

# Small-Area Estimation based on Survey Data from a Left-Censored Fay-Herriot Model

Eric V. Slud and Tapabrata Maiti  
University of Maryland College Park & Iowa State University

July 28, 2005

**Abstract.** We study Small Area Estimation based on data obtained by left-censored responses from a Fay-Herriot (1979) normal-error model. The problem is motivated by the Census Bureau's ongoing Small Area Income and Poverty Estimation (SAIPE) project, where a FH model is fitted to a logarithmically transformed response variable (count of sampled poor children within a CPS-sampled county), with PSU's providing responses of 0 being discarded. Alternative small area estimates and associated mean-squared error formulas are provided and supported by a simulation study, and applied to a SAIPE data analysis.

**Key words:** bias correction, Fay-Herriot model, left-censoring, left-truncation, mean-squared error, misspecified model, survey estimation.

*This paper describes research and analysis of its authors, and is released to inform interested parties and encourage discussion. Results and conclusions are the authors' and have not been endorsed by the Census Bureau.*

## 1 Introduction

Small-area estimation is becoming increasingly important in survey applications, particularly in those fields of official statistics where legislative mandates require socioeconomic estimates within narrower jurisdictions than can accurately be described by direct estimates from national surveys. An especially prominent and successful small-area project of this sort is the US Census Bureau's Small Area Income and Poverty Estimation (SAIPE) program, an ongoing effort — mandated by Title I of the US federal Code and currently funded under the No Child Left Behind Act — to estimate (among other things) the numbers of poor school-age children by state, county, and ultimately school district, based upon data from the Current

Population Survey (CPS), Internal Revenue Service (IRS), Food Stamps, and the latest decennial census. The estimation methodology used at the county level — the level to which we restrict attention here — is to fit a linear regression model to logarithms of counts of poor school-age children related to the householders in CPS-sampled households, a regression model of a form introduced by Fay and Herriot (1979) in which sampling error and a random effect at the small-area level are separated and one (usually the sampling error) is assumed known, based either on direct survey estimates or a generalized variance estimation model.

In the SAIPE and many other applications of Fay-Herriot models, area-level Small Area models are specified using a transformation (most often, the logarithm) of the original sampled data or of weighted estimators derived from it. PSU's where sampled responses fall below a (possibly PSU-dependent) threshold may be dropped, e.g. in SAIPE, counties with 0 sampled poor children are dropped from the estimating equations defining national parameter values. In this paper, we develop formulas showing the top-order effect of such left-truncation in biasing model parameters and SAE's. We show how the bias could be corrected approximately, starting from estimates based on a Fay-Herriot model ignoring censoring, and how more accurate parameter estimates and SAE's could be derived by treating the data as left-censored.

The problem treated here is generic in small-area estimation because of the wide importance of Fay-Herriot models. Left-censoring arises either because of transformation and excluded zeroes, as in SAIPE, or for structural reasons, e.g. due to an establishment survey's imposition of lesser reporting requirements on small units. For example, the Energy Information Administration's monthly crude oil report is based on a survey (EIA-813, <http://www.eia.doe.gov/oss/forms.html>) in which respondent companies that carry or store more than 1000 barrels of crude oil are required to file data monthly. Another example is in the US National Resource Inventory Survey (Nusser and Goebel 1997) data on wind erosion, collected annually to produce average wind erosion at the national level. State and county authorities are interested in county-level estimates of wind erosion, yet some counties are typically discarded from the data analysis because their observed wind erosion remains stable and small over time. The methods of this paper would apply to such surveys, possibly after nonlinear transformation of the sampled measurements, when small-area estimates are needed if there are good linear-model predictors for the response.

## 2 FH Model for SAIPE Small Area Estimation

One of the most prevalent mixed-effect linear models used in small-area estimations (Ghosh & Rao 1994), including SAIPE, is the Fay-Herriot (1979) model (FH) form, described as follows. For each PSU indexed by  $i = 1, \dots, m$ , assume that a sample-size  $n_i$  and  $p$ -dimensional vector  $x_i$  of predictor variables are known, and that response-variables satisfying

$$y_i = x_i^{tr} \beta_0 + u_i + e_i \quad , \quad u_i \sim \mathcal{N}(0, \sigma_0^2) \quad , \quad e_i \sim \mathcal{N}(0, s_i) \quad (1)$$

are observed (whenever  $n_i > 0$ ), where  $\beta_0 \in \mathbf{R}^p$  is a vector of unknown fixed-effect coefficients, and  $u_i, e_i$  are respectively PSU random effects and sampling errors, independent of each other within and across PSU's. The variances  $s_i$  are assumed to be known functions of  $n_i$  except possibly for a constant  $v_e$  of proportionality: the usual form is  $s_i \equiv v_e/n_i$ . Ordinarily, the parameter  $\sigma^2$  is unknown and estimated, with  $\sigma_0^2$  denoting its true value for the observed data, while  $v_e$  is known. In SAIPE it also makes sense to treat  $\sigma_0^2$  as known (estimated from an auxiliary model fitted to the most recent decennial census data, *cf.* Citro and Kalton 2000, App. A) and  $v_e$  as unknown, and a parallel treatment of the issues treated here can be given for this case.

*Small area estimates* (SAE's) based on such FH models are statistics designed to estimate with small mean squared error (MSE) the parameters

$$\vartheta_i = x_i^{tr} \beta_0 + u_i \quad , \quad i = 1, \dots, m$$

The values  $y_i$  are generally direct survey estimators of the target small-area parameters  $\vartheta_i$  in the sampled PSU's but may be unacceptably variable because of small sample-size  $n_i$ . In the SAIPE log-count FH models,  $y_i$  is the observed log number of poor children in the  $i$ 'th PSU (county), with the small-area parameter for the count itself defined by exponentiating:

$$\vartheta_i^* = \exp(\vartheta_i) \equiv \exp(x_i^{tr} \beta_0 + u_i) \quad (2)$$

### 2.1 SAE Formulas

In the FH model, the estimators we consider for  $\vartheta_i$  based on the data  $\{y_i, n_i : n_i > 0, 1 \leq i \leq m\}$  above are the EBLUP estimators (*cf.* Prasad and Rao 1990, Ghosh and Rao 1994, Rao 2003)

$$\hat{\vartheta}_i = x_i^{tr} \hat{\beta} + \hat{\gamma}_i (y_i - x_i^{tr} \hat{\beta}) \quad (3)$$

where  $(\hat{\beta}, \hat{\sigma}^2)$  or  $(\hat{\beta}, \hat{v}_e)$  are the maximum likelihood (ML) estimators in the model (1), and  $\hat{\gamma}_i = \hat{\sigma}^2/(\hat{\sigma}^2 + s_i)$ . (Note that while the papers of Prasad and Rao (1990) and Lahiri and Rao (1995) treating large-sample properties of EBLUP estimators restricted attention to moment-based estimators, an analogous theory for ML estimators has been provided by Datta and Lahiri (2001).) We follow the convention that  $\hat{\gamma}_i \equiv 0$  (so that  $\hat{v}_i = x_i^{tr} \hat{\beta}$ ) when  $n_i = 0$ . In addition, we define for future reference the notations

$$\tau_i = \sigma_0^2 + s_i, \quad \gamma_i = \frac{\sigma_0^2}{\tau_i}, \quad \eta_i = x_i^{tr} \beta_0, \quad \hat{\eta}_i = x_i^{tr} \hat{\beta}$$

and let  $\phi(z), \Phi(z)$  respectively denote the standard normal density and distribution function.

In later sections, we consider estimation of parameters under model (1) based on different mechanisms causing the survey data to be incomplete. We study several likelihoods and estimators for  $\nu_0 \equiv (\beta_0, \sigma_0^2)$ . In each case, those estimators could be substituted into (3) to create small-area predictors, but we also explore and compare other likelihood-based SAE's.

