COMPOSITE ESTIMATORS FOR MULTIPLE DOMAINS AND INFORMATION SOURCES VIA HIERARCHICAL BAYES LINEAR MODELS AND GIBBS SAMPLING

Alan M. Zaslavsky, Harvard University
Department of Statistics, 1 Oxford Street, Cambridge, MA 02138

1 Introduction

In Empirical Bayes estimation problems, estimates are desired for some quantities (domain-specific parameters) measured with error in a number of domains or for a number of units. The usual specification involves some exchangeable model for the domain-specific parameters of interest, and a sampling model for the observable measurements given the parameters. Auxiliary information on the domains is incorporated through regression modeling, yielding a mixed model with fixed effects for auxiliary variables and random effects for domains. The observables are assumed to be unbiased but noisy estimates of the parameters, while the Empirical Bayes procedure yields estimates that are biased but less noisy. Empirical Bayes estimates have been shown to be superior to unbiased estimates in many situations by loss measures such as mean squared error as well as for ranking of domains on a measure.

In this paper, we consider problems in which there are a number of sources of information pertaining to the same parameters, such as measurements of the parameters and estimates of the bias of these measurements.

Example 1. Nurses in the surgical wards of a hospital evaluate the severity of the condition of patients daily. Each ward reports the distribution of severities; annual averages are used in determining staffing and budgeting for the next year. A team from the central nursing administration does an independent evaluation of severity for a small sample of patients from each ward in order to detect bias in the statistics reported by the ward nurses. □

Example 2. The Census of Population produces measures of population for domains; these census counts are known to be biased, differentially undercounting different racial groups and geographical areas. The Post Enumeration Survey (PES), conducted in a sample of census blocks after the census, is combined with census data to generate corrected "Dual System Estimates" (DSE) of population in particular domains of interest. The DSE is also biased although probably less so than the census. Information from a number of evaluation programs and sensitivity analyses, conducted after the PES and generally involving still smaller samples, is combined to generate estimates of DSE bias (Zaslavsky 1991; Zaslavsky 1992a). □

Example 3. Federally-funded welfare programs are administered by states through local welfare offices. The local offices determine the amount of payments to which applicants are entitled in accordance with state and federal guidelines. The state welfare administrations are required to audit a sample of local determinations and calculate an estimated error rate, i.e. bias of the local determinations. Auditors from federal agencies then re-audit a small sample of the state audit cases to estimate the bias of the state audit relative to the federal interpretation of the eligibility requirements. It is desired to obtain estimates of state error rates and particularly of the probability that a state's error rate in a year exceeded a penalty threshold (Fairley, Izenman and Bagchi 1990; Hansen and Tepping 1990). □

In each of these problems, the data may be arranged along two distinct dimensions. There are two or more information sources measuring a quantity of interest or the bias of other measurements of these quantities, and there are a number of different domains for which estimates are desired, which may be defined spatially or, as in Example 3, both spatially and by time. Direct estimates of parameters may be subject to substantial biases, while estimates of the biases are based on relatively small evaluation or audit samples and therefore are subject to large sampling errors. Hierarchical Bayes methods therefore may yield composite estimators which are superior to both the biased uncorrected estimates and the unbiased but noisy estimates obtained by subtracting sample estimates of bias from the uncorrected estimates.

In this paper, we develop a general modeling framework for problems such as these, using normal hierarchical models. Methods are described for drawing inferences from these models that are applicable to a wide range of cases likely to appear in practice.

2 Model specification

We begin by specifying the overall hierarchical model in general terms, followed by a discussion of the details of the specification of level of the model, in which much of the substance of the methodology

resides. We then show how the examples presented above can be represented in terms of this model.

