^T \tau^{-1} (\mathbf{I}_J - \rho\mathbf{W}) \right) \mathbf{b} \right. \right. \\ &\quad \left. \left. - 2 \frac{1}{\sigma^2} (\mathbf{Y} - \mathbf{X}\gamma)^T \mathbf{V}\mathbf{b} \right] \right] \end{aligned} \quad (\text{A11})$$

So by completing the square,

$$\Sigma = \left(\frac{1}{\sigma^2} \mathbf{V}^T \mathbf{V} + (\mathbf{I}_J - \rho \mathbf{W})^T \tau^{-1} (\mathbf{I}_J - \rho \mathbf{W}) \right)^{-1} \quad (\text{A12})$$

$$\mu^T = \frac{1}{\sigma^2} (\mathbf{Y} - \mathbf{X}\gamma)^T \mathbf{V} \Sigma \quad (\text{A13})$$

$$\mathbf{E}(\mathbf{b}|\mathbf{Y}) = \mu \quad (\text{A14})$$

$$\mathbf{E}(\mathbf{b}\mathbf{b}^T|\mathbf{Y}) = \mu\mu^T + \Sigma. \quad (\text{A15})$$

A.3. Observed-Data Log-Likelihood

Using Bayes' theorem, the observed-data likelihood $L(\mathbf{Y}|\gamma, \sigma^2, \rho, \tau)$ can be written as

$$\begin{aligned} L(\mathbf{Y}|\gamma, \sigma^2, \rho, \tau) &\propto \frac{f(\mathbf{Y}, \mathbf{b}|\gamma, \sigma^2, \rho, \tau)}{f(\mathbf{b}|\mathbf{Y}, \gamma, \sigma^2, \rho, \tau)} \\ &\propto \frac{g(\mathbf{Y}|\mathbf{b}, \gamma, \sigma^2) * h(\mathbf{b}|\rho, \tau)}{f(\mathbf{b}|\mathbf{Y}, \gamma, \sigma^2, \rho, \tau)}, \end{aligned} \quad (\text{A16})$$

where

$$\begin{aligned} g(\mathbf{Y}|\mathbf{b}, \gamma, \sigma^2) &= (2\pi)^{-\frac{N}{2}} (\sigma^2)^{-\frac{N}{2}} \\ &\quad \exp \left[-\frac{1}{2\sigma^2} (\mathbf{Y} - \mathbf{X}\gamma - \mathbf{V}\mathbf{b})^T (\mathbf{Y} - \mathbf{X}\gamma - \mathbf{V}\mathbf{b}) \right] \\ h(\mathbf{b}|\rho, \tau) &= (2\pi)^{-\frac{J}{2}} |(\mathbf{I}_J - \rho \mathbf{W})^{-1} \tau (\mathbf{I}_J - \rho \mathbf{W})^{-1T}|^{-\frac{1}{2}} \\ &\quad \exp \left[-\frac{1}{2} \mathbf{b}^T (\mathbf{I}_J - \rho \mathbf{W})^T \tau^{-1} (\mathbf{I}_J - \rho \mathbf{W}) \mathbf{b} \right] \\ f(\mathbf{b}|\mathbf{Y}, \gamma, \sigma^2, \rho, \tau) &= (2\pi)^{-\frac{J}{2}} |\Sigma|^{-\frac{1}{2}} \exp \left[-\frac{1}{2} (\mathbf{b} - \mu)^T \Sigma^{-1} (\mathbf{b} - \mu) \right] \end{aligned}$$

μ and Σ as in the E-step. So, the observed-data log-likelihood l^* is

$$\begin{aligned} l^* &= -\frac{N}{2} \ln(2\pi) - \frac{N}{2} \ln(\sigma^2) - \frac{1}{2} \left(-2 * \ln |(\mathbf{I}_J - \rho \mathbf{W})| + J * \ln \tau \right) \\ &\quad + \frac{1}{2} \ln |\Sigma| - \frac{1}{2\sigma^2} (\mathbf{Y} - \mathbf{X}\gamma)^T (\mathbf{Y} - \mathbf{X}\gamma - \mathbf{V}\mu) \end{aligned} \quad (\text{A17})$$

APPENDIX B: EXPECTED MEAN SQUARED ERROR
DERIVATION

Denote mean squared error as MSE and expected mean squared error as EMSE.

