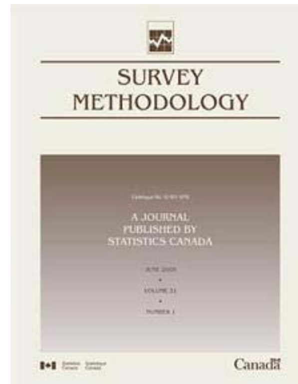
Component of Statistics Canada
Catalogue no. 12-001-X Business Survey Methods Division

Article

Estimation of census adjustment factors



by C.T. Isaki, J.H.Tsay and W.A. Fuller

June 2000

Statistics Canada Statistique Canada

Canadä

Survey Methodology, June 2000
Vol. 26, No. 1, pp. 31-42
Statistics Canada, Catalogue No. 12-001

Estimation of census adjustment factors

C.T. Isaki, J.H. Tsay and W.A. Fuller [1]

Abstract

A components-of-variance approach and an estimated covariance error structure were used in constructing predictors of adjustment factors for the 1990 Decennial Census. The Variability of the estimated covariance matrix is the suspected cause of certain anomalies that appeared in the regression estimation and in the estimated adjustment factors. We investigate alternative prediction methods and propose a procedure that is less influenced by variability in the estimated covariance matrix. The proposed methodology is applied to a data set composed of 336 adjustment factors from the 1990 Post Enumeration Survey.

Key Words: Components-of-variance; Small area estimation; Undercount; Decennial Census; Smoothing.

1. Introduction

While the objective of a population census is to record data for all individuals, it has long been recognized that this goal is not achieved in practice. Post enumeration studies associated with the U.S. Census of 1970 and 1980 suggested that the coverage rate was different for different demographic groups. See U.S. Bureau of the Census (1988).

In 1990, a post enumeration survey (PES), using dual system (or capture-recapture) estimation, was used to produce estimates for 1,392 subdivisions of the total population of the United States at the time of the 1990 Census. The PES sample contained approximately 377,000 persons in about 5,200 sample blocks. Sample persons were divided into post-strata defined by geographic divisions of the country, tenure, size-of-place, race, sex, and age, where the two tenure classes are owners and renters of homes, and size-of-place is a measure of urbanization. The subdivisions were called poststrata. The ratio of the PES estimate to the Census total, called the adjustment factor, was produced for each poststratum. An adjustment factor greater than one is associated with an estimated undercount and a factor less than one is associated with an estimated overcount.

Because relatively large sampling variances were anticipated for individual ratios, a smoothing technique based on components-of-variance and a regression model was used to create the final estimated adjustment factors. The elements of the error covariance matrix used in the prediction model were estimated with a jackknife algorithm, see Fay (1990).

The explanatory variables in the regression model were chosen using a best subsets selection algorithm. Some explanatory variables were forced into the model. For example, in the Midwest region, the ten explanatory variables forced into the model were Black, Hispanic, renter, age group 0-9, age group 10-19, age group 20-29, age group 30-44, age group 45-64, male 10-19 and male 20-64. Most variables were indicator variables, but some were proportions. For example, a variable "percent Black" was used when Black and Hispanic were grouped into a single post-stratum. Nine other variables were selected for inclusion in the model based on a best subsets regression algorithm. The variables included mail return rate, substitution rate, type-of-place and six race-by-age and race-by-tenure interaction variables. The mail return rate is the fraction of Census questionnaires returned from the mail distribution, the substitution rate is the fraction of Census households that were entirely replaced with responding households.

The smoothing technique was applied to poststrata ratios by regions of the country. The adjustment factors were designed to be applied to Census counts in the appropriate poststrata to create population estimates adjusted for undercount or overcount. Hogan (1992) contains an overview of the PES. Isaki, Huang and Tsay (1991) provide a detailed description of the results of the smoothing of the poststratum ratios.

Fay (1992) in a manuscript discussing the adjustment factors constructed from the 1990 PES, identified some disturbing results. He noted that some of the estimated regression coefficients in the model differed considerably depending on the form of the estimated covariance matrix used to construct the estimated generalized least squares estimator. Fay conjectured that large differences in coefficients could arise because of an unstable estimator of the error covariance matrix. Although the estimated error variances were smoothed, it was felt that estimated variances of linear combinations might still have large variances. He felt that the estimated variances had large variances because the direct estimates for many blocks were zero.

The Secretary of Commerce ultimately decided to use the unadjusted counts in the Decennial Census. The possible use of adjusted counts for other purposes, such as the Bureau's postcensal estimation program, was left for additional study.

1. C.T. Isaki and J.H. Tsay, Statistical Research Division, Bureau of the Census, Washington, D.C. 20233, U.S.A.; W.A. Fuller, Department of Statistics, Iowa State University, Ames, IA 50010, U.S.A.

We explore alternative smoothed estimators for the adjustment factors, focusing on the effect of estimating the covariance matrix of the vector of the estimated adjustment factors. In the empirical part of our study, we construct estimates based on the 1990 Census data.

2. Smoothing model

The model chosen for the construction of predictors is the multivariate components-of-variance model. Closely related models that lead to smoothed estimators for a set of unknowns, have been studied by a number of authors. Fay and Herriot (1979) suggested the use of the model in a small area estimation procedure. Battese, Harter and Fuller (1988) applied the components-of-variance model to crop area estimation. Ericksen and Kadane (1985), Cressie (1992), and Ericksen, Kadane and Tukey (1989) suggested smoothing procedures for census adjustment. Singh, Gambino and Mantel (1994) discuss a range of small area procedures. Efron and Morris (1972) and Morris (1983) contain good discussions of some of the basic theory. Kackar and Harville (1984), Peixoto and Harville (1986), Fay (1987), Fuller and Harter (1987), Hulting and Harville (1991), Ghosh (1992), and Prasad and Rao (1990) discuss estimation and variance estimation for such procedures. Ghosh and Rao (1994) is a review article.

Under the multivariate components-of-variance model, the vector of true values to be predicted is

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{w}, \tag{1}$$

where $\mathbf{y}$ is an n-dimensional column vector, $\mathbf{X}$ is an $n \times k$ matrix of observable characteristics, $\mathbf{w}$ is an n-dimensional column vector of random effects and $\boldsymbol{\beta}$ is a k-dimensional unknown column vector. The vector $\mathbf{Y}$ is observed, where

$$\mathbf{Y} = \mathbf{y} + \mathbf{e}, \tag{2}$$

$\mathbf{Y}$ is an n-dimensional column vector and $\mathbf{e}$ is the n-dimensional column vector of estimation errors. In our application $\mathbf{Y}$ is the vector of estimated adjustment factors. It is assumed that

$$(\mathbf{w}', \mathbf{e}')' \sim (0, \text{block diag } \{\mathbf{I}\sigma^2, \boldsymbol{\Sigma}_{ee}\}), \tag{3}$$

where $\boldsymbol{\Sigma}_{ee}$ is the covariance matrix of the estimation errors, and σ^2 is the unknown variance of the random effects.

A class of predictors of $\mathbf{y}$ is defined by

$$\tilde{\mathbf{y}} = \mathbf{XB} + \mathbf{G}'(\mathbf{Y} - \mathbf{XB}), \tag{4}$$

where $\mathbf{B}$ is a k-dimensional vector and $\mathbf{G}$ is an $n \times n$ matrix. Under model (1) with

$$(\mathbf{w}', \mathbf{e}')' \sim N(\mathbf{0}, \text{block diag}\{\mathbf{I}\sigma^2, \boldsymbol{\Sigma}_{ee}\}), \tag{5}$$

the conditional expected value of $\mathbf{y}$ given $\mathbf{Y}$ is

$$E\{\mathbf{y} \mid \mathbf{Y}\} = \mathbf{X}\beta + \mathbf{G}'_{zz}(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}), \tag{6}$$

where $\mathbf{G}_{zz} = \boldsymbol{\Sigma}_{zz}^{-1}\sigma^2$ and $\boldsymbol{\Sigma}_{zz} = \mathbf{I}\sigma^2 + \boldsymbol{\Sigma}_{ee}$ is the $n \times n$ covariance matrix of $\mathbf{z} = \mathbf{w} + \mathbf{e}$. Under the normal distribution model defined by (1), (2), and (5) and with the parameters σ^2, $\boldsymbol{\Sigma}_{ee}$, $\boldsymbol{\beta}$ known, the minimum mean square error predictor of $\mathbf{y}$ is given by the right side of equation (6).

Generally, some of the parameters are unknown. Consider first the case in which $\boldsymbol{\beta}$ is unknown. Let $\hat{\boldsymbol{\beta}}$ be an estimator of $\boldsymbol{\beta}$, where

$$\hat{\boldsymbol{\beta}} = \left(\mathbf{X}'\mathbf{M}^{-1}\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{M}^{-1}\mathbf{Y}, \tag{7}$$

and $\mathbf{M}$ is an $n \times n$ matrix. If $\mathbf{M}$ is fixed

$$\begin{aligned}\tilde{\mathbf{y}} - \mathbf{y} &= (\mathbf{I} - \mathbf{G})\mathbf{X}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) - (\mathbf{I} - \mathbf{G})\mathbf{w} + \mathbf{Ge} \\ &= (\mathbf{K} - \mathbf{I})\mathbf{w} + \mathbf{Ke},\end{aligned}$$

where $\mathbf{K} = (\mathbf{I} - \mathbf{G}')\mathbf{X}(\mathbf{X}'\mathbf{M}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{M}^{-1} + \mathbf{G}'$. Thus, if $\mathbf{M}$ and $\mathbf{G}$ are fixed,

$$\mathbf{V}\{\tilde{\mathbf{y}} - \mathbf{y}\} = (\mathbf{K} - \mathbf{I})(\mathbf{K} - \mathbf{I})'\sigma^2 + \mathbf{K}\boldsymbol{\Sigma}_{ee}\mathbf{K}'. \tag{8}$$

If model (1), (2), and (3) holds, and if $\boldsymbol{\Sigma}_{ee}$ and σ^2 are known, then replacing $\mathbf{B}$ with

$$\tilde{\boldsymbol{\beta}} = (\mathbf{X}'\boldsymbol{\Sigma}_{zz}^{-1}\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\Sigma}_{zz}^{-1}\mathbf{Y} \tag{9}$$

and replacing $\mathbf{G}$ with

$$\mathbf{G}_{zz} = \boldsymbol{\Sigma}_{zz}^{-1}\sigma^2 \tag{10}$$

in (4) defines the best linear unbiased predictor of $\mathbf{y}$. See Henerson (1950), Harville (1976), and Robinson (1991). If $\boldsymbol{\Sigma}_{ee}$ and σ^2 are also unknown, it is natural to use estimators of $\boldsymbol{\Sigma}_{ee}$ and σ^2 to construct an estimated best linear unbiased predictor. Very often, an estimator of $\boldsymbol{\Sigma}_{ee}$ is associated with the procedure used to construct the estimator $\mathbf{Y}$. Then σ^2 is estimated from model (1), (2), and (5), treating the estimator of $\boldsymbol{\Sigma}_{ee}$ as the true $\boldsymbol{\Sigma}_{ee}$.

