EMPIRICAL BAYES ESTIMATION OF DEMOGRAPHIC SCHEDULES FOR SMALL AREAS*

RENATO M. ASSUNÇÃO, CARL P. SCHMERTMANN, JOSEPH E. POTTER, AND SUZANA M. CAVENAGHI

In this article, we analyze empirical Bayes (EB) methods for estimating small-area rate schedules. We develop EB methods that treat schedules as vectors and use adaptive neighborhoods to keep estimates appropriately local. This method estimates demographic rates for local subpopulations by borrowing strength not only from similar individuals elsewhere but also from other groups in the same area and from regularities in schedules across locations. EB is substantially better than standard methods when rates have strong spatial and age patterns. We illustrate this method with estimates of age-specific fertility schedules for over 3,800 Brazilian municipalities.

Accurate estimation of rate schedules for small areas is an essential part of many demographic analyses, and it has become more important as demographers have gained greater access to geocoded data. However, even with very large samples and censuses, small areas often have small risk populations that produce unstable estimates. This is particularly true when local populations are further disaggregated by age. With common estimation techniques, the least-populated areas often produce extreme estimates, dominated by sampling noise, that may have little relationship to underlying local risks (Bernadinelli and Montomoli 1992). Over a large number of small areas, one may observe high variability in estimated rates that poorly reflects the true level of geographic heterogeneity.

One way to deal with this problem is through empirical Bayes (EB) techniques, also known as "shrinkage" methods. EB applications span a variety of disciplines, including epidemiology and public health (Clayton and Kaldor 1987; Manton et al. 1989; Marshall 1991), digital image processing (Lee 1980), genetics (Efron et al. 2001; Kendziorski et al. 2003; Kitada, Hayashi, and Kishino 2000; Newton et al. 2001), econometrics (Kamakura and Wedel 2004; Lamm-Tennant, Starks, and Stokes 1992), and geology (Solow 2001). There have been interesting uses of EB techniques in demography, notably from the U.S. Census Bureau's SAIPE program for small-area income estimation (e.g., Fay and Herriot 1979; Fay and Train 1995), but also in other areas (Ferguson et al. 2004; Longford 1999; Stolzenberg and Relles 1989).

EB methods improve estimates by "borrowing strength" in various ways across related observational units. They are especially well suited to problems in which an analyst wants to estimate many similar parameters from many small samples. In this article, we pursue one example of such a problem: estimation of age-specific fertility schedules for over 3,800 small areas in Brazil.

Two developments in small-area EB estimation are important for our analysis. First, Marshall (1991) proposed a spatial approach to shrinkage that uses adaptively defined,

*Renato M. Assunção, Department of Statistics, Universidade Federal de Minas Gerais, Brazil. Carl P. Schmertmann, Center for Demography and Population Health, Florida State University. Joseph E. Potter, Population Research Center, University of Texas. Suzana M. Cavenaghi, National School of Statistical Sciences (ENCE/IBGE), Brazil. Address correspondence to Carl P. Schmertmann, Center for Demography and Population Health, Florida State University, Tallahassee, FL 32306-2240; E-mail: schmertmann@fsu.edu. We gratefully acknowledge support from NIH research Grant R01HD041528, and from CNPq—Conselho Nacional de Desenvolvimento Científico e Tecnológico, Brazil.

moving neighborhoods. Marshall's technique borrows strength selectively, using information from nearby locations more intensively than information from distant, presumably less-similar, areas. The moving-neighborhood approach improves demographic estimates when rates vary substantially across the map, but at scales larger than the small areas under study. Second, Longford (1999) developed an EB method that treats sets of related local rates as vectors of parameters to be estimated simultaneously, rather than element by element in piecewise fashion. Longford's method improves estimates by borrowing strength from related rates within small areas; for example, if fertility rates in age groups 20–24 and 25–29 (f_{20-24} and f_{25-29}, respectively) have a strong positive correlation across locations, then a high sample estimate for f_{20-24} in area a provides information that the true value of f_{25-29} in a is likely to be higher than average. Longford's approach exploits such cross-component correlations to improve the overall accuracy of a large set of vector estimates.

We propose a general method for estimating small-area rate schedules that uses both Marshall's moving neighborhoods and Longford's vector approach. We then investigate the performance of this estimator and of several variants using census fertility data from Brazilian municipalities. The procedures that we describe are not limited to the Brazilian context or to fertility estimation. The ideas of treating small-area rate schedules as sets of parameters and of using spatial relationships to increase the signal-to-noise ratio in estimated schedules are relevant for many problems in demography.

This article proceeds as follows. We first introduce EB by means of an intuitive example and then explain our synthesis of the Marshall and Longford approaches. Next, we describe microdata on current fertility in the 1991 Brazilian census and present examples from EB estimation of small-area fertility schedules. Finally, we assess the relative performance of several variants of EB methods.

EB (SHRINKAGE) ESTIMATION

Because Bayesian inference is unfamiliar to many demographers, we begin with some basics. A *fully Bayesian* approach to parameter estimation has four steps:

1. Specify a *likelihood model* $L(x|\theta)$ for data x, given parameters θ;
2. Specify a *prior distribution* $f(\theta)$ for the parameters, which are treated as random variables (in a small-area estimation problem, one can think of $f(\theta)$ as the distribution of parameters across areas);
3. Calculate a *posterior distribution* $P(\theta|x)$ that describes the probabilities of different parameter values, conditional on observed data, using Bayes' Theorem:

$$P(\theta|x) \propto L(x|\theta) \cdot f(\theta); \text{ and}$$

4. Draw inferences about the parameters from the posterior distribution.

An EB approach is similar, but it uses observed data to infer important features of the prior distribution, rather than specifying it fully in Step 2. Carlin and Louis (2000) discussed and compared the two approaches thoroughly.

EB estimation is a valuable statistical tool when the analyst must estimate many parameters from separate samples but each sample comes from some fundamentally similar process. The main ideas behind EB "shrinkage" are to examine individual estimates in the context of their overall distribution and to recognize that outliers are more likely to contain unusual sampling errors. For example, the lowest estimates in a set are the most likely to be underestimates, as in the introductory example that follows. Thus, even when individual parameters are seemingly unrelated, one can improve estimation by considering their distribution.

An Introductory Example

As an instructive example, suppose that there are many areas $a = 1 \ldots A$. For some demographic process of interest, area-specific rates are $\theta_1 \ldots \theta_A$. To keep the example

simple, assume that across areas, the θ values are independent, with a uniform distribution on the interval [0,2μ] and a mean of μ. This is a prior distribution for parameters, as in Step 2: rates vary from place to place, they are spread evenly between 0 and 2μ, and the rate in one location is unrelated to rates elsewhere.

Now suppose that the researcher does not observe θ_a directly, but instead has local estimates from samples of n individuals in each area. For reasons that we elaborate later, we model the number of events in the past year to these sampled individuals (x_a) as a Poisson random variable:

$$x_a \left| \theta_a \right. \sim \text{Poisson}\left(n\theta_a\right).$$

This defines a likelihood function, $L(x|\theta)$, as in Step 1. Notice that in this model, x and θ have a joint distribution across areas, with some (x,θ) pairs more likely than others. The maximum likelihood (ML) estimator identifies the most likely θ value, given the observed x_a:

$$\hat{\theta}_a = x_a / n,$$

which also corresponds to events in a per person-year.

Because $\hat{\theta}_a$ is an unbiased estimator of θ_a within each area, true parameters θ and ML estimates $\hat{\theta}$ have identical means when averaged over all areas:[1]

$$E\left(\hat{\theta}_a\right) = E_\theta\left[E_x\left(\hat{\theta}_a \left| \theta_a \right.\right)\right] = E_\theta\left[\theta_a\right] = \mu.$$

ML estimates have other good statistical properties, such as efficiency, but they can be further improved, as we explain next.

To estimate true rates $\{\theta_1 \ldots \theta_A\}$ from estimates $\{\hat{\theta}_1 \ldots \hat{\theta}_A\}$, consider the conditional distribution of $(\theta \mid x)$ or equivalently $(\theta \mid \hat{\theta})$. That is, consider the possible values and relative frequencies of various *true* rates θ within a subset of areas that all have identical *estimated* rates $\hat{\theta}$. This question has a Bayesian flavor, because it involves posterior probabilities

$$P\left(\theta \left| x \right.\right) \propto L\left(x \left| \theta \right.\right) \cdot f\left(\theta\right)$$

that depend on the prior distribution of θ.