$$\text{EMSE}(\hat{\beta}) = \text{E}[\text{MSE}(\hat{\beta}|\beta)] \quad (\text{B1})$$

$$\begin{aligned} \text{MSE}(\hat{\beta}|\beta) &= \text{E}[(\hat{\beta} - \beta)(\hat{\beta} - \beta)^T|\beta] \\ &= (\text{Bias}[\hat{\beta}|\beta])(\text{Bias}[\hat{\beta}|\beta])^T + \text{Var}[\hat{\beta}|\beta] \end{aligned} \quad (\text{B2})$$

B.1. *EMSE of EBS*

$$\text{E}[\hat{\beta}_{EBS}|\beta] = \text{E}[\Lambda_S \bar{\mathbf{y}} + (\mathbf{I}_J - \Lambda_S)\gamma|\beta] = \Lambda_S \beta + (\mathbf{I}_J - \Lambda_S)\gamma \quad (\text{B3})$$

$$\text{Bias}[\hat{\beta}_{EBS}|\beta] = \text{E}[\hat{\beta}_{EBS}|\beta] - \beta = -(\mathbf{I}_J - \Lambda_S)(\beta - \gamma \mathbf{1}_J) \quad (\text{B4})$$

$$\begin{aligned} \text{Var}[\hat{\beta}_{EBS}|\beta] &= \text{E}[(\Lambda_S \bar{\mathbf{y}} + (\mathbf{I}_J - \Lambda_S)\gamma - \Lambda_S \beta - (\mathbf{I}_J - \Lambda_S)\gamma) \\ &\quad (\Lambda_S \bar{\mathbf{y}} + (\mathbf{I}_J - \Lambda_S)\gamma - \Lambda_S \beta - (\mathbf{I}_J - \Lambda_S)\gamma)^T|\beta] \\ &= \text{E}[(\Lambda_S \bar{\mathbf{y}} - \Lambda_S \beta)(\Lambda_S \bar{\mathbf{y}} - \Lambda_S \beta)^T|\beta] \\ &= \Lambda_S \text{Var}[\bar{\mathbf{y}}|\beta] \Lambda_S^T \\ &= \sigma^2 \Lambda_S \mathbf{N}^{-1} \Lambda_S^T \end{aligned} \quad (\text{B5})$$

where $\mathbf{N} = \text{diag}[n_1, n_2, \dots, n_J]$ is a $J \times J$ diagonal matrix.

$$\begin{aligned} \text{trEMSE}(\hat{\beta}_{EBS}) &= \text{trE}[\text{Bias}[\hat{\beta}_{EBS}|\beta]\text{Bias}[\hat{\beta}_{EBS}|\beta]^T] \\ &\quad + \text{trE}[\text{Var}[\hat{\beta}_{EBS}|\beta]] \\ &= \text{trE}[(\mathbf{I}_J - \Lambda_S)(\beta - \gamma \mathbf{1}_J)(\beta - \gamma \mathbf{1}_J)^T(\mathbf{I}_J - \Lambda_S)^T] \\ &\quad + \text{tr}\sigma^2 \Lambda_S \mathbf{N}^{-1} \Lambda_S^T \\ &= \tau \text{tr}(\mathbf{I}_J - \Lambda_S)^T (\mathbf{I}_J - \Lambda_S)(\mathbf{I}_J - \rho \mathbf{W})^{-1} (\mathbf{I}_J - \rho \mathbf{W})^{-1T} \\ &\quad + \sigma^2 \text{tr} \mathbf{N}^{-1} \Lambda_S^T \Lambda_S \end{aligned} \quad (\text{B6})$$

B.2. EMSE of EBE

Similarly,

$$E[\hat{\beta}_{EBE}|\beta] = E[\Lambda_E \bar{\mathbf{y}} + (\mathbf{I}_J - \Lambda_E)\gamma|\beta] = \Lambda_E \beta + (\mathbf{I}_J - \Lambda_E)\gamma \quad (\text{B7})$$

$$\text{Bias}[\hat{\beta}_{EBE}|\beta] = E[\hat{\beta}_{EBE}|\beta] - \beta = -(\mathbf{I}_J - \Lambda_E)(\beta - \gamma \mathbf{1}_J) \quad (\text{B8})$$

