# Hierarchical and Empirical Bayes Methods for Adjustment of Census Undercount: The 1988 Missouri Dress Rehearsal Data

## G.S. DATTA, M. GHOSH, E.T. HUANG, C.T. ISAKI, L.K. SCHULTZ and J.H. TSAY[1]

## ABSTRACT

The present article discusses a model-based approach towards adjustment of the 1988 Census Dress Rehearsal Data collected from test sites in Missouri. The primary objective is to develop procedures that can be used to model data from the 1990 Census Post Enumeration Survey in April, 1991 and smooth survey-based estimates of the adjustment factors. We have proposed in this paper hierarchical Bayes (HB) and empirical Bayes (EB) procedures which meet this objective. The resulting estimators seem to improve consistently on the estimators of the adjustment factors based on dual system estimation (DSE) as well as the smoothed regression estimators.

KEY WORDS: Post Enumeration Survey; Adjustment factors; Dual system estimation; Hierarchical Bayes; Empirical Bayes; Variance components; EBLUP's; Regression estimates; Standard errors.

## 1. INTRODUCTION

The present article discusses a model-based approach towards adjustment of the 1988 Census dress rehearsal data collected from test sites in Missouri. The main objective behind this exercise is to develop procedures that can be used to model data from the 1990 Census Post Enumeration Survey (PES) in April, 1991, and smooth survey-based estimates of the so-called "raw adjustment factors". These raw adjustment factors which are ratios of estimates of the unknown total population to the corresponding 1990 Census count, are computed at various levels of aggregation (geographic areas such as cities, suburbs, *etc.*) crossed by various demographic categories (such as age, sex, race, *etc.*). The cross-classified categories are called poststrata.

Before proceeding further, a brief historical anecdote is in order. Adjustment of 1980 decennial census counts in the United States has been a topic of heated debate for nearly a decade. Despite the intensive efforts and the massive expenditure incurred by the U.S. Bureau of the Census to achieve near-complete coverage in the 1980 Census, there have been many lawsuits against the Bureau by individual states and cities demanding revision of the reported counts. In one such instance of litigation, by now well-publicized to the Statistics community in the articles of Ericksen and Kadane (1985) and Freedman and Navidi (1986), New York City among others sued the Census Bureau, and many reputed statisticians appeared as expert witnesses on either side. In particular Ericksen and Kadane appeared on the plaintiff's side, and proposed a model-based approach towards the adjustment of census counts. They advocated shrinking the adjustment factors calculated on the basis of the PES data towards some suitable regression model. This approach documented in Ericksen and Kadane (1985) is similar to the one considered in Fay and Herriot (1979) or Morris (1983). Despite criticism of the Ericksen-Kadane approach by some statisticians (most severely by Freedman and

---

[1] G.S. Datta, University of Georgia, Athens, GA 30602; M. Ghosh, University of Florida, Gainesville, FL 32611; E.T. Huang, C.T. Isaki, L.K. Schultz and J.H. Tsay, U.S. Bureau of the Census, Washington, DC 20233.

Navidi (1986)), most people recognize the importance of the model-based approach for adjustment. Indeed, in this article, barring a few differences in the assumptions, to be pointed out later in section 2, we use the Fay-Herriot or the Ericksen-Kadane model for the analysis of the 1988 Missouri Dress Rehearsal data. A different model-based approach which does not include co-variates is given in Cressie (1989).

A good description of the PES conducted as part of the 1988 Missouri Dress Rehearsal can be found in Childers and Hogan (1990). Hogan and Wolter (1988) discuss the categories of error that occur in a PES and a means of their evaluation. Basically, the PES design consists of a single stage stratified sample of blocks and dual system estimation of the number of persons by poststrata.

In the present article, we begin at the point where a set of estimated raw adjustment factors and their covariances from the PES are available for modelling based on the 1988 Census Dress Rehearsal Data from the Missouri test sites. It is also assumed that a set of possible explanatory variables defined at the poststrata level and to be used in regression are also available. There are two geographic areas under consideration: the city of St. Louis which is a large central city, and East Central Missouri, which is a collection of areas of moderate population size. In defining the poststrata in St. Louis, persons were classified into the following demographic categories: (i) race: white non-hispanic and others, (ii) owners and non-owners (renters) of dwellings, (iii) sex: male and female, (iv) age groups: 0-9, 10-19, 20-29, 30-44, 45-64 and 65 + . This led to a total of $2 \times 2 \times 2 \times 6 = 48$ adjustment factors for St. Louis. In East Central Missouri, the sex and the age-group categories remained the same as in St. Louis, but instead of (i) and (ii), a new category (i)' classifying persons as (a) White non-Hispanic in Tape Address Register (TAR) areas, (b) White non-Hispanic in non-TAR areas, and (c) others in all areas were introduced. For East Central Missouri, a total of $3 \times 2 \times 6 = 36$ adjustment factors were calculated. Thus, a total of 84 adjustment factors were used for modelling. Within each area, estimated adjustment factors were correlated due to the use of a block cluster sampling scheme. This led to a block-diagonal sample covariance matrix of the adjustment factors of dimensions $48 \times 48$ and $36 \times 36$ corresponding to St. Louis and East Central Missouri, respectively.

In Section 2 of this article, we describe a general model-based method for obtaining smoothed adjustment factors, and the associated standard errors. Both the hierarchical and empirical Bayes methods are used. The EB method can also be regarded as a variance components method (see for example Harville (1985)). The formulas for posterior standard errors associated with the HB estimators are also provided. We may point out here that an EB method when employed naively can lead to serious underestimates of the associated standard errors. This is due to the fact that a naive EB method does not take into account the uncertainty due to estimation of the unknown variance components. However, Kackar and Harville (1984), and Prasad and Rao (1990) have suggested interesting approximations to the estimated mean squared errors (MSE's) of the EB estimators. Following their principle, we have derived formulas for the estimated MSE's in the present context. We have also pointed out in this section how some (though not all) of the criticisms levelled against the Ericksen-Kadane (1985) procedure by Freedman and Navidi (1986) can be avoided in the present context.

In Section 3, we have analyzed the actual data. The sample estimates, the HB estimates, the EB estimates and the regression estimates of the adjustment factors are all provided. Also, the associated standard errors are given. Both the HB method and the EB methods which take into account the uncertainty due to unknown prior parameters stand on par in their performance, and enjoy a clear-cut superiority over the raw estimates as well as the regression estimates in reducing the estimated standard errors.