2.1 The hierarchical model

Our hierarchical Bayes model consists of three parts. The sampling (or measurement) model specifies the distribution of the data given the parameters. The structural model expresses the relationships among the parameters; assumptions of exchangeability or of varying degrees of similarity among different domains are represented in this part of the model by joint distributions of the parameters conditional on hyperparameters whose values are not known *a priori*. The parameters may include quantities of substantive interest that are not directly observable, as well as latent variables that are not of intrinsic interest but that facilitate specification of the model. Finally, the hyperparameter model specifies the prior distribution of the hyperparameters, which are variance components of the structural model.

The sampling model: Conditional on parameters, the data are normally distributed with known covariance and with mean related linearly to the parameters: $Y \mid X, \tau \sim N(AX, V)$.

The structural model: Conditional on variance component hyperparameters τ, the parameters are normally distributed with mean 0 and a patterned covariance matrix determined by τ: $X \mid \tau \sim N(0, S(\tau))$, $S(\tau) = \text{diag}(S^{(1)}(\tau^{(1)}), \ldots, S^{(G)}(\tau^{(G)}))$. Details of this model are discussed in Section 2.3.

The hyperparameter model: The variance component hyperparameters have a non-informative prior distribution, and the components $\tau^{(g)}$ corresponding to each block $S^{(g)}$ are independent: $\tau \sim \pi$, $\pi(\tau) = \prod \pi_g(\tau^{(g)})$. The specification of this distribution is discussed in Section 2.4.

Inferential targets: The targets of interest are linear combinations CX of the parameters X (possibly including components of X themselves).

The following notation is used above and throughout the paper.

$X = (X_1', \ldots, X_G')' =$ a vector of parameters, where $g = 1, \ldots, G$ indexes different "groups" of parameters. Furthermore, $X_g = (X_{g1}', \ldots, X_{gM_g}')'$, where K_g-vector or scalar X_{gm} is a *component* of X_g, $m = 1, \ldots, M_g$.

$Y = (Y_1', \ldots, Y_P')' =$ a data vector, where $p = 1, \ldots, P$ indexes different types of data; Y_p is a *component* of Y.

$A =$ known coefficient matrix relating X and Y.

$V =$ sampling variance-covariance matrix of Y, assumed known.

$\tau = (\tau^{(1)}, \tau^{(2)}, \ldots, \tau^{(G)}) = G$-tuple of variance components scale factors, each of which may be a scalar, possibly infinite, or a symmetric positive (semi)definite matrix.

$S^{(g)}(\tau^{(g)}) =$ structural variance-covariance matrices, known except for scale factors from $\tau^{(g)}$, $g = 1, \ldots, G$.

$\pi_g(\tau^{(g)}) =$ prior density of $\tau^{(g)}$.

$\text{diag}(S_1, \ldots, S_G) =$ block diagonal matrix.

$X_{gm(k)} =$ scalar element of component X_{gm} corresponding to domain k, $k = 1, \ldots, K_g$. (Ordinary subscripts refer to components of X or Y or corresponding blocks of matrices, while subscripts in parentheses refer to individual scalar elements within a component or block.)

Note 1: If there are K domains, some or all parameter components may be K-vectors corresponding to the K domains. Similarly, Y may contain observations corresponding to domains so that some or all of the Y_p are K-vectors. In Example 3 below, on the other hand, parameter components in different groups are of different lengths corresponding to the dimensions of a cross-classification of domains. In Example 2(b), Y_2 consists of observations for a more highly aggregated set of domains than those for Y_1, so each element is a weighted mean corresponding to a combination of the original domains.

Note 2: If the components of both X and Y are K-vectors as described in Note 1, a typical specification of A is $A = A_0 \otimes I_K$, where the operation $\otimes$ is defined by $D \otimes E = \begin{pmatrix} d_{11}E & \cdots & d_{1J}E \\ \vdots & \ddots & \vdots \\ d_{I1}E & \cdots & d_{IJ}E \end{pmatrix}$. Then $Y_{(k)} = A_0 X_{(k)}$ where $X_{(k)}, Y_{(k)} =$ vectors of elements of X, Y respectively that correspond to domain k. In words, the expectations of the observations combine the components in the same way in every domain. If not all parameters are of this form, the block of A corresponding to the K-vectors of X and Y may take this form.