One substitution predictor is

$$\hat{\mathbf{y}} = X\hat{\boldsymbol{\beta}} + \hat{\sigma}^2\hat{\boldsymbol{\Sigma}}_{zz}^{1-}(\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}}), \tag{11}$$

where

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\hat{\boldsymbol{\Sigma}}_{zz}^{-1}\mathbf{X})^{-1}\mathbf{X}'\hat{\boldsymbol{\Sigma}}_{zz}^{-1}\mathbf{Y} \tag{12}$$

is the estimated generalized least squares estimator of $\boldsymbol{\beta}$,

$$\hat{\boldsymbol{\Sigma}}_{zz} = \mathbf{I}\hat{\sigma}^2 + \hat{\boldsymbol{\Sigma}}_{ee}, \tag{13}$$

$\hat{\boldsymbol{\Sigma}}_{ee}$ is an estimator of $\boldsymbol{\Sigma}_{ee}$, and $\hat{\sigma}^2$ is an estimator of σ^2. The estimator of σ^2 can be based on likelihood or analysis of variance procedures. Retaining only the terms in the Taylor expansion of the error in (11) that are errors in the basic estimators, we have

$$\begin{aligned}\hat{\mathbf{y}} - \mathbf{y} &\doteq \mathbf{e} - \mathbf{H}'\mathbf{z} + \mathbf{H}'\mathbf{X}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \\ &\quad + (\hat{\sigma}^2 - \sigma^2)\mathbf{H}'\boldsymbol{\Sigma}_{zz}^{-1}\mathbf{z} \\ &\quad - \mathbf{G}'(\hat{\boldsymbol{\Sigma}}_{ee} - \boldsymbol{\Sigma}_{ee})\boldsymbol{\Sigma}_{zz}^{-1}\mathbf{z},\end{aligned} \tag{14}$$

where $\mathbf{H}' = \boldsymbol{\Sigma}_{ee}\boldsymbol{\Sigma}_{zz}^{-1}$ and $\mathbf{G}' = \mathbf{I} - \mathbf{H}' = \sigma^2\boldsymbol{\Sigma}_{zz}^{-1}$. If it is assumed that $\hat{\boldsymbol{\Sigma}}_{ee}$ is distributed as a multiple of a Wishart matrix and d_e degrees of freedom, if the covariance between $\hat{\sigma}^2$ and $\hat{\boldsymbol{\Sigma}}_{ee}$ is ignored, if expectations are computed as if $\hat{\sigma}^2$ and $\mathbf{z}$ are independent, and if expectations are computed as if $\mathbf{z}$ and $\hat{\boldsymbol{\Sigma}}_{ee}$ are independent, an approximation to the variance of $\hat{\mathbf{y}} - \mathbf{y}$ obtained from (14) is

$$\mathbf{V}\{\hat{\mathbf{y}} - \mathbf{y}\} \doteq \boldsymbol{\Sigma}_{ee}\mathbf{G} + \mathbf{H}'\mathbf{X}\mathbf{V}_{\beta\beta}\mathbf{X}'\mathbf{H} + \boldsymbol{\Gamma}_{33} + \boldsymbol{\Gamma}_{44}, \quad (15)$$

where

$$\begin{aligned}
\mathbf{V}_{\beta\beta} &= \mathbf{V}\{\hat{\boldsymbol{\beta}}\} = (\mathbf{X}'\boldsymbol{\Sigma}_{zz}^{-1}\mathbf{X})^{-1} + d_e^{-1}\,\mathrm{tr}\{\boldsymbol{\Sigma}_{zz}^{-1}\boldsymbol{\Sigma}_{ee}\}\mathbf{L}\boldsymbol{\Sigma}_{ee}\mathbf{L}',\\
\boldsymbol{\Gamma}_{33} &= \mathbf{H}'\boldsymbol{\Sigma}_{zz}^{-1}\mathbf{H}V_{\sigma\sigma},\\
\boldsymbol{\Gamma}_{44} &= d_e^{-1}\sigma^4\boldsymbol{\Sigma}_{zz}^{-1}\boldsymbol{\Sigma}_{ee}\boldsymbol{\Sigma}_{zz}^{-1}[\mathrm{tr}\{\boldsymbol{\Sigma}_{zz}^{-1}\boldsymbol{\Sigma}_{ee}\}],\\
\mathbf{L} &= (\mathbf{X}'\boldsymbol{\Sigma}_{zz}^{-1}\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\Sigma}_{zz}^{-1}
\end{aligned}$$

and $V_{\sigma\sigma} = V\{\hat{\sigma}^2\}$ is the variance of $\hat{\sigma}^2$. The term $\boldsymbol{\Sigma}_{ee}\mathbf{G}$ is the prediction covariance matrix if all parameters are known. The remaining three terms of (15) are the contributions to the variance due to estimating $\boldsymbol{\beta}$, σ^2, and $\boldsymbol{\Sigma}_{ee}$, respectively. The second term in $\mathbf{V}\{\hat{\boldsymbol{\beta}}\}$ is a crude approximation for the increase in the variance of $\hat{\boldsymbol{\beta}}$ due to using an estimator of $\boldsymbol{\Sigma}_{zz}$ in place of $\boldsymbol{\Sigma}_{zz}$ in constructing $\hat{\boldsymbol{\beta}}$.

If the dimension of $\boldsymbol{\Sigma}_{zz}$ is large and the degrees of freedom, d_e, only slightly larger than the dimension, then the second part of the variance of $\hat{\boldsymbol{\beta}}$ and the term $\boldsymbol{\Gamma}_{44}$ can make important contributions to the variance. This is particularly true if σ^2 is small relative to the diagonal elements of $\boldsymbol{\Sigma}_{ee}$. The Monte Carlo study of the next section demonstrates that the contribution to variance approximated by these terms can be important.

A predictor that reduces the effect of the estimation error in $\hat{\boldsymbol{\Sigma}}_{ee}$ uses only diagonal elements of $\boldsymbol{\Sigma}_{ee}$ in the shrinkage component. Let

$$\hat{\mathbf{y}}_d = \mathbf{X}\hat{\boldsymbol{\beta}}_d + \hat{\sigma}^2\hat{\mathbf{D}}_{zz}^{-1}(\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}}_d), \quad (16)$$

where

$$\begin{aligned}
\hat{\boldsymbol{\beta}}_d &= (\mathbf{X}'\hat{\mathbf{D}}_{zz}^{-1}\mathbf{X})^{-1}\mathbf{X}'\hat{\mathbf{D}}_{zz}^{-1}Y,\\
\hat{\mathbf{D}}_{zz} &= \mathrm{diag}(\hat{\boldsymbol{\Sigma}}_{ee} + \mathbf{I}\hat{\sigma}^2),
\end{aligned}$$

$\hat{\sigma}^2$ is an estimator of σ^2 and diag $(\mathbf{A})$ is the diagonal matrix composed of the diagonal elements of $\mathbf{A}$. Retaining only the leading terms in the Taylor expansion of the error in (16) gives

$$\begin{aligned}
\hat{\mathbf{y}}_d - \mathbf{y} \doteq{}& -(\mathbf{w} - \mathbf{G}_d'\mathbf{z}) + \mathbf{H}_d'\mathbf{X}(\hat{\boldsymbol{\beta}}_d - \boldsymbol{\beta})\\
&+ (\hat{\sigma}^2 - \sigma^2)\mathbf{H}_d'\mathbf{D}_{zz}^{-1}\mathbf{z} - \mathbf{G}_d'(\hat{\mathbf{D}}_{ee} - \mathbf{D}_{ee})\mathbf{D}_{zz}^{-1}\mathbf{z}, \quad (17)
\end{aligned}$$

where $\mathbf{D}_{zz} = \mathrm{diag}\{\boldsymbol{\Sigma}_{zz}\}$, $\mathbf{G}_d = \mathbf{D}_{zz}^{-1}\sigma^2$, $\mathbf{H}_d = \mathbf{I} - \mathbf{G}_d$, and $\mathbf{D}_{ee} = \mathrm{diag}\{\boldsymbol{\Sigma}_{ee}\}$. If $\mathbf{w}$ and $\mathbf{e}$ are normally distributed, and if $\hat{\sigma}^2$ and $\hat{\mathbf{D}}_{zz}$ are quadratic estimators, then $\hat{\sigma}^2$ and $\hat{\mathbf{D}}_{zz}$ are uncorrelated with $\mathbf{z}$. The i^{th} element of $\mathbf{w} - \sigma^2\mathbf{D}_{zz}^{-1}\mathbf{z}$ is uncorrelated with the i^{th} element of $\mathbf{z}$, but is not necessarily uncorrelated with the vector $\mathbf{z}$. If this possible correlation is ignored, if it is assumed that $\hat{\boldsymbol{\Sigma}}_{ee}$ is a Wishart matrix with d_e degrees of freedom, and if the correlation between $\hat{\sigma}^2$ and $\hat{\boldsymbol{\Sigma}}_{ee}$ is ignored, an approximation to the variance of $\hat{\mathbf{y}}_d - \mathbf{y}$ obtained from (17) is

$$\begin{aligned}
\mathbf{V}\{\hat{\mathbf{y}}_d - \mathbf{y}\} ={}& \mathbf{H}_d'\mathbf{H}_d\sigma^2 + \mathbf{G}_d'\boldsymbol{\Sigma}_{ee}\mathbf{G}_d\\
&+ \mathbf{H}_d'\mathbf{X}\mathbf{V}_{\beta\beta}\mathbf{X}'\mathbf{H}_d + \boldsymbol{\Gamma}_{33dd} + \boldsymbol{\Gamma}_{44dd}, \quad (18)
\end{aligned}$$

where $\mathbf{G}_d = \mathbf{D}_{zz}^{-1}\sigma^2$, $\mathbf{H}_d = \mathbf{I} - \mathbf{G}_d$,

$$\mathbf{V}_{\beta\beta} = (\mathbf{X}'\mathbf{D}_{zz}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{D}_{zz}^{-1}\boldsymbol{\Sigma}_{zz}\mathbf{D}_{zz}^{-1}\mathbf{X}(\mathbf{X}'\mathbf{D}_{zz}^{-1}\mathbf{X})^{-1}, \quad (19)$$

$$\boldsymbol{\Gamma}_{33dd} = \mathbf{H}_d'\mathbf{D}_{zz}^{-1}\boldsymbol{\Sigma}_{zz}\mathbf{D}_{zz}^{-1}\mathbf{H}_dV_{\sigma\sigma}, \quad (20)$$

$$\boldsymbol{\Gamma}_{44dd} = d_e^{-1}\mathbf{G}_d'\boldsymbol{\Omega}\mathbf{G}_d'$$

and the ij^{th} element of $\boldsymbol{\Omega}$ is

$$\omega_{ij} = 2\sigma_{eeij}^2\sigma_{zzii}^{-1}\sigma_{zzjj}^{-1}\sigma_{zzij}.$$

The term in $\boldsymbol{\Gamma}_{44dd}$ is an estimator of the contribution to the variance due to using $\hat{\boldsymbol{\Sigma}}_{ee}$ to estimate the covariance matrix. Expression (19) assumes that the contribution of the error in $\hat{\mathbf{D}}_{zz}$ to the variance of $\hat{\boldsymbol{\beta}}$ can be ignored for large d_e. The difference between (15) and (18) is that the multipliers in (19) and (20) do not depend on the dimension of $\boldsymbol{\Sigma}_{zz}$. Therefore, the error in estimating $\boldsymbol{\Sigma}_{zz}$ makes a smaller contribution to the variance. On the other hand, the variance of $\mathbf{w} - \mathbf{G}_d'\mathbf{z}$, the order one term of (17), will be larger than the corresponding term of the error in (14), unless $\boldsymbol{\Sigma}_{zz}$ is diagonal. The first two terms on the right of (18) are the variance of $\mathbf{w} - \mathbf{G}_d'\mathbf{z}$.