Finally, some of the technical details of this paper are given in the Appendix.

## 2. HB AND EB ESTIMATION

This section describes the general HB and EB estimation procedures for certain hierarchical models. The specific application to estimation of adjustment factors is considered in Section 3.

The following hierarchical model is considered:

I.    $Y \mid \Theta, \beta, \sigma^2 \sim N(\Theta, V)$, where $V$ is a known $m \times m$ positive definite matrix;

II.   $\Theta \mid \beta, \sigma^2 \sim N(X\beta, \sigma^2 I)$;

III.  $\beta$ and $\sigma^2$ are marginally independent with $\beta$ uniform $(R^p)$ and $\sigma^2$ uniform $(0,\infty)$.

The HB analysis is based on I-III. In the absence of precise prior information on $\beta$ and $\sigma^2$, we prefer the use of diffuse priors in III. We also analyzed the data with the prior *pdf* of $\sigma^2$ proportional to $\sigma^{-2}$ on $(0, \infty)$. The results were quite similar and are not reported. The following theorem is proved.

**Theorem 1.** Consider the model given in (I) – (III). Write $\Sigma = V + \sigma^2 I$. Suppose $m \geq p + 3$. Then (i) the conditional pdf of $\Theta$ given $\sigma^2$ and $Y = y$ is $N(GV^{-1}y, G)$, where

$$G = V - V\Sigma^{-1}V + V\Sigma^{-1}X[X^T\Sigma^{-1}X]^{-1}X^T\Sigma^{-1}V; \qquad (2.1)$$

(ii) the conditional pdf of $\sigma^2$ given $Y = y$ is

$$f(\sigma^2 \mid y) \propto \mid \Sigma \mid^{-1/2} \mid X^T\Sigma^{-1}X \mid^{-1/2} \exp(-1/2\, y^T F y), \qquad (2.2)$$

where

$$F = \Sigma^{-1} - \Sigma^{-1}X[X^T\Sigma^{-1}X]^{-1}X^T\Sigma^{-1}. \qquad (2.3)$$

The proof of the theorem is deferred to the appendix. Using formulas for conditional expectations and variances, one then gets

$$E(\Theta \mid y) = E[E(\Theta \mid \sigma^2, y) \mid y] = (E(GV^{-1} \mid y))\, y; \qquad (2.4)$$

$$V(\Theta \mid y) = V[E(\Theta \mid \sigma^2, y) \mid y] + E[V(\Theta \mid \sigma^2, y) \mid y] = V(GV^{-1}y \mid y) + E(G \mid y). \qquad (2.5)$$

Using (2.2) and (2.3), one obtains $E(\Theta \mid y)$ and $V(\Theta \mid y)$ from (2.4) and (2.5) via numerical integration.

The calculations involved in (2.1) – (2.3) can be somewhat simplified when one uses the spectral decomposition theorem for $V$. Thus, $V = PDP^T$, where $D = \text{Diag}(d_1, \ldots, d_m)$, $d_i$ being the eigenvalues of $V$, and $P = (\xi_1 \ldots, \xi_m)$, $\xi_i$ being the corresponding orthonormal eigenvectors. Using the orthogonality of $P$, one now gets

$$| \Sigma | = | \sigma^2 I + D | = \prod_{i=1}^{m} (\sigma^2 + d_i);$$

$$\Sigma^{-1} = P(\sigma^2 I + D)^{-1} P^T;$$

$$X^T \Sigma^{-1} X = (P^T X)^T (\sigma^2 I + D)^{-1} (P^T X);$$

$$F = P(\sigma^2 I + D)^{-1} P^T - P(\sigma^2 I + D)^{-1} (P^T X) \times$$

$$\left[ (P^T X)^T (\sigma^2 I + D)^{-1} (P^T X) \right]^{-1} (P^T X) (\sigma^2 I + D)^{-1}.$$

The actual numerical integration over $\sigma^2$ which needs evaluation of the integrand at different values of $\sigma^2$, is somewhat simplified since $P$ and $X$ are known and $\sigma^2 I + D$ is a diagonal matrix.

Next we consider EB estimation. Then, one does not use III. First a Bayes estimator, *i.e.* the posterior mean of $\Theta$ is obtained from I and II assuming $\beta$ and $\sigma^2$ to be known. This estimator is given by

$$\hat{\Theta}^B = E(\Theta \mid Y, \beta, \sigma^2)$$

$$= (V^{-1} + \sigma^{-2} I)^{-1} (V^{-1} Y + \sigma^{-2} X \beta)$$

$$= \Sigma^{-1} (\sigma^2 Y + V X \beta). \tag{2.6}$$

The corresponding posterior variance is given by

$$V(\Theta \mid Y, \beta, \sigma^2) = (V^{-1} + \sigma^{-2} I)^{-1} = V - V \Sigma^{-1} V.$$

However, in practice, $\beta$ and $\sigma^2$ are unknown, and are estimated via the maximum likelihood method from the marginal distribution of $Y$ which is $N(X\beta, \Sigma)$. These MLE's are denoted by $\hat{\beta}$ and $\hat{\sigma}^2$, where $\hat{\beta} = (X^T \hat{\Sigma}^{-1} X)^{-1} X^T \hat{\Sigma}^{-1} Y$, $\hat{\Sigma} = V + \hat{\sigma}^2 I$. Substituting such estimators of $\Sigma$, $\sigma^2$ and $\beta$ in (2.6), an EB estimator of $\Theta$ is found as

$$\hat{\Theta}^{EB} = \hat{\Sigma}^{-1} (\hat{\sigma}^2 Y + V X \hat{\beta}) = X \hat{\beta} + \hat{\sigma}^2 \hat{\Sigma}^{-1} (Y - X \hat{\beta}). \tag{2.7}$$

The estimator given in (2.7) is also obtainable as an estimated best linear unbiased predictor (EBLUP). First assume that $\sigma^2$ is known, and find the BLUP $\hat{\Theta}^{BLUP} = X \tilde{\beta} + \sigma^2 \Sigma^{-1} (Y - X \tilde{\beta})$ of $\Theta$ where $\tilde{\beta} = (X^T \Sigma^{-1} X)^{-1} X^T \Sigma^{-1} Y$. Next estimate $\sigma^2$ by $\hat{\sigma}^2$, its MLE and correspondingly $\Sigma$ by $\hat{\Sigma}$. Substitution of $\hat{\sigma}^2$, and $\hat{\Sigma}$ in place of $\sigma^2$ and $\Sigma$ in $\hat{\Theta}^{BLUP}$ results in the EBLUP $\hat{\Theta}^{EB}$.