Note 3: In most applications, V can be estimated from knowledge of the sampling or measurement process, from sample variances, or by standard survey sampling variance estimation techniques such as the bootstrap or jackknife. In this paper V is assumed to be known with sufficient precision that it may be treated as fixed; this is a standard assumption in Empirical Bayes modeling although not always entirely realistic. V will have a block-diagonal structure if some or all of the components Y_p are measured independently of each other.

2.2 Parametrization and sampling model

We now turn to the specification of the parametrization and sampling model for each of our examples.

Example 1 (continued): Let $Y_{1(k)}$ = average patient condition severity reported by nursing staff in ward k, $Y_{2(k)}$ = estimate of the bias in $Y_{1(k)}$ based on evaluations by nursing administration staff, X_{11} = average severity across all wards, X_{12} = average bias in severity ratings across all wards, $X_{21(k)}$ = deviation from average severity in ward k, $X_{22(k)}$ = deviation from average bias in ward k staff ratings. Then $Y_1 \sim N((X_{11}+X_{12})\mathbf{1}+X_{21}+X_{22}, V_{11})$, $Y_2 \sim N(X_{12}\mathbf{1}+X_{22}, V_{22})$, $V_{12} = \mathrm{Cov}(Y_1, Y_2) \approx 0$ if the second sample is small. In matrix notation, $A = (A_0 \otimes \mathbf{1}_K \quad A_0 \otimes I_K)$, where $A_0 = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$. The quantities of interest are the true ward severities $X_{11}+X_{21(k)}$ (for assessment of staffing requirements) and the ward biases $X_{12}+X_{22(k)}$ (for feedback to the ward on the accuracy of its ratings), so $C = \begin{pmatrix} \mathbf{1} & 0 & I & 0 \\ 0 & \mathbf{1} & 0 & I \end{pmatrix}$. □

Example 2(a) (continued): Suppose that the same domains are used for calculation of estimated undercount rates and of estimated biases of the DSE. This is true if the domains are the Evaluation Poststrata (EPS), which are large domains defined for purposes of summarizing and presenting the relatively sparse evaluation data.

The following specification focuses on relative undercount rates, defined as the ratio of error to true population share for each domain and measurement procedure. The data are Y_1 = estimated census undercount rates as estimated using the DSE as the standard, and Y_2 = estimated relative bias of the DSE estimates from evaluation programs. Let X_{g1} = effects for census relative undercount rates, X_{g2} = effects for expectation of DSE undercount rates (i.e. the relative bias of the DSE for domain shares), $K_g = K$, $g = 1, \ldots, G$, so the census undercount rates are $\sum_g X_{g1}$ and the DSE biases are $\sum_g X_{g2}$.

Then (assuming that the undercount rates are all small) $Y_1 \sim N(\sum_g (X_{g1} - X_{g2}), V_d)$, $Y_2 \sim N(\sum_g X_{g2}, V_b)$. Furthermore, $A = \mathbf{1}_G' \otimes A_0 \otimes I_K$, where $A_0 = \begin{pmatrix} 1 & -1 \\ 0 & 1 \end{pmatrix}$, and $V = \mathrm{diag}(V_d, V_b)$, where V_d, V_b are sampling covariance matrices for estimated census undercount rates and for estimated biases. The quantities of interest are the census undercount rates relative to true population shares, $\sum_g X_{g1} = CX$, $C = \mathbf{1}_G' \otimes (1\ 0) \otimes I_K$. □