3. Monte Carlo study

To examine the variability in the predictors associated with variability in the estimation of $\boldsymbol{\Sigma}_{ee}$ we conducted a small Monte Carlo study. The model for the study is

$$\mathbf{Y}_j = \mu\mathbf{J} + \mathbf{w} + \mathbf{e}_j, \qquad j = 1, 2, ..., r \quad (21)$$

$$\begin{aligned}
\mathbf{w} &\sim (\mathbf{0}, \mathbf{I}\sigma^2),\\
\mathbf{e}_j &\sim \mathrm{ind}(\mathbf{0}, \boldsymbol{\Sigma}_{ee}),
\end{aligned}$$

where $\mathbf{J}$ is the k-dimensional column vector of ones, $\mathbf{J} = (1, 1, ..., 1)'$, $\mathbf{w}$ is the k-dimensional vector of random small area effects, $\mathbf{e}_j$ is a vector of errors, and $\mathbf{w}$ and $\mathbf{e}_j$ are independent. The model is a simplified version of the model defined in (1), (2), and (3). The mean is the constant function and, hence, we use μ in place of β. To create a vector of correlated variables, we define, for $k = 8$,

$$\begin{bmatrix} e_{1j} \\ e_{2j} \\ e_{3j} \\ e_{4j} \\ e_{5j} \\ e_{6j} \\ e_{7j} \\ e_{8j} \end{bmatrix} = \begin{bmatrix} 1.3u_{1j} \\ 1.5u_{1j} + 0.4u_{2j} \\ 0.9u_{1j} + 0.9u_{3j} \\ 0.9u_{3j} + 1.6u_{4j} \\ 1.6u_{4j} + 0.6u_{5j} \\ 1.0u_{4j} + 1.6u_{6j} \\ 1.0u_{7j} \\ 2.83u_{8j} \end{bmatrix},$$

where u_{ij} are independent random variables. The w_i, $i = 1, 2, ..., 8$, are NI(0, 0.36) random variables, where $\mathrm{NI}(\mu, \sigma^2)$ denotes normal independent random variables with mean μ and variance σ^2. This configuration gives a range of error variances and a range of correlations between estimates.

The estimator of σ^2 used in the Monte Carlo study is

$$\hat{\sigma}^2 = \max\{(k-1)^{-1} \times [(\bar{\mathbf{y}} - \mathbf{J}\hat{\mu}_{(0)})'(\bar{\mathbf{y}} - \mathbf{J}\hat{\mu}_{(0)}) - \mathrm{tr}\{r^{-1}\hat{\mathbf{\Sigma}}_{ee}\mathbf{A}_0\}], 0\} \quad (22)$$

where $\mathrm{tr}\{\mathbf{A}\}$ is the trace of the matrix $\mathbf{A}$,

$$\mathbf{A}_0 = \mathbf{I} - k^{-1}\mathbf{J}\mathbf{J}'$$

$$\hat{\mathbf{\Sigma}}_{ee} = (r-1)^{-1}\sum_{j=1}^{r}(\mathbf{Y}_j - \bar{\mathbf{y}})(\mathbf{Y}_j - \bar{\mathbf{y}})', \quad (23)$$

and

$$\hat{\mu}_0 = k^{-1}\mathbf{J}'\bar{\mathbf{y}}. \quad (24)$$

The estimator $\hat{\sigma}^2$ is a quadratic estimator closely related to the analysis of variance estimator.

Two predictors were compared in the Monte Carlo study. Both are of the form

$$\hat{\mathbf{y}} = \bar{\mathbf{y}} - \hat{\mathbf{H}}'(\bar{\mathbf{y}} - \hat{\mu}\mathbf{J}), \quad (25)$$

where

$$\bar{\mathbf{y}} = r^{-1}\sum_{j=1}^{r} Y_j.$$

They differ in the construction of $\hat{\mathbf{H}}$ and $\hat{\mu}$. The first predictor is of the form (11) and uses the full estimated $\hat{\mathbf{\Sigma}}_{ee}$ in $\hat{\mathbf{H}}$ and in the estimator of μ. The predictor is called the general predictor as an abbreviation for estimated generalized least squares predictor. The general predictor is

$$\hat{\mathbf{y}}_g = \bar{\mathbf{y}} - \hat{\mathbf{H}}'_g(\bar{\mathbf{y}} - \hat{\mu}_g\mathbf{J}), \quad (26)$$

where

$$\hat{\mathbf{H}}'_g = r^{-1}\hat{\mathbf{\Sigma}}_{ee}\hat{\mathbf{\Sigma}}_{zz}^{-1},$$

$$\hat{\mu}_g = (\mathbf{J}'\hat{\mathbf{\Sigma}}_{zz}^{1}\mathbf{J})^{-1}\mathbf{J}'\mathbf{\Sigma}_{zz}^{-1}\bar{\mathbf{y}}, \quad (27)$$

$$\hat{\mathbf{\Sigma}}_{zz} = r^{-1}\hat{\mathbf{\Sigma}}_{ee} + \mathbf{I}\hat{\sigma}^2, \quad (28)$$

and $\hat{\mu}_g$ is the estimated generalized least squares estimator of μ.

The second predictor is

$$\hat{\mathbf{y}}_d = \bar{\mathbf{y}} - \hat{\mathbf{H}}'_d(\bar{\mathbf{y}} - \hat{\mu}_d\mathbf{J}), \quad (29)$$

where

$$\hat{\mathbf{H}}'_d = r^{-1}\mathbf{M}_{ee}\hat{\mathbf{D}}_{zz}^{-1},$$

$\mathbf{M}_{ee} = \mathrm{diag}\,\hat{\mathbf{\Sigma}}_{ee}$, $\hat{\mathbf{D}}_{zz} = \mathrm{diag}\,\hat{\mathbf{\Sigma}}_{zz}$, and the estimated μ is

$$\hat{\mu}_d = [\mathbf{J}'\hat{\mathbf{D}}_{zz}^{-1}\mathbf{J}]^{-1}\mathbf{J}'\hat{\mathbf{D}}_{zz}^{-1}\bar{\mathbf{y}}.$$

This predictor might be called the diagonal predictor because only the diagonal elements of $\hat{\mathbf{\Sigma}}_{ee}$ are used in the construction.

The entries in Table 1 are for $r = 14$. Each sample is composed of a random selection of $\mathbf{w}$ and a random sample of 14 **e**-vectors. Results are given for errors $\mu_{ij} \sim \mathrm{NI}(0, 2)$ and errors that are centered one-degree-of-freedom chi-square random variables. Thus, in both cases the errors have zero means and variances equal to two. The mean μ was set equal to zero. The second column of Table 1 contains the variance of the sample mean as an estimator of the w_i. Column three of Table 1 contains the ratio of the Monte Carlo variance of an element of $\hat{\mathbf{y}}_g$, where $\hat{\mathbf{y}}_g$ is defined by (28), to the Monte Carlo variance of the corresponding element of $\bar{\mathbf{y}}$ for normal errors. The ratios for elements one through four and element 7 are greater than one. The last two elements of $\mathbf{Y}_j$ are uncorrelated with other elements. Element seven has a small variance and element eight has a large variance. There is a large loss for the predictor relative to the simple mean for element seven and a large gain for element eight.

The fourth column of Table 1 contains the ratios of the variance of the predictor of (29) to the variance of the mean for normal errors. In all cases the diagonal predictor is superior to the general predictor defined in (28). The difference is relatively constant at about 30%. The diagonal predictor is not always superior to the simple mean but the loss is small for elements one, three, and seven. On the other hand, the gains relative to the simple mean are large for elements six and eight. The Monte Carlo variances for both predictors are larger than the approximations associated with equations (15) and (18) except for element 8.

It is somewhat surprising that the diagonal procedure did better relative to the simple mean for chi-square errors than for normal errors. With the chi-square error, the estimated mean and estimated variance are correlated. Hence, on the average, the large positive mean deviations are pulled toward the mean by a larger amount than the smaller negative deviation. The Associate Editor conjectured, and we concur, that this is one reason for the superior performance of the diagonal predictor. On the other hand, the general prediction procedure is poorer relative to the simple mean for chi-square errors than for normal errors. As the last column of Table 1 demonstrates, the diagonal predictor procedure uniformly dominates both the mean and the general prediction procedure for this parametric configuration with chi-square errors.

Table 1
Monte Carlo variance ratios for alternative small area predictors (10,000 samples, $r = 14$)

		Normal Errors		Chi-square Errors	
i	$V\{\bar{y}_i - w_i\}$	$\frac{\hat{V}\{\hat{y}_{gi} - w_i\}}{\hat{V}\{\bar{y}_i - w_i\}}$	$\frac{\hat{V}\{\hat{y}_{di} - w_i\}}{\hat{V}\{\bar{y}_i - w_i\}}$	$\frac{\hat{V}\{\hat{y}_{gi} - w_i\}}{\hat{V}\{\bar{y}_i - w_i\}}$	$\frac{\hat{V}\{\hat{y}_{di} - w_i\}}{\hat{V}\{\bar{y}_i - w_i\}}$
1	0.2414	1.277	1.025	1.430	0.899
2	0.3445	1.252	0.875	1.371	0.768
3	0.2268	1.351	1.019	1.480	0.954
4	0.4771	1.003	0.735	1.099	0.686
5	0.4113	0.926	0.876	1.016	0.699
6	0.5121	0.913	0.677	0.975	0.618
7	0.1449	1.366	1.006	2.261	0.896
8	1.1214	0.520	0.384	0.725	0.371

The Monte Carlo variances of $\hat{\mu}_0$, $\hat{\mu}_g$, and $\hat{\mu}_d$ as estimators of μ are 0.150, 0.273, and 0.146, respectively. If $\mathbf{\Sigma}_{ee}$ and σ^2 are known, the variances of $\hat{\mu}_0$, $\hat{\mu}_g$, and $\hat{\mu}_d$ are 0.149, 0.122, and 0.140, respectively. The use of an estimated covariance matrix for $\hat{\mu}_g$ produced an estimator with larger variance than that of the simple mean.

The predictors are unbiased under the model when the errors are normally distributed. The predictors are biased with chi-square errors because the sample mean is correlated with the sample variance. Table 2 contains the Monte Carlo bias divided by the Monte Carlo standard error of the mean. The bias of the general procedure is 20% to 50% larger than that of the diagonal procedure. In both cases, the squared bias added to the variance produces a mean square error for the procedure that is about 4% to 10% larger than the variance.

This small study demonstrates that use of an estimated covariance matrix with large variability can lead to predictors that are less efficient than the simple mean.

Table 2
Monte Carlo relative bias of alternative small area predictors (10,000 samples, r=14, chi-square errors)

i	$\frac{\text{Ave.}(\hat{y}_{gi} - w_i)}{[\hat{V}\{\bar{y}_i - w_i\}]^{1/2}}$	$\frac{\text{Ave.}\{\hat{y}_{di} - w_i)}{[\hat{V}\{\bar{y}_i - w_i\}]^{1/2}}$
1	-0.28	-0.19
2	-0.27	-0.18
3	-0.30	-0.17
4	-0.27	-0.18
5	-0.26	-0.21
6	-0.29	-0.20
7	-0.24	-0.20
8	-0.24	-0.21

4. Application to PES data for postcensal estimation

4.1 Postcensal estimation

The U.S. Bureau of the Census provides annual estimates of the population of small areas based on the decennial censuses and on other sources of information. To consider the possible use of adjusted 1990 Census counts in the postcensal estimation process, the Bureau examined the PES data and defined a new set of 357 poststrata.