$$\begin{aligned} \text{Var}[\hat{\beta}_{EBE}|\beta] &= E[(\Lambda_E \bar{\mathbf{y}} + (\mathbf{I}_J - \Lambda_E)\gamma - \Lambda_E \beta - (\mathbf{I}_J - \Lambda_E)\gamma) \\ &\quad (\Lambda_E \bar{\mathbf{y}} + (\mathbf{I}_J - \Lambda_E)\gamma - \Lambda_E \beta - (\mathbf{I}_J - \Lambda_E)\gamma)^T|\beta] \quad (\text{B9}) \end{aligned}$$

$$\begin{aligned} &= \Lambda_E \text{Var}[\bar{\mathbf{y}}|\beta] \Lambda_E^T \\ &= \sigma^2 \Lambda_E \mathbf{N}^{-1} \Lambda_E^T \quad (\text{B10}) \end{aligned}$$

So,

$$\begin{aligned} \text{trEMSE}(\hat{\beta}_{EBE}) &= \text{tr}E[\text{Bias}[\hat{\beta}_{EBE}|\beta]\text{Bias}[\hat{\beta}_{EBE}|\beta]^T] \\ &\quad + \text{tr}E[\text{Var}[\hat{\beta}_{EBE}|\beta]] \\ &= \text{tr}E[(\mathbf{I}_J - \Lambda_E)(\beta - \gamma \mathbf{1}_J)(\beta - \gamma \mathbf{1}_J)^T(\mathbf{I}_J - \Lambda_E)^T \\ &\quad + \text{tr}\sigma^2 \Lambda_E \mathbf{N}^{-1} \Lambda_E^T] \\ &= \tau \text{tr}[(\mathbf{I}_J - \Lambda_E)(\mathbf{I}_J - \Lambda_E)(\mathbf{I}_J - \rho \mathbf{W})^{-1}(\mathbf{I}_J - \rho \mathbf{W})^{-1T} \\ &\quad + \sigma^2 \sum_{j=1}^J \lambda_{Ej}^2/n_j] \quad (\text{B11}) \end{aligned}$$

B.3. EMSE of OLS

$$\text{Bias}[\hat{\beta}_{OLS}|\beta] = 0 \quad (\text{B12})$$

$$\text{Var}[\hat{\beta}_{OLS}|\beta] = \sigma^2 \mathbf{N}^{-1} \quad (\text{B13})$$

$$\text{trEMSE}[\hat{\beta}_{OLS}] = \text{tr}\sigma^2 \mathbf{N}^{-1} = \sigma^2 \sum_{j=1}^J n_j^{-1} \quad (\text{B14})$$

REFERENCES

- Anselin, Luc. 1988. *Spatial Econometrics: Methods and Models (Studies in Operational Regional Science)*. Dordrecht, Netherlands: Kluwer.
- . 2003. "Spatial Externalities, Spatial Multipliers, and Spatial Econometrics." *International Regional Science Review* 26:153–66.
- Assunção, Renato M., Carl P. Schmertmann, Joseph E. Potter, and Suzana M. Cavenaghi. 2005. "Empirical Bayes Estimation of Demographic Schedules for Small Areas." *Demography* 42:537–58.
- Banerjee, Sudipto, Bradley P. Carlin, and Alan E. Gelfand. 2004. *Hierarchical Modeling and Analysis for Spatial Data*. Boca Raton, FL: Chapman and Hall.
- Besag, Julian. 1974. "Spatial Interaction and the Statistical Analysis of Lattice Systems." *Journal of the Royal Statistical Society, Series B*, 36:192–236.
- Bingenheimer, Jeffrey B., and Stephen W. Raudenbush. 2004. "Statistical and Substantive Inferences in Public Health: Issues in the Application of Multilevel Models." *Annual Review of Public Health* 25:53–77.
- Britt, Heather R., Bradley P. Carlin, Traci L. Toomey, and Alexander C. Wagenaar. 2005. "Neighborhood Level Spatial Analysis of the Relationship Between Alcohol Outlet Density and Criminal Violence." *Environmental and Ecological Statistics* 12:411–26.
- Browning, Christopher R. 2002. "The Span of Collective Efficacy: Extending Social Disorganization Theory to Partner Violence." *Journal of Marriage and the Family* 64(4):833–50.
- Browning, Christopher R., Tama Leventhal, and Jeanne Brooks-Gunn. 2005. "Sexual Initiation in Early Adolescence: The Nexus of Parental and Community Control." *American Sociological Review* 70(5):758–78.
- Buka, Stephen L., Robert T. Brennan, Janet W. Rich-Edwards, Stephen W. Raudenbush, and Felton Earls. 2003. "Neighborhood Support and the Birth Weight of Urban Infants." *American Journal of Epidemiology* 157(1):1–8.
- Carter, Grace M., and John E. Rolph. 1974. "Empirical Bayes Methods Applied to Estimating Fire Alarm Probabilities." *Journal of the American Statistical Association* 69:880–85.
- Clayton, David, and John Kaldor. 1987. "Empirical Bayes Estimates of Age-Standardized Relative Risks for Use in Disease Mapping." *Biometrics* 43:671–81.
- Dempster, A. P., N. M. Laird, and D. B. Rubin. 1977. "Maximum Likelihood from Incomplete Data Via the EM Algorithm." *Journal of the Royal Statistical Society, Series B*, 39:1–38.
- Diez-Roux, Ana V. 2004. "The Study of Group-Level Factors in Epidemiology: Rethinking Variables, Study Designs, and Analytical Approaches." *Epidemiologic Reviews* 26:104–11.
- Duncan, Greg J., and Stephen W. Raudenbush. 1999. "Assessing the Effects of Context in Studies of Child and Youth Development." *Educational Psychologist* 34(1):29–41.