Example 2(b) (continued): The DSE's are also calculated for more refined domains, the poststrata, but evaluation data are not generated at this level of detail. Extending the specification of Example 2(a), Y_p is a K_p-vector, $p = 1, 2$, where K_1 is the number of poststrata and K_2 is the number of EPSs. Define X as in Example 2(a), except that the domains are now poststrata. Then $A = \mathbf{1}_G \otimes \begin{pmatrix} I & -I \\ 0 & B \end{pmatrix}$, where B is the matrix mapping poststratum to EPS undercount rates, b_{jk} = fraction of EPS j that comes from poststratum k. The specification of the quantities of interest is unchanged. □

Example 3 (continued): A natural decomposition of effects for a cross classification is into an intercept, main effects, and interaction. Let X_1 = grand mean (intercept) effect, X_2 = state effect, X_3 = year effect, X_4 = residual effect. For each group $g = 1, \ldots, 4$, let X_{g1} = component for relative error in payments (error/total payments), X_{g2} = component for bias of state estimates of relative error. Then $K_1 = 1, K_2 = K_S$ = number of states, $K_3 = K_Y$ = number of years, and $K_4 = K_S K_Y$.

Let $Y_{1(sy)}$ = state estimate of mean relative error in payments in state s, year y, and $Y_{2(sy)}$ = corresponding federal estimate of bias of state estimates of mean relative error. (Note that the components of X have four different lengths (1, K_S, K_Y, $K_S K_Y$), and group 4 components are doubly indexed.)

Then under an additive model for these effects, and assuming independent sampling in each state and year, $Y_{1(sy)} \sim N(X_{11} + X_{21(s)} + X_{31(y)} + X_{41(sy)} + X_{12} + X_{22(s)} + X_{32(y)} + X_{42(sy)}, V_{1(sy)})$ and $Y_{2(sy)} \sim N(X_{12} + X_{22(s)} + X_{32(y)} + X_{42(sy)}, V_{2(sy)})$. If the doubly-indexed vectors $\{X_{41(sy)}\}, \{X_{42(sy)}\}, \{Y_{1(sy)}\}, \{Y_{2(sy)}\}$ are ordered columnwise in the component vectors, e.g. $X_{41(11)}, \ldots, X_{41(K_S 1)}, \ldots, X_{41(1K_Y)}, X_{41(K_S K_Y)}$, then $A = (A_0 \otimes \mathbf{1}_{K_s K_y} \quad A_0 \otimes \mathbf{1}_{K_s} \otimes I_{K_y} \quad A_0 \otimes I_{K_y} \otimes \mathbf{1}_{K_s} \quad A_0 \otimes I_{K_s K_y})$, where $A_0 = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$. The parameters of interest are the annual error rates of the states, $X_{11} + X_{21(s)} + X_{31(y)} + X_{41(sy)}$, so $C = (\mathbf{1}_{K_s K_y} \ 0 \ \mathbf{1}_{K_s} \otimes I_{K_y} \ 0 \ I_{K_y} \otimes \mathbf{1}_{K_s} \ 0 \ I_{K_s K_y} \ 0)$. □

2.3 Specification of the structural model

While the sampling model (particularly the sampling variances V) describes the statistical properties of the data-collection procedures, the structural model embodies the substantive scientific models underlying the analysis. The key ideas motivating the specification of $S(\tau)$ are those of proximity or interrelatedness of domains and of exchangeability of parameters (Lindley and Smith 1972), ideas that must be rooted in some prior knowledge about these

domains and parameters (or a choice not to reflect such knowledge). In this section we consider principles relevant to generating scientifically sensible specifications for $S(\tau)$.

The parameter vector X is a concatenation of components $X_{11}, \ldots, X_{GM_G}$. The specification of $S(\tau)$ implies that components $X_{gm}, X_{g'm'}, g \neq g'$ belonging to different groups are conditionally independent, given τ, in the structural model. Components in the same group g have conditional variance-covariance $\mathrm{Cov}(X_{gm}, X_{gm'} \mid \tau) = S^{(g)}_{mm'}(\tau^{(g)})$.