### 3 Left-Censored FH-Model Data

Consider now the data  $y_i$  satisfying model (1) reported in such a way that the exact response-value  $y_i$  is observed only if  $y_i \geq \kappa_i$ , where  $\kappa_i \in \mathbf{R}$  is a known real threshold. For count data, the threshold  $\kappa_i = 1$  applies, while if  $y_i$  denotes the logarithm of a rate,  $y_i = \log(\text{count}_i/n_i)$ , then a finite transformed value is observed only if  $e^{y_i} \geq 1/n_i$ , i.e.,  $\kappa_i \equiv -\log(n_i)$ .

Under the assumption that the predictor variables (and sample sizes  $n_i$ ) are available for all PSU's, it is clear that the threshold-based sampling framework described above results in the classic left-censored data structure (Klein and Moeschberger 2003):

$$(x_i, n_i, \max(y_i, \kappa_i), I_{[y_i \geq \kappa_i]} : i = 1, \dots, m) \quad (4)$$

with corresponding log-likelihood

$$l_{cens}(\beta, \sigma^2) = - \sum_{i=1}^m \left\{ \frac{1}{2} I_{[y_i \geq \kappa_i]} \left( \log(2\pi(\sigma^2 + s_i)) + \frac{(y_i - x_i^{tr} \beta)^2}{\sigma^2 + s_i} \right) - I_{[y_i < \kappa_i]} \log\left(\Phi\left(\frac{\kappa_i - x_i^{tr} \beta}{\sqrt{\sigma^2 + s_i}}\right)\right) \right\} \quad (5)$$

In this setting, the value  $y_i$  is replaced by an indicator whenever  $y_i < \kappa_i$ . Although that information has not always been used in surveys, the data do include the indices of which sampled PSU's have below-threshold responses.

In some surveys, there may instead be no information whatever on PSU's within the sampling frame which would have been sampled but would have produced below-threshold responses. In such cases,  $m$  itself (as an overall characteristic of the sampling frame) would be unknown, and the data would consist only of

$$(x_i, n_i, y_i : 1 \leq i \leq m, y_i \geq \kappa_i) \quad (6)$$

which is a *left-truncated* sample (Klein and Moeschberger 2003), with log-likelihood

$$l_{trunc}(\beta, \sigma^2) = - \sum_{i: y_i \geq \kappa_i} \left\{ \frac{1}{2} \left( \log(2\pi(\sigma^2 + s_i)) + \frac{(y_i - x_i^{tr}\beta)^2}{\sigma^2 + s_i} \right) - \log(1 - \Phi(\frac{\kappa_i - x_i^{tr}\beta}{\sqrt{\sigma^2 + s_i}})) \right\} \quad (7)$$

For survey data from a sampling frame assumed to satisfy (1), there are at least four ways the statistician could analyze the data to provide parameter estimators  $\hat{\nu} = (\hat{\beta}, \hat{\sigma}^2)$  to be substituted into small-area estimators.

(A) One can drop the PSU's with below-threshold responses, which is to say estimate parameters  $\nu = (\beta, \sigma^2)$  from the dataset (6), using standard FH methodology based on complete data. The underlying log-likelihood here is

$$l_{compl}(\beta, \sigma) = - \frac{1}{2} \sum_{i: y_i \geq \kappa_i} \left\{ \log(2\pi(\sigma^2 + s_i)) + \frac{(y_i - x_i^{tr}\beta)^2}{\sigma^2 + s_i} \right\} \quad (8)$$

(B) One can recognize that the FH complete-data log-likelihood (8) is misspecified under the sampling framework (6), and correct the parameter estimators for bias and robustly estimate their large-sample variances.

(C) One can drop the PSU's with below-threshold responses, but analyze the left-truncated dataset (6) using the log-likelihood (7).

(D) Finally, one can analyze the left-censored dataset (4) using the left-censored regression model log-likelihood.

The Census Bureau’s SAIPE methodology has essentially been option (A) above. The SAIPE program has historically (up through its 1997 county-level estimates of numbers of poor children) treated its log-transformed county-level counts of poor related sampled children in the CPS (ASEC) through a slightly modified FH model (1) treated as a complete-data model. Strictly speaking, the FH model was used only in the variant form where  $\sigma^2$  is taken as known – fitted within a generalized-variance framework from decennial census data — and  $s_i = v_e/n_i^\alpha$  with  $v_e$  treated as unknown. The value  $\alpha$  used in practice has been  $\alpha = 1/2$  up through the 1995 estimates and  $\alpha = 1/4$  starting with the SAIPE 1997 production estimates.

Our objective in this paper is first to study option (B) as a way of deriving ‘correct’ large-sample estimators of parameters from left-censored data, with a view to assessing, approximating and correcting for the biases which arise in (A). Next, since the truncated dataset (6) contains strictly less information (both in the vernacular and technical senses) than the censored dataset (4), there is no good motivation for option (C) above, even though its model is correctly specified: when the left-censored form (4) of the data is available, a fully informative likelihood-based treatment (D) will provide better large-sample estimators than those based on the reduced data (6).

The remainder of the paper is organized as follows. First, in Section 4, we consider option (A) of estimating parameters using the complete-data FH parameter estimates on (6), ignoring the left-truncation. We exhibit a top-order approximation of the estimation biases, numerically assess the quality of that approximation under several scenarios, and study the consequences of the biases for small-area estimation. Next, in Section 5, we derive maximum-likelihood estimators based on a parametric left-censored-data likelihood, and compare the large-sample behavior and asymptotic variances of those estimators with the ones found in (B). All methods are compared first in a realistic finite-sample simulation study, in Section 6, and then in a real-data SAIPE example, in Section 7. Overall conclusions are drawn in Section 8.

## 4 Misspecified Analysis of Censored Data as Complete

For the data-set observed in the form (6), a naive estimator  $\tilde{\nu} = (\tilde{\beta}, \tilde{\sigma}^2)$  of  $\nu_0 = (\beta_0, \sigma_0^2)$  obtained by maximizing  $l_{cens}(\beta, \sigma^2)$ , equivalent to the one actually used in SAIPE for the log-count model, is given by using precisely the same formulas as before restricted to the observed data, namely:

$$\tilde{\beta}(t) = \left( \sum_{i=1}^m \frac{x_i^{\otimes 2}}{t + s_i} I_{[y_i \geq \kappa_i]} \right)^{-1} \sum_{i=1}^m \frac{x_i y_i}{t + s_i} I_{[y_i \geq \kappa_i]}$$

and

$$\tilde{\beta} \equiv \tilde{\beta}(\tilde{\sigma}^2), \quad \tilde{\sigma}^2 = \arg \min_t \sum_{i=1}^m I_{[y_i \geq \kappa_i]} \left( \log(t + s_i) + \frac{(y_i - x_i^{\text{tr}} \tilde{\beta}(t))^2}{t + s_i} \right)$$

Since these estimators are not derived from a likelihood or moment criterion based on the left-truncated data model, one might expect them to be somewhat biased. It turns out that they are, and we calculate the large-sample limits  $\nu_* = (\beta_*, \sigma_*^2)$  of  $\tilde{\nu} = (\tilde{\beta}, \tilde{\sigma}^2)$ , as follows. First,

$$\tilde{\beta} - \beta_0 = \left( \sum_{i=1}^m I_{[y_i \geq \kappa_i]} \frac{x_i^{\otimes 2}}{\tilde{\sigma}^2 + s_i} \right)^{-1} \left[ \sum_{i=1}^m I_{[y_i \geq \kappa_i]} \frac{x_i(u_i + e_i)}{\tilde{\sigma}^2 + s_i} \right]$$

Next, denoting the variable of integration for the standard normal deviate  $(u_i + e_i)/\sqrt{\tau_i}$  as  $z$ , and defining

$$\xi_i \equiv (\kappa_i - \eta_i)/\sqrt{\tau_i} = (\kappa_i - x_i^{\text{tr}} \beta_0)/\sqrt{\sigma_0^2 + s_i}$$

note that  $y_i = \eta_i + (u_i + e_i)$ , and

$$E\left(I_{[y_i \geq \kappa_i]} (u_i + e_i)\right) = \sqrt{\tau_i} \int_{\xi_i}^{\infty} \frac{z e^{-z^2/2}}{\sqrt{2\pi}} dz = \frac{\sqrt{\tau_i}}{\sqrt{2\pi}} e^{-\xi_i^2/2}$$