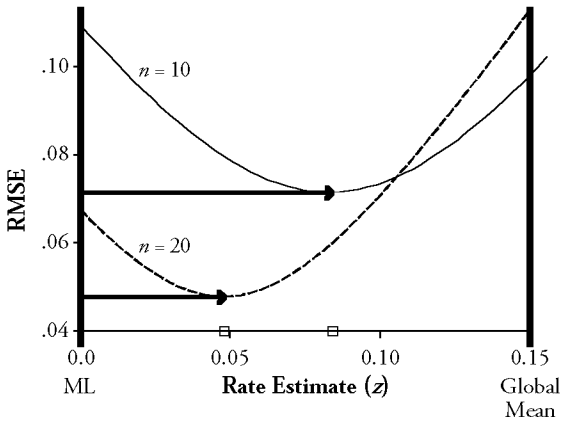
As a specific example, if there were no sample events recorded in the past year in area 6, then $x_6 = 0$ and $\hat{\theta}_6 = 0$. Many possible values of θ_6 could lead to local samples with $x_6 = 0$, but the ML estimate of 0 is almost certainly too low. A common criterion for selecting from alternative estimators is mean squared error (MSE), often expressed in terms of its positive square root (RMSE). If we used a rate estimate z in all areas with $\hat{\theta}_a = 0$, its RMSE would be

$$RMSE(z) = +\sqrt{E_\theta\left[\left(\theta_a - z\right)^2 \left| x = 0 \right.\right]}.$$

---

1. Throughout this article, $E(\ )$ and $V(\ )$ denote expectations and variances over the joint distribution of (x,θ). There are several ways to interpret this joint distribution, but in the small-area estimation setting it is useful to think of $E[g(\mathbf{x},\boldsymbol{\theta})]$ as an answer to the question, If we select an area at random (all areas equally likely) and then select a sample from that area, what is the expected value of the function g? For conditional moments, we add subscripts. For example, $E_x(\hat{\boldsymbol{\theta}} \mid \boldsymbol{\theta})$ is the expected value of the ML estimator $\hat{\boldsymbol{\theta}}$ in areas with a specific value of $\boldsymbol{\theta}$, averaged over possible sample values of $\mathbf{x}$. Similarly $E_\theta[V_x(\hat{\boldsymbol{\theta}} \mid \boldsymbol{\theta})]$ is the expected sample variance of the ML estimator, averaged over possible values of $\boldsymbol{\theta}$.

Figure 1. RMSE of Estimators z in Areas With $x / n = 0$



In our model, $P(\theta_a \mid x_a = 0) \propto e^{-n\theta_a}$ for θ values between 0 and 2μ.[2] If we take $\mu = .15$ as an example, calculating RMSEs produces the information in Figure 1. Figure 1 illustrates several important points. First, the ML estimator (represented by the vertical line at $z = 0$) does not minimize MSE. The solid curve shows that for areas in which 10 sampled people all report zero events, estimates of $z = .084$ (not zero) will minimize the average squared difference from true local rates. Second, the minimum-MSE estimate is a shrinkage estimate: the ML estimate pulled partway toward the global mean μ. Third, the optimal amount of shrinkage depends on the local sample size. In areas with $\hat{\theta}_a = 0$ and $n = 20$ (dashed curve), the minimum-MSE estimator is .049, considerably farther from the global mean than the optimum when $\hat{\theta}_a = 0$ and $n = 10$. This makes intuitive sense: when local samples are larger, local estimates are more informative, and less shrinkage is necessary to minimize MSE.

Finding the optimal estimate requires knowledge of the marginal distribution of true rates θ across areas. In our simple example, this is equivalent to knowing μ. In a fully Bayesian analysis, parameters like μ are chosen in advance when the researcher specifies the prior distribution of θ. Minimum-MSE estimates of quantities such as θ_6 then depend on the prior distribution chosen by the researcher.

In EB analysis, the researcher does not specify the prior distribution of θ. Instead, she or he uses the empirical (marginal) distribution of estimates $\{\hat{\theta}_1 \ldots \hat{\theta}_A\}$ to *estimate* relevant characteristics of $f(\theta)$. Our simple example has an obvious estimator for the single distributional parameter, namely, $\hat{\mu} = \Sigma \hat{\theta}_a / A$. Given this empirical estimate, the researcher can approximate the distribution of $(\theta \mid \hat{\theta})$ and construct (approximately) minimum-MSE estimators. For overall means $\hat{\mu} = .05$, .15, or .25 in our example, the minimum-MSE estimator when $\hat{\theta}_6 = 0$ and $n = 10$ would be .042, .084, or .097, respectively. That is, a higher global mean for the ML estimates implies higher EB estimates in areas with $\hat{\theta} = 0$. In this manner, the distribution of estimates across all areas informs the estimate in each particular location, changing a local estimate more when it is farther from the global mean.

2. This equation arises from Bayes' Theorem, using a uniform prior distribution of θ_a over $[0,2\mu]$ and the assumed Poisson likelihood $L(x_a = 0 \mid \theta_a) = e^{-n\theta_a}$.

EB Estimation in the General Scalar Case

We simplified the preceding example by assuming particular forms for the distributions of local parameters (θ_a) and estimators ($\hat{\theta}_a$), but EB estimation remains possible with much weaker mathematical assumptions. For any distribution of (x, θ) in which local estimators are unbiased [$E_x(\hat{\theta}_a \mid \theta_a) = \theta_a$], the minimum-MSE estimator among all linear functions of the form $\alpha + \beta\hat{\theta}_a$ (a class that includes $\hat{\theta}_a$ itself, when $\alpha = 0$, and $\beta = 1$) is a shrinkage estimator:

$$\tilde{\theta}_a = \hat{\theta}_a + S_a \cdot \left[E(\theta_a) - \hat{\theta}_a \right],$$

with shrinkage adjustment factor (cf. Robbins 1983:714)

$$S_a = \frac{E_\theta\left[V_x\left(\hat{\theta}_a \middle| \theta_a\right)\right]}{V\left(\hat{\theta}_a\right)} = \frac{E_\theta\left[V_x\left(\hat{\theta}_a \middle| \theta_a\right)\right]}{E_\theta\left[V_x\left(\hat{\theta}_a \middle| \theta_a\right)\right] + V\left(\theta_a\right)}.$$

The numerator in this formula is the expected sampling variance in location a; the denominator is the variance of ML estimates, which depends both on expected sampling variance and on the underlying variance of true parameters θ. The $\tilde{\theta}$ formula shrinks local estimates $\hat{\theta}$ toward a global mean, with more shrinkage when local sampling noise is expected to be large relative to the variability of estimates across areas. Thus, as in our introductory example, the optimal amount of shrinkage varies inversely with local sample size.

Restricting estimators to linear functions of $\hat{\theta}_a$ has a big advantage: for *any* distribution of true rates θ, the minimum-MSE solution depends only on easily estimated means and variances (Robbins 1983). For many problems, including fertility estimation, EB estimation is possible because the unknown moments in the linear shrinkage formula—$E_\theta[V_x(\hat{\theta}_a \mid \theta_a)], V(\hat{\theta}_a), E(\theta_a)$—can be estimated from the empirical distribution of $\{\hat{\theta}_1 \ldots \hat{\theta}_A\}$ and from model assumptions about the sampling process.

ADAPTING EB FOR DEMOGRAPHIC SCHEDULES IN SMALL AREAS

Vector Estimation and Smoothing

Frequently, the researcher's interest is a *vector* of area-specific parameters $\boldsymbol{\theta}_a$. For example, $\boldsymbol{\theta}_a$ may represent an area's labor force participation rates in two ethnic groups (Longford 1999) or its fertility rates in seven age groups (our application). When local parameters are K-dimensional vectors, the multivariate analog to the minimum-MSE shrinkage estimator is (cf. Efron and Morris 1972; Longford 1999; O'Hagan 1994:163):

$$\tilde{\boldsymbol{\theta}}_a = \hat{\boldsymbol{\theta}}_a + \mathbf{S}_a \cdot \left[E(\boldsymbol{\theta}_a) - \hat{\boldsymbol{\theta}}_a \right],$$

with matrix shrinkage factor

$$\mathbf{S}_a = E_\theta\left\{\left[V_x\left(\hat{\boldsymbol{\theta}}_a \middle| \boldsymbol{\theta}_a\right)\right]\right\}\left[V\left(\hat{\boldsymbol{\theta}}_a\right)\right]^{-1},$$

where the $V(\,)$ terms are now $K \times K$ covariance matrices and $E(\boldsymbol{\theta}_a)$ is a $K \times 1$ vector of means. As before, small shrinkage adjustments are made by the $\tilde{\boldsymbol{\theta}}_a$ estimator in areas where local rates are expected to be estimated precisely, and vice versa. As in the scalar case, EB estimation is possible because the unknown moments in the shrinkage formula can be estimated from the set of vector estimates $\{\hat{\boldsymbol{\theta}}_1 \ldots \hat{\boldsymbol{\theta}}_A\}$. We discuss this part of the estimation problem more fully in subsequent sections.

Vector shrinkage is "minimum-MSE" in the sense that it minimizes the MSE of any linear combination of the K elements in $\boldsymbol{\theta}_a$. For local fertility-rate estimation, this implies

that the vector-shrinkage formula will minimize the MSE of each age-specific fertility rate separately, as well as the MSE of demographically interesting quantities, such as local total fertility rates (TFR).

Shrinkage has a more complicated meaning in the vector case because the matrix formula may move individual components of the vector away from their global (univariate) means even as it moves a vector of estimates toward a more typical pattern in the sample of empirically estimated schedules. We return to this point later when we discuss the results of the application of the method to Brazilian data.

Moving Neighborhoods

Longford's (1999) EB approach shrinks all local estimates toward a national mean vector. Small-area estimation problems are inherently spatial, however, and in many demographic applications, shrinkage toward a single national schedule may be undesirable. In particular, if demographic schedules exhibit regional variations in shapes and levels, then moving all small-area estimates toward the mean of the national vector may oversmooth, obscuring important and interesting regional differences. There is a considerable literature on EB estimation of local rates by selective borrowing of information from nearby areas on the map. Clayton and Kaldor (1987), Marshall (1991), Mollié and Richardson (1991), Assunção and Reis (1999), and Meza (2003) have used spatial EB methods to estimate small-area mortality rates.