- Efron, Bradley, and Carl Morris. 1972a. "Empirical Bayes on Vector Observations: An Extension of Stein's Method." *Biometrika* 59:335–47.
- . 1972b. "Limiting the Risk of Bayes and Empirical Bayes Estimators—Part II: The Empirical Bayes Case." *Journal of the American Statistical Association* 67:130–39.
- . 1973. "Stein's Estimation Rule and Its Competitors—An Empirical Bayes Approach." *Journal of the American Statistical Association* 68:117–30.
- . 1975. "Data Analysis Using Stein's Estimator and Its Generalizations." *Journal of the American Statistical Association* 70:311–19.
- . 1977. "Stein's Paradox in Statistics." *Scientific American* 36:119–27.
- Galster, George, Dave E. Marcotte, Marv Mandell, Hal Wolman, and Nancy Augustine. 2007. "The Influence of Neighborhood Poverty During Childhood on Fertility, Education, and Earnings Outcomes." *Housing Studies* 22(5):723–51.
- Goldstein, Harvey. 2002. *Multilevel Statistical Models*. 3rd ed. London, UK: A Hodder Arnold Publication.
- Jackson, Paul H., Melvin R. Novick, and Dorothy T. Thayer. 1971. "Estimating Regressions in m Groups." *British Journal of Mathematical and Statistical Psychology* 24:129–53.
- James, W., and Charles Stein. 1961. "Estimation with Quadratic Loss." *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability* 1:361–79.
- Leventhal, Tama, and Jeanne Brooks-Gunn. 2000. "The Neighborhoods They Live in: The Effects of Neighborhood Residence on Child and Adolescent Outcomes." *Psychological Bulletin* 126:309–37.
- Lindley, D. V. 1971. "The Estimation of Many Parameters." Pp. 435–55 in *Foundations of Statistical Inference*, edited by V. P. Godambe and D. A. Sprott. Toronto: Holt, Rinehart and Winston of Canada.
- Longford, Nicholas T. 1999. "Multivariate Shrinkage Estimation of Small Area Means and Proportions." *Journal of the Royal Statistical Society, Series A*, 162(2):227–45.
- Marshall, Roger J. 1991. "Mapping Disease and Mortality Rates Using Empirical Bayes Estimators." *Applied Statistics* 40:283–94.
- McCrea, Rod, Tung-Kai Shyy, John Western, and Robert J. Stimson. 2005. "Fear of Crime in Brisbane: Individual, Social and Neighborhood Factors in Perspective." *Journal of Sociology* 41(1):7–27.
- Molnar, Beth E., Steven L. Gortmaker, Fiona C. Bull, and Stephen L. Buka. 2004. "Unsafe to Play? Neighborhood Disorder and Lack of Safety Predict Reduced Physical Activity Among Urban Children and Adolescents." *American Journal of Health Promotion* 18(5):378–86.
- Morenoff, Jeffrey D. 2003. "Neighborhood Mechanisms and the Spatial Dynamics of Birth Weight." *American Journal of Sociology* 108(5):976–1017.
- Morenoff, Jeffrey D., and Robert J. Sampson. 1997. "Violent Crime and the Spatial Dynamics of Neighborhood Transition: Chicago 1970–1990." *Social Forces* 76(1):31–64.