Therefore, applying a nonidentical-summand Law of Large Numbers (uniform over the parameter  $t = \tilde{\sigma}^2$ ) within the previous expression for  $\tilde{\beta} - \beta_0$ , the limit  $\beta_* - \beta_0$  differs by an amount asymptotically negligible in probability from

$$\left( \sum_{i=1}^m \frac{x_i^{\otimes 2}}{\tilde{\sigma}^2 + s_i} (1 - \Phi(\xi_i)) \right)^{-1} \frac{1}{\sqrt{2\pi}} \sum_{i=1}^m \frac{x_i \sqrt{\tau_i}}{\tilde{\sigma}^2 + s_i} e^{-\xi_i^2/2}$$

In this same spirit, we deduce that for large  $m$  the estimator  $\tilde{\beta}(t)$  differs asymptotically negligibly in probability from  $\beta_*(t)$  defined by

$$\beta_*(t) = \beta_0 + \left( \sum_{i=1}^m \frac{x_i^{\otimes 2}}{t + s_i} (1 - \Phi(\xi_i)) \right)^{-1} \frac{1}{\sqrt{2\pi}} \sum_{i=1}^m \frac{x_i \sqrt{\tau_i}}{t + s_i} e^{-\xi_i^2/2} \quad (9)$$

Similarly,  $\tilde{\sigma}^2$  differs asymptotically negligibly, as  $m \rightarrow \infty$ , from the argument  $t = \sigma_*^2$  minimizing

$$\begin{aligned} & \sum_{i=1}^m \left( (1 - \Phi(\xi_i)) \log(t + s_i) + \int_{\xi_i}^{\infty} \frac{(z\sqrt{\tau_i} - x_i^{tr}(\beta_*(t) - \beta_0))^2}{t + s_i} e^{-z^2/2} \frac{dz}{\sqrt{2\pi}} \right) \\ &= \sum_{i=1}^m \left( (1 - \Phi(\xi_i)) \left\{ \log(t + s_i) + \frac{1}{t + s_i} (\tau_i + (x_i^{tr}(\beta_*(t) - \beta_0))^2) \right\} \right. \\ & \quad \left. + \frac{1}{\sqrt{2\pi}} e^{-\xi_i^2/2} \frac{1}{t + s_i} \{-2\sqrt{\tau_i}(x_i^{tr}(\beta_*(t) - \beta_0)) + \tau_i \xi_i\} \right) \end{aligned}$$

and after substituting the formula (9) left-multiplied by  $(\beta_*(t) - \beta_0)^{tr} \cdot \sum_i (t + s_i)^{-1} x_i^{\otimes 2} (1 - \Phi(\xi_i))$ , we obtain

$$\begin{aligned} \sigma_*^2 &= \arg \min_t \sum_{i=1}^m \left( (1 - \Phi(\xi_i)) \left( \log(t + s_i) + \frac{\tau_i}{t + s_i} \right) \right. \\ & \quad \left. + e^{-\xi_i^2/2} \sqrt{\frac{\tau_i}{2\pi}} \cdot \frac{\kappa_i - x_i^{tr} \beta_*(t)}{t + s_i} \right) \end{aligned} \quad (10)$$

We collect our conclusions in the following Theorem. Further justifications of the steps can be found in the Appendix.

**Theorem 4.1** *Under the assumptions (a)–(c), as  $m \rightarrow \infty$ , the estimators  $\hat{\beta}$ ,  $\tilde{\sigma}^2$  maximizing (8) differ asymptotically negligibly respectively from  $\beta_*(\sigma_*^2)$  and  $\sigma_*^2$  defined above in formulas (9) and (10). Moreover, in the limiting case where as  $m \rightarrow \infty$ ,  $m^{-1} \sum_{i=1}^m I_{[y_i \geq \kappa_i]} \approx 1$ , to top order*

$$\begin{aligned} \sigma_*^2 - \sigma_0^2 &\approx \frac{1}{\sqrt{2\pi}} \sum_{i=1}^m \frac{\xi_i}{\tau_i} e^{-\xi_i^2/2} / \sum_{i=1}^m \frac{1}{\tau_i} (1 - \Phi(\xi_i)) \equiv \delta_{\sigma^2} \\ \beta_* - \beta_0 &\approx \left( \sum_{i=1}^m \frac{x_i^{\otimes 2}}{\tau_i} (1 - \Phi(\xi_i)) \right)^{-1} \frac{1}{\sqrt{2\pi}} \sum_{i=1}^m \frac{x_i}{\sqrt{\tau_i}} e^{-\xi_i^2/2} \equiv \delta_{\beta} \end{aligned}$$

It is not hard to check that the calculation just completed is essentially the same as finding the arguments  $\nu_* = (\beta_*, \sigma_*^2)$  minimizing the Kullback-Leibler distance between the model (A) with log-likelihood (8) for the data on  $\{i : y_i \geq \kappa_i\}$  versus the correct log-likelihood (5). Thus our derivation of  $\nu_*$  follows the well-established lines of the asymptotic misspecified-model theory in White (1982), regarding the estimator  $\hat{\nu}$  as the maximum-likelihood estimator under the misspecified complete-data likelihood for (6).

We continue by finding the form of the asymptotic variances of these estimators, as follows. Under suitable regularity conditions, described in the Appendix, we have as  $m \rightarrow \infty$ ,

$$\sqrt{m} \begin{pmatrix} \tilde{\beta} & - & \beta_* \\ \tilde{\sigma}^2 & - & \sigma_*^2 \end{pmatrix} \xrightarrow{\mathcal{D}} \mathcal{N}(0, A^{-1}\Sigma A^{-1}) \quad (11)$$

where the  $(p+1) \times (p+1)$  matrices  $A, \Sigma$  are defined by nonrandom limits

$$A = \lim_{m \rightarrow \infty} \frac{1}{2m} \sum_{i=1}^m \nabla_{b,t}^{\otimes 2} E \left( I_{[y_i \geq \kappa_i]} \left[ \log(t + s_i) + \frac{(y_i - x_i^{tr}b)^2}{t + s_i} \right] \right) \Big|_{(b,t)=a_*}$$

and

$$\Sigma = \lim_{m \rightarrow \infty} \frac{1}{4m} \sum_{i=1}^m \text{Var} \left( I_{[y_i \geq \kappa_i]} \nabla_{b,t} \left[ \log(t + s_i) + \frac{(y_i - x_i^{tr}b)^2}{t + s_i} \right] \right)^{\otimes 2} \Big|_{(b,t)=a_*}$$

and  $A, \Sigma$  are given in block-decomposed form

$$A = \begin{pmatrix} A_{11} & A_{12} \\ A_{12}^{tr} & A_{22} \end{pmatrix}, \quad \Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{12}^{tr} & \Sigma_{22} \end{pmatrix}$$

with  $p \times p$  upper-left blocks  $A_{11}, \Sigma_{11}$ , in formulas (26)-(31) in the Appendix.