One simple alternative to global shrinkage is to move local estimates toward subnational (e.g., state-level) means. In this scenario, a country is partitioned in advance into presumably homogeneous regions. This method is problematic, however, because of its treatment of areas on regional frontiers: two adjacent areas falling on opposite sides of a boundary could produce radically different vector-shrinkage estimates, even if their local ML estimates and sample sizes are identical.

In this article, we adopt another solution that overcomes the boundary problem by using the moving-neighborhood approach proposed by Marshall (1991). For each small area a, we define a set of areas H_a that includes a and some of its nearest spatial neighbors. We then shrink a's estimates toward the empirical average of schedules in H_a, using neighborhood-specific moments in the EB formula. Because adjacent areas will have partially overlapping neighborhoods, their local estimates will be shrunk toward similar target means. Ideal neighborhood sizes will be large enough for EB shrinkage to produce demographically plausible schedules at the local level, but small enough to allow real variability at intermediate geographic scales between local and national.

By combining vector shrinkage with spatial methods, we estimate local age-specific rates by "borrowing strength" not only from other age groups within each area but also from individuals of the same age in neighboring areas and from observed regularities in demographic schedules across locations. The combined method promises significant improvements in local-area rate estimates when both spatial and age-related patterns are strong.

Poisson Model and Shrinkage Estimator

Each area $a = 1 \ldots A$ and population group $k = 1 \ldots K$ has a true rate θ_{ak}. Population groups do not overlap. The *schedule* for location a is the vector $\boldsymbol{\theta}_a = [\theta_{a1} \ldots \theta_{aK}]'$. Estimates of θ_{ak} come from data on events in the previous year to a sample of n_{ak} individuals.

Given θ_{ak}, we assume that

$$x_{ak} \sim Poisson\left(n_{ak}\theta_{ak}\right),$$

where x_{ak} is the number of sample events observed in area a and group k. We prefer a Poisson distribution to a binomial, even if there is a maximum of one observed event per

sampled individual, because a Poisson distribution has a larger variance than does a binomial with the same mean (np, rather than $np(1-p)$, where p is the event probability). This additional variance in $\mathbf{x}|\boldsymbol{\theta}$ crudely captures the effects of population heterogeneity and unmeasured covariates. Brillinger (1986) suggested additional reasons for adopting a Poisson distribution by showing how vital rates appear as statistics that are calculated from a Poisson process, generating events on the Lexis diagram.

Conditional on the true local schedule $\boldsymbol{\theta}_a$, the ML estimator

$$\hat{\boldsymbol{\theta}}_a = \left[x_{a1}/n_{a1} \ldots x_{aK}/n_{aK} \right]'$$

is unbiased [$E_x(\hat{\boldsymbol{\theta}}_a \mid \boldsymbol{\theta}_a) = \boldsymbol{\theta}_a$] and its sampling covariance is a $K \times K$ diagonal matrix,

$$V_x\left(\hat{\boldsymbol{\theta}}_a \middle| \boldsymbol{\theta}_a\right) = diag\left\{\theta_{ak}/n_{ak}, k = 1 \ldots K\right\}.$$

In practice, the true local schedule is unknown, so rather than condition on $\boldsymbol{\theta}_a$, we must consider expected values of moments over possible local schedules. For the rest of this section, we use an asterisk to indicate a statistical calculation or value specific to H_a. This distinction is important because the distribution of $\boldsymbol{\theta}_a$ within a neighborhood will generally differ from its distribution across the entire map; this is the reason for taking a more local approach to estimation. Thus, operators such as $E^*(\ )$ and $V^*(\ )$ refer to the joint distribution of $(\mathbf{x},\boldsymbol{\theta})$ over H_a only.

The ML estimator is centered on a true neighborhood-specific mean vector,

$$E^*\left[\hat{\boldsymbol{\theta}}_a\right] = E^*_{\theta}\left[E_x\left(\hat{\boldsymbol{\theta}}_a \middle| \boldsymbol{\theta}_a\right)\right] = E^*_{\theta}\left[\boldsymbol{\theta}_a\right] = \boldsymbol{\mu}^*,$$

with covariance

$$V^*\left[\hat{\boldsymbol{\theta}}_a\right] = E^*\left[\left(\hat{\boldsymbol{\theta}}_a - \boldsymbol{\mu}^*\right)\left(\hat{\boldsymbol{\theta}}_a - \boldsymbol{\mu}^*\right)'\right] = \boldsymbol{\Sigma}^* + \boldsymbol{\Omega}^*_a,$$

where $\boldsymbol{\Sigma}^* = E^*[(\boldsymbol{\theta}_a - \boldsymbol{\mu}^*)(\boldsymbol{\theta}_a - \boldsymbol{\mu}^*)']$ denotes the covariance of true schedules across H_a and $\boldsymbol{\Omega}^*_a = E^*_{\theta}[V_x(\hat{\boldsymbol{\theta}}_a \mid \boldsymbol{\theta}_a)]$ is a diagonal matrix representing expected sampling noise in area a itself.

In this framework, the EB shrinkage estimator for the demographic schedule in location a, given $\boldsymbol{\mu}^*$, $\boldsymbol{\Sigma}^*$, and $\boldsymbol{\Omega}^*_a$, is

$$\tilde{\boldsymbol{\theta}}_a = \hat{\boldsymbol{\theta}}_a + \mathbf{S}_a\left(\boldsymbol{\mu}^* - \hat{\boldsymbol{\theta}}_a\right),$$

with adjustment factor

$$\mathbf{S}_a = \boldsymbol{\Omega}^*_a\left[\boldsymbol{\Sigma}^* + \boldsymbol{\Omega}^*_a\right]^{-1}.$$

Shrinkage toward the neighborhood average schedule is more desirable when the neighborhood is homogeneous (small between-area variation $\boldsymbol{\Sigma}^*$) or samples in the target area are small (big within-area variation $\boldsymbol{\Omega}^*_a$).

Method-of-Moments Estimation of Shrinkage Factors

The matrix shrinkage formula requires values for $\boldsymbol{\mu}^*$, $\boldsymbol{\Sigma}^*$, and $\boldsymbol{\Omega}^*_a$. In the EB approach, one uses the model and the sample distribution of ML estimators to estimate these moments. Averages weighted by sample sizes are generally more efficient than are simple averages for estimating moments, and we use weighted averages throughout.

Neighborhood mean vectors are estimated as weighted averages (cf. Marshall 1991:285):

$$\hat{\boldsymbol{\mu}}^* = \sum_{s \in H_a} \mathbf{P}_s \hat{\boldsymbol{\theta}}_s,$$

where s indexes areas in H_a, and $\mathbf{P}_s$ is a (neighborhood-specific) diagonal matrix containing the proportions of neighborhood samples in each group that come from location s:

$$\mathbf{P}_s = diag\left\{ p_{sk} = n_{sk} \Big/ \sum_{r \in H_a} n_{rk}, \quad k = 1 \ldots K \right\}.$$

The moment estimate for means has a simple interpretation: element $\hat{\mu}_k^*$ equals the ratio of total neighborhood events to total neighborhood exposure for group k.

An estimate of the expected sampling covariance matrix for location a depends on average neighborhood rates and local sample sizes:

$$\hat{\boldsymbol{\Omega}}_a^* = diag\left\{ \hat{\mu}_k^* \Big/ n_{ak}, \quad k = 1 \ldots K \right\}.$$

Estimating the neighborhood covariance matrix of true rates ($\boldsymbol{\Sigma}^*$) also involves equating population moments to weighted sample averages, but the procedure is more complex than that for calculations for $\boldsymbol{\mu}^*$ and $\boldsymbol{\Omega}_a^*$. Begin by defining an empirical estimate of the covariance of ML estimates (as opposed to true rates) in neighborhood H_a:

$$\mathbf{Q} = \sum_{s \in H_a} \mathbf{P}_s^{1/2} \left(\hat{\boldsymbol{\theta}}_s - \boldsymbol{\mu}^* \right) \left(\hat{\boldsymbol{\theta}}_s - \boldsymbol{\mu}^* \right)' \mathbf{P}_s^{1/2}.$$

The neighborhood mean μ^* is unknown, but $\mathbf{Q}$ can be approximated by

$$\hat{\mathbf{Q}} = \sum_{s \in H_a} \mathbf{P}_s^{1/2} \left(\hat{\boldsymbol{\theta}}_s - \hat{\boldsymbol{\mu}}^* \right) \left(\hat{\boldsymbol{\theta}}_s - \hat{\boldsymbol{\mu}}^* \right)' \mathbf{P}_s^{1/2}.$$

The expected value of $\mathbf{Q}$ is

$$E^*\left(\mathbf{Q}\right) = \sum_{s \in H_a} \mathbf{P}_s^{1/2} \left(\boldsymbol{\Sigma}^* - \boldsymbol{\Omega}_s^* \right) \mathbf{P}_s^{1/2} = \sum_{s \in H_a} \mathbf{P}_s^{1/2} \boldsymbol{\Sigma}^* \mathbf{P}_s^{1/2} + \sum_{s \in H_a} \mathbf{P}_s \boldsymbol{\Omega}_s^*.$$