The 357 poststrata are composed of 51 poststratum groups, each of which is subdivided into 7 age-sex categories. The seven age-sex categories were (1) both sexes 0-17, (2) males 18-29, (3) males 30-49, (4) females 18-29, (5) females 30-49, (6) males 50+ and (7) females 50+. The factors that define the 51 poststratum groups are race/ethnicity (Non-Hispanic White, Black, Non-Black Hispanic, Asian, American Indian); tenure (owner, renter); type of area (urbanized area of population greater than 250,000, other urbanized area, non-urbanized area) and region (west, South, Midwest, Northeast). Due to sample size limitations, American Indians comprised a single poststratum group and Asians were dichotomized into two poststratum groups - owners and renters. Of the remaining 48 poststratum groups, the first 24 groups reflect a full cross classification of categories for Non-Hispanic White. The next 12 groups are for Black and provide a full cross classification of tenure by region for urbanized areas of population greater than 250,000 but otherwise do not provide regional detail. The same 12 poststratum groups were used for Non-Black Hispanics as were used for Blacks.

A 357×357 covariance matrix was obtained with the same jackknife algorithm used for the 1,392 poststrata of the 1990 PES. We denote this raw covariance matrix by $\hat{\mathbf{\Sigma}}_{ee}$. Hogan (1993) provides a detailed description of the 357 poststrata and gives the motivation for their construction.

4.2 Regression model

We eliminated Asian and American Indian data from the smoothing process. Hence, minority refers to the combination of Black and Non-Black Hispanic. The data set of interest contains 336 adjustment factors and their estimated raw covariances. The minority by age-sex interaction was included in the regression model after examination of the 1990 data indicated that the net undercount differential between Black and non-Black varied by sex and age-group. The regression model (1) contains 21 explanatory variables. They are:

1. X_0 = intercept
2. X_j = indicator variable for age-sex categories: $j = 1, 2, ..., 6$ in the order; ages 0-17, male 18-29, male 30-49, *etc.* (female 50+ is the class with no variable)

3. X_7 = indicator variable for renter
4. X_8 = indicator variable for Black
5. X_9 = indicator variable for Non-Black Hispanic
6. X_j = indicator variable for type of place: $j = 10, 11$ for urbanized area 250,000+ and other urban, respectively
7. X_j = indicator variable for region: $j = 12, 13, 14$ for Northeast, South and West, respectively
8. X_j = indicator variable for minority by age-sex interaction: $j = 15, \ldots, 20$ for minority 0-17, minority male (18-29), *etc*.

The variables X_{12}, X_{13} and X_{14} were the 1990 census proportions of persons in the poststratum group in the particular region for the Black and Non-Black Hispanic poststratum groups that were combined over regions.

A refinement was made in model (3) for the empirical application. On the basis of preliminary analysis, the specified error structure of $\mathbf{w}$, the model error, was changed from $\mathbf{\Sigma}_{ww} = \sigma^2 \mathbf{I}$ to

$$\mathbf{\Sigma}_{ww} = \mathbf{K}_1 \sigma_1^2 + \mathbf{K}_2 \sigma_2^2, \tag{30}$$

where $\mathbf{K}_1$ is an $n \times n$ diagonal matrix with ones for minority poststrata and zeros elsewhere and $\mathbf{K}_2$ is an $n \times n$ diagonal matrix with ones for nonminority poststrata and zeros elsewhere. The estimated variances are $\hat{\sigma}_1^2 = 0.000506$ (0.000140) and $\hat{\sigma}_2^2 = 0.000112$ (0.000030), where the numbers in parentheses are standard errors. The standard error of the difference is (0.000141). Hence there is evidence that the variances are different for the two groups.

In our discussion of predictors, we considered two predictors, the substitution predictor of (11) and the diagonal predictor of (16). It is natural to consider a compromise predictor of the form

$$\hat{\mathbf{y}}_\varphi = \mathbf{X}\hat{\boldsymbol{\beta}}_\varphi + \hat{\mathbf{G}}'_\varphi(\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}}_\varphi)$$

$$= \mathbf{Y} - \hat{\mathbf{H}}'_\varphi(\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}}_\varphi), \tag{31}$$

where $0 \le \varphi \le 1$,

$$\hat{\mathbf{G}}'_\varphi = \hat{\mathbf{\Sigma}}_{\varphi\varphi}^{-1} \hat{\mathbf{\Sigma}}_{ww},$$

$$\hat{\mathbf{H}}'_\varphi = \mathbf{I} - \hat{\mathbf{G}}'_\varphi = \hat{\mathbf{\Sigma}}_{\varphi\varphi}^{-1}[\varphi \hat{\mathbf{D}}_{ee} + (1-\varphi)\hat{\mathbf{\Sigma}}_{ee}],$$

$$\hat{\mathbf{\Sigma}}_{\varphi\varphi} = \hat{\mathbf{\Sigma}}_{ww} + \varphi \hat{\mathbf{D}}_{ee} + (1-\varphi)\hat{\mathbf{\Sigma}}_{ee},$$

$$\hat{\mathbf{D}}_{ee} = \operatorname{diag}\{\hat{\mathbf{\Sigma}}_{ee}\},$$

$$\hat{\boldsymbol{\beta}}_\varphi = (\mathbf{X}'\hat{\mathbf{\Sigma}}_{\varphi\varphi}^{-1}\mathbf{X})^{-1}\mathbf{X}'\hat{\mathbf{\Sigma}}_{\varphi\varphi}^{-1}\mathbf{Y},$$

and

$$\hat{\mathbf{\Sigma}}_{ww} = \mathbf{K}_1 \hat{\sigma}_1^2 + \mathbf{K}_2 \hat{\sigma}_2^2.$$

The predictor (31) with $\varphi = 0$ is the substitution predictor and the predictor (31) with $\varphi = 1$ is the diagonal predictor. There should be some φ, $0 < \varphi < 1$, that gives a predictor with smaller prediction variance than either of the extremes.

The PES direct estimate of the total number of persons is the weighted sum of the adjustment factors, where the weights are the census counts in the post strata. The standard error of the direct estimator of the total is relatively small and the direct estimator is judged to be the preferred estimator of the total. Therefore, the model predictors are constructed subject to the constraint that the weighted sum of the predictors is equal to the direct estimate of the total. Thus, the restriction is

$$\hat{Y}_T = \sum_{i=1}^{336} a_i y_i = \sum_{i=1}^{336} a_i \tilde{y}_i,$$

where $\hat{Y}_T$ is PES direct estimator of the total, a_i is the census count in the i^{th} post stratum, and $\tilde{y}_i$ is the final predictor. In the actual computations the a_i were normalized to sum to one. Battese, Harter and Fuller (1988) made an adjustment in the predictions to create estimators to meet the restriction. Ghosh and Rao (1994) discuss such adjustments. We use a procedure that permits direct estimation of the variance of the restricted predictions.

We imposed the restriction on the initial predictors by a procedure that, approximately, constructed the best predictors of 335 quantities that are estimated to be uncorrelated with $\hat{Y}_T$. Let $\hat{\mathbf{\Sigma}}_{zz}$ be the estimated covariance matrix of $\mathbf{Y} = (Y_1, Y_2, \ldots, Y_{336})'$ and define

$$\mathbf{CY} = (\hat{Y}_T, Y_2 - b_2\hat{Y}_T, \ldots, Y_{336} - b_{336}\hat{Y}_T)',$$

where

$$\mathbf{C} = \mathbf{BT},$$

$$\mathbf{T} = \begin{pmatrix} & \mathbf{a} & \\ \mathbf{0} & & \mathbf{I}_{335} \end{pmatrix},$$

$$\mathbf{a} = (a_1, a_2, \ldots, a_{336}),$$

$$\mathbf{B} = \begin{pmatrix} 1 & \mathbf{0}' \\ -\mathbf{b}_{335} & \mathbf{I}_{335} \end{pmatrix},$$

$$\mathbf{b}_{335} = \begin{pmatrix} \mathbf{0}' \\ \mathbf{I}_{335} \end{pmatrix}' \mathbf{\Sigma}_{zz}\, \mathbf{a}' (\mathbf{a}\hat{\mathbf{\Sigma}}_{zz}\mathbf{a}')^{-1},$$

$\mathbf{I}_k$ is the $k \times k$ identity matrix, and $\mathbf{0}$ is a column vector containing all zeros. The elements of $\mathbf{CY}$ are uncorrelated with $\hat{Y}_T$.

If we let $\hat{\mathbf{y}}$ be the model predictor of $\mathbf{y}$, then the model predictor of $\mathbf{Cy}$ is $\mathbf{C}\hat{\mathbf{y}}$. If we use the model predictor for the last 335 elements of $\mathbf{Cy}$ and use $\hat{\mathbf{Y}}_T$ as the estimator for the first element of $\mathbf{Cy}$, the predictor of $\mathbf{y}$ is

$$\tilde{\mathbf{y}} = \mathbf{Y} - \mathbf{C}^{-1}\mathbf{AC}\hat{\mathbf{H}}'_{\varphi}(\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}}_{\varphi}), \qquad (32)$$

where

$$\mathbf{A} = \begin{pmatrix} 0 & \mathbf{0}' \\ \mathbf{0} & \mathbf{I}_{335} \end{pmatrix}.$$

The estimated variance of $\tilde{\mathbf{y}}$ is

$$\begin{aligned} &\hat{\mathbf{V}}\{\tilde{\mathbf{y}} - \mathbf{y}\} \\ &= (\mathbf{I} - \ddot{\mathbf{H}}'_{\varphi})\hat{\boldsymbol{\Sigma}}_{ee}(\mathbf{I} - \ddot{\mathbf{H}}'_{\varphi})' + \ddot{\mathbf{H}}'_{\varphi}\hat{\boldsymbol{\Sigma}}_{ww}\ddot{\mathbf{H}}_{\varphi} \\ &+ \mathbf{C}^{-1}\mathbf{AC}[\hat{\mathbf{H}}'_{\varphi}\mathbf{X}\hat{\mathbf{V}}_{\beta\beta}\mathbf{X}'\hat{\mathbf{H}}_{\varphi} + \hat{\boldsymbol{\Gamma}}_{33} + \hat{\boldsymbol{\Gamma}}_{44}]\,\mathbf{C}'\mathbf{AC}^{-1\prime}, \qquad (33) \end{aligned}$$

where $\ddot{\mathbf{H}}'_{\varphi} = \mathbf{C}^{-1}\mathbf{AC}\hat{\mathbf{H}}'_{\varphi}$, and $\hat{\mathbf{V}}_{\beta\beta}$, $\hat{\boldsymbol{\Gamma}}_{33}$ and $\hat{\boldsymbol{\Gamma}}_{44}$ are defined in Appendix B. The sum of the first two terms on the right of (33) is an estimator of the variance treating $\ddot{\mathbf{H}}_{\varphi}$ as a fixed matrix. The final term on the right of (33) estimates the increase in variance due to estimating the variance.