A naive EB estimator of the variance matrix of $\hat{\Theta}^{EB}$ is $V - V \hat{\Sigma}^{-1} V$. This is a gross underestimation of the variance matrix since uncertainty due to estimation of $\beta$ and $\sigma^2$ is not taken into account. If $\sigma^2$ is assumed known, and $\beta$ is assigned a uniform prior on $R^p (m \geq p + 3)$, then the HB estimator of $\Theta$ is the same as $\hat{\Theta}^{BLUP}$, and the posterior variance matrix is then $M = V - V \Sigma^{-1} V + V \Sigma^{-1} X (X^T \Sigma^{-1} X)^{-1} X^T \Sigma^{-1} V$. This implies immediately that $E[(\hat{\Theta}^{BLUP} - \Theta)(\hat{\Theta}^{BLUP} - \Theta)^T] = M$, where expectation is taken over the

joint distribution of $Y$ and $\Theta$ given in I and II. Thus, in the Bayesian language, $V\Sigma^{-1}$ $X(X^T\Sigma^{-1}X)^{-1}X^T\Sigma^{-1}V$ can be interpreted as the excess in the posterior variability due to the uncertainty involved in $\beta$, while using the classical terminology, the same phenomenon can be interpreted as the excess in the MSE due to the same uncertainty.

We have the additional problem of tackling unknown $\sigma^2$. The Bayesian method enables us to find the posterior distribution of $\sigma^2$ given $Y = y$, while even without introducing a prior for $\Theta$, it is still possible to find an approximation to the MSE of $\hat{\Theta}^{EB}$ by adapting an argument of Kackar and Harville (1984) or Prasad and Rao (1990).

The necessary theorem whose proof is deferred to the Appendix is given below.

**Theorem 2.** An approximate estimate of MSE of $\hat{\Theta}^{EB}$ is given by

$$\widehat{\text{MSE}}(\Theta^{EB}) \doteq V - V\hat{K}V + \left(V\hat{K}^3V\right)\left[2(\text{tr}\hat{\Sigma}^{-2})^{-1}\right], \tag{2.8}$$

where

$$\hat{K} = \hat{\Sigma}^{-1} - \hat{\Sigma}^{-1}X\left(X^T\hat{\Sigma}^{-1}X\right)^{-1}X^T\hat{\Sigma}^{-1}. \tag{2.9}$$

The third term in the right hand side of (2.8) can be interpreted as the excess in the mean squared error due to uncertainty in estimating $\sigma^2$. A general decomposition of the prediction error is given in Harville (1985).

Although the posterior variances $V(\Theta \mid y)$ associated with the HB estimator $\hat{\Theta}^{HB}$ of $\Theta$ and the estimated MSE of the EB estimator $\hat{\Theta}^{EB}$ of $\Theta$ are motivated from two distinct inferential philosophies, one common thread tying the two is that they both attempt to incorporate the uncertainty due to estimation of the model variance. For a better understanding of this, note that writing $K = \Sigma^{-1} - \Sigma^{-1}X(X^T\Sigma^{-1}X)^{-1}X^T\Sigma^{-1}$,

$$E\left[V(\Theta \mid \sigma^2, y)\right] = G = V - VKV \tag{2.10}$$

and $E(G \mid y)$ is approximated by $V - V\hat{K}V$ which is one of the two terms given in (2.8). Also, $E(\Theta \mid \sigma^2, y) = GV^{-1}y$, and it can be shown after some simplification that $GV^{-1} = I - VK$. Thus, $V(GV^{-1}y \mid y) = VV(K \mid y)V$, and $V(K \mid y)$ is apparently approximated by $\hat{K}^3[2(\text{tr}\hat{\Sigma}^{-2})^{-1}]$. However, as evidenced later in the numerical calculations of Section 3, MSE approximation of $\hat{\Theta}^{EB}$ need not match $V(\Theta \mid y)$ perfectly.

In Ericksen and Kadane (1985) one assumption involved was that of known $\sigma^2$. Freedman and Navidi (1986) insisted on estimation of $\sigma^2$, and we have in Theorems 1 and 2 accounted for this source of uncertainty both in a Bayesian and frequentist way. It should be noted that unlike previous work that addressed the estimation of net undercount of total population at the city and balance of state level, our interests lay in the estimation of adjustment factors at finer levels of detail. Operationally, adjustment at the finer levels allows for considerable savings in terms of time and computer costs as census files need to be used only once. Adjustment models using higher levels of geography would require several passes through the census data because they would require a method of distributing the undercount to lower levels of geography. Finally, correlation in the error structure allows the possibility of a non-diagonal $V$, another important generalization of the Fay-Herriot (1979) or Ericksen-Kadane (1985) model. Thus, the Freedman-Navidi criticism of lack of correlation across estimated adjustment factors does not hold against the present set up. The remaining main criticism of assuming the components of $V$ to be known, whereas in reality these are sample based estimates, is yet to be resolved. Efforts are now being made to model the components of $V$ as a function of

variables such as the number of sample persons, the initial regression predictor, *etc.* It is hoped that such models will stabilize the estimated variances by reducing their variance.

Along with the HB and EB estimators of $\Theta$, there are also the regression estimators given by $\hat{\Theta}^{REG} = X(X^T\hat{\Sigma}^{-1}X)^{-1}X^T\hat{\Sigma}^{-1}Y$. The associated variance-covariance matrix is given by $M_o - \hat{\sigma}^2(M_o\hat{\Sigma}^{-1} + \hat{\Sigma}^{-1}M_o^T - I)$, where $M_o = X(X^T\hat{\Sigma}^{-1}X)^{-1}X^T$.

## 3. DATA ANALYSIS

Let $Y_i = DSE_i/Census_i =$ adjustment factor $i$, $i = 1, \ldots, 84$, and $Y = (Y_i, \ldots, Y_{84})^T$. The set of explanatory variables $X$ is quite large when all possible interactions are considered. To simplify the analysis, experts at the Census Bureau were consulted and a reduced set of 22 potential explanatory variables were considered for modelling purposes. (See Huang *et al.* 1991). The number of potential explanatory variables was also limited by the capability of the computer. The present model was selected using a best subset regression procedure with minimum Mallows' $C_p$ as the criterion over a set of 22 possible explanatory variables. Because the computer software required the input data to be in the ordinary least squares situation, we transformed the dependent and explanatory variables in the usual manner. Also, because $\sigma^2$ is unknown, an interative procedure was used.