Rearranging this matrix equation and taking the (j,k) element yields

$$\left[E^*\left(\mathbf{Q}\right) - \sum_{s \in H_a} \mathbf{P}_s \boldsymbol{\Omega}_s^* \right]_{jk} = \boldsymbol{\Sigma}_{jk}^* \cdot \sum_{s \in H_a} \sqrt{p_{sj} p_{sk}},$$

so an approximate method-of-moments estimator for $\boldsymbol{\Sigma}$ is

$$\hat{\boldsymbol{\Sigma}}_{jk}^* = \left[\hat{\mathbf{Q}} - \sum_{s \in H_a} \mathbf{P}_s \hat{\boldsymbol{\Omega}}_s^* \right]_{jk} \Big/ \sum_{s \in H_a} \sqrt{p_{sj} p_{sk}}, \quad j \in 1 \ldots K \quad k \in 1 \ldots K.$$

In small neighborhoods, this procedure can produce $\boldsymbol{\Sigma}^*$ estimates that are not positive definite, that is, matrix estimates such that some linear combinations of elements of $\hat{\boldsymbol{\theta}}_a$ would have negative estimated variances. In such cases, we a used matrix version of Marshall's (1991:285) convention of replacing negative variance estimates with zero. We adapted a matrix-truncation procedure suggested by Efron and Morris (1972), replacing the offending $\hat{\boldsymbol{\Sigma}}^*$ with the "nearest" positive semidefinite matrix of the same dimensions. Specifically, we did a standard eigen-decomposition $\hat{\boldsymbol{\Sigma}}^* = \mathbf{XDX}'$, replaced any negative eigenvalues in $\mathbf{D}$ with zeroes to form a new matrix $\mathbf{D}_+$, and replaced $\hat{\boldsymbol{\Sigma}}^*$ with $\hat{\boldsymbol{\Sigma}}_+^* = \mathbf{XD}_+\mathbf{X}'$.

Table 1. Distribution of 1991 Census Sample Sizes Across 3,829 Municipalities

Sample Characteristics	Number of Municipalities
Number of Women Aged 15–49	
0–49	42
50–59	330
100–499	2,028
500–999	725
1,000+	704
Number of Births Last Year	
0	19
1–9	600
10–49	1,808
50–59	720
100+	682

With empirical estimates of the required moments in hand, we constructed the EB estimator for each schedule. For location a the estimated schedule is

$$\underset{K\times 1}{\tilde{\boldsymbol{\theta}}_a} = \underset{K\times 1}{\hat{\boldsymbol{\theta}}_a} + \underset{K\times K}{\hat{\boldsymbol{\Omega}}_a^*}\left[\underset{K\times K}{\hat{\boldsymbol{\Sigma}}^*} + \hat{\boldsymbol{\Omega}}_a^*\right]^{-1}\left(\underset{K\times 1}{\hat{\boldsymbol{\mu}}^*} - \underset{K\times 1}{\hat{\boldsymbol{\theta}}_a}\right).$$

APPLICATION AND DATA

We investigated the performance of the foregoing EB vector estimator by applying it to microdata from Brazil's 1991 Demographic Census. This data set includes fertility information for 3,829 small areas that we refer to here as "municipalities," although in some cases, we clustered 1991 *municípios* into larger spatial units to preserve comparability with geographic coding in earlier Brazilian censuses. (Clustering is a technical detail with no important effect on our results.) Fertility levels and socioeconomic conditions vary widely across Brazilian municipalities; this diversity is a primary motivation for using localized, as opposed to national, means as targets for vector shrinkage.

Fertility data come from long-form questionnaires that are given to a sample of census households. We used data on women aged 15–49, aggregated into standard five-year age groups. The 1991 census included questions on the date of each woman's last live birth (from which we derived births in the past year and current fertility), as well as a count of children ever born (parity). As in many countries with low average educational levels, Brazilian data on births in the year preceding a census suffer from reporting errors, which has been amply demonstrated in previous analyses of the 1991 and earlier Brazilian censuses (e.g., Wong 1986). Demographers often use parity data to correct for incomplete or inaccurate reporting of birth dates in such cases (Coale and Demeny 1967; United Nations 1983). Such corrections are possible, but they are beyond the scope of this article. Here, we focus on estimation issues, recognizing that any complete demographic analysis also requires a careful assessment of data quality.

Census sampling fractions are 10% and 20% for municipalities with estimated populations larger and smaller than 15,000 inhabitants, respectively. Most municipal samples are quite small, as illustrated in Table 1.

ML-estimated schedules for the 3,829 municipalities are noisy, but they suggest clear patterns in true fertility rates across age and space. These patterns imply, in turn, that

Table 2. Correlation of ML Rate Estimates Across Areas With Large Samples

	Age						
Age	15–19	20–24	25–29	30–34	35–39	40–44	45–49
15–19	1						
20–24	.75	1					
25–29	.53	.72	1				
30–34	.33	.55	.65	1			
35–39	.39	.57	.67	.72	1		
40–44	.42	.60	.62	.70	.75	1	
45–49	.38	.50	.48	.52	.57	.53	1

borrowing strength via EB methods may lead to large reductions in estimation errors. Table 2 illustrates the (weighted) correlation matrix of ML estimates across the 568 areas that had at least 100 women sampled in every age group, which we hereafter call the *large-n* areas. Because ML estimates in large-*n* areas have low sampling variances, this matrix may be taken as a crude but effective estimate of the correlations between true rates. (Several other calculation methods lead to similar results; see Assunção et al. 2003.) These data, particularly the high values for adjacent age groups along the subdiagonal, indicate potentially large gains in efficiency from using vector EB methods that exploit cross-component correlations.

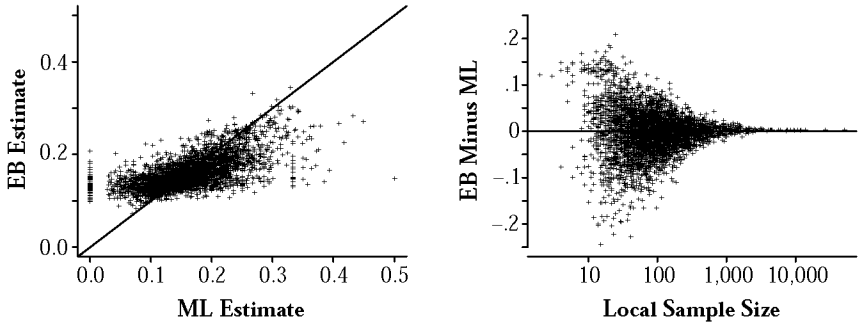
ML fertility-schedule estimates also exhibit spatial patterns that increase the utility of a localized approach to EB estimation. Using the variance-weighted Moran I index proposed by Assunção and Reis (1999), we can reject the hypothesis of no first-order spatial autocorrelation in each age group at $p < .0005$. This finding is not surprising, since Brazil is a large country with well-known regional variations in demographic rates.

We defined neighborhoods H_a as follows. For each area a, we first ranked all 3,829 areas by the distance between their geographic centroids and the centroid of a. We constructed an initial neighborhood as the union of a and its six nearest municipalities. If there were fewer than 21,000 women sampled in this initial neighborhood or if there were no neighborhood births in one of the age groups, then we added the seventh-nearest municipality, the eighth-nearest, and so on. We stopped expanding the neighborhood as soon as the cumulative sample size reached at least 21,000 and there was at least one sampled birth in every age group. Under this default definition, which we denote *Local21*, the number of municipalities in neighborhoods varies from 7 to 112, with a median of 33.

We chose 21,000 as the smallest neighborhood size. Under this definition, neighborhoods have about three times as many women as a typical sample from a Demographic and Health Survey (Lê and Verma 1997), and the mean sample size in an [age, neighborhood] cell is 3,000. To get a sense of the likely precision of ML estimates for a Local21 neighborhood, note that with $n = 3{,}000$ women and $\theta = .250$ (a fairly high value in the 1991 Brazilian context), estimates would fall within .015 of the true rate in approximately 95% of samples; we decided that this level of precision would be adequate to yield stable targets for shrinkage estimation.

Area-specific schedules are 7×1 vectors $[f_{15-19} \ldots f_{45-49}]'$. All EB results presented in the next section use the vector-estimation method described in the third section, with the Local21 neighborhood definition. In the sixth section, we experiment with several other EB variants.

Figure 2. ML and EB Estimates for 20- to 24-Year-Olds in 3,829 Brazilian Municipalities



EB FERTILITY SCHEDULES FOR BRAZILIAN MUNICIPALITIES

We illustrate vector EB shrinkage of the raw ML estimates using examples with a single rate, a pair of rates, and finally entire local schedules. To begin, Figure 2 displays estimates for a single age group. The results are similar in all seven age groups, so we use 20- to 24-year-olds as a representative example. In the left panel, each municipality is represented by a point, with ML estimates (births per 20- to 24-year-old woman) on the horizontal axis and EB estimates on the vertical axis. Points fall on the 45-degree line if local EB and ML estimates are identical. The right panel shows the EB–ML difference for each area as a function of the (log) number of 20- to 24-year-olds in the local sample.