Our objective in this Section has been to contrast the actual large-sample behavior (11) of the complete-data estimators with the nominal behavior that would be expected under complete-data FH formulas:

$$\sqrt{m} \begin{pmatrix} \tilde{\beta} & - & \beta_0 \\ \tilde{\sigma}^2 & - & \sigma_0^2 \end{pmatrix} \longrightarrow \mathcal{N} \left( 0, \begin{pmatrix} \Sigma_\beta & 0 \\ 0 & \Sigma_{\sigma^2} \end{pmatrix} \right) \quad (12)$$

where

$$\Sigma_\beta^{-1} = \frac{1}{m} \sum_{i=1}^m \frac{x_i^{\otimes 2}}{\sigma_0^2 + s_i} I_{[y_i \geq \kappa_i]}, \quad \Sigma_{\sigma^2}^{-1} = \frac{1}{m} \sum_{i=1}^m \frac{1}{2(\sigma_0^2 + s_i)^2} I_{[y_i \geq \kappa_i]}$$

The major result of this section is that, in the setting where left-censoring exists but is relatively light, the extent of bias in parameter estimation can be well approximated and estimated, and mean-squared estimation error can be estimated robustly. To accomplish this, we want essentially to find estimators of  $\nu_0 = (\beta_0, \sigma_0^2)$  to substitute into the terms  $\tau_i, \xi_i$  within

the formulas for  $\delta_\beta, \delta_{\sigma^2}$  in Theorem 4.1. Now, we will see below that the differences between  $\nu_* = (\beta_*, \sigma_*^2)$  and  $\nu_0$  are often large enough that direct substitution of  $\tilde{\nu}$  for  $\nu_0$  is a bad idea. On the other hand, this substitution does give a preliminary estimator for  $(\delta_\beta, \delta_{\sigma^2})$  which corrects  $\tilde{\nu}$  in the right direction. Then  $\tilde{\nu}$  corrected by the preliminary-estimated biases  $(\delta_\beta, \delta_{\sigma^2})$  provided an improved estimator of  $\nu_0$  which can in turn be substituted into an improved estimator of  $(\delta_\beta, \delta_{\sigma^2})$ , and this correction, bias-estimation, and re-substitution can be iterated to obtain a sequence of estimators  $\nu^{(k)} = (\tilde{\beta}^{(k)}, (\tilde{\sigma}^{(k)})^2)$ ,  $k \geq 0$ , according to the scheme

$$(\tilde{\beta}^{(0)}, (\tilde{\sigma}^{(0)})^2) \equiv (\tilde{\beta}, \tilde{\sigma}^2) \ , \quad \tau_i^{(k)} = (\tilde{\sigma}^{(k)})^2 + s_i \ , \quad \xi_i^{(k)} = \frac{\kappa_i - x_i^{tr} \tilde{\beta}^{(k)}}{\sqrt{\tilde{\tau}_i^{(k)}}}$$

$$\tilde{\delta}_\beta^{(k)} \equiv \left( \sum_{i=1}^m \frac{x_i^{\otimes 2}}{\tilde{\tau}_i^{(k)}} (1 - \Phi(\tilde{\xi}_i^{(k)})) \right)^{-1} \sum_{i=1}^m \frac{x_i}{\sqrt{\tilde{\tau}_i^{(k)}}} \phi(\tilde{\xi}_i^{(k)})$$

$$\tilde{\delta}_{\sigma^2}^{(k)} = \sum_{i=1}^m \frac{\xi_i}{\tilde{\tau}_i^{(k)}} \phi(\tilde{\xi}_i^{(k)}) / \sum_{i=1}^m \frac{1}{(\tilde{\tau}_i^{(k)})^2} (1 - \Phi(\tilde{\xi}_i^{(k)}))$$

$$\tilde{\beta}^{(k+1)} \equiv \tilde{\beta} - \tilde{\delta}_\beta^{(k)} \ , \quad (\tilde{\sigma}^{(k+1)})^2 \equiv \tilde{\sigma}^2 - \tilde{\delta}_{\sigma^2}^{(k)}$$

For a fixed dataset, the limit of this sequence of estimators as  $k \rightarrow \infty$  can equivalently be given by the solution (assumed unique, which appears to be the case in practice) of the estimating equations

$$\hat{\beta} = \tilde{\beta} - \left( \sum_{i=1}^m \frac{x_i^{\otimes 2}}{\hat{\tau}_i} (1 - \Phi(\hat{\xi}_i)) \right)^{-1} \sum_{i=1}^m \frac{x_i}{\sqrt{\hat{\tau}_i}} \phi(\hat{\xi}_i) \quad (13)$$

$$\hat{\sigma}^2 = \tilde{\sigma}^2 - \sum_{i=1}^m \frac{\xi_i}{\hat{\tau}_i} \phi(\hat{\xi}_i) / \sum_{i=1}^m \frac{1}{\hat{\tau}_i^2} (1 - \Phi(\hat{\xi}_i)) \quad (14)$$

where

$$\hat{\tau}_i = \hat{\sigma}^2 + s_i \ , \quad \hat{\xi}_i = (\kappa_i - x_i^{tr} \hat{\beta}) / \sqrt{\hat{\tau}_i} \quad (15)$$

Note that the estimators given by equations (13)–(15) are the method-B estimators  $\hat{\nu} \equiv \hat{\nu}^B$ , but we omit the  $B$  subscript for simplicity. Although the solution  $\hat{\nu}$  to these equations cannot be asserted to be consistent but only approximately so, we will see that this estimator performs very well. We summarize our result in:

**Theorem 4.2** *Under the hypotheses of Theorem 4.1, the estimators  $(\hat{\beta}, \hat{\sigma}^2)$  defined as solutions of the estimating equations (13)–(15) are approximately consistent estimators of  $(\beta_0, \sigma_0^2)$ , in the sense that the errors are of smaller order than  $\delta_\beta + \delta_{\sigma^2}$  with  $(\delta_\beta, \delta_{\sigma^2})$  given in Theorem 4.1, when censoring is rare in the sense that  $m^{-1} \sum_{i=1}^m I_{[y_i \geq \kappa_i]} \rightarrow 1$  in probability as  $m \rightarrow \infty$ .*

We shall examine finite-sample aspects of this approximation in the simulation study of Section 6 below. At this point, we briefly consider the exact calculation of large-sample differences, within specific scenarios, between  $a_0, a_*$ , and the bias approximations given by Theorem 4.1.

Suppose for illustration that  $x_i = (1, w_i)^{tr} \in \mathbf{R}^2$  with  $w_i \sim \mathcal{N}(0, \sigma_w^2)$  and that the sample sizes  $n_i$  are integer-valued random variables distributed independently of  $w_i$ , with  $n_i \sim \text{Unif}\{10, 11, \dots, 50\}$  (discrete-uniform), and with  $\kappa_i = \kappa$  the same for all  $i$ . Then in a large- $m$  limit, formulas (9)–(10) and (26)–(31) become expectations, and  $\beta_*, \sigma_*^2, \Sigma_\beta, \Sigma_{\sigma^2}$  and the diagonal elements of  $A^{-1}\Sigma A^{-1}$ , can be calculated numerically, as can the differences  $\delta_{\sigma^2}$  and  $\delta_\beta$  between the left and right hand sides of the approximate equalities in Theorem 4.1. We did this by coding a function in R, and we exhibit the results in Table 1 for several combinations of the parameter values  $\beta_0, \sigma_0^2, \sigma_w^2, v_e$ . In each case, we quantify the degree of censoring through the parameter  $c$  (proportion censored), where

$$c = \lim_{m \rightarrow \infty} \frac{1}{m} \sum_{i=1}^m P(y_i < \kappa_i) = E\left(\Phi\left(\frac{\kappa_1 - x_1^{tr}\beta_0}{\sigma_0^2 + v_e/n_1}\right)\right)$$

There are two immediate conclusions from the tabulated results. First, the corrections  $\delta_\beta, \delta_{\sigma^2}$  approximate the corresponding biases  $\beta_* - \beta_0$  and  $\sigma_*^2 - \sigma_0^2$  remarkably well, with some discrepancies visible at 20% censoring and progressively smaller errors as the degree of censoring decreases. Secondly, the ‘robust’ or sandwich-formula variances given by the diagonal elements of  $A^{-1}\Sigma A^{-1}$  are generally very similar to the simpler ‘nominal’ variances given by the diagonal elements of  $\Sigma_\beta$  together with  $\Sigma_{\sigma^2}$ .