As an aside, in selecting explanatory variables in the modelling process of adjustment factors for the 1990 Census, a slightly different procedure was used. In 1990, several explanatory variables were forced into the model and a best subset procedure was used to select additional explanatory variables. The change in procedure was made to counteract the potential for understating $\sigma^2$. (See Isaki *et al.* 1991).

The $X$ matrix obtained via best subsets regression is of the form $X = (1_{84}, X_2, X_3, X_4, X_5, X_6, X_7, X_8, X_9, X_{10})$. All of the explanatory variables in $X$ are obtained from the 1988 Dress Rehearsal Census and defined at the poststrata level, the unit of analysis. $1_{84}$ is a unit vector; $X_2$ is the indicator variable for St. Louis; $X_3$ is the indicator variable for renters or is the proportion of renters for the East Central Missouri poststrata; $X_4$ through $X_7$ are indicator variables for age groups 0-9, 10-19, 20-29 and 30-44, respectively; $X_8$ is an indicator or proportion variable for males aged 20-64 that rent; $X_9$ is an indicator variable for other males aged 20-64; and $X_{10}$ is an indicator variable for other persons in St. Louis.

Using the above design matrix, we obtained $\hat{\beta} = (.9812, -.0271, .0485, .0699, .0695, .0533, .0386, .0628, .0475, .0778)^T$ and $\hat{\sigma}^2 = .000574$. The EB's or the EBLUP's and the associated approximate standard errors can now be computed using formulas derived in Section 2. For consistency, the HB analysis was also performed with the same $X$ matrix (we do not require $\hat{\beta}$ or $\hat{\sigma}^2$ for that analysis).

In Figures 1 and 2 we plot the estimated adjustment factors and standard errors by poststrata. The first 12 poststrata refer to white non-Hispanic non-owners in St. Louis; poststrata 13-24 refer to all other non-owners in St. Louis; poststrata 25-36 refer to white non-Hispanic owners in St. Louis and poststrata 37-48 refer to all other owners in St. Louis. Poststrata 49-60 refer to white non-Hispanic persons in Tape Address Register (TAR) areas in East Central Missouri; poststrata 61-72 refer to white non-Hispanic persons in non-TAR areas in East Central Missouri; poststrata 73-84 refer to all other persons in East Central Missouri.