The shrinkage effect is obvious in the left panel of Figure 2. EB estimates have a much tighter distribution than ML estimates, and the cloud of points has a lower slope than the 45-degree line. This represents a shrinkage effect, in which low ML estimates are pulled up and high estimates are pushed down. Another important point concerns ML estimates that are equal to zero. In 93 municipalities, there were no births to sampled 20- to 24-year-old women in the year preceding the census, and the births/woman estimate is consequently zero in these areas. Clearly $f_{20\text{–}24} = 0$ is unlikely to be close to the true rate in most of these municipalities, and (as in this article's introductory example) a shrinkage estimator that pulls these zeroes toward neighborhood means for $f_{20\text{–}24}$ is likely to have lower average errors. Figure 2 shows that the EB transformations of zero-valued ML estimates for $f_{20\text{–}24}$ range from .097 to .206, depending on the sample sizes, estimates in other age groups (especially ages 25–29) in the same area, and the mean values for ML rates in other nearby areas.

The funnel-shaped cloud of points in the right panel of Figure 2 shows how EB changes estimates most in the areas with the smallest populations and sample sizes. The EB procedure leaves births/woman estimates almost unchanged when local sample sizes are large, but alters estimates much more freely when there is little area-specific information in the sample.

Vector EB shrinkage operates on all age groups simultaneously, using estimated correlation patterns to adjust ML estimates. Figure 3 provides a partial, bivariate view of this process for two age groups, 20–24 and 25–29. In the left panel, each municipality's ML estimates appear as a point, with the 20–24 estimate on the horizontal axis and the 25–29 estimate on the vertical axis; the right panel contains analogous data for EB shrinkage estimates. The ellipse in the left panel contains approximately 90% of the ML estimates for the large-n areas. An identical ellipse appears for reference in the right panel.

Figure 3. ML and EB Estimates of 20- to 24-Year-Olds and 25- to 29-Year-Olds in 3,829 Brazilian Municipalities

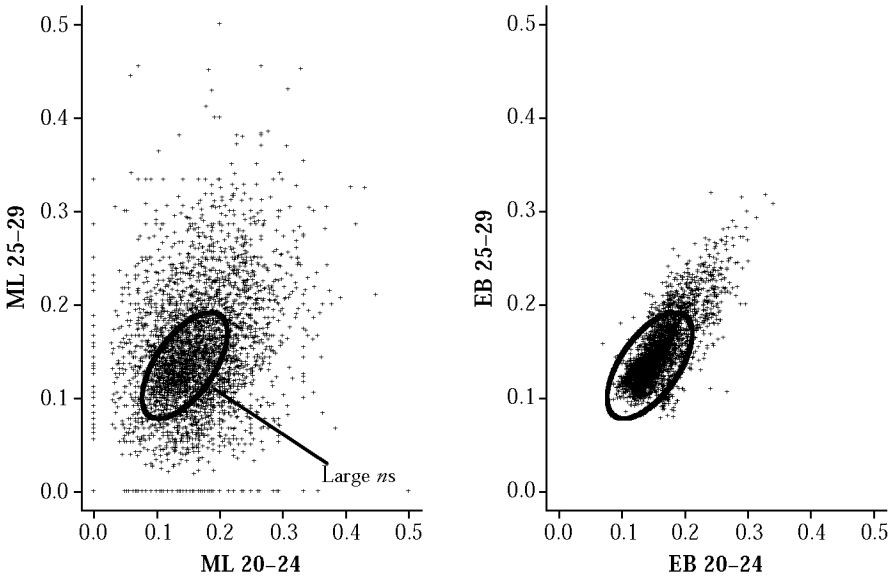
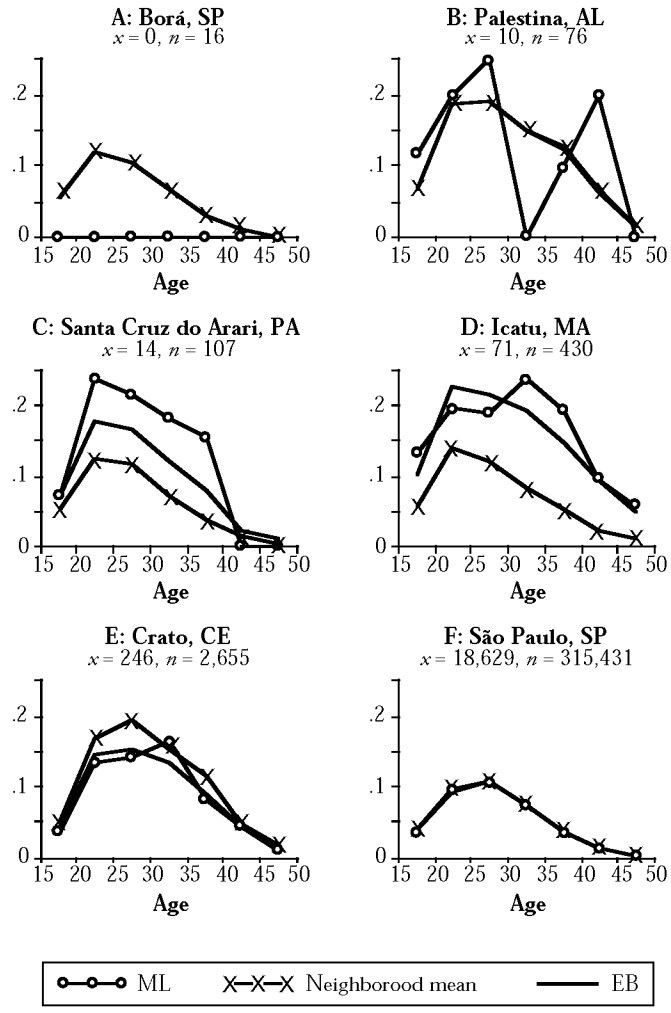


Figure 3 illustrates the effects of shrinkage on vector data. The cloud of EB points on the right is much more tightly concentrated and positively tilted. The ML estimates on the left, particularly in areas with large samples, suggest a moderately strong positive correlation between f_{20-24} and f_{25-29}: areas with high f_{20-24} also tend to have high f_{25-29}. The EB method shrinks each area's pair of ML estimates toward a point that conforms to this pattern. As always, more shrinkage occurs when local samples are smaller and local ML estimates are less reliable. The ML pairs that are the most likely to be altered substantially by EB are those from sparsely populated areas *and* with unusual bivariate deviations from the global mean. For example, the ML pairs at the northwestern edge of the left-hand scatter in Figure 3 look like extreme outliers because it is unlikely (given a positive correlation between f_{20-24} and f_{25-29}) that the true f_{20-24} for an area would be far below average while the true f_{25-29} would be far above average. In the absence of reliable local evidence to the contrary, EB insists that pairs of estimates conform more to the typical pattern, and it consequently squeezes the ML cloud in the left panel into the much tighter, steeper shape on the right.

Bivariate scatters provide important insights into EB estimation, but the vector shrinkage actually occurs in K-dimensional space. To understand the full effect of vector shrinkage on local-schedule estimates, one must examine all K components. In our application, we have $K = 7$ age groups, and it is natural to view them as line graphs with age on the horizontal axis.

Figure 4 illustrates schedule estimates for six selected areas, which include the municipalities with the smallest sample (Panel A: Borá, in São Paulo state, with only 16 long-form questionnaires for women aged 15–49) and the largest sample (Panel F: the

Figure 4. ML, Neighborhood Mean, and EB Schedules in Six Selected Municipalities



city of São Paulo itself, with over 300,000 women sampled). Each panel shows the local ML schedule (o), the mean schedule for the neighborhood surrounding the municipality (X), and the EB schedule (solid line) that is an approximately optimal combination of the Xs and os.

To read Panel A, for example, note that the ML schedule for Borá is a flat line along the horizontal axis, with fertility estimated at zero for every age group because none of the 16 sampled women had a birth in the year preceding the census ($x = 0$, $n = 16$). The Local21 neighborhood around Borá contains 50 municipalities, with an aggregate births/woman schedule given by the Xs. There is almost no local fertility information for

this municipality, and EB estimation (the solid line) shrinks the local ML schedule virtually all the way back to the neighborhood mean vector. Read Panels B–F similarly.

Figure 4 illustrates several interesting points:

- There is no underlying model forcing the EB estimates into particular shapes. Patterns come directly from data in the neighborhood around a municipality.
- EB schedules generally conform better than ML schedules to known regularities in fertility. Unlike the ML schedules, for example, all EB schedules in the figure have a single mode in the second or third age group, followed by a monotonic decline in rates.
- There is substantial variation in the shapes and levels of neighborhood mean schedules, indicating that a moving-neighborhood approach may produce better estimates than an EB procedure that shrinks all area schedules toward the national mean.
- The EB procedure shrinks local ML estimates almost all the way back to the neighborhood mean vector ($\mathbf{S} \approx \mathbf{I}_K$) when there are few local data and when such data either lead to implausible ML schedules (such as the zeroes in Panel A) or provide inconsistent evidence about differences between the local schedule and the neighborhood schedule (such as the odd ML schedule in Panel B).
- If local data are scarce but tell a consistent story about differences between local and neighborhood rates (Panels C and D), then EB treats the ML schedule as a less extreme outlier in $\mathbb{R}^7$ and shrinks it less. The resulting EB schedule combines local and neighborhood information more equally than in other areas with small samples. As always, a smaller amount of shrinkage is optimal when there are more local data (compare Panels C and D).
- If local samples are large (Panels E and F), the optimal EB mixture tilts strongly toward the local schedule ($\mathbf{S} \approx \mathbf{0}$). With an extremely large sample, such as that for São Paulo (Panel F), the local area contains almost all its neighborhood's population, and all three schedules overlap.
- EB schedules combine local and neighborhood curves, but the vector EB procedure does not always produce age-specific rate estimates that lie between local and neighborhood average values. Of the 42 rate estimates in Figure 4, 15 have EB values outside the range defined by ML and neighborhood rates for the same age. The most significant cases are in age groups 2 and 3 in Icatu, Maranhão (Panel D), and in age group 4 in Crato, Ceará (Panel E). In Icatu, local data suggest above-neighborhood fertility levels (Icatu's neighborhood includes a large urban area, São Luís, with much lower fertility), but vector shrinkage moves f_{20-24} and f_{25-29} estimates even farther from neighborhood levels to make the local schedule's shape more like the neighborhood pattern. In Crato, EB smooths the local schedule mainly by reducing the high ML estimate for 30- to 34-year-olds, so that it conforms to the pattern of below-average rates in the other six age groups.