4.3 Smoothed factors

For the vector of 336 observations, we produced smoothed factors using the generalized predictor (32) for several values of φ. Note that $\varphi = 0$ corresponds to the substitution predictor and $\varphi = 1$ corresponds to the diagonal predictor.

The estimated standard errors of the predictors were calculated using the crude variance approximation of Appendix B. The average of the ratios of the standard error of $\tilde{y}_{\varphi}$ to $\tilde{y}_{0.6}$ for some selected values of φ are given in Table 3. The ordering of the ratios is approximately the same for the 48 stratum groups as for the original 336 poststrata. A poststratum group is formed by combining the seven age-sex cells within a given race-by-tenure-by-urbanity-by-region classification. On the basis of these calculations, a φ of 0.5 or 0.6 is the preferred estimator, although the estimated differences in efficiencies are not large. Any member of the φ-class is much superior to the original Y-estimator. The average estimated variance efficiency is about 400% for the φ-predictors, relative to the original poststratum estimators.

Table 3
Average of ratio of standard error of $\tilde{y}_{\varphi}$ and of Y to standard error $\tilde{y}_{0.6}$

Predictor	336 Poststrata	48 Poststratum groups
$\varphi = 0$	1.014	1.045
$\varphi = 0.5$	0.995	1.001
$\varphi = 0.6$	1.000	1.000
$\varphi = 0.7$	1.006	1.001
$\varphi = 0.8$	1.014	1.005
$\varphi = 1.0$	1.046	1.037
Original Y	2.235	2.294

Table 4 presents the raw PES estimates, $\mathbf{Y}$, and the $\tilde{y}_{0.6}$ estimates of net undercount for each of 48 poststratum groups. The net undercount is the difference between the estimated total population in the poststratum and the census count divided by the census count.

We chose $\varphi = 0.6$ as the preferred estimator on the basis of the crude standard error ratios of Table 3. The predictions and standard errors are very similar for $\varphi = 0.5$, 0.6 and 0.7. A φ greater than zero has advantages over a φ of zero. The accuracy of the numerical calculations should be better with φ greater than zero because $\hat{\boldsymbol{\Sigma}}_{\varphi\varphi}$ has larger eigenvalues with $\varphi > 0$ than with $\varphi = 0$. One could make a case for using $\varphi = 1.0$ because of the simplicity of the calculations and of the good estimated relative efficiency.

The estimated standard errors of the predictors are considerably smaller than those of the raw estimates. In addition, the set of predictors contains fewer extreme estimates. For example, for poststratum groups 34, 39 and 48, the $\tilde{y}_{0.6}$ estimates of the percent net undercount are 6.04, 0.17 and 7.51 while the raw estimates are 11.06, -4.14 and 18.76, respectively. Most smoothed estimates differ from the direct estimate by less than one direct estimated standard error. The three largest standardized differences are for Black Owner-Large Urban in the West, Black Renter-Larger Urban in the Northeast, and Non-Black Hispanic Owner-Large Urban in the Midwest. In the three cases, the difference between the direct estimate and the smoothed estimate divided by the direct standard error is about 1.8.

Acknowledgements

This research was partly supported by Cooperative Agreement 43-3AEU-3-80088 between Iowa State University, the National Agricultural Statistics Service and the U.S. Bureau of the Census. This paper reports the results of research and analysis undertaken by staff of the Bureau of the Census and Iowa State University. It has undergone a more limited review than official Census Bureau publications. This paper is released to inform interested parties of research and to encourage discussion. We thank the referees and editors for many comments that led to improvements in the manuscript.

Table 4
Estimated percent net undercount by poststratum group

Poststratum Group		Y	s.e. (Y)	$\tilde{y}_{0.6}$	s.e. $(\tilde{y}_{0.6})$
Non-Hispanic White Owner Large Urban					
1.	N.E.	-2.08	1.04	-0.63	0.60
2.	South	0.69	0.72	0.38	0.44
3.	Midwest	-0.26	0.39	-0.13	0.31
4.	West	-0.34	0.64	-0.02	0.44
Non-Hispanic White Owner Other Urban					
5.	N.E.	-1.07	0.48	-0.73	0.35
6.	South	0.52	0.43	0.53	0.33
7.	Midwest	-0.10	0.40	0.01	0.31
8.	West	0.63	0.58	0.30	0.40
Non-Hispanic White Owner Non-Urban					
9.	N.E.	-0.53	0.69	-0.28	0.47
10.	South	0.18	0.69	0.58	0.45
11	Midwest	-0.70	1.16	0.16	0.64
12.	West	0.29	0.69	0.38	0.46
Non-Hispanic White Renter Large Urban					
13.	N.E.	1.17	1.43	2.07	0.61
14.	South	2.62	1.56	3.53	0.64
15.	Midwest	2.39	1.70	2.53	0.60
16.	West	3.28	1.72	3.10	0.58
Non-Hispanic White Renter Other Urban					
17.	N.E.	3.53	1.62	2.29	0.61
18.	South	3.30	1.86	3.67	0.67
19.	Midwest	1.24	1.13	2.39	0.53
20.	West	4.70	1.47	3.20	0.57
Non-Hispanic White Renter Non-Urban					
21.	N.E.	6.97	4.67	3.54	0.92
22.	South	6.65	1.93	3.60	0.66
23.	Midwest	2.93	1.60	2.36	0.66
24.	West	6.48	2.06	3.48	0.67
Black Owner Large Urban					
25.	N.E.	1.65	1.96	0.97	0.91
26.	South	2.20	0.94	2.30	0.70
27.	Midwest	0.82	0.88	1.13	0.67
28.	West	6.49	2.16	2.54	0.96
Black Owner Other Urban					
29.	U.S.	1.36	1.01	2.05	0.72
Black Owner Non-Urban					
30.	U.S.	3.64	2.03	2.85	0.98
Black Renter Large Urban					
31.	N.E.	9.13	1.93	5.57	0.96
32.	South	6.69	2.17	6.42	1.10
33.	Midwest	6.38	1.91	5.43	1.03
34.	West	11.06	3.35	6.04	1.12
Black Renter Other Urban					
35.	U.S.	4.33	1.28	4.99	0.82
Black Renter Non-Urban					
36.	U.S.	4.84	5.95	5.90	1.24
Non-Black Hispanic Owner Large Urban					
37.	N.E.	0.68	4.44	3.00	1.18
38.	South	2.59	0.95	2.52	0.72
39.	Midwest	-4.14	2.38	0.17	0.97
40.	West	2.98	0.92	2.89	0.68
Non-Black Hispanic Owner Other Urban					
41.	U.S.	0.95	1.70	2.32	0.87
Non-Black Hispanic Owner Non-Urban					
42.	U.S.	2.80	2.83	2.88	1.16
Non-Black Hispanic Renter Large Urban					
43.	N.E.	7.21	4.04	5.85	1.27
44.	South	10.30	3.11	7.35	1.15
45.	Midwest	7.11	3.74	5.71	1.21
46.	West	6.29	2.09	6.45	0.98
Non-Black Hispanic Renter Other Urban					
47.	U.S.	7.07	3.10	6.26	1.09
Non-Black Hispanic Renter Non-Urban					
48.	U.S.	18.76	7.24	7.51	1.38

Appendix A

Estimation of $\boldsymbol{\Sigma}_{ww}$

The estimators of σ_1^2 and σ_2^2 of $\boldsymbol{\Sigma}_{ww}$ are patterned after analysis of variance estimators. The estimation process contains several steps using improved estimators from one step in the next step. We partition the regression problem as

$$\begin{pmatrix}\mathbf{Y}_1\\ \mathbf{Y}_2\end{pmatrix} = \begin{pmatrix}\mathbf{X}_1 & \mathbf{0}\\ \mathbf{0} & \mathbf{X}_2\end{pmatrix}\begin{pmatrix}\boldsymbol{\beta}_1\\ \boldsymbol{\beta}_2\end{pmatrix} + \begin{pmatrix}\mathbf{z}_1\\ \mathbf{z}_2\end{pmatrix},$$

where $(\mathbf{Y}_1, \mathbf{X}_1)$ contains the observations for minorities and $(\mathbf{Y}_2, \mathbf{X}_2)$ contains the remaining observations. Let $\mathbf{Y}_1$ be an n_1-dimensional column vector and $\mathbf{Y}_2$ be an n_2-dimensional column vector observations. An initial estimator of $(\boldsymbol{\beta}_1', \boldsymbol{\beta}_2')'$ is

$$\begin{pmatrix}\tilde{\boldsymbol{\beta}}_1\\ \tilde{\boldsymbol{\beta}}_2\end{pmatrix} = \begin{pmatrix}\left(\mathbf{X}_1'\hat{\boldsymbol{\Sigma}}_{ee11}^{-1}\mathbf{X}_1\right)^{-1} & \mathbf{X}_1'\hat{\boldsymbol{\Sigma}}_{ee11}\mathbf{Y}_1\\ \left(\mathbf{X}_2'\hat{\boldsymbol{\Sigma}}_{ee22}^{-1}\mathbf{X}_2\right)^{-2} & \mathbf{X}_2'\hat{\boldsymbol{\Sigma}}_{ee22}\mathbf{Y}_2\end{pmatrix},$$

where

$$\hat{\boldsymbol{\Sigma}}_{ee} = \begin{pmatrix}\hat{\boldsymbol{\Sigma}}_{ee11} & \hat{\boldsymbol{\Sigma}}_{ee12}\\ \hat{\boldsymbol{\Sigma}}_{ee21} & \hat{\boldsymbol{\Sigma}}_{ee22}\end{pmatrix}$$

is partitioned to conform to the partition of $\mathbf{Y}$.

Initial estimators of σ_1^2 and σ_2^2 are

$$\tilde{\sigma}_i^2 = \max\{[(\mathbf{Y}_i - \mathbf{X}_i\tilde{\boldsymbol{\beta}}_i)'\hat{\boldsymbol{\Sigma}}_{eeii}^{-1}(\mathbf{Y}_i - \mathbf{X}_i\tilde{\boldsymbol{\beta}}_i) - g_{1i}]\, g_{2i}^{-1},\ 0\},$$

for $i = 1, 2,$ where

$$g_{1i} = \mathrm{tr}\{\hat{\boldsymbol{\Sigma}}_{ee11}(\mathbf{I}_{ni} - \mathbf{X}_i\mathbf{A}_{Mii})'\hat{\boldsymbol{\Sigma}}_{eeii}^{-1}(\mathbf{I}_{ni} - \mathbf{X}_i\mathbf{A}_{Mii}),$$

$$g_{2i} = \mathrm{tr}\{(\mathbf{I}_{ni} - \mathbf{X}_i\mathbf{A}_{Mii})'\hat{\boldsymbol{\Sigma}}_{eeii}^{-1}(\mathbf{I}_{ni} - \mathbf{X}_i\mathbf{A}_{Mii}),$$

$$\mathbf{A}_{Mii} = (\mathbf{X}_i'\hat{\boldsymbol{\Sigma}}_{eeii}^{-1}\mathbf{X}_i)^{-1}\mathbf{X}_i'\hat{\boldsymbol{\Sigma}}_{eeii}^{-1},$$

and $\mathbf{I}_{ni}$ is the $n_i \times n_i$ identity matrix.