- Morenoff, Jeffrey D., Robert J. Sampson, and Stephen W. Raudenbush. 2001. "Neighborhood Inequality, Collective Efficacy, and the Spatial Dynamics of Urban Violence." *Criminology* 39(3):517–59.
- Morris, Carl N. 1983. "Parametric Empirical Bayes Inference: Theory and Applications." *Journal of the American Statistical Association* 78(381):47–55.
- Novick, Melvin R., Paul H. Jackson, Dorothy T. Thayer, and Nancy S. Cole. 1972. "Estimating Multiple Regressions in m Groups: A Cross-Validation Study." *British Journal of Mathematical and Statistical Psychology* 25:33–50.
- Oakes, J. Michael. 2004. "The (Mis)estimation of Neighborhood Effects: Causal Inference for a Practicable Social Epidemiology." *Social Science and Medicine* 58(10):1929–52.
- Ord, Keith. 1975. "Estimation Methods for Models of Spatial Interaction." *Journal of the American Statistical Association* 70:120–26.
- Raudenbush, Stephen W. 1988. "Educational Applications of Hierarchical Linear Models: A Review." *Journal of Educational Statistics* 13:85–116.
- Raudenbush, Stephen W., and Anthony S. Bryk. 2002. *Hierarchical Linear Models: Applications and Data Analysis Methods*. 2nd ed. Thousand Oaks, CA: Sage.
- Raudenbush, Stephen W., Brian Rowan, and Sang Jin Kang. 1991. "A Multilevel, Multivariate Model for Studying School Climate in Secondary Schools with Estimation via the EM Algorithm." *Journal of Educational Statistics* 16(4):295–330.
- Raudenbush, Stephen W., and Robert J. Sampson. 1999a. "Ecometrics: Toward a Science of Assessing Ecological Settings, with Application to the Systematic Social Observation of Neighborhoods." *Sociological Methodology* 29:1–41.
- . 1999b. "Assessing Direct and Indirect Effects in Multilevel Designs with Latent Variables." *Sociological Methods & Research* 28(2):123–53.
- Sampson, Robert J., Jeffrey D. Morenoff, and Felton Earls. 1999. "Beyond Social Capital: Spatial Dynamics of Collective Efficacy for Children." *American Sociological Review* 64:633–60.
- Sampson, Robert J., and Stephen W. Raudenbush. 1999. "Systematic Social Observation of Public Spaces: A New Look at Disorder in Urban Neighborhoods." *American Journal of Sociology* 105:603–51.
- . 2003. "Seeing Disorder: Neighborhood Stigma and the Social Construction of 'Broken Windows'." *Social Psychology Quarterly* 67(4):319–42.
- Sampson, Robert J., Stephen W. Raudenbush, and Felton Earls. 1997. "Neighborhoods and Violent Crime: A Multilevel Study of Collective Efficacy." *Science* 277:918–24.
- Seltzer, Michael, Wing Hung Wong, and Anthony S. Bryk. 1996. "Bayesian Analysis in the Applications of Hierarchical Models: Issues and Methods." *Journal of Educational and Behavioral Statistics* 21:131–67.
- Shaw, Clifford, and Henry D. McKay. 1942. *Juvenile Delinquency and Urban Areas: A Study of Delinquents in Relation to Differential Characteristics of Local Communities*. Chicago, IL: University of Chicago Press.
- Stein, Charles. 1956. "Inadmissibility of the Usual Estimator for the Mean of a Multivariate Normal Distribution." *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability* 1:197–206.

- Tsutakawa, Robert K., Gary L. Shoop, and Carl J. Marienfeld. 1985. "Empirical Bayes Estimation of Cancer Mortality Rates." *Statistics in Medicine* 4:201–12.
- Verbitsky, Natalya. 2007. "Associational and Causal Inference in Spatial Hierarchical Settings : Theory and Applications." PhD dissertation, Department of Statistics, University of Michigan, Ann Arbor, MI.
- Zubrick, Stephen R. 2007. "Commentary: Area Social Cohesion, Deprivation and Mental Health—Does Misery Love Company?" *International Journal of Epidemiology* 36(2):345–47.