EVALUATING EB ESTIMATION PROCEDURES

Alternative Estimation Methods

We have so far compared ML estimation to only one EB procedure. Specifically, our EB estimator used vector shrinkage (rather than shrinking each schedule component separately) and neighborhoods defined by the Local21 criterion (rather than national or some other intermediate-sized neighborhoods). Both choices are subject to question. Furthermore, comparisons in the previous section were rather informal. In this section, we analyze a wider variety of EB estimators in a more formal setting.

We compare ML to six alternative EB methods, defined by a choice of neighborhood size (Local21, Local84, or national) and by whether shrinkage operates on entire vectors or on vector components separately (vector or scalar). The Local84 neighborhood

definition is identical to Local21, except that it uses 84,000 as the minimum neighborhood sample size. With the national definition, the "neighborhood" is the same for every area—namely, all of Brazil.

Measures

In our application, any one estimation method produces 3,829 × 7 = 26,803 parameters. Evaluating the "goodness" of such a large set is a difficult task, particularly because we do not know the true rate schedules at the municipal level. We can, however, establish basic principles for comparing different estimation rules. The first principle is *demographic plausibility*. A good estimation method should generate small-area schedules with demographically sensible shapes and levels. It should be an effective smoother that purges local estimates of sampling noise, while producing estimates with the "look and feel" of real fertility schedules. The second principle is *regional accuracy*. Although we do not know true rates at the small-area level, sample sizes are often sufficient for the accurate estimation of schedules at higher levels of subnational geography. A good method should preserve spatial patterns at these larger geographic scales. That is, it should avoid oversmoothing as it tries to produce demographically plausible schedules for small areas.

To operationalize comparisons along the demographic and geographic dimensions, we define the following functions. For any $K \times 1$ schedule $\boldsymbol{\theta}$, its *shape* is defined as the $K \times 1$ standardized vector

$$shape(\boldsymbol{\theta}) = \boldsymbol{\theta} \Big/ \sum_k \theta_k,$$

and its *dissimilarity* to a reference schedule **Ref** is a percentage relative error measure:

$$DISS(\boldsymbol{\theta}, \mathbf{Ref}) = 100 \cdot \left(\sum_k \left| \theta_k - Ref_k \right| \Big/ \sum_k Ref_k \right).$$

Note that *DISS* = 0 (the minimum possible value) when its two arguments are identical and that *DISS* could exceed 100 if its two arguments are very dissimilar.

Demographic Comparisons

Like beauty, demographic plausibility is largely in the eye of the beholder. Period fertility schedules depend on many factors (local marriage customs, contraceptive practices, disease prevalence, cohort timing effects, and so forth) and come in many shapes and sizes. This diversity makes it difficult to establish a single standard of plausibility. We therefore opted to compare area schedules to 226 schedules from the U.S. Census Bureau's International Data Base (IDB), and to define plausibility in terms of a match between the shape of the area schedules and the shape of *any* of the 226 IDB schedules. (For details and links to these IDB schedules, see Schmertmann 2003.) Specifically, we measured the demographic plausibility of a schedule $\boldsymbol{\theta} \in \mathbb{R}^K$ by the best match to the shape of any of the schedules $\mathbf{IDB}_1 \ldots \mathbf{IDB}_{226}$:

$$D(\boldsymbol{\theta}) = \min_{j=1 \ldots 226} \left\{ DISS\left[shape(\boldsymbol{\theta}), shape(\mathbf{IDB}_j) \right] \right\}.$$

D is a measure of demographic implausibility. It is close to zero for schedules that have a shape that is similar to at least one schedule in the IDB and has large values for schedules with no close IDB counterpart. To get a sense of *D*'s scale, note that the ML schedules (o) in Panels B–F of Figure 4 have rounded *D* values of 41, 14, 10, 9, and 4, respectively. (The IDB schedules that are the most similar to B–F are, in order, Grenada, Nauru, Mauritania, Pakistan, and French Guiana.)

Figure 5. Distribution of the Demographic Implausibility Index (*D*) Across 3,829 Municipalities, for Various Estimation Methods

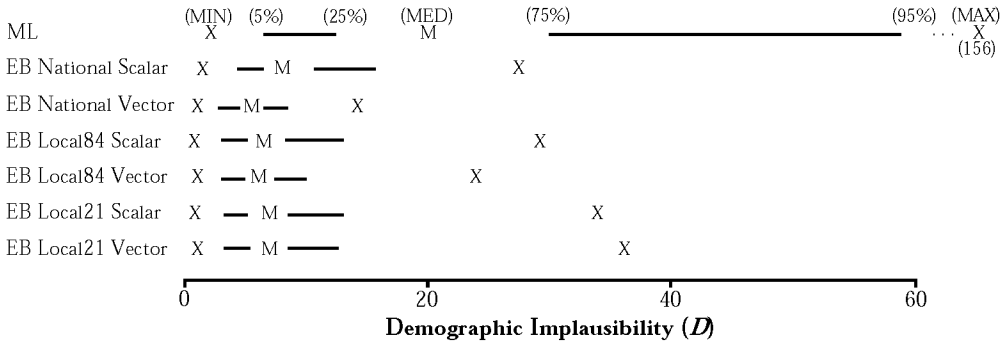


Figure 5 shows summary information on the distribution of D for ML estimates (the top line) and for the six EB variants. The horizontal scale at the bottom of the figures represents D values. The row for each estimation method contains Xs marking the minimum and maximum levels of D in the 3,829 municipalities (note that the maximum is off the scale for the ML estimates), solid bars extending from the 5–25 and 75–95 percentiles, and an M marking the median D value. The solid bars are the key to interpreting Figure 5: from end to end, they span 90% of the distribution, with 20% of the D values in the ranges for each bar separately, 50% in the gap between the bars, 5% to the left, and 5% to the right.

The main result in Figure 5 is a marked improvement in the demographic plausibility of municipal-level schedules when we switched from ML to any of the EB procedures. If we define *plausible* as $D < 20$, for example, then only half the ML schedules have plausible shapes, compared to virtually all the EB schedules. As one may expect, EB estimation with larger neighborhoods and more smoothing appears to produce more-plausible schedules, but that improvement is small.

Compared with scalar methods that operate separately on each age group, vector EB shrinkage improves the plausibility of the estimates of the small-area schedules for all three neighborhood sizes. The improvement from vectorizing is the largest when shrinking toward national-level targets and almost negligible with Local21 neighborhoods. This pattern suggests that, to some extent, Longford's (1999) vector approach and Marshall's (1991) local approach can produce similar effects. The total covariance of fertility rates over the map can be partitioned into within- and between-neighborhood components as

$$V(\boldsymbol{\theta}) = \underset{\text{within}}{E_H\left[V(\boldsymbol{\theta}|H)\right]} + \underset{\text{between}}{V_H\left[E(\boldsymbol{\theta}|H)\right]}.$$

As neighborhoods H become larger, the *between* component inevitably decreases, and more of the overall variability in true rates occurs within the neighborhoods. Vector EB methods exploit *within* correlations, so they improve estimates more when neighborhoods are large.

Regional Comparisons

Brazilian census samples are large enough to make ML estimates accurate for geographic aggregates, such as states or subnational regions, and fertility patterns are quite diverse across larger-than-municipal areas. Parts of Brazil, notably cities in the more industrially developed Southeast and South, had experienced a nearly complete transition to near- or below-replacement fertility by 1991, while other regions and areas still had very high fertility levels (Potter, Schmertmann, and Cavenaghi 2002).

To test how accurately an estimation method replicates Brazil's spatial fertility patterns, we would like to identify distinct regions of truly high and low fertility on the map. These regions must be large enough to ensure that birth/woman ratios from census samples are accurate, so that we have a baseline for testing. Ideally the regions should be internally homogeneous, but also very different from one another. Political units like Brazil's 27 states may not be good for this purpose because of possible internal heterogeneity in fertility levels and patterns.

For geographic testing, we created an empirically defined set of 44 regions, based on the principle of maximum interregional and minimal intraregional variance in fertility levels. Specifically, for the large-n municipalities defined earlier, we fit a regression tree (Breiman et al. 1984) for estimated TFR, $F_a = 5\sum_k \hat{\theta}_{ak}$, using as predictors the latitude and longitude of the municipality's centroid.[3] The resulting tree splits Brazil into 44 regions with maximal TFR differences and minimal internal heterogeneity. The new regions are large enough to produce stable maximum-likelihood estimates (all 308 [region, age] cells have at least 1,000 sampled women), and they capture spatial heterogeneity in fertility better than states (74% of the variance in F across large-n municipalities is between the new regions, compared with 53% between the states).