#### 4.1 Small Area Estimates and MSE

We now consider alternative SAE’s and their MSE’s according to the analysis options (A)–(D) listed in Section 3. Recall that in all cases, we view the data as being governed by the left-censored FH model with data (4) and log-likelihood (5).

Table 1: Limits  $(\beta_*, \sigma_*^2)$ , asymptotic variance parameters, and proportion  $c$  of PSU's censored, for misspecified (complete-data) estimators of  $(\beta_0, \sigma_0^2)$  when FH model data are actually left-censored at a constant threshold  $\kappa$ . In upper portion of Table (above double-line),  $\beta_0 = (1, -.5)$ ,  $\sigma_0^2 = 2$ , and sample-sizes  $n_i$  are discrete-uniform over  $\{10, \dots, 50\}$ ; in lower portion,  $\beta_0 = (1, 1)$ ,  $\sigma_0^2 = 1$ , and  $n_i$  are uniform in  $\{30, 70, 100\}$ . Asymptotic biases  $\beta_* - \beta_0$  and  $\sigma_*^2 - \sigma_0^2$  should respectively be compared with approximations  $\delta_\beta, \delta_{\sigma^2}$  from Theorem 4.1. Diagonal elements of  $\Sigma_\beta$  are given in the column headed  $\Sigma_\beta$ , and the diagonal elements of  $A^{-1}\Sigma A^{-1}$  in the columns  $\Sigma_\beta^R$  and  $\Sigma_{\sigma^2}^R$  ( $R$  for *Robust*). Columns involving  $\beta$  have two entries corresponding to first and second components.

$\kappa$	$\sigma_0^2$	$v_e$	$c$	$\beta_*$	$\sigma_*^2$	$\delta_\beta$	$\delta_{\sigma^2}$	$\Sigma_\beta$	$\Sigma_{\sigma^2}$	$\Sigma_\beta^R$	$\Sigma_{\sigma^2}^R$
0	.5	2	.168	1.2037	.3578	.2040	-.0896	.55	.45	.49	.44
				-.3760		.1241		.32		.31	
-.5			.075	1.0953	.4160	.0954	-.0677	.54	.52	.52	.50
				-.4269		.0732		.30		.28	
-.75			.046	1.0615	.4393	.0616	-.0526	.55	.56	.53	.53
				-.4472		.0528		.30		.27	
-1			.027	1.0381	.4581	.0381	-.0382	.55	.59	.55	.56
				-.4636		.0364		.29		.27	
-1.2			.017	1.0252	.4700	.0252	-.0282	.56	.61	.56	.59
				-.4739		.0261		.29		.27	
-1.4			.011	1.0161	.4792	.0161	-.0200	.56	.63	.56	.61
				-.4819		.0181		.29		.27	
0	.2	5	.146	1.1326	.1457	.1338	-.0440	.38	.22	.35	.20
				-.4066		.0939		.23		.22	
-.5			.057	1.0543	.1649	.0547	-.0296	.37	.24	.36	.23
				-.4511		.0491		.21		.19	
-.75			.033	1.0326	.1763	.0327	-.0214	.37	.25	.37	.24
				-.4672		.0330		.20		.19	
-1			.018	1.0186	.1848	.0187	-.0143	.37	.26	.37	.25
				-.4791		.0210		.20		.18	
0	1	21	.259	1.4798	.5552	.3870	-.2188	1.36	2.17	1.14	2.17
				.6611		-.1848		1.54		1.55	
-.5			.166	1.3125	.6597	.2646	-.2239	1.28	2.43	1.16	2.33
				.7421		-.1897		1.46		1.41	
-1			.098	1.1914	.7556	.1696	-.1909	1.27	2.72	1.20	2.56
				.8158		-.1584		1.41		1.33	

The simplest option is the EBLUP estimator (3) arising in the misspecified complete-data analysis (A), which — after treating censored PSU's as wholly unobserved — becomes

$$\tilde{\vartheta}_i^A = \frac{\tilde{\sigma}^2}{\tilde{\sigma}^2 + s_i} (y_i - x_i^{tr} \tilde{\beta}) I_{[y_i \geq \kappa_i]} + x_i^{tr} \tilde{\beta}$$

Recall that the actual asymptotic behavior of the usual ML estimators  $(\tilde{\beta}, \tilde{\sigma}^2)$  is given by (11). To top order, we find under option (A) using small-area predictors  $\tilde{\vartheta}_i^A$  :

$$MSE_i^A = E\left(\frac{\tilde{\sigma}^2}{\tilde{\sigma}^2 + s_i} (y_i - x_i^{tr} \tilde{\beta}) I_{[y_i \geq \kappa_i]} + x_i^{tr} (\tilde{\beta} - \beta_0) - u_i\right)^2$$

Now substitute of  $(\beta_*, \sigma_*^2)$  for  $(\tilde{\beta}, \tilde{\sigma}^2)$ , with  $o_P(1)$  error, and recall that  $u_i \sim \mathcal{N}(\frac{\sigma_0^2}{\tau_i} y_i, s_i \frac{\sigma_0^2}{\tau_i})$  conditionally given  $y_i$ . Then  $MSE_i^A = o_P(1) +$

$$\frac{\sigma_0^2 s_i}{\tau_i} + E\left((y_i - x_i^{tr} \beta_0) \left\{ \frac{\sigma_*^2}{\sigma_*^2 + s_i} I_{[y_i \geq \kappa_i]} - \frac{\sigma_0^2}{\tau_i} \right\} + \Delta_i \left\{ 1 - \frac{\sigma_*^2}{\sigma_*^2 + s_i} I_{[y_i \geq \kappa_i]} \right\}\right)^2$$

where as in Section 9.3 we denote  $\Delta_i = x_i^{tr} (\beta_* - \beta_0)$ . Next substitute formula for  $\mu_{1i}$  from Section 9.3, to find

$$\begin{aligned} MSE_i^A &= \frac{\sigma_0^2 s_i}{\tau_i} + \left( \frac{\sigma_*^2 \sqrt{\tau_i}}{\sigma_*^2 + s_i} \phi(\xi_i) + \Delta_i \left( 1 - \frac{\sigma_*^2}{\sigma_*^2 + s_i} (1 - \Phi(\xi_i)) \right) \right)^2 \\ &+ Var\left( \frac{\sigma_*^2}{\sigma_*^2 + s_i} (y_i - x_i^{tr} \beta_*) I_{[y_i \geq \kappa_i]} - \frac{\sigma_0^2}{\tau_i} (y_i - x_i^{tr} \beta_0) \right) + o_P(1) \end{aligned}$$

Next, substitute formulas for  $\mu_{1i}, \mu_{2i}$  from Section 9.3 to obtain, after some further algebraic reductions,

$$\begin{aligned} MSE_i^A &= \sigma_0^2 + (1 - \Phi(\xi_i) + \xi_i \phi(\xi_i)) \left( \tau_i \left( \frac{\sigma_*^2}{\sigma_*^2 + s_i} \right)^2 - 2\sigma_0^2 \right) \\ &+ \Delta_i^2 \left\{ 1 - 2(1 - 2\Phi(\xi_i)) \frac{\sigma_*^2}{\sigma_*^2 + s_i} \right\} + 2\Delta_i \sqrt{\tau_i} \phi(\xi_i) \left( \frac{\sigma_0^2}{\tau_i} + \frac{\sigma_*^2}{\sigma_*^2 + s_i} \right) \end{aligned}$$

This MSE formula includes a bias term due to the discrepancy between the large-sample limit  $\beta_*$  of the misspecified ML estimator  $\tilde{\beta}$ , but *not* (nominal or robust) asymptotic-variance terms for  $\tilde{\beta}$  or  $\tilde{\sigma}^2$  (since these are  $\mathcal{O}(1/m)$ ). The magnitude of the bias has been approximated in Theorem 4.1, and we have seen that under realistic parameter combinations with left-censoring of