The specification of the diagonal component blocks $S^{(g)}_{mm}(\tau^{(g)})$ represents some notion of exchangeability of elements of X_m, and similarly the off-diagonal blocks $S^{(g)}_{mm'}(\tau^{(g)})$, $m \neq m'$, represent some notion of exchangeable degree of relationship between X_{gm} and $X_{gm'}$. Exchangeability is relevant because $X_m, m \in M_g$ are all vectors of parameters for the same set of domains.

We specify the blocks as $S^{(g)}_{mm'}(\tau^{(g)}) = \tau^{(g)}_{mm'} U^{(g)}_{mm'}$ where $U^{(g)}_{mm'}$ is a prespecified, constant matrix and $\tau^{(g)}_{mm'}$ is a scalar to be estimated. We assume that $U^{(g)}_{mm'} = U^{(g)}$, a single matrix for all blocks, so $S^{(g)}(\tau^{(g)}) = \tau^{(g)} \otimes U^{(g)}$. This specification guarantees that if $\tau^{(g)}$ is positive (semi)definite then $S^{(g)}(\tau^{(g)})$ is as well, and it is further supported by arguments given below. (A more general specification meeting the definiteness condition allows $U^{(g)}_{mm} \neq U^{(g)}_{m'm'}$ and sets $U^{(g)}_{mm'} = (U^{(g)}_{mm})^{1/2}(U^{(g)}_{m'm'})^{1/2}$; we will not pursue this generalization further.)

We now turn to considerations to be taken into account in the specification of $U^{(g)}$.

Flat priors. A uniform prior on X_m is a reasonable specification for a grand mean effect (such as the overall mean state error in Example 3), for domain-specific parameters that are estimable from the data but cannot way be regarded as drawn from a distribution with known mean, or for other parameters that can in no way be regarded as part of an exchangeable collection (such as a single regression parameter – but see below, "Smoothing Models").

Interdomain covariances for a single parameter component. In many cases the parameters for individual domains within the same group of parameters may be regarded as *a priori* varying around 0. The following briefly describes some approaches to elicitation of $U^{(g)}$, dealt with in more detail in Zaslavsky (1992b).

Independent or exchangeable domains may be modeled as independent, so $U^{(g)}$ is diagonal.

Subjective similarity: subjective notions of similarity of domains may be represented by writing down a valid correlation matrix representing the degree of perceived similarity.

Formal models of proximity may be appropriate when there is spatial or temporal ordering of domains.

Random effects models may be appropriate when domains are cross-classified on some categorical variables.

Smoothing models extend this construction to continuous or mixed covariates.

Examples. We continue here the discussion of the specification of the models for the examples, in particular the specification of $U^{(g)}$.

Example 1, continued: We treat all wards as exchangeable and have two groups of parameters, intercepts (*a priori* uniformly distributed) and ward effects (exchangeable).

Example 2(a,b), continued: The poststrata or EPSs are not fully exchangeable, but a structural covariance matrix U may be elicited as outlined above, incorporating ideas on random effects and smoothing models and constrained sums. Since we do not know *a priori* the relative contributions of dividing the components of interest (census undercount and DSE bias) into two or more components as described above, it would be appropriate to specify several groups, corresponding to various specifications of $U^{(g)}$.

Example 3, continued: First a structural covariance U_S for states may be elicited using characteristics of the states as described above, or we might choose $U_S = I$ (for policy reasons, if we do not wish to make a state's estimate particularly dependent on those for nearby or similar states). Define U_Y as described above to correspond to an autocorrelation model. Then there are four groups of components, for mean, state, year, and residual effects, $U^{(1)} = 1$, $U^{(2)} = U_S$, $U^{(3)} = U_Y$, $U^{(4)} = U_S \otimes U_Y$.