For neighborhoods defined by these regions, such as H = {municipalities with centroids in region 1}, we can calculate an average of municipal-level schedules θ as

$$\mathbf{m}(\boldsymbol{\theta}) = \sum_{s\in H} n_{sk}\theta_{sk} \Big/ \sum_{s\in H} n_{sk}, \qquad k = 1\ldots 7.$$

The ML mean vector is $\hat{\boldsymbol{\mu}} = \mathbf{m}(\hat{\boldsymbol{\theta}})$, and its components are simply events-to-exposure ratios for the region in each age group. At the regional level, $\mathbf{m}(\hat{\boldsymbol{\theta}})$ will be an extremely accurate estimator of the true rate vector $\mathbf{m}(\boldsymbol{\theta})$ because ML sampling variance is negligible with large regional samples. Consequently, we can define the *regional error* of an EB estimation procedure as $R = DISS[\mathbf{m}(\tilde{\boldsymbol{\theta}}), \mathbf{m}(\hat{\boldsymbol{\theta}})]$. That is, we measure the accuracy of the EB procedure at the regional level by aggregating EB schedule estimates from each member municipality and then comparing that schedule to the ML schedule for the region as a whole. Notice that under this definition, $R = 0$ for the ML procedure itself and will be near zero for other procedures that produce local estimates that approximate the region-level ML schedule well.

Figure 6 displays the distribution of our regional error measure (R) for each estimation method, in a format identical to that of Figure 5. By definition, the ML estimation method scores $R = 0$ for every region, so its distribution is concentrated completely at the left edge of the plot.

---

3. The regression tree produces a sequence of binary decision rules (much like playing 20 Questions) that classify large-n municipalities into fertility regions based on centroid coordinates. For example, to distinguish high and low municipal TFR, the best single question is, Is the centroid north of –16.3° latitude? The average TFR is 3.4 to the north of this line, and 2.4 to the south. Among municipalities that are north of –16.3°, the sharpest dividing line is at –40.1° longitude (the average TFR is 4.0 in municipalities with centroids to the west versus 3.1in municipalities with centroids to the east), and so on. Recursive splitting based on latitude and longitude lines yields 44 rectangular regions, and each municipality belongs to the region containing its centroid.

Figure 6. Distribution of the Regional Error Index (R) Across 44 Fertility Regions, for Various Estimation Methods

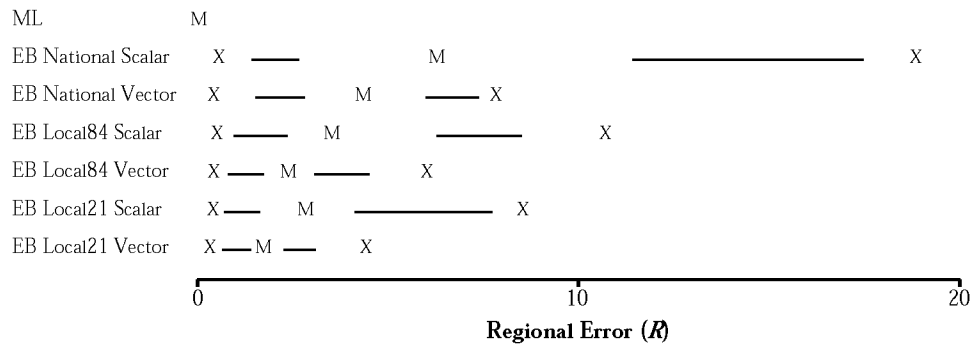


Figure 6 illustrates three key findings. First, all EB methods generally produce municipal estimates that match regional schedules well when aggregated, with the possible exception of national-level scalar shrinkage.

Second, smaller neighborhoods improve the regional accuracy of EB estimates. Some of the real heterogeneity in Brazilian regional fertility patterns is smoothed away when EB neighborhoods expand. An additional analysis (not shown) suggests that the highest R values for the national and Local84 procedures are in small, high-fertility pockets in extreme eastern Brazil (especially in Rio Grande do Norte, Paraíba, and Pernambuco). EB methods with large neighborhoods tend to underestimate fertility in such places; they shrink municipal schedules too far toward lower, larger-area means.

Third, vector procedures produce more-accurate regional-level schedules than scalar procedures. Stated broadly, vector EB is "more local" than scalar EB because it borrows more strength locally (from local data for other groups). More precisely, vector EB outperforms scalar EB in sparsely populated areas near large cities. Icatu (Figure 4, Panel D) is a good example, with ML rates consistently above a neighborhood average dominated by a low-fertility city (São Luís). Scalar EB (not visible in Figure 4) shrinks ML estimates farther toward the mean schedule than the vector method (approximately halfway toward the Xs at each age) because it ignores the [+ . . . +] pattern of deviations across age groups. Scalar EB tends to overshrink and underestimate fertility in areas like Icatu. Consequently, small-area scalar EB schedules tend to underestimate regional fertility when they are aggregated over regions with many "Icatus."

Discussion

Demographic and regional analyses for Brazil suggest that vector EB methods outperform scalar methods. At all neighborhood sizes, the vector approach produced small-area schedule estimates that were, demographically speaking, either more plausible or (for Local21 neighborhoods) approximately the same as with the scalar approach. At all neighborhood sizes, vector estimates conformed better to known regional patterns in fertility. There appears to be no significant trade-off in the vector versus scalar choice: vector methods win.

Our results are less definitive about the appropriate choice of neighborhood size. Localizing estimation by making neighborhoods smaller improves regional accuracy.

Table 3. Median Values of Demographic Implausibility (D) and Regional Error (R) Indices for Alternative Estimation Methods

		Scalar EB			Vector EB	
ML	National	Local84	Local21	National	Local84	Local21
19.5	8.2	6.7	6.7	5.6	6.0	6.8
0	6.2	3.5	2.8	4.3	2.4	1.7

However, for vector EB methods, there may be a small demographic cost because schedules estimated from smaller neighborhoods have slightly less plausible shapes. This is a classic bias-versus-variance trade-off: smaller neighborhoods reduce bias in the EB method, but they increase the vulnerability of local schedule estimates to sampling noise. Table 3 summarizes the choice between methods by showing median values of D and R for each method, from the more complete information in Figures 5 and 6. In our application, we decided that regional accuracy was the more important objective. After we examined Table 3 and the distributions in Figures 5 and 6, we selected vector EB with Local21 neighborhoods as the best compromise between small-area bias ($\sim R$) and variance ($\sim D$). However, the main message from the comparative analysis is that all the vector EB methods that we tried work reasonably well, with localized versions offering important spatial improvements over shrinkage to national targets. The optimal size of moving neighborhoods for localized vector shrinkage is likely to be problem- and data specific.

CONCLUSION

EB is a powerful tool for many demographic estimation problems. In this article, we have presented an integrated EB technique that combines vector methods with adaptively defined local neighborhoods to produce estimates that are both demographically plausible and appropriately sensitive to spatial patterns. The application to Brazilian fertility data demonstrates that exploiting spatial and age-related correlation in rates can substantially reduce estimation errors for small-area schedules. Our method is particularly appropriate in cases with many areas, small local samples, substantial regional variations, strong correlations between schedule components, and known spatial relationships between areas.

Fully Bayesian or hierarchical Bayesian approaches (see Ghosh and Rao 1994), which we have not discussed in this article, are computationally intensive alternatives to EB methods. In these models, quantities such as our μs and Σs are modeled as hyperparameters, with their own prior distributions that, in turn, affect posterior distributions of demographic rates. The more complex Bayesian models have theoretical appeal, and recent advances in estimation algorithms have led to the wider use of fully Bayesian models for small-area estimation of scalar rates (e.g., Besag, York, and Mollié 1991 on the incidence of cancer; Assunção, Reis, and Oliveira 2001; Bernadinelli et al. 2000; Fisher and Asher 2000; Gelfand, Zhu, and Carlin 2001). There has also been recent research on fully Bayesian modeling of spatial vector data (Kim, Sun, and Tsutakawa 2001; Knorr-Held and Best 2001; Wang and Wall 2001). However, the latter models are even more computationally demanding than the scalar versions, and making them operational remains a serious obstacle for most researchers.

In contrast to fully Bayesian estimation, EB methods are already computationally feasible for demographers with access to standard statistical software. For example, we performed all calculations presented in this article using the publicly available R statistical environment, and the code is available on request. As another option, it would not be difficult to program macros in SAS using its matrix language features to run our procedure.

EB also shares with its complex cousins an ability to adapt to many types of estimation models. In this article, we shrank small-area estimates $\hat{\boldsymbol{\theta}}_a$ toward nonparametric neighborhood schedules $\boldsymbol{\mu}^*$, but that is only one of many settings in which shrinkage is useful. The same approach can be used for parametric model schedules, with shrinkage in parameter, rather than rate, space. Perhaps less obviously, EB can use parametric models and local variables $\mathbf{X}_a$ to define shrinkage targets [$\boldsymbol{\mu} = f(\mathbf{X}_a, \boldsymbol{\beta})$]. EB estimators that shrink local census estimates $\hat{\boldsymbol{\theta}}_a$ toward model predictions $\hat{\boldsymbol{\mu}} = f(\mathbf{X}_a, \hat{\boldsymbol{\beta}})$ borrow strength primarily from areas with similar socioeconomic conditions $\mathbf{X}$, rather than similar geographic locations. (The U.S. Census Bureau's SAIPE estimation procedure works in this manner.) Such flexibility makes EB a rich set of techniques for many types of demographic analysis. We recommend that demographers take note and consider using EB methods more frequently.