20% or less, this approximation is very good, correcting more than 95% of the bias. A natural first approach to improving the small-area predictions is to modify  $\tilde{\beta}$  by the bias-estimator within the SAE formula, using the estimators  $(\hat{\beta}, \hat{\sigma}^2)$  defined by (13)-(14), as follows:

$$\tilde{\vartheta}_i^B = \frac{\hat{\sigma}^2}{\hat{\sigma}^2 + s_i} (y_i - x_i^{tr} \hat{\beta}) I_{[y_i \geq \kappa_i]} + x_i^{tr} \hat{\beta}$$

An expression for the mean-squared error  $MSE_i^B$  of this SAE can be derived along the same lines as the formula for  $MSE_i^A$  above. One main difference is that the term  $\tilde{\Delta}_i$  corresponding to  $\Delta_i$  in the new formula is much smaller than  $\Delta_i$ . However, as can be verified from simulations,  $MSE_i^B$  is often quite a bit larger than  $MSE_i^A$ : the main point is that in many examples  $\hat{\beta}_i$  estimates  $\beta_0$  quite accurately, but there is intrinsic positive bias in  $\tilde{\vartheta}_i^B$  due to the fact that the correction to  $x_i^{tr} \hat{\beta}$  by residuals is employed only when those residuals are above the threshold  $\kappa_i - x_i^{tr} \hat{\beta}$ . Therefore, a slightly improved small-area estimator using the adjustment (B) involves correcting  $\tilde{\vartheta}_i^B$  by its estimated bias, namely

$$\tilde{\vartheta}_i^{B*} = \frac{\hat{\sigma}^2}{\hat{\sigma}^2 + s_i} (y_i - x_i^{tr} \hat{\beta}) I_{[y_i \geq \kappa_i]} + x_i^{tr} \hat{\beta} - \frac{\hat{\sigma}^2}{\sqrt{\hat{\sigma}^2 + s_i}} \phi\left(\frac{\kappa_i - x_i^{tr} \hat{\beta}}{\sqrt{\hat{\sigma}^2 + s_i}}\right) \quad (16)$$

## 5 Analysis using Left-Censored Data Likelihood

The method of analysis proposed above as method (D) is a left-censored parametric linear regression model with the slightly unusual PSU-dependent variance term  $\sigma^2 + s_i$  of Fay and Herriot. With the  $s_i$  terms absent, such analyses have appeared in older survival analysis literature: recent survival literature has emphasized instead the *semiparametric* censored linear regression model with ( $e_i$  absent and) the distribution of  $u_i$  unknown (with mean 0), *cf.* Buckley and James (1979), Tsiatis (1990), Ritov (1990), and Ying (1993). As in Klein and Moeschberger (2003), the maximum-likelihood estimation of  $(\beta_0, \sigma_0^2)$  using the parametric log-likelihood  $l_{cens}$  is straightforward but does not lead to explicit formulas.

The likelihood equations determining the ML estimates, expressed in terms of the notation  $z_i \equiv z_i(\beta, \sigma^2) \equiv (\kappa_i - x_i^{tr} \beta) / \sqrt{\sigma^2 + s_i}$ , are:

$$\sum_{i=1}^m x_i \left\{ I_{[y_i \geq \kappa_i]} \frac{y_i - x_i^{tr} \beta}{\sigma^2 + s_i} - I_{[y_i < \kappa_i]} \frac{1}{\sqrt{\sigma^2 + s_i}} \cdot \frac{\phi(z_i)}{\Phi(z_i)} \right\} = 0 \quad (17)$$

$$\sum_{i=1}^m \frac{1}{\sigma^2 + s_i} \left\{ I_{[y_i \geq \kappa_i]} \left( 1 - \frac{(y_i - x_i^{tr} \beta)^2}{\sigma^2 + s_i} \right) + I_{[y_i < \kappa_i]} z_i \frac{\phi(z_i)}{\Phi(z_i)} \right\} = 0 \quad (18)$$

and the asymptotic variance-covariance matrix for the resulting maximum likelihood estimates of  $(\beta, \sigma^2)$  is readily computed and estimated as the inverse of the per-PSU information matrix

$$I(\beta_0, \sigma_0^2) = \begin{pmatrix} I_{11} & I_{12} \\ I_{12}^{tr} & I_{22} \end{pmatrix} \equiv -E(\nabla_{\beta, \sigma^2}^{\otimes 2} l_{cens}(\beta_0, \sigma_0^2)) \quad (19)$$

$$I_{11} = \frac{1}{m} \sum_{i=1}^m \frac{x_i^{\otimes 2}}{\tau_i} \left[ 1 - \Phi(\xi_i) + \xi_i \phi(\xi_i) + \phi^2(\xi_i)/\Phi(\xi_i) \right]$$

$$I_{12} = \frac{1}{m} \sum_{i=1}^m \frac{x_i}{2\tau_i^{3/2}} \phi(\xi_i) \left( 1 + \xi_i^2 + \xi_i \phi(\xi_i)/\Phi(\xi_i) \right)$$

$$I_{22} = \frac{1}{m} \sum_{i=1}^m \frac{1}{4\tau_i^2} \left[ 2(1 - \Phi(\xi_i)) + \xi_i \phi(\xi_i) (1 + \xi_i^2 + \xi_i \phi(\xi_i)/\Phi(\xi_i)) \right]$$

Letting  $(\hat{\beta}^D, (\hat{\sigma}^D)^2)$  denote the censored-data ML estimators obtained by solving the likelihood equations (17)-(18), we know from standard maximum-likelihood estimation theory (the regularity conditions for which are easily satisfied in the present setting) that

$$\sqrt{m} \begin{pmatrix} \hat{\beta}^D - \beta_0 \\ (\hat{\sigma}^D)^2 - \sigma_0^2 \end{pmatrix} \xrightarrow{\mathcal{D}} \mathcal{N}\left(0, (I(\beta_0, \sigma_0^2))^{-1}\right)$$

We briefly compare the asymptotic variances of the estimators  $(\hat{\beta}^D, (\hat{\sigma}^D)^2)$  versus  $(\tilde{\beta}, \tilde{\sigma}^2)$  (normalized by  $\sqrt{m}$ ) in the first illustrative case considered in calculating Table 1: this is the case where  $x_i = (1, w_i)^{tr}$ ,  $w_i \sim \mathcal{N}(0, 2)$ ,  $\beta_0 = (1, -.5)$ ,  $\sigma_0^2 = 0.5$ , with  $n_i$  discrete-uniform in  $\{10, 11, \dots, 50\}$ . The nominal (complete-data) Fay-Herriot variances for the two components of  $\beta$  and  $\sigma^2$  are respectively .550, .318, .453, while the diagonal elements of the correct asymptotic variance matrix in (11) for the Fay-Herriot estimators are .490, .307, .442. By contrast, the corresponding variances for the censored-data ML estimators are .623, .340, .843. At first sight, it is puzzling that the ML component estimators do not all have smaller variances. But recall that the FH and censored-ML estimators are not directly comparable because the former are biased, and the asymptotic variability of the bias estimators has so far not been taken into account.