The final estimators are

$$\ddot{\sigma}_i^2 = \max\{[(\mathbf{Y}_i - \mathbf{X}_i\ddot{\boldsymbol{\beta}}_i)'\tilde{\boldsymbol{\Sigma}}_{zzii}^{-1}(\mathbf{Y}_i - \mathbf{X}_i\ddot{\boldsymbol{\beta}}_i) - \ddot{g}_{1i}]\,\ddot{g}_{2i}^{-1},\ 0\},$$

for $i = 1, 2,$ where

$$\ddot{g}_{1i} = \mathrm{tr}\{\hat{\boldsymbol{\Sigma}}_{eeii}(\mathbf{I}_{ni} - \mathbf{X}_i\ddot{\mathbf{A}}_{Mii})\tilde{\boldsymbol{\Sigma}}_{zzii}^{-1}(\mathbf{I}_{ni} - \mathbf{X}_i\ddot{\mathbf{A}}_{Mii})\}$$

$$\ddot{g}_{2i} = \mathrm{tr}\{(\mathbf{I}_{ni} - \mathbf{X}_i\ddot{\mathbf{A}}_{Mii})\tilde{\boldsymbol{\Sigma}}_{zzii}^{-1}(\mathbf{I}_{ni} - \mathbf{X}_i\ddot{\mathbf{A}}_{Mii})\}$$

$$\tilde{\boldsymbol{\Sigma}}_{zzii} = \hat{\boldsymbol{\Sigma}}_{eeii} + \tilde{\sigma}_i^2\mathbf{I}_{ni}$$

$$\ddot{\boldsymbol{\beta}}_i = (\mathbf{X}_i'\hat{\boldsymbol{\Sigma}}_{zzii}^{-1}\mathbf{X}_i)^{-1}\mathbf{X}_i'\hat{\boldsymbol{\Sigma}}_{zzii}^{-1}\mathbf{Y}_i = \ddot{\mathbf{A}}_{Mii}^{-1}\mathbf{Y}_i.$$

Estimators of the variance are

$$\hat{V}\{\ddot{\sigma}_i^2\}$$

$$= 2\ddot{g}_{2i}^{-2}$$

$$\times\ \mathrm{tr}\{[\tilde{\boldsymbol{\Sigma}}_{zzii}(\mathbf{I}_{ni} - \mathbf{X}_i\ddot{\mathbf{A}}_{Mii})'\tilde{\boldsymbol{\Sigma}}_{zzii}^{-1}(\mathbf{I}_{ni} - \mathbf{X}_i\ddot{\mathbf{A}}_{Mii})]^2\}$$

$$+\ 2\ddot{g}_{2i}^{-2}d_e^{-1}$$

$$\times\ \mathrm{tr}\{[\hat{\boldsymbol{\Sigma}}_{eeii}(\mathbf{I}_{ni} - \mathbf{X}_i\ddot{\mathbf{A}}_{Mii})'\tilde{\boldsymbol{\Sigma}}_{zzii}^{-1}(\mathbf{I}_{ni} - \mathbf{X}_i\ddot{\mathbf{A}}_{Mii})]^2\},$$

for $i = 1, 2.$ The estimated covariance is

$$\hat{C}\{\ddot{\sigma}_1^2, \ddot{\sigma}_2^2\} = 2\,\mathrm{tr}\{\tilde{\boldsymbol{\Sigma}}_{zz}\mathbf{M}_{11}\tilde{\boldsymbol{\Sigma}}_{zz}\mathbf{M}_{22}\}$$

$$+\ 2d_e^{-1}\,\mathrm{tr}\{\hat{\boldsymbol{\Sigma}}_{ee}\mathbf{M}_{11}\hat{\boldsymbol{\Sigma}}_{ee}\mathbf{M}_{22}\},$$

where

$$\mathbf{M}_{11} = \begin{pmatrix} g_{21}^{-1}(\mathbf{I}_{n_1} - \mathbf{X}_1\mathbf{A}_{M11})'\boldsymbol{\Sigma}_{zz11}^{-1}(\mathbf{I}_{n_1} - \mathbf{X}_1\mathbf{A}_{M11}) & \mathbf{0}\\ \mathbf{0}' & \mathbf{0}\end{pmatrix}$$

and

$$\mathbf{M}_{22} = \begin{pmatrix}\mathbf{0} & \mathbf{0}'\\ \mathbf{0} & g_{21}^{-1}(\mathbf{I}_{n_2} - \mathbf{X}_2\mathbf{A}_{M22})'\boldsymbol{\Sigma}_{zz22}^{-1}(\mathbf{I}_{n_2} - \mathbf{X}_2\mathbf{A}_{M22})\end{pmatrix}.$$

See Searle (1971, Chapter 2 and page 435).

Appendix B

Approximations for the variance of predictors

Our model is

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{w} + \mathbf{e}, \tag{B.1}$$

where $\mathbf{Y}$ is an n-dimensional column vector, $\mathbf{X}$ is an $n \times k$ fixed matrix,

$$\begin{pmatrix}\mathbf{w}\\ \mathbf{e}\end{pmatrix} \sim N\left(\begin{pmatrix}\mathbf{0}\\ \mathbf{0}\end{pmatrix}, \begin{pmatrix}\boldsymbol{\Sigma}_{ww} & \mathbf{0}\\ \mathbf{0} & \boldsymbol{\Sigma}_{ee}\end{pmatrix}\right), \tag{B.2}$$

and $\boldsymbol{\Sigma}_{ww}$ is defined in (30) of the text.

For purpose of variance estimation, we assume $\hat{\boldsymbol{\Sigma}}_{ee}$ is an unbiased estimator of $\boldsymbol{\Sigma}_{ee}$ distributed as a multiple of a Wishart matrix with d_e degrees of freedom independent of $(\mathbf{w}, \mathbf{e})$. We let $\mathbf{y}$ be the unknown true vector to be predicted and write

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{w} \text{ and } \mathbf{z} = \mathbf{w} + \mathbf{e}.$$

By a Taylor expansion

$$\hat{\mathbf{y}}_\varphi - \mathbf{y} = \mathbf{e} - \hat{\mathbf{H}}_\varphi'(\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}}_\varphi)$$

$$= \mathbf{e} - \mathbf{H}_\varphi'\mathbf{z} + \mathbf{H}_\varphi'\mathbf{X}(\hat{\boldsymbol{\beta}}_\varphi' - \boldsymbol{\beta}) - (\hat{\mathbf{H}}_\varphi' - \mathbf{H}_\varphi')\,\mathbf{z} + O_p(n^{-1}), \tag{B.3}$$

where

$$\mathbf{H}_{\varphi} = \boldsymbol{\Sigma}_{\varphi\varphi}^{-1}[\varphi\,\mathbf{D}_{ee} + (1-\varphi)\,\boldsymbol{\Sigma}_{ee}]$$

and $\hat{\mathbf{H}} = \hat{\mathbf{H}}_{\varphi}$ is defined in (31). The error in $\hat{\boldsymbol{\beta}}_{\varphi}$ is

$$\hat{\boldsymbol{\beta}}_{\varphi} - \boldsymbol{\beta} = \left(\mathbf{X}'\hat{\boldsymbol{\Sigma}}_{\varphi\varphi}^{-1}\mathbf{X}\right)^{-1}\mathbf{X}'\hat{\boldsymbol{\Sigma}}_{\varphi\varphi}^{-1}\mathbf{z} \tag{B.4}$$

Now $\hat{\boldsymbol{\Sigma}}_{ee}$ is independent of $\mathbf{z}$ and $\mathbf{Y} - \mathbf{X}\tilde{\boldsymbol{\beta}}$ is uncorrelated with $\tilde{\boldsymbol{\beta}} - \boldsymbol{\beta}$ if the true $\boldsymbol{\Sigma}_{zz}$ is used in place of $\hat{\boldsymbol{\Sigma}}_{\varphi\varphi}$. Therefore

$$E\{\mathbf{H}_0'\,\mathbf{X}(\tilde{\boldsymbol{\beta}}_0 - \boldsymbol{\beta})\,\mathbf{z}'(\hat{\mathbf{H}}_0 - \mathbf{H}_0)\} = \mathbf{0}, \tag{B.5}$$

where $\hat{\mathbf{H}}_0$ is constructed using $\mathbf{Y} - \mathbf{X}\tilde{\boldsymbol{\beta}}$ in the estimators of the elements of $\hat{\boldsymbol{\Sigma}}_{ww}$ defined in Appendix A and $\mathbf{H}_0 = \boldsymbol{\Sigma}_{zz}^{-1}\boldsymbol{\Sigma}_{ee}$. We set the covariance between $\hat{\boldsymbol{\beta}}$ and $\hat{\mathbf{H}}_{\varphi}$ equal to zero for all φ. Now

$$\begin{aligned}\hat{\mathbf{H}}_{\varphi} &= \hat{\boldsymbol{\Sigma}}_{\varphi\varphi}^{-1}[\varphi\,\hat{\mathbf{D}}_{ee} + (1-\varphi)\,\hat{\boldsymbol{\Sigma}}_{ee}]\\ &= [\hat{\boldsymbol{\Sigma}}_{ww} + \varphi\,\hat{\mathbf{D}}_{ee} + (1-\varphi)\,\hat{\boldsymbol{\Sigma}}_{ee}]^{-1}[\varphi\,\hat{\mathbf{D}}_{ee} + (1-\varphi)\,\hat{\boldsymbol{\Sigma}}_{ee}]\end{aligned}$$

and

$$\begin{aligned}\hat{\mathbf{H}}_{\varphi} - \mathbf{H}_{\varphi} &= \boldsymbol{\Sigma}_{\varphi\varphi}^{-1}[\varphi(\hat{\mathbf{D}}_{ee} - \mathbf{D}_{ee}) + (1-\varphi)(\hat{\boldsymbol{\Sigma}}_{ee} - \boldsymbol{\Sigma}_{ee})]\\ &\quad - \boldsymbol{\Sigma}_{\varphi\varphi}^{-1}[\hat{\boldsymbol{\Sigma}}_{ww} - \boldsymbol{\Sigma}_{ww} + \varphi(\hat{\mathbf{D}}_{ee} - \mathbf{D}_{ee})\\ &\quad + (1-\varphi)(\hat{\boldsymbol{\Sigma}}_{ee} - \boldsymbol{\Sigma}_{ee})]\,\mathbf{H}_{\varphi}\\ &= \boldsymbol{\Sigma}_{\varphi\varphi}^{-1}[\varphi(\hat{\mathbf{D}}_{ee} - \mathbf{D}_{ee} + (1-\varphi)(\hat{\boldsymbol{\Sigma}}_{ee} - \boldsymbol{\Sigma}_{ee})]\,\mathbf{G}_{\varphi}\\ &\quad - \boldsymbol{\Sigma}_{\varphi\varphi}^{-1}[\hat{\boldsymbol{\Sigma}}_{ww} - \boldsymbol{\Sigma}_{ww}]\,\mathbf{H}_{\varphi},\end{aligned}$$

where $\mathbf{G}_{\varphi} = \mathbf{I} - \mathbf{H}_{\varphi}$. The contribution of $\hat{\mathbf{D}}_{ee} - \mathbf{D}_{ee}$ to the variance of $\hat{\mathbf{H}}_{\varphi}$ is small relative to the contribution of $\hat{\boldsymbol{\Sigma}}_{ee} - \boldsymbol{\Sigma}_{ee}$. Therefore, we omit $\hat{\mathbf{D}}_{ee} - \mathbf{D}_{ee}$ in our variance approximation. Then the expectation