Within each group of 12 poststrata, the first six refer to males by age 0-9, 10-19, 20-29, 30-44, 45-64 and 65 + . We note in Figure 1 that the raw adjustment factors for the other group tend to be higher than those for the white non-Hispanic except for TAR area in East Central Missouri. The same observation nearly holds in Figure 2 concerning the raw standard errors. In Figure 3 a plot of the estimated standard errors versus the adjustment factors is provided.
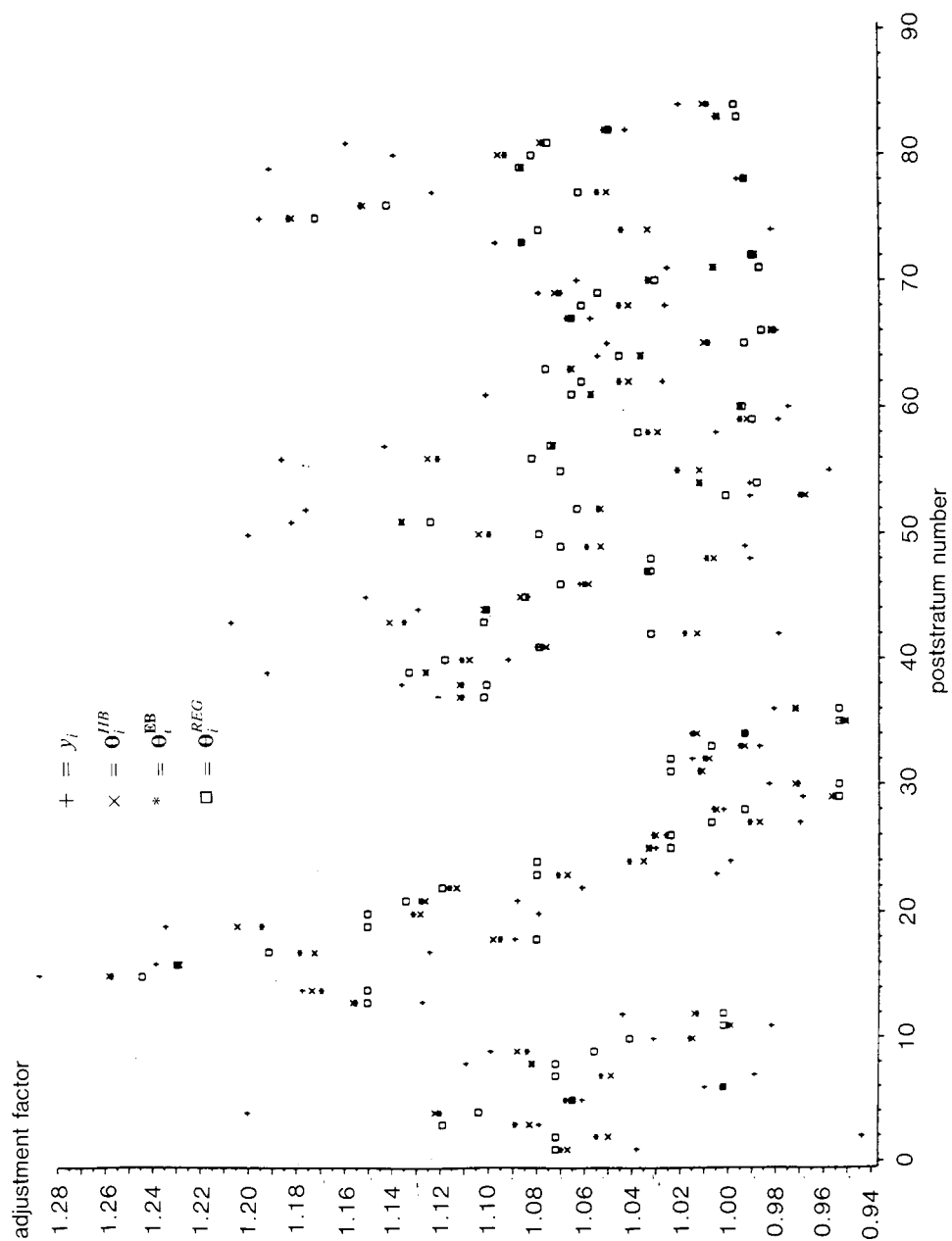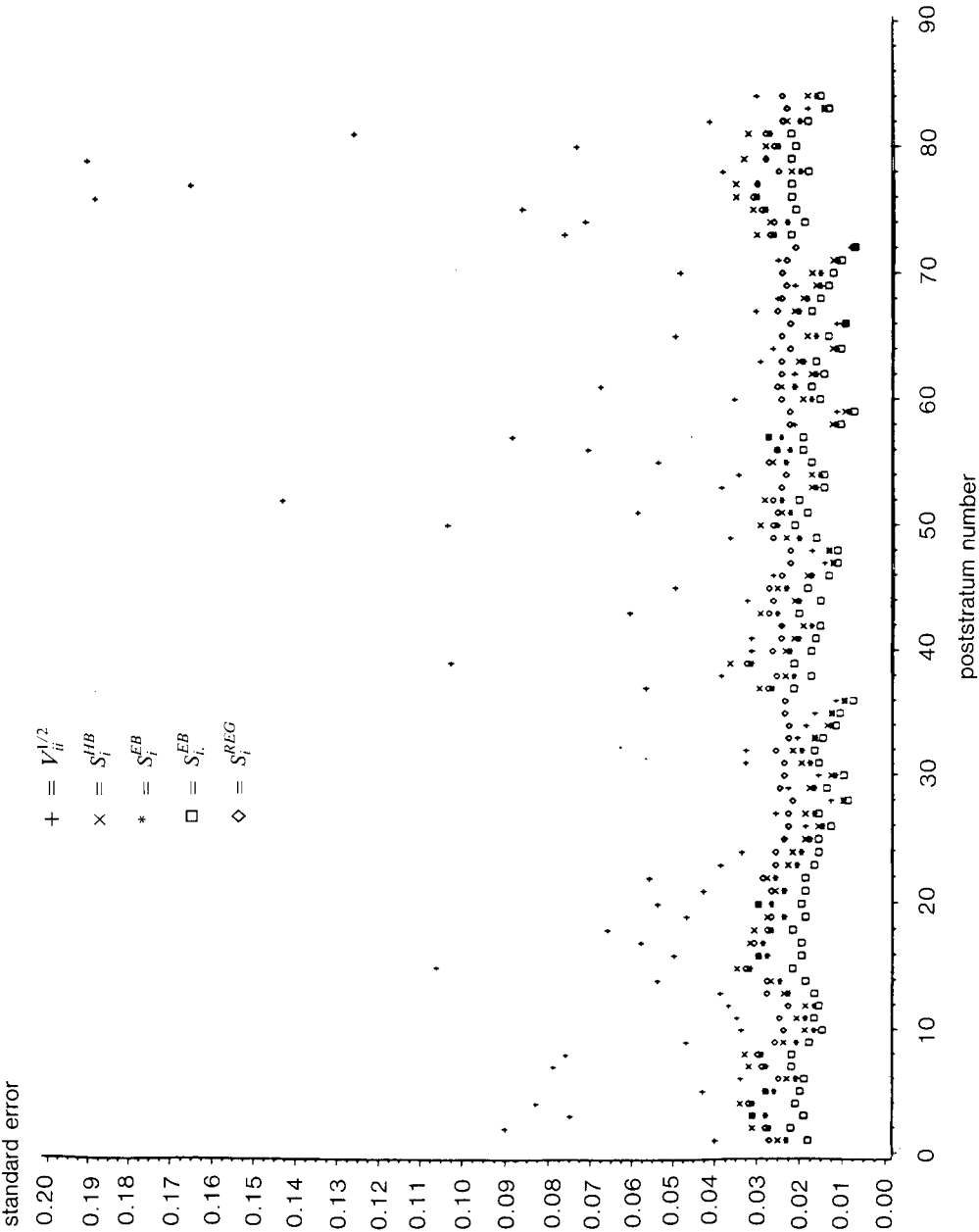
**Figure 1.** Adjustment Factors by Poststrata.

standard error

poststratum number

$+ \;=\; V_{ii}^{1/2}$

$\times \;=\; S_i^{HB}$

$* \;=\; S_i^{EB}$

$\square \;=\; S_{i.}^{EB}$

$\diamond \;=\; S_i^{REG}$

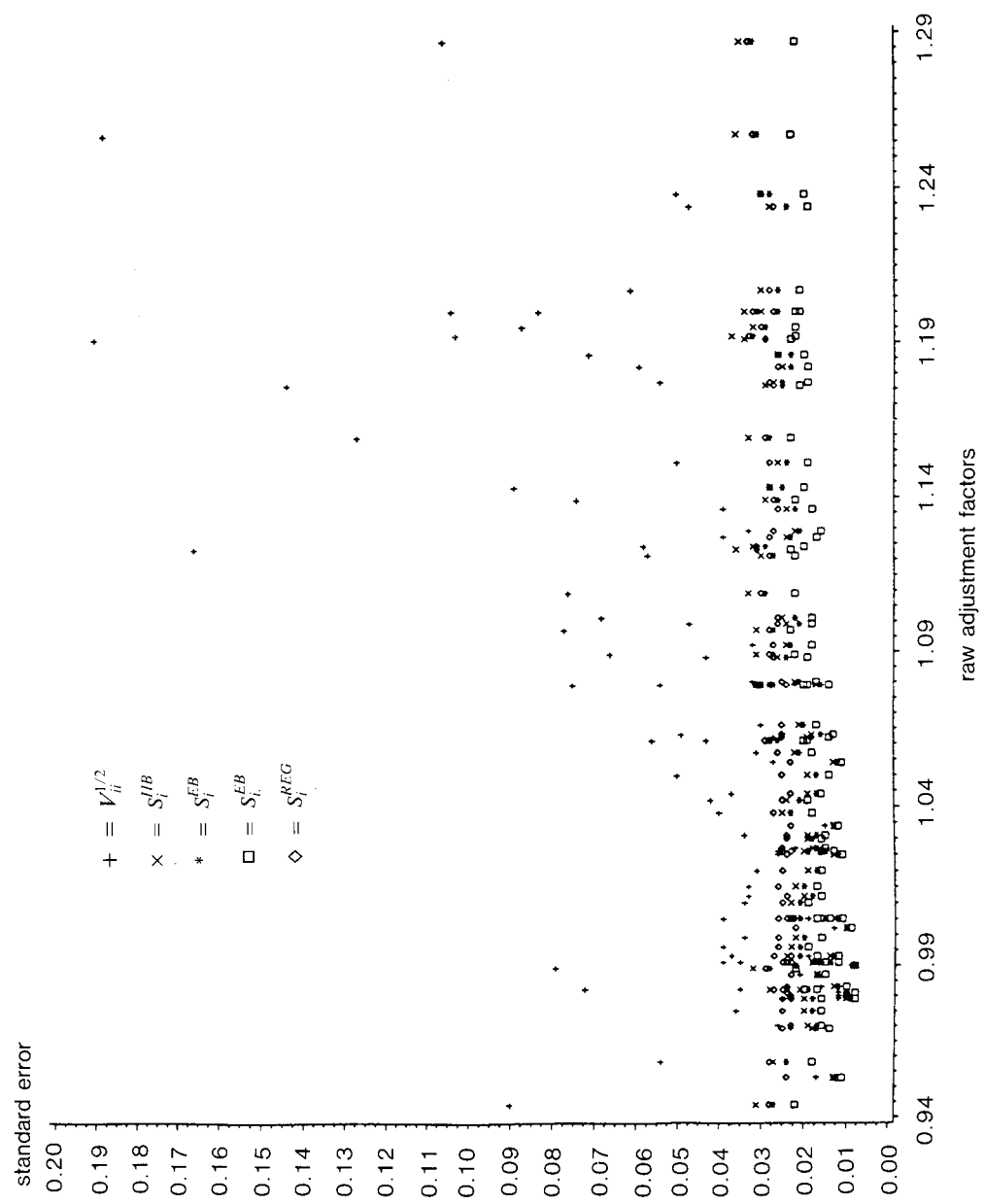**Figure 2.** SE of Adjustment Factors by Poststrata.