### 5.1 Small Area Estimates and MSE

How would the censored-data likelihood be used to generate small-area estimators? The most natural generalization of the EBLUP idea is to estimate  $\vartheta_i$  in PSU's without sample data as  $x_i^{tr} \hat{\beta}^D$ , and to estimate in PSU's *with* sampled data by  $\hat{\vartheta}_i^D \equiv$

$$\left( x_i^{tr} \beta + I_{[y_i \geq \kappa_i]} \cdot E(u_i | y_i) + I_{[y_i < \kappa_i]} \cdot E(u_i | y_i < \kappa_i) \right) \Big|_{(\beta, \sigma^2) = (\hat{\beta}^D, (\hat{\sigma}^D)^2)} \quad (20)$$

Here the superscript  $D$  refers to the censored-data ML as method (D) discussed in Section 3. Note that the SAE  $\hat{\vartheta}_i^D$  necessarily takes a different form according to whether the response  $y_i$  in the  $i$ 'th PSU is left-censored. The conditional expectations within  $\vartheta_i^D$  are  $E(u_i | y_i) = (y_i - x_i^{tr} \beta_0) \sigma_0^2 / \tau_i$  and

$$E(u_i | y_i < \kappa_i) = \frac{\sigma_0^2}{\tau_i} \int_{-\infty}^{\xi_i} z \phi(z) dz \frac{\sqrt{\tau_i}}{\Phi(\xi_i)} = -\frac{\phi(\xi_i) \sigma_0^2}{\Phi(\xi_i) \sqrt{\tau_i}}$$

Therefore,

$$\hat{\vartheta}_i^D = x_i^{tr} \hat{\beta}^D + I_{[y_i \geq \kappa_i]} \frac{(\hat{\sigma}^D)^2 (y_i - x_i^{tr} \hat{\beta}^D)}{(\hat{\sigma}^D)^2 + s_i} - I_{[y_i < \kappa_i]} \frac{\phi(\hat{\xi}_i^D) (\hat{\sigma}^D)^2}{\Phi(\hat{\xi}_i^D) \sqrt{\hat{\tau}_i^D}} \quad (21)$$

where

$$\hat{\tau}_i^D \equiv (\hat{\sigma}^D)^2 + s_i \quad , \quad \hat{\xi}_i^D \equiv (\kappa_i - x_i^{tr} \hat{\beta}^D) / \sqrt{\hat{\tau}_i^D}$$

To calculate MSE's, we need to take into account the probability of the observation in the  $i$ 'th PSU having been left-censored, where  $x_i, \kappa_i, n_i$  are fixed. The calculation (for  $n_i > 0$ , as in the case of  $MSE_i^A, MSE_i^B$ ) is as follows.

$$MSE_i^D = E\left( I_{[y_i \geq \kappa_i]} (\hat{\vartheta}_i^D - \vartheta_i)^2 + I_{[y_i < \kappa_i]} (\hat{\vartheta}_i^D - \vartheta_i)^2 \right)$$

Then, using the consistency of the ML estimators, we find to top order

$$\begin{aligned} MSE_i^D &\approx E\left( I_{[y_i \geq \kappa_i]} \left\{ \frac{\sigma_0^2}{\tau_i} (y_i - x_i^{tr} \beta_0) - u_i \right\}^2 + I_{[y_i < \kappa_i]} \left\{ \frac{\phi(\xi_i) \sigma_0^2}{\Phi(\xi_i) \sqrt{\tau_i}} + u_i \right\}^2 \right) \\ &= E\left( I_{[y_i \geq \kappa_i]} \left( \frac{\sigma_0^2}{\tau_i} (u_i + e_i) - u_i \right)^2 + \right. \\ &\quad \left. + I_{[y_i < \kappa_i]} \cdot \left[ (u_i - \frac{\sigma_0^2}{\tau_i} (u_i + e_i))^2 + \frac{\sigma_0^4}{\tau_i^2} (u_i + e_i + \sqrt{\tau_i} \frac{\phi(\xi_i)}{\Phi(\xi_i)})^2 \right] \right) \end{aligned}$$

since the cross-term in the expanded square term multiplying  $I_{[y_i < \kappa_i]}$  has expectation 0. Evaluating the final expectations now shows

$$\begin{aligned} MSE_i^D &\approx \frac{s_i \sigma_0^2}{\tau_i} + \frac{\sigma_0^4}{\tau_i} \Phi(\xi_i) \left( \frac{1}{\Phi(\xi_i)} \int_{-\infty}^{\xi_i} z^2 \phi(z) dz - \frac{\phi^2(\xi_i)}{\Phi^2(\xi_i)} \right) \\ &= \frac{s_i \sigma_0^2}{\tau_i} + \frac{\sigma_0^4}{\tau_i} \left( \Phi(\xi_i) - \xi_i \phi(\xi_i) - \frac{\phi^2(\xi_i)}{\Phi(\xi_i)} \right) \end{aligned} \quad (22)$$

The formula (22) for  $MSE_i^D$  naturally indicates that this MSE is close to the (top-order) nominal Fay-Herriot MSE of  $s_i \sigma_0^2 / \tau_i$ . When censoring is moderate, i.e. when most of the values  $\xi_i = (\kappa_i - x_i^{tr} \beta_0) / \sqrt{\tau_i}$  are less than  $-0.5$ , say, we note from our numerical experience in Table 1 that the bias-term  $\tilde{\Delta}_i$  in  $MSE_i^B$  is quite small, while the function  $\Phi(\xi) - \xi \phi(\xi) - \phi^2(\xi) / \Phi(\xi)$  is bounded between 0 and .083 for  $\xi \leq -0.5$ . Thus in moderate-censoring settings, the comparison of  $MSE_i^D$  versus nominal MSE is effectively between  $s_i \sigma_0^2 / \tau_i$  and  $s_i (s_i \sigma_0^2 + \sigma_*^4) / (\sigma_*^2 + s_i)^2$ , and it is easily checked that the first of these is always larger than the second, with noticeable differences when  $\sigma_*^2$  is far from  $\sigma_0^2$ . Thus, since we found in our numerical calculations of Table 1 that  $\sigma_*^2$  can easily differ by more than 15% when the censoring-proportion  $c$  is no more than 10%, we conclude that the mean-squared-errors for small-area estimators  $\hat{\vartheta}_i^D$  are likely to be much better than for  $\hat{\vartheta}_i^B$  or  $\hat{\vartheta}_i^A$ . We proceed, in the following Section, to test these theoretical predictions in terms of finite-sample SAE and MSE behavior from a simulation study using design matrix and parameters as in the SAIPE 1993 data (Citro and Kalton 2000). We also compare there the top-order formula (22), with remainders  $o(1)$ , to a more elaborate formula (33) derived in the Appendix to have remainders  $O(1/m)$ .

## 6 Comparative Simulation Study

We conducted a simulation study to check the performance of the estimation methodology presented in the previous sections. Our simulation design closely imitates the situation encountered in the U.S. Census Bureau's ongoing SAIPE project described in greater detail below in Section 7. For simplicity and confidentiality, the covariates used in the simulation are pseudo values simulated (once only) from a multivariate normal distribution with the same means and variances as the original covariates for all US counties which were used in the SAIPE 1993 log-rate model for poverty among school-age children related to sampled householders. The covariates are as

described in Section 7 and the coefficients used to generate the response variables are  $\beta = (-1.860, .236, .313, -.119, .393)$ . The sample sizes  $n_i$  are the actual US Current Population Survey (CPS) numbers of sampled households in 1992-94, as in the SAIPE project, for subsets of the first  $m$  alphabetically ordered US counties, after deleting Los Angeles county (by far the largest one). This was done because, in simulations not reported here, with fixed values of  $\sigma^2$  as small as .014, we found that the L.A. county SAE was very erratic and distorted the summary measures of MSE.

Fixing  $\beta$ , we generated values  $\{y_i\}_{i=1}^m$  in simulations with  $N = 1000$  iterations, according to model (1) with  $s_i = v_e/n_i$ , but we left-censored the generated values as  $y_i^* \equiv \max(\kappa_i, y_i)$ , where  $\kappa_i = -\log n_i$ . We have explored various combinations of simulation parameter values  $(m, \sigma^2, v_e)$ , but display results only for  $m = 100, 500, 1000$  cross-classified with four labelled combinations for  $(\sigma_0^2, v_e)$ :

Par	1	2	3	4
$(\sigma_0^2, v_e)$	(.5, 30)	(.5, 17)	(1, 30)	(1, 17)

These  $\sigma_0^2$  values are somewhat larger than the values fixed in the SAIPE log-count and log-rate models described in Citro & Kalton (2000), but they are reasonable, being similar to the values  $(\sigma^2, v_e)$  jointly fitted by maximum likelihood to the SAIPE log-rate FH model on 1993 data. Table 2 shows the averages over strata  $i$  of the small area parameters, the simulated SAE biases, and the true MSE under methods A= FH, B=Badj and D=Cens. The first set of A, B, D columns displays SAE bias and the second set MSE. In all cases, method D yields the smallest bias by far, as well as the smallest range. The bias for methods A and B is generally positive, much less so for method D. Both in terms of bias and MSE of SAE, method B appears always inferior to A.