REFERENCES

Assunção, R.M. and I.A. Reis. 1999. "A New Proposal to Adjust Moran's I for Population Density." *Statistics in Medicine* 18:2147–62.

Assunção, R.M., I.A. Reis, and C.L. Oliveira. 2001. "Diffusion and Prediction of Leishmaniasis in a Large Metropolitan Area in Brazil With a Bayesian Space-Time Model." *Statistics in Medicine* 20:2319–35.

Assunção, R.M., C.P. Schmertmann, J.E. Potter, and S.M. Cavenaghi. 2003. "Bayesian Multivariate Spatial Estimation of Small-Area Fertility Schedules." Paper presented at the 2003 meeting of the Population Association of America, Minneapolis, MN.

Bernadinelli, L. and C. Montomoli. 1992. "Empirical Bayes Versus Fully Bayesian Analysis of Geographical Variation in Disease Risk." *Statistics in Medicine* 11:983–1007.

Bernadinelli, L., C. Pascutto, C. Montomoli, and W. Gilks. 2000. "Investigating the Genetic Association Between Diabetes and Malaria: An Application of Bayesian Ecological Regression Models With Errors in Covariates." Pp. 286–301 in *Spatial Epidemiology,* edited by P. Elliott, J.C. Wakefield, N.G. Best, and D.J. Briggs. New York: Oxford University Press.

Besag, J., J. York, and A. Mollié. 1991. "Bayesian Image Restoration With Two Applications in Spatial Statistics." *Annals of the Institute of Statistics and Mathematics* 43:1–59.

Breiman, L., J.H. Friedman, R.A. Olshen, and C.J. Stone. 1984. *Classification and Regression Trees*. Monterey CA: Wadsworth and Brooks/Cole.

Brillinger, D. 1986. "The Natural Variability of Vital Rates and Associated Statistics." *Biometrics* 42:693–734.

Carlin, B. and T. Louis. 2000. *Bayes and Empirical Bayes Methods for Data Analysis,* 2nd ed. New York: Chapman and Hall.

Clayton, D.G. and J. Kaldor. 1987. "Empirical Bayes Estimates of Age-Standardised Relative Risks for Use in Disease Mapping." *Biometrics* 43:671–81.

Coale, A.J. and P.D. Demeny. 1967. *Manual IV: Methods of Estimating Basic Demographic Measures From Incomplete Data*. New York: United Nations.

Efron, B. and C. Morris. 1972. "Empirical Bayes on Vector Observations: An Extension of Stein's Method." *Biometrika* 59:335–47.

Efron, B., R.S. Tibshirani, J.D. Storey, and V. Tusher. 2001. "Empirical Bayes Analysis of a Microarray Experiment." *Journal of the American Statistical Association* 96(456):1151–60.

Fay, R.E. and R.A. Herriot. 1979. "Estimates of Income for Small Places: An Application of James-Stein Procedure to Census Data." *Journal of the American Statistical Association* 74:269–77.

Fay, R.E. and G.F. Train. 1995. "Aspects of Survey and Model-Based Postcensal Estimation of Income and Poverty Characteristics for States and Counties." Paper presented at the 1995 meeting of the American Statistical Association, Orlando, FL. Available on-line at http://www.census.gov/hhes/www/saipe/asapaper/FayTrain95.pdf

Ferguson, B., G. Reniers, T. Araya, J.H. Jones, and E. Sanders. 2004. "Empirical Bayes Estimation of Small Area Adult Mortality Risk in Addis Ababa, Ethiopia." Paper presented at the 2004 meeting of the Population Association of America, Boston.

Fisher, R. and J. Asher. 2000. "Bayesian Hierarchical Modeling of U.S. County Poverty Rates." Paper presented at the 2000 meeting of the American Statistical Association, Indianapolis, IN. Available on-line at http://www.census.gov/hhes/www/saipe/asapaper/FisherAsher00.pdf

Gelfand, A.E., L. Zhu, and B.P. Carlin. 2001. "On the Change of Support Problem for Spatio-Temporal Data." *Biostatistics* 2:31–45.

Ghosh, M. and J.N.K. Rao. 1994. "Small Area Estimation: An Appraisal." *Statistical Science* 9(1):55–76.

Kamakura, W.A. and M. Wedel. 2004. "An Empirical Bayes Procedure for Improving Individual-Level Estimates and Predictions From Finite Mixtures of Multinomial Logit Models." *Journal of Business and Economic Statistics* 22(1):121–25.

Kendziorski, C.M., M.A. Newton, H. Lan, and M.N. Gould. 2003. "On Parametric Empirical Bayes Methods for Comparing Multiple Groups Using Replicated Gene Expression Profiles." *Statistics in Medicine* 22:3899-3914.

Kim, H., D. Sun, and R.K. Tsutakawa. 2001. "A Bivariate Bayesian Method for Improving Estimators of Mortality Rates With 2-fold CAR Model." *Journal of the American Statistical Association* 96:1506–21.

Kitada, S., T. Hayashi, and H. Kishino. 2000. "Empirical Bayes Procedure for Estimating Genetic Distance Between Populations and Effective Population Size." *Genetics* 156:2063–79.

Knorr-Held, L. and N.G. Best. 2001. A Shared Component Model for Detecting Joint and Selective Clustering of Two Diseases." *Journal of the Royal Statistical Society A* 164:73–85.

Lamm-Tennant, J., L.T. Starks, and L. Stokes. 1992. "An Empirical Bayes Approach to Estimating Loss Ratios." *Journal of Risk and Insurance* 59:426–42.

Lê, T.N. and V.K. Verma. 1997. "An Analysis of Sample Designs and Sampling Errors of the Demographic and Health Surveys." Demographic and Health Surveys Analytical Report 3. Calverton, MD: Macro International.

Lee, J.S. 1980. "Digital Image Enhancement and Noise Filtering by Use of Local Statistics." *IEEE Transactions on Pattern Analysis and Machine Intelligence* PAMI-2:165–68.

Longford, N.T. 1999. "Multivariate Shrinkage Estimation of Small Area Means and Proportions." *Journal of the Royal Statistical Society A* 162:227–45.

Manton, K.G., M.A. Woodbury, E. Stallard, W.B. Riggan, J.P. Creason, and A.C. Pellom. 1989. "Empirical Bayes Procedures for Stabilizing Maps of U.S. Cancer Mortality Rates." *Journal of the American Statistical Association* 84:637–50.

Marshall, R.J. 1991. "Mapping Disease and Mortality Rates Using Empirical Bayes Estimators." *Applied Statistics* 40:283–94.

Meza, J.L. 2003. "Empirical Bayes Estimation Smoothing of Relative Risks in Disease Mapping." *Journal of Statistical Planning and Inference* 112:43–62.

Mollié, A. and S. Richardson. 1991. "Empirical Bayes Estimates of Cancer Mortality Rates Using Spatial Models." *Statistics in Medicine* 10:95–112.

Newton, M.A., C.M. Kendziorski, C.S. Richmond, F.R. Blattner, and K.W. Tsui. 2001. "On Differential Variability of Expression Ratios: Improving Statistical Inference About Gene Expression Changes From Microarray Data." *Journal of Computational Biology* 8:37–52.

O'Hagan, A. 1994. *Kendall's Advanced Theory of Statistics Volume 2B: Bayesian Inference*. London: Edward Arnold.

Potter, J.E., C.P. Schmertmann, and S.M. Cavenaghi. 2002. "Fertility and Development: Evidence From Brazil. *Demography* 39:739–61.

Robbins, H. 1983. "Some Thoughts on Empirical Bayes Estimation." *Annals of Statistics* 11: 713–23.

Schmertmann, C.P. 2003. "A System of Model Fertility Schedules With Graphically Intuitive Parameters." *Demographic Research* 9:81–110. Available online at http://www.demographic-research.org/volumes/vol9/5

Solow, A.R. 2001. "An Empirical Bayes Analysis of Volcanic Eruptions." *Mathematical Geology* 33(1):95–102.

Stolzenberg, R.M. and D. Relles. 1989. "The Utility of Empirical Bayes Methods for Comparing Regression Structures in Small Subsamples." *Sociological Methodology* 19:183–211.

United Nations. 1983. *Manual X: Indirect Techniques for Demographic Estimation*. New York: United Nations.

Wang, F. and M.M. Wall. 2001. "Modeling Multivariate Data With a Common Spatial Factor." Research Report No. 2001-008 (2001), Division of Biostatistics, University of Minnesota, Minneapolis, MN.

Wong, L.R. 1986. "A diminuição dos nascimentos e a queda de fecudidade no Brasil dos anos pós 80" [The reduction in births and the fall of fertility in Brazil in the post-1980 period]. *Annals of the ABEP Encontro Nacional de Estudos Populacionais,* Águas de São Pedro SP.