$$\begin{aligned}&E\{(\mathbf{I} - \mathbf{H}_{\varphi})'(\hat{\boldsymbol{\Sigma}}_{ee} - \boldsymbol{\Sigma}_{ee})\boldsymbol{\Sigma}_{\varphi\varphi}^{-1}\mathbf{z}\mathbf{z}'\boldsymbol{\Sigma}_{\varphi\varphi}^{-1}\\ &\qquad(\hat{\boldsymbol{\Sigma}}_{ee} - \boldsymbol{\Sigma}_{ee})(\mathbf{I} - \mathbf{H}_{\varphi})\}\\ &= d_e^{-1}\mathbf{G}_{\varphi}'[\boldsymbol{\Sigma}_{ee}(\mathrm{tr}\{\boldsymbol{\Lambda}\boldsymbol{\Sigma}_{ee}\}) + \boldsymbol{\Sigma}_{ee}\boldsymbol{\Lambda}\boldsymbol{\Sigma}_{ee}]\,\mathbf{G}_{\varphi},\end{aligned} \tag{B.6}$$

where $\boldsymbol{\Lambda} = \boldsymbol{\Sigma}_{\varphi\varphi}^{-1}\boldsymbol{\Sigma}_{zz}\boldsymbol{\Sigma}_{\varphi\varphi}^{-1}$, because $\mathbf{z}$ is independent of $\hat{\boldsymbol{\Sigma}}_{ee}$. We also omit the term $d_e^{-1}\mathbf{G}_{\varphi}'\boldsymbol{\Sigma}_{ee}\boldsymbol{\Sigma}_{\varphi\varphi}^{-1}\boldsymbol{\Sigma}_{zz}\boldsymbol{\Sigma}_{\varphi\varphi}^{-1}\boldsymbol{\Sigma}_{ee}\mathbf{G}_{\varphi}$ in our variance approximation.

The expectation for the term containing $(\hat{\boldsymbol{\Sigma}}_{ww} - \boldsymbol{\Sigma}_{ww})$ is

$$E\{\mathbf{H}_{\varphi}'(\hat{\boldsymbol{\Sigma}}_{ww} - \boldsymbol{\Sigma}_{ww})\boldsymbol{\Sigma}_{\varphi\varphi}^{-1}\mathbf{z}\mathbf{z}'\boldsymbol{\Sigma}_{\varphi\varphi}^{-1}(\hat{\boldsymbol{\Sigma}}_{ww} - \boldsymbol{\Sigma}_{ww})\,\mathbf{H}_{\varphi}\},$$

where

$$\hat{\boldsymbol{\Sigma}}_{ww} - \boldsymbol{\Sigma}_{ww} = \begin{pmatrix}\mathbf{I}_{n1}(\hat{\sigma}_1^2 - \sigma_1^2) & \mathbf{0}\\ \mathbf{0} & \mathbf{I}_{n2}(\hat{\sigma}_2^2 - \sigma_2^2)\end{pmatrix}.$$

Approximating the expectation by treating $\mathbf{z}$ as independent of $\hat{\boldsymbol{\Sigma}}_{ww}$, we obtain

$$\mathbf{H}_{\varphi}'\begin{pmatrix}\boldsymbol{\Lambda}_{11}V\{\hat{\sigma}_1^2\} & \boldsymbol{\Lambda}_{12}C\{\hat{\sigma}_1^2,\ \hat{\sigma}_2^2\}\\ \boldsymbol{\Lambda}_{21}C\{\hat{\sigma}_1^2,\ \hat{\sigma}_2^2\} & \boldsymbol{\Lambda}_{22}V\{\hat{\sigma}_2^2\}\end{pmatrix}\mathbf{H}_{\varphi}, \tag{B.7}$$

where

$$\boldsymbol{\Lambda} = \begin{pmatrix}\boldsymbol{\Lambda}_{11} & \boldsymbol{\Lambda}_{12}\\ \boldsymbol{\Lambda}_{21} & \boldsymbol{\Lambda}_{22}\end{pmatrix} = \boldsymbol{\Sigma}_{\varphi\varphi}^{-1}\boldsymbol{\Sigma}_{zz}\boldsymbol{\Sigma}_{\varphi\varphi}^{-1}.$$

The Taylor expansion of $\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}$ is

$$\begin{aligned}\hat{\boldsymbol{\beta}} - \boldsymbol{\beta} &= (\mathbf{X}'\boldsymbol{\Sigma}_{\varphi\varphi}^{-1}\mathbf{X})^{-1}\mathbf{X}'\hat{\boldsymbol{\Sigma}}_{\varphi\varphi}^{-1}\mathbf{z}\\ &= (\mathbf{X}'\boldsymbol{\Sigma}_{\varphi\varphi}^{-1}\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\Sigma}_{\varphi\varphi}^{-1}\mathbf{z}\\ &\quad + (\mathbf{X}'\boldsymbol{\Sigma}_{\varphi\varphi}^{-1}\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\Sigma}_{\varphi\varphi}^{-1}(\hat{\boldsymbol{\Sigma}}_{\varphi\varphi} - \boldsymbol{\Sigma}_{\varphi\varphi})\\ &\quad \times \boldsymbol{\Sigma}_{\varphi\varphi}^{-1}\mathbf{X}(\mathbf{X}'\boldsymbol{\Sigma}_{\varphi\varphi}^{-1}\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\Sigma}_{\varphi\varphi}^{-1}\mathbf{z}\\ &\quad - (\mathbf{X}'\boldsymbol{\Sigma}_{\varphi\varphi}^{-1}\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\Sigma}_{\varphi\varphi}^{-1}\\ &\quad \times (\hat{\boldsymbol{\Sigma}}_{\varphi\varphi} - \boldsymbol{\Sigma}_{\varphi\varphi})\boldsymbol{\Sigma}_{\varphi\varphi}^{-1}\mathbf{z} + \text{Remainder.}\\ &= (\mathbf{X}'\boldsymbol{\Sigma}_{\varphi\varphi}^{-1}\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\Sigma}_{\varphi\varphi}^{-1}\mathbf{z}\\ &\quad + \mathbf{L}(\hat{\boldsymbol{\Sigma}}_{\varphi\varphi} - \boldsymbol{\Sigma}_{\varphi\varphi})\boldsymbol{\Sigma}_{\varphi\varphi}^{-1}\mathbf{Q}\boldsymbol{\Sigma}_{\varphi\varphi}^{-1}\mathbf{z}\\ &\quad - \mathbf{L}(\hat{\boldsymbol{\Sigma}}_{\varphi\varphi} - \boldsymbol{\Sigma}_{\varphi\varphi})\boldsymbol{\Sigma}_{\varphi\varphi}^{-1}\mathbf{z} + \text{Remainder}\end{aligned} \tag{B.8}$$

where $\mathbf{Q} = \mathbf{X}(\mathbf{X}'\boldsymbol{\Sigma}_{\varphi\varphi}^{-1}\mathbf{X})^{-1}\mathbf{X}'$ and $\mathbf{L} = (\mathbf{X}'\boldsymbol{\Sigma}_{\varphi\varphi}^{-1}\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\Sigma}_{\varphi\varphi}^{-1}$.

If $\boldsymbol{\Sigma}_{\varphi\varphi} = \boldsymbol{\Sigma}_{zz}$ and if $\hat{\boldsymbol{\Sigma}}_{zz}$ is distributed as a multiple of a Wishart with d_e degrees of freedom, independent of $\mathbf{z}$, then

$$\begin{aligned}&E\{\mathbf{L}(\hat{\boldsymbol{\Sigma}}_{zz} - \boldsymbol{\Sigma}_{zz})\boldsymbol{\Sigma}_{zz}^{-1}\mathbf{Q}\boldsymbol{\Sigma}_{zz}^{-1}\mathbf{z}\mathbf{z}'\boldsymbol{\Sigma}_{zz}^{-1}\mathbf{Q}\boldsymbol{\Sigma}_{zz}^{-1}\\ &\qquad \times (\hat{\boldsymbol{\Sigma}}_{zz} - \boldsymbol{\Sigma}_{zz})\,\mathbf{L}'\}\\ &\qquad = d_e^{-1}\,\mathbf{L}[\boldsymbol{\Sigma}_{zz}\,\mathrm{tr}\{\boldsymbol{\Sigma}_{zz}^{-1}\mathbf{Q}\} + \mathbf{Q}]\,\mathbf{L}'\\ &\qquad = d_e^{-1}(\mathbf{X}'\boldsymbol{\Sigma}_{zz}^{-1}\mathbf{X})^{-1}(k+1).\end{aligned}$$

Using a similar approximation

$$\begin{aligned}&E\{\mathbf{L}(\hat{\boldsymbol{\Sigma}}_{zz} - \boldsymbol{\Sigma}_{zz})\boldsymbol{\Sigma}_{zz}^{-1}\mathbf{X}\mathbf{L}\mathbf{z}\mathbf{z}'\boldsymbol{\Sigma}_{zz}^{-1}\big|\,\hat{\boldsymbol{\Sigma}}_{zz} - \boldsymbol{\Sigma}_{zz}\,\mathbf{L}'\}\\ &\quad \doteq E\{\mathbf{L}(\hat{\boldsymbol{\Sigma}}_{zz} - \boldsymbol{\Sigma}_{zz})\boldsymbol{\Sigma}_{zz}^{-1}\mathbf{X}\mathbf{L}\boldsymbol{\Sigma}_{zz}\boldsymbol{\Sigma}_{zz}^{-1}(\hat{\boldsymbol{\Sigma}}_{zz} - \boldsymbol{\Sigma}_{zz})\,\mathbf{L}'\}\\ &\quad = E\{\mathbf{L}(\hat{\boldsymbol{\Sigma}}_{zz} - \boldsymbol{\Sigma}_{zz})\boldsymbol{\Sigma}_{zz}^{-1}\mathbf{Q}\boldsymbol{\Sigma}_{zz}^{-1}(\hat{\boldsymbol{\Sigma}}_{zz} - \boldsymbol{\Sigma}_{zz})\,\mathbf{L}'\}\\ &\quad = d_e^{-1}(\mathbf{X}'\boldsymbol{\Sigma}_{zz}^{-1}\mathbf{X})^{-1}(k+1).\end{aligned}$$

On the basis of this result, we use the approximation

$$\hat{\boldsymbol{\beta}} - \boldsymbol{\beta} \doteq (\mathbf{X}'\boldsymbol{\Sigma}_{\varphi\varphi}^{-1}\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\Sigma}_{\varphi\varphi}^{-1}\mathbf{z} - \mathbf{L}\,(\hat{\boldsymbol{\Sigma}}_{\varphi\varphi} - \boldsymbol{\Sigma}_{\varphi\varphi})\boldsymbol{\Sigma}_{\varphi\varphi}^{-1}\mathbf{z}.$$