2.4 Prior distributions of variance component hyperparameters

The choice of prior distribution π is limited by the requirement that the posterior distribution $\tau \mid Y$ be proper (except in the case of *a priori* infinite components signifying flat prior distributions) and nondegenerate. A degenerate posterior would imply certain knowledge of the values of some variance components. This claim of certainty cannot be justified from the data, since any observed Y are consistent with all values of any variance component (except possibly zero). If the posterior distribution of a variance component is improper for all possible

data, putting infinite mass at infinity (i.e. above any given upper bound), then no borrowing of strength takes place across domains, regardless of the degree of similarity between domains demonstrated by the data.

We will consider priors of the form

$$\pi_g(\tau^{(g)}) = \mid \tau^{(g)} \mid^{\alpha_g} \mid R(\tau^{(g)}) \mid^{\gamma_g} \qquad (1)$$

where $R(\tau^{(g)})$ is the correlation matrix corresponding to $\tau^{(g)}$. These priors are scale-invariant; we will find in Section 3 that they are tractable as well. Zaslavsky (1992b) develops principles for selecting the parameters of these priors.

3 Model fitting and posterior inference via Gibbs sampling

In much of the literature on hierarchical models, starting with Lindley and Smith (1972), a semi-Bayesian inference is conducted by fixing hyperparameters at maximum-likelihood or method-of-moments estimates and basing inference on the posterior distributions of parameters. When the hyperparameters can be fairly precisely estimated from the data, this procedure is adequate. Even when there is some uncertainty in hyperparameter estimates, there are approximate methods for estimating the resulting additional variance in parameters distributions using resampling (Laird and Louis 1987) or asymptotic analytical approximations for special cases (Prasad and Rao 1990).

However, these methods may be difficult to apply to the models proposed here. First, these models lack the i.i.d. structure required for some bootstrap methods (although "resampled" data could be simulated by sampling $Y^* \sim N(Y, V)$). The complexity of the models make general asymptotic approximations difficult. Most important, the hierarchical structure is deliberately designed to permit overparametrization of the model, in the sense that there may be many more parameters than data points, and therefore there may be random effects that cannot be readily distinguished using the data. Under these circumstances, asymptotic assumptions are not reliable.

Instead, we prefer to pursue a fully Bayesian strategy that draws from posterior distributions, conducting inferences based on samples from the posterior. While this approach is computationally expensive, it yields an inference whose conceptual basis is unambiguous.

Bayesian inferences are statements about the posterior distributions of functions of the parameters (CX) given the data (Y). Bayesian simulations require draws from these posterior distributions, which may be summarized for a variety of inferences.

A simple and general computational procedure for drawing from posterior distributions is the method of alternating conditional distributions or Gibbs sampling (Carlin and Gelfand 1990; Gelfand and Smith 1990; Geman and Geman 1984). In this application, the vector of unknowns (X, τ) partitions naturally into two components X and τ. In the Gibbs sampling procedure, we alternately draw $X^{[t]}$ from the conditional distribution $X \mid \tau = \tau^{[t]}, Y$ and $\tau^{[t+1]}$ from the conditional distribution $\tau \mid X = X^{[t]}$, where t indexes iterations. With successive iterations the simulation distribution of $(X^{[t]}, \tau^{[t]})$ converges to the joint posterior distribution.

The series of draws from the posterior distributions of τ and X may be summarized to provide the desired inferences. Summary statistics may be accumulated simultaneously with sampling of $X^{(t)}$.

4 Results: Estimation of Census Undercount for Poststrata

The data and models for estimation of undercount in the census have been described in Examples 2(a)–(b) of the preceding sections. Zaslavsky (1992a) fits a model equivalent to those proposed in this paper, with a single parameter group, to data of the form described in Example 2(a), i.e. assuming that all data sources are available for the same set of domains (EPSs). In an extensive simulation experiment, the hierarchical Bayes estimators were found to have good coverage properties and to have smaller squared-error loss (on the average across conditions) than several competing non-Bayesian estimators. The Bayesian and non-Bayesian estimators were applied to two Census Bureau data sets, as well.