**Figure 3.** SE of Adjustment Factors by Raw Adjustment Factors.

Figures 1 to 3 lead to several interesting conclusions.

(1)   For every stratum, the estimated standard errors of the HB and the EB estimators of the adjustment factors are much smaller than the standard errors of the raw adjustment factors when compared to the unadjusted DSE's.

(2)   The EB estimators improve on the regression estimators for all the 84 strata by providing reduced estimated standard errors. Although the HB estimators do not improve on the regression estimators for all the strata, the improvement is substantial for most of the strata.

(3)   The data plots demonstrate that the difference between the point estimates $\hat{\theta}_i^{EB}s$ and $\hat{\theta}_i^{HB}s$ is quite small. Indeed, the percentage difference is always less than (and most often far less than) 1%.

(4)   The posterior standard errors associated with the HB estimates $(s_i^{HB})$ are always bigger than the approximate MSE's of the EB estimates $(s_i^{EB})$. As discussed earlier, the two need not be the same. It is our feeling that the approximate standard errors of the EB estimates are often slight underestimates. However, a comparison of $s_i^{EB}$ and $s_{i.}^{EB}$ reveals that a naive EB procedure (with associated estimated standard errors $s_{i.}^{EB}$) can grossly underestimate the estimated standard errors by failing to incorporate uncertainty due to estimation of $\sigma^2$. This deficiency is largely rectified by $s_i^{EB}$ which is based on second order approximations.

At the time of revision of this article, adjustment of the 1990 Decennial Census was completed. The EB estimation procedure was used. Basically, most of the same steps followed in modelling the adjustment factors in the 1988 Dress Rehearsal Census were used. However, there were several differences. In 1990 adjustment, the estimated adjustment factors were modelled by each of four census regions and a special set for Indian reservations. The number of adjustment factors ranged from 12 for the Indian set to 456 in one of the regions. In addition, estimated variances of the raw adjustment factors were smoothed via regression models. Smoothing of the estimated variances tended to reduce large estimated variances and increase small estimated variances. The net effect was an increase in the contribution of the associated adjustment factors with large estimated variances to the EB estimates and vice versa. Other differences were that outlier detection procedures were used in both the variance and adjustment factor smoothing. Finally, the EB estimates at the poststratum level were ratio adjusted to regional total population estimates derived from the raw adjustment factors. The ratio adjusted smoothed factors were then applied to related census population counts at the census block level. The results were then integer rounded by collection of blocks in such a manner that each cell within a block is rounded up or down to an integer and that control totals are off by at most one person.

The procedures used to adjust the 1990 Census counts were pre-specified and the entire operation was conducted under a very tight time schedule. The Bureau of the Census recommended that the 1990 Census adjusted counts be used. A special panel selected by the Secretary of Commerce was evenly divided in this issue. Upon weighing the evidence, the Secretary decided against using the adjusted counts. The issue is now subject to litigation. A current issue is the possible use of adjusted counts for use in postcensal estimation. Research in obtaining better adjusted counts for use in postcensal estimation is currently underway.

## APPENDIX – PROOFS OF THE THEOREMS

**Proof of Theorem 1.** We provide only an outline of the proof. The details appear in Datta *et al.* (1991). The joint (improper) pdf of $Y$, $\Theta$, $\beta$ and $\sigma^2$ is given by:

$$f(y,\Theta,\beta,\sigma^2) \propto \exp\left[-1/2(y-\Theta)^T V^{-1}(y-\Theta)\right]\sigma^{-m}\exp\left[-1/(2\sigma^2)\|\Theta-X\beta\|^2\right], \quad \text{(A.1)}$$

where $\|\cdot\|$ denotes the Euclidean norm. Writing $P_X = X(X^TX)^{-1}X^T$, $\|\Theta-X\beta\|^2 = \left[\beta-(X^TX)^{-1}X^T\Theta\right]^T(X^TX)\left[\beta-(X^TX)^{-1}X^T\Theta\right] + \Theta^T(I-P_x)\Theta$.