We assume $\hat{\boldsymbol{\Sigma}}_{ee}$ is a multiple of a Wishart matrix with d_e-degrees of freedom and approximate $\hat{\boldsymbol{\Sigma}}_{\varphi\varphi} - \boldsymbol{\Sigma}_{\varphi\varphi}$ with $(1-\varphi)(\hat{\boldsymbol{\Sigma}}_{ee} - \boldsymbol{\Sigma}_{ee})$. We have

$$(1-\varphi)^2 E\{\mathbf{L}(\hat{\boldsymbol{\Sigma}}_{ee} - \boldsymbol{\Sigma}_{ee})\boldsymbol{\Sigma}_{\varphi\varphi}^{-1}\boldsymbol{\Sigma}_{zz}\boldsymbol{\Sigma}_{\varphi\varphi}^{-1}(\hat{\boldsymbol{\Sigma}}_{ee} - \boldsymbol{\Sigma}_{ee})\,\mathbf{L}'\}$$

$$= (1-\varphi)^2 d_e^{-1}\,\mathbf{L}[\boldsymbol{\Sigma}_{ee}\,\mathrm{tr}\{\boldsymbol{\Sigma}_{\varphi\varphi}^{-1}\boldsymbol{\Sigma}_{zz}\boldsymbol{\Sigma}_{\varphi\varphi}^{-1}\boldsymbol{\Sigma}_{ee}\}$$

$$+ \boldsymbol{\Sigma}_{ee}\boldsymbol{\Sigma}_{\varphi\varphi}^{-1}\boldsymbol{\Sigma}_{zz}\boldsymbol{\Sigma}_{\varphi\varphi}^{-1}\boldsymbol{\Sigma}_{ee}]\,\mathbf{L}'. \tag{B.9}$$

The dominant term is that associated with the trace and we retain only that term in our approximation. Thus, an approximation to the variance of $\hat{\boldsymbol{\beta}}$ is

$$\mathbf{V}_{\beta\beta} = \mathbf{L}\boldsymbol{\Sigma}_{zz}\mathbf{L}'$$

$$+ d_e^{-1}(1-\varphi)^2\,\mathrm{tr}\{\boldsymbol{\Sigma}_{\varphi\varphi}^{-1}\boldsymbol{\Sigma}_{zz}\boldsymbol{\Sigma}_{\varphi\varphi}^{-1}\boldsymbol{\Sigma}_{ee}\}\,\mathbf{L}\boldsymbol{\Sigma}_{ee}\mathbf{L}' \tag{B.10}$$

Combining results (B.6), (B.7), and (B.9), a crude estimator of the variance of the predictor (31) is

$$\hat{\mathbf{V}}\{\hat{\mathbf{y}}_\varphi\} = \hat{\mathbf{H}}_\varphi'\hat{\boldsymbol{\Sigma}}_{ww}\hat{\mathbf{H}}_\varphi + \hat{\mathbf{G}}_\varphi'\hat{\boldsymbol{\Sigma}}_{ee}\hat{\mathbf{G}}_\varphi$$

$$+ \hat{\mathbf{H}}_\varphi'\mathbf{X}\hat{\mathbf{V}}_{\beta\beta}\mathbf{X}'\hat{\mathbf{H}}_\varphi' + \hat{\boldsymbol{\Gamma}}_{44} + \boldsymbol{\Gamma}_{33}, \tag{B.11}$$

where

$$\hat{\mathbf{H}}_\varphi = \mathbf{I} - \hat{\mathbf{G}}_\varphi,$$

$$\hat{\mathbf{V}}_{\beta\beta} = \hat{\mathbf{L}}_\varphi\hat{\boldsymbol{\Sigma}}_{zz}\hat{\mathbf{L}}_\varphi' + d_e^{-1}(1-\varphi)^2$$

$$\times\,\mathrm{tr}\{\hat{\boldsymbol{\Sigma}}_{\varphi\varphi}^{-1}\hat{\boldsymbol{\Sigma}}_{zz}\hat{\boldsymbol{\Sigma}}_{\varphi\varphi}^{-1}\hat{\boldsymbol{\Sigma}}_{ee}\}\,\hat{\mathbf{L}}_\kappa\hat{\boldsymbol{\Sigma}}_{ee}(1+\delta_\varphi)\hat{\mathbf{L}}_\varphi',$$

$$\hat{\mathbf{L}}_\varphi = (\mathbf{X}'\hat{\mathbf{S}}_{\varphi\varphi}^{-1}\mathbf{X}')^{-1}\mathbf{X}'\hat{\mathbf{S}}_{\varphi\varphi}^{-1},$$

$$\hat{\mathbf{S}}_{\varphi\varphi}^{-1} = (\hat{\boldsymbol{\Sigma}}_{\varphi\varphi} + \delta_\varphi\hat{\boldsymbol{\Sigma}}_{\varphi\varphi})^{-1},$$

$$\hat{\boldsymbol{\Sigma}}_{zz} = \hat{\boldsymbol{\Sigma}}_{ww} + \hat{\boldsymbol{\Sigma}}_{ee},$$

$$\delta_\varphi = [d_e - \mathrm{tr}\{\hat{\boldsymbol{\Sigma}}_{zz}^{-1}\hat{\boldsymbol{\Sigma}}_{ee}\}]^{-1}\,\mathrm{tr}\{\hat{\boldsymbol{\Sigma}}_{zz}^{-1}\hat{\boldsymbol{\Sigma}}_{ee}\},$$

$$\hat{\boldsymbol{\Gamma}}_{44} = d_e^{-1}(1-\varphi)^2\,\mathrm{tr}\{\hat{\boldsymbol{\Sigma}}_{\varphi\varphi}^{-1}\hat{\boldsymbol{\Sigma}}_{zz}\hat{\boldsymbol{\Sigma}}_{\varphi\varphi}^{-1}\hat{\boldsymbol{\Sigma}}_{ee}\}\,\hat{\mathbf{G}}'\hat{\boldsymbol{\Sigma}}_{ee}\hat{\mathbf{G}},$$

$$\hat{\boldsymbol{\Gamma}}_{33} = \hat{\mathbf{H}}_\varphi'\begin{pmatrix}\hat{\boldsymbol{\Lambda}}_{11}\hat{V}\{\hat{\sigma}_1^2\} & \hat{\boldsymbol{\Lambda}}_{12}\hat{C}\{\hat{\sigma}_1^2,\,\hat{\sigma}_2^2\}\\ \hat{\boldsymbol{\Lambda}}_{21}\hat{C}\{\hat{\sigma}_1^2,\,\hat{\sigma}_2^2\} & \hat{\boldsymbol{\Lambda}}_{22}\hat{V}\{\hat{\sigma}_2^2\}\end{pmatrix}\hat{\mathbf{H}}_\varphi,$$

$$\hat{\boldsymbol{\Lambda}} = \begin{pmatrix}\hat{\boldsymbol{\Lambda}}_{11} & \hat{\boldsymbol{\Lambda}}_{12}\\ \hat{\boldsymbol{\Lambda}}_{21} & \hat{\boldsymbol{\Lambda}}_{22}\end{pmatrix} = \hat{\boldsymbol{\Sigma}}_{\varphi\varphi}^{-1}\hat{\boldsymbol{\Sigma}}_{zz}\hat{\boldsymbol{\Sigma}}_{\varphi\varphi}^{-1},$$

$\hat{V}\{\hat{\sigma}_j^2\}$, $j = 1, 2$, is the estimated variance of $\hat{\sigma}_j^2$, and $\hat{C}\{\hat{\sigma}_1^2, \sigma_2^2\}$ is the estimated covariance between $\hat{\sigma}_1^2$ and $\hat{\sigma}_1^2$. See Appendix A. The estimator of the variance of $\hat{\boldsymbol{\beta}}$ contains an adjustment for the fact that $(\mathbf{X}'\hat{\boldsymbol{\Sigma}}_{zz}^{-1}\mathbf{X})^{-1}$ is a biased estimator of $(\mathbf{X}'\boldsymbol{\Sigma}_{zz}^{-1}\mathbf{X})^{-1}$.

References

Battese, G.E., Harter, R.M., and Fuller, W.A. (1988). An error components model for prediction of county crop areas using survey and satellite data. *Journal of the American Statistical Association*, 83, 28-36.

Cressie, N. (1992). REML Estimation in Empirical Bayes smoothing of census undercount. *Survey Methodology*, 18, 75-94.

Efron, B., and Morris, C. (1972). Limiting the risk of Bayes and Empirical Bayes estimates – Part II: The Empirical Bayes case. *Journal of the American Statistical Association*, 67, 130-139.

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

Ericksen, E.P., Kadane, J.B., and Tukey, J.W. (1989). Adjusting the 1980 Census of Population and Housing (with discussion). *Journal of the American Statistical Association*, 84, 927-944.

Fay, R.E. (1987). Application of multivariate regression to small domain estimation. In *Small Area Statistics*, (Eds. R. Platek, J.N.K. Rao, C.-E. Särndal and M.P. Singh). New York: John Wiley & Sons, Inc., 91-102.

Fay, R.E. (1990). VPLX: Variance estimates for complex samples. *Proceedings of the Section on Survey Research Method, American Statistical Association*, 266-271.

Fay, R.E. (1992). Inferences for Small Domain Estimates From the 1990 Post Enumeration Survey. Unpublished manuscript, U.S. Bureau of the Census.

Fay, R.E., and Herriott, 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.

Fuller, W.A., and Harter, R.M. (1987). The multivariate components of variance model for small area estimation. In *Small Area Statistics*, (Eds., R. Platek, J.N.K. Rao, C.-E. Särndal and M.P. Singh), New York: John Wiley & Sons, Inc., 103-123.

Ghosh, M. (1992). Hierarchical and Empirical Bayes multivariate estimation. In *Current Issues in Statistical Inference*: *Essays in Honor of D. Basu*, (Eds. M. Ghosh and P.K. Pathak), IMS Lecture Notes monograph Series, 17, 151-177.

Ghosh, M., and RAO, J.N.K. (1994). Small area estimation: An appraisal. *Statistical Science*, 9, 55-93.

Harville, D.A. (1976). Extension of the Gauss-Markov Theorem to include estimation of random effects. *Annals of Statistics*, 4, 384-395.

Henderson, C.R. (1950). Estimation of genetic parameters (abstract). *Annals of Mathematical Statistics*, 21, 309-310.

Hogan, H. (1992). The 1990 Post Enumeration Survey: An overview. *The American Statistician*, 46, 261-269.

Hogan, H. (1993). The 1990 Post Enumeration Survey: operations and results. *Journal of the American Statistical Association*, 88, 1047-1060.

Hulting, F.L., and Harville, D.A. (1991). Some Baysian and non-Bayesian procedures for the analysis of comparative experiments and small area estimation: Computational aspects, frequentist properties, and relationships. *Journal of the American Statistical Association*, 86, 557-568.

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, 338-343.

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

Morris, C. (1983). Parametric Empirical Bayes inference: Theory and applications (with discussions). *Journal of the American Statistical Association*, 78, 47-65.

Peixoto, J.L., and Harville, D.A. (1986). Comparisons of alternative predictors under the balanced one-way random model. *Journal of the American Statistical Association*, 81, 431-436.

Prasad, N.G.N., and Rao, J.N.K. (1990). The estimation of mean squared errors of small-area estimators. *Journal of the American Statistical Association*, 85, 163-171.

Robinson, G.K. (1991). That BLUP is a good thing: The estimation of random effects. *Statistical Science*, 6, 15-51.

Searle, S.R. (1971). *Linear Models*. New York: John Wiley & Sons, Inc.

Singh, M.P., Gambino, J., and Mantel, H.J. (1994). Issues and strategies for small area data. *Survey Methodology*, 20, 3-14.

U.S. bureau of the census (1988). The Coverage of Population in the 1980 Census, Evaluation and Research Program, PHC(E)-4.