Zaslavsky (1992b) extends those empirical results to more complicated models and data sets; the application and results are outlined in this section. The data are of the form of Example 2(b), and are derived from a repoststratified data set prepared in 1992 at the Census Bureau, based on corrected data from the 1990 census and PES. Undercount estimates are provided for 48 poststratum groups (PSG), i.e. poststrata collapsed acress age and sex, excluding three for Asians, Pacific Islanders and Native Americans for whom data were too sparse to compile bias estimates. Bias estimates are only available for ten EPSs, a more aggregated unit.

The models we consider allow for several groups in the X vector, each with its corresponding variance components $\tau^{(g)}$ and structural matrix $U^{(g)}$.

We will fit a series of models, each incorporating different specifications of $U^{(g)}$, including the following:

I: independent homoskedastic errors;

$U_H = \text{diag}(Y_1) + .04$: independent errors with variance proportional to (estimated) undercount;

U_S: a "similarity" matrix with equal weights for tenure, race, urbanicity, region, and independent error, the variables (excluding the independent errors) define the poststratum groups.

Fitted values (posterior means) for the true relative undercount rate for each PSG may be plotted against the raw estimates under each of several different models. Regressions of fitted values on raw estimates were calculated (excluding one outlier, PSG 48). Their slopes may be regarded as average shrinkage factors, i.e. average factors by which raw data are multiplied to obtain fitted values. (These coefficients are only calculated as a summary and are not part of the model used for smoothing.)

As the models are made more complicated (with addition of groups of components), the slopes of the regression lines become larger, indicating that there is less smoothing of the adjustment factors. We would expect to find less smoothing under a model which detects systematic patterns in the data. The regression coefficients (and average posterior standard errors of adjustment factors) are: I only, 0.56 (.019); I and U_H, 0.60 (.021); I and U_S, 0.64 (.013); I, U_H, and U_S, 0.66 (.014). The large reduction in posterior standard errors when U_S is added to the specification suggests that the model finds this a good fit and when it can "pull" the estimates toward it, relatively little residual variance is left.

References

Carlin, B. P. and Gelfand, A. E.(1990), "Approaches for Empirical Bayes Confidence Intervals," *Journal of the American Statistical Association*, 85, 105–114.

Fairley, W. B., Izenman, A. J. and Bagchi, P.(1990), "Inference for Welfare Quality Control Programs," *Journal of the American Statistical Association*, 85, 874–890.

Gelfand, A. E. and Smith, A. F. M.(1990), "Sampling-Based Approaches to Calculating Marginal Densities," *Journal of the American Statistical Association*, 85, 398–409.

Geman, S. and Geman, D.(1984), "Stochastic Relaxation, Gibbs Distributions, and the Bayesian Restoration of Images," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 6, 721–741.

Hansen, M. H. and Tepping, B. H.(1990), "Regression Estimates in Federal Welfare Quality Control Programs," *Journal of the American Statistical Association*, 85, 856–863, 871-873.

Laird, N. A. and Louis, T. A.(1987), "Empirical Bayes Confidence Intervals Based on Bootstrap Samples," *Journal of the American Statistical Association*, 82, 739–750.

Lindley, D. V. and Smith, A. F. M.(1972), "Bayes Estimates for the Linear Model," *Journal of the Royal Statistical Society, Series B*, 34, 1–41.

Prasad, N. G. N. and Rao, J. N. K.(1990), "The Estimation of the Mean Squared Error of Small-area Estimators," *Journal of the American Statistical Association*, 85, 163–171.

Zaslavsky, A. M.(1991), "Combining Census and Dual-system Estimates of Population," *Proceedings, Section on Survey Research Methods, American Statistical Association.*

_____(1992a), "Combining Census, Dual-system, and Evaluation Study Data to Estimate Population Shares," submitted to *Journal of the American Statistical Association.*

_____(1992b), "Composite Estimators for Multiple Domains and Information Sources via Hierarchical Bayes Models and Gibbs Sampling," Technical Report, Harvard University Department of Statistics .