Now, integrating with respect to $\beta$ in (A.1), it follows that the joint improper pdf of $Y$, $\Theta$ and $\sigma^2$ is

$$f(y,\Theta,\sigma^2) \propto \sigma^{-(m-p)}\exp\left[-1/2(y-\Theta)^T V^{-1}(y-\Theta) - 1/(2\sigma^2)\Theta^T(I-P_x)\Theta\right]. \quad \text{(A.2)}$$

Next writing $E = V^{-1} + \sigma^{-2}(I-P_x)$, it follows after some simplifications that

$$(y-\Theta)^T V^{-1}(y-\Theta) + \sigma^{-2}\Theta^T(I-P_x)\Theta =$$

$$(\Theta-E^{-1}V^{-1}y)^T E(\Theta-E^{-1}V^{-1}y) + y^T(V^{-1} - V^{-1}E^{-1}V^{-1})y. \quad \text{(A.3)}$$

Hence, the posterior distribution of $\Theta$ given $\sigma^2$ and $Y = y$ is $N(E^{-1}V^{-1}y, E^{-1})$. Using the familiar matrix inversion formula $(A + BDB^T)^{-1} = A^{-1} - A^{-1}B(D^{-1} + B^TA^{-1}B)^{-1}B^TA^{-1}$ (see for example Exercise 2.9, p. 33 of Rao (1973)), one gets $E^{-1} = G$. This completes the proof of the first part of the Theorem. Next, using (A.3) and integrating with respect to $\Theta$ in (A.2), one gets the joint (improper) pdf of $Y$ and $\sigma^2$ is

$$f(y,\sigma^2) \propto \sigma^{-(m-p)}|E|^{-1/2}\exp\left[-(1/2)y^T(V^{-1} - V^{-1}E^1V^{-1})y\right]. \quad \text{(A.4)}$$

Using Exercise 2.4, p. 32 of Rao (1973), it follows that

$$|E| = |V^{-1} + \sigma^2(I-P_x)| = |\Delta| \div |\sigma^2 X^TX|$$

which on simplification reduces to

$$|V^{-1}||I + \sigma^{-2}V||X^T(I + \sigma^{-2}V)^{-1}X| \div |X^TX| \propto |I + \sigma^{-2}V||X^T\Sigma^{-1}X|. \quad \text{(A.5)}$$

Also, after some calculations, it follows that

$$V^{-1} - V^{-1}E^{-1}V^{-1} = F. \quad \text{(A.6)}$$

The proof of part (ii) of Theorem 1 follows now from (A.4) – (A.6) and noting that $f(\sigma^2 \mid y) \propto f(\sigma^2, y)$. Note, however that the posterior pdf of $\sigma^2$ given $Y = y$ is proper.

**Proof of Theorem 2.** Once again, only a sketch of the proof is given. The details are available in Datta *et al.* (1991).

Recall

$$\tilde{\beta} = (X^T\Sigma^{-1}X)^{-1}X^T\Sigma^{-1}Y.$$

Define,

$$\hat{\Theta} = X\tilde{\beta} + \sigma^2 \Sigma^{-1}(Y - X\tilde{\beta}).$$

Now, observe that (i) $\hat{\Theta}$ is the best unbiased predictor of $\Theta$ (due to normality) for every fixed $\sigma^2$, and (ii) $E(\hat{\Theta}^{EB} - \hat{\Theta}) = 0$ since $\hat{\sigma}^2$ is the MLE of $\sigma^2$ (*cf* Kackar and Harville (1984)). Now using Lemma 3.3.1 of Datta (1990), $\hat{\Theta}^{EB} - \hat{\Theta}$ is uncorrelated with $\hat{\Theta}$. Hence,

$$E\left[(\hat{\Theta}^{EB} - \hat{\Theta})(\hat{\Theta}^{EB} - \hat{\Theta})^T\right] =$$

$$E\left[(\hat{\Theta}^{EB} - \hat{\Theta})(\hat{\Theta}^{EB} - \hat{\Theta})^T\right] + E\left[(\hat{\Theta} - \Theta)(\hat{\Theta} - \Theta)^T\right]. \qquad (A.7)$$

Next, write $\hat{\Theta}^{B} = X\beta + \sigma^2 \Sigma^{-1}(Y - X\beta)$. Then standard arguments give

$$E\left[(\hat{\Theta} - \Theta)(\hat{\Theta} - \Theta)^T\right] = E\left[(\hat{\Theta} - \hat{\Theta}^B)(\hat{\Theta} - \hat{\Theta}^B)^T\right] + E\left[(\hat{\Theta}^B - \Theta)(\hat{\Theta}^B - \Theta)^T\right]. \qquad (A.8)$$

Our previous calculations yield

$$E\left[(\hat{\Theta}^{B} - \Theta)(\hat{\Theta}^{B} - \Theta)^T\right] = V - V\Sigma^{-1}V. \qquad (A.9)$$

Further,

$$E\left[(\hat{\Theta} - \hat{\Theta}^B)(\hat{\Theta} - \hat{\Theta}^B)^T\right] = V\Sigma^{-1}X(X^T\Sigma^{-1}X)^{-1}X^T\Sigma^{-1}V. \qquad (A.10)$$

Finally, write $\hat{\Theta} = g(\sigma^2)$ and $\hat{\Theta}^{EB} = g(\hat{\sigma}^2)$. Using first order Taylor approximation, one gets

$$E\left[(\hat{\Theta}^{EB} - \hat{\Theta})(\hat{\Theta}^{EB} - \hat{\Theta})^T\right] \doteq E\left[(\hat{\sigma}^2 - \sigma^2)^2 \frac{dg(\sigma^2)}{d\sigma^2} \frac{dg(\sigma^2)^T}{d\sigma^2}\right]. \qquad (A.11)$$

Since $g(\sigma^2) = Y - V\Sigma^{-1}[Y - X(X^T\Sigma^{-1}X)^{-1}X^T\Sigma^{-1}Y]$, using matrix differentiation, techniques, one gets

$$\frac{dg}{d\sigma^2} = V\left[\Sigma^{-1} - \Sigma^{-1}X(X^T\Sigma^{-1}X)^{-1}X^T\Sigma^{-1}\right]\Sigma^{-1}(Y - X\hat{\beta}). \qquad (A.12)$$

$$E\left[\frac{dg}{d\sigma^2} \frac{dg^T}{d\sigma^2}\right] = VK\Sigma^{-1}E\left[(Y - X\hat{\beta})(Y - X\hat{\beta})^T\right]\Sigma^{-1}KV. \qquad (A.13)$$

But, simple algebra gives

$$E\left[(Y - X\hat{\beta})(Y - X\hat{\beta})^T\right] = \Sigma - X(X^T\Sigma^{-1}X)^{-1}X^T = \Sigma K\Sigma. \qquad (A.14)$$

Hence, from (A.13),

$$E\left[\frac{dg}{d\sigma^2} \frac{dg^T}{d\sigma^2}\right] = VK^3V. \qquad (A.15)$$

Using, one more approximation, it follows from (A.11) and (A.15) that

$$\mathrm{E}\big[(\hat{\boldsymbol{\theta}}^{EB} - \hat{\boldsymbol{\theta}})(\hat{\boldsymbol{\theta}}^{EB} - \hat{\boldsymbol{\theta}})^T\big] \doteq \mathrm{E}(\hat{\sigma}^2 - \sigma^2)^2 VK^3V. \tag{A.16}$$

To estimate $\mathrm{E}(\hat{\sigma}^2 - \sigma^2)^2 = \mathrm{MSE}(\hat{\sigma}^2)$, we proceed as follows.

Since $Y \sim \mathrm{N}(X\beta, \Sigma)$, write the likelihood function as

$$L(\sigma^2) \propto |\Sigma|^{-1/2} \exp\big[-1/2(Y - X\beta)^T\Sigma^{-1}(Y - X\beta)\big]. \tag{A.17}$$

Hence,

$$\frac{d\log L}{d\sigma^2} = -1/2\,\frac{d}{d\sigma^2}\log|\Sigma| - 1/2\,\frac{d}{d\sigma^2}\big[(Y - X\beta)^T\Sigma^{-1}(Y - X\beta)\big]; \tag{A.18}$$

$$\frac{d^2\log L}{d(\sigma^2)^2} = -1/2\,\frac{d^2}{d(\sigma^2)^2}\log|\Sigma| - 1/2\,\frac{d^2}{d(\sigma^2)^2}\big[(Y - X\beta)^T\Sigma^{-1}(Y - X\beta)\big]. \tag{A.19}$$

As before, denote by $d_1, \ldots, d_m$ the eigenvalues of $V$.

Then, $\log|\Sigma| = \Sigma_{i=1}^m \log(\sigma^2 + d_i)$. Hence

$$\frac{d^2}{d(\sigma^2)^2}\log|\Sigma| = -\Sigma_{i=1}^m (\sigma^2 + d_i)^{-2} = -\mathrm{tr}(\Sigma^{-2}). \tag{A.20}$$

Using (A.20) and matrix differentiation, it follows from (A.19) that

$$\frac{d^2\log L}{d(\sigma^2)^2} = 1/2\,\mathrm{tr}(\Sigma^{-2}) - (Y - X\beta)^T\Sigma^{-3}(Y - X\beta). \tag{A.21}$$

Thus,

$$\mathrm{E}\bigg[-\frac{d^2\log L}{d(\sigma^2)^2}\bigg] = -1/2\,\mathrm{tr}(\Sigma^{-2}) + \mathrm{tr}(\Sigma^{-2}) = 1/2\,\mathrm{tr}(\Sigma^{-2}).$$

Approximating $\mathrm{E}[(\hat{\sigma}^2 - \sigma^2)^2]$ by

$$\bigg(\mathrm{E}\bigg[-\frac{d^2\log L}{d(\sigma^2)^2}\bigg]\bigg)^{-1},$$

justifiable by the asymptotic theory of maximum likelihood, one gets, from (A.16),

$$\mathrm{E}\big[(\hat{\boldsymbol{\theta}}^{EB} - \hat{\boldsymbol{\theta}})(\hat{\boldsymbol{\theta}}^{EB} - \hat{\boldsymbol{\theta}})^T\big] \doteq 2(\mathrm{tr}(\Sigma^{-2}))^{-1}VK^3V. \tag{A.22}$$

Combining (A.7) – (A.10) and (A.22), one gets

$$\mathrm{MSE}(\hat{\boldsymbol{\theta}}^{EB}) \doteq V - V\Sigma^{-1}V + V\Sigma^{-1}X(X^T\Sigma^{-1}X)^{-1}X^T\Sigma^{-1}V + VK^3V\big[2(\mathrm{tr}\Sigma^{-2})^{-1}\big] \tag{A.23}$$

Substitution of $\hat{\Sigma}$ for $\Sigma$ yields the approximation given in (2.8). This completes the proof of Theorem 2.

## ACKNOWLEDGEMENTS

## REFERENCES

CHILDERS, D.R., and HOGAN, H. (1990). Results of the 1988 dress rehearsal post enumeration survey. *Proceedings of the Section on Survey Research Methods, American Statistical Association*, 547-552.

DATTA, G.S. (1990). Bayesian prediction in mixed linear models with applications in small area estimation. Unpublished Ph.D. dissertation. University of Florida.

DATTA, G.S., GHOSH, M., HUANG, E.T., ISAKI, C.T., and SCHULTZ, L.K. (1990). Hierarchical and Empirical Bayes methods for adjustment of census undercount: The 1988 Missouri Dress-Rehearsal data. Technical Report No. 376. Department of Statistics, University of Florida.

ERICKSEN, E.P., and KADANE, J.B. (1985). Estimating the population in a census year: 1980 and beyond (with discussion). *Journal of the American Statistical Association*, 80, 98-131.

FAY, R.E., and HERRIOT, R.A. (1979). Estimates of income for Small Places: An application of James-Stein procedures to census data. *Journal of the American Statistical Association*, 74, 269-277.

FREEDMAN, D.A., and NAVIDI, W.C. (1986). Regression models for adjusting the 1980 census (with discussion). *Statistical Science*, 1, 3-39.

HARVILLE, D.A. (1985). Decomposition of prediction error. *Journal of the American Statistical Association*, 80, 132-138.

HOGAN, H., and WOLTER, K. (1988). Measuring accuracy in a Post-Enumeration Survey. *Survey Methodology*, 14, 99-116.

HUANG, E.T., ISAKI, C.T., and TSAY, J.H. (1991). Modelling PES adjustment factors using 1988 dress rehearsal data. *Proceedings of the Social Statistics Section, American Statistical Association*, to appear.

ISAKI, C.T., HUANG, E.T., and TSAY, J.H. (1991). Smoothing adjustment factors from the 1990 post enumeration survey. *Proceedings of the Social Statistics Section, American Statistical Association*, to appear.

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

MORRIS, C. (1983). Parametric Empirical Bayes inference and applications. *Journal of the American Statistical Association*, 78, 47-65.

PRASAD, N.G.N., and RAO, J.N.K. (1990). The estimation of mean squared error of small area predictors. *Journal of the American Statistical Association*, 85, 163-171.

RAO, C.R. (1973). *Linear Statistical Inference and Its Applications*, (2nd. Edition). New York: Wiley.