Table 3 summarizes the mean and standard-error behavior of the maximum likelihood parameter estimates of  $(\beta, \sigma^2)$  obtained by methods A, B and D, over the same range of simulations as in Table 2, for parameter-combination  $Par = 2$ . (We calculated but do not display the corresponding results for the other parameter combinations, since the results were very similar.) Here also, method D performs best. As  $m$  increases, clearly the MLE converges to the correct values. In our simulation, the censoring rate varies between 5% to 14%. Method A produces SAE's biased considerably above their targets, with bias-squared as a percentage of MSE ranging from 4 to 12 %, and Method B did not materially improve the results. However, method D results in very accurate parameter estimates, at least for large samples.

Table 2: Average small-area parameters, censoring proportions, SAE biases and MSE's over 1000 simulation iterations, for various  $(m, \sigma^2, v_e)$  combinations, with  $\beta$  fixed =  $(-1.860, .236, .313, -.119, .393)$  as described in text.

m	Par	%Cens	$\bar{\theta}_i$	SAE Bias			MSE		
				A	B	D	A	B	D
100	1	6.6	-3.581	0.056	0.083	0.009	0.295	0.296	0.275
	2	5.4	-3.582	0.043	0.065	0.006	0.226	0.227	0.206
	3	8.0	-3.584	0.108	0.175	0.021	0.488	0.495	0.400
	4	7.0	-3.585	0.092	0.154	0.016	0.378	0.385	0.285
500	1	12.3	-3.770	0.105	0.150	0.016	0.321	0.328	0.286
	2	10.9	-3.767	0.085	0.126	0.011	0.265	0.268	0.229
	3	13.9	-3.773	0.172	0.272	0.027	0.567	0.590	0.443
	4	12.7	-3.772	0.156	0.254	0.023	0.476	0.493	0.338
1000	1	10.0	-3.782	0.087	0.127	0.015	0.295	0.300	0.266
	2	8.8	-3.783	0.070	0.105	0.011	0.240	0.242	0.209
	3	11.5	-3.781	0.150	0.240	0.030	0.519	0.537	0.408
	4	10.5	-3.784	0.133	0.220	0.024	0.426	0.439	0.305

A striking feature of Table 2 is that the MSE's under method B are systematically larger than those of method A, despite the clear indication of Table 3 that the parameter estimators under method B are systematically closer than those of method A to the truth [or to those of method D]. This apparent paradox is resolved by recalling that the MSE formulas for methods A and method D: the one in D explicitly adjusts the fixed-effect predictor downward for each unobserved PSU known to be below-threshold, while the one in methods A and B takes the fixed-effect predictor as is. So the finding is that the MSE and bias become worse if one uses the method-A SAE formula with the method-B or method-D parameter estimators.

Clearly the current practice of the Census Bureau in SAIPE, method A, would produce serious bias in SAE's, with MSE's too large by 10-20%, if the FH model simulated were the correct one. Note however that these biases and MSE's are expressed on the measurement scale of the underlying FH model, which in the case of SAIPE is logarithmic. Comparisons between the performance of Methods A, B, and D on real SAIPE data are given in Section 7 below.

Table 3: Parameter Estimates and their SE's, for  $Par = 2$  case.

m	Method	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_4$	$\hat{\sigma}^2$
100	A	-1.909 (0.405)	0.252 (0.464)	0.249 (0.367)	-0.113 (0.917)	0.410 (0.558)	0.427 (0.098)
	B	-1.873 (0.405)	0.222 (0.459)	0.295 (0.360)	-0.103 (0.913)	0.420 (0.546)	0.457 (0.097)
	D	-1.874 (0.410)	0.224 (0.461)	0.293 (0.363)	-0.113 (0.927)	0.420 (0.540)	0.458 (0.105)
500	A	-1.932 (0.187)	0.249 (0.192)	0.219 (0.148)	-0.151 (0.426)	0.419 (0.225)	0.445 (0.048)
	B	-1.857 (0.185)	0.241 (0.193)	0.313 (0.144)	-0.103 (0.420)	0.391 (0.222)	0.493 (0.047)
	D	-1.858 (0.187)	0.238 (0.195)	0.314 (0.144)	-0.108 (0.425)	0.390 (0.222)	0.492 (0.052)
1000	A	-1.909 (0.127)	0.276 (0.136)	0.222 (0.100)	-0.170 (0.322)	0.412 (0.147)	0.455 (0.033)
	B	-1.860 (0.127)	0.228 (0.136)	0.315 (0.097)	-0.113 (0.318)	0.398 (0.144)	0.497 (0.032)
	D	-1.862 (0.128)	0.229 (0.137)	0.314 (0.098)	-0.113 (0.322)	0.397 (0.146)	0.496 (0.035)

Table 4: Empirical  $MSE_i^D$  values averaged over  $i$  (MSED), in simulations of size  $m = 100, 500, 1000$ , along with relative differences between theoretical MSE and MSED (RD), and between estimated MSE and MSED (RB).

m	Par	MSED	RD1	RB1	RD2	RB2
100	1	0.275	-0.098	-0.175	-0.023	-0.100
	2	0.206	-0.072	-0.120	-0.017	-0.064
	3	0.400	-0.095	-0.135	-0.044	-0.083
	4	0.285	-0.079	-0.110	-0.041	-0.071
500	1	0.286	-0.042	-0.059	-0.021	-0.038
	2	0.229	-0.038	-0.049	-0.024	-0.034
	3	0.443	-0.053	-0.065	-0.039	-0.051
	4	0.338	-0.055	-0.062	-0.044	-0.051
1000	1	0.266	-0.033	-0.040	-0.024	-0.031
	2	0.209	-0.033	-0.038	-0.026	-0.031
	3	0.408	-0.057	-0.062	-0.051	-0.056
	4	0.305	-0.056	-0.059	-0.052	-0.055

Table 4 compares the empirical, theoretical, and estimated quantities  $MSE_i^D$  within the same SAIPE-style simulation as in the other Tables. The empirical MSE's using small-area estimators  $\hat{v}_i^D$  are calculated directly, over the 1000 simulation iterations in each of three simulations, with  $m = 100, 500$  and  $1000$  strata. The theoretical MSE's are calculated using formulas (22) and (33), with true parameters  $(\beta, \sigma^2, v_e)$  substituted, and the estimated MSE's from the same formulas with method-D parameter estimators substituted (based on fixed  $v_e$ ). Each of these MSE's is then averaged over the  $m$  strata of the simulation. Column RD1 [respectively RD2] gives the relative difference between theoretical MSE based on (22) [resp. (33)] and  $MSE_i^D$ ; and column RB1 [resp. RB2] give relative differences between MSE estimators, based on the corresponding formulas, and  $MSE_i^D$ .

Table 4 shows that the MSE formulas (22) and (33) under-estimate the actual empirical  $MSE_i^D$  on average, by an amount which is no more than 10% for (22) and 5% for (33), for all combinations  $(m, Par)$  tried. The relative errors for estimated MSE's — based on plugged-in rather than true parameter values — are somewhat worse, but substantially so only for  $m = 100$ . None of the theoretical or estimated MSE's show a clear decrease with  $m$  in the Table, but this is because the simulations with different  $m$  are considerably different. (For example, see the differences in censoring percentages across  $m$  displayed in Table 2.)

## 7 Real-data Comparisons

### 7.1 Log-rate model for SAIPE 1993 data

As described in the Introduction, our motivation for this paper came from the small-area estimation method of the SAIPE program based on a Fay-Herriot model for log-transformed county child-poverty response data. We present model-fitting results using methods A, B, and D on the SAIPE data for income-year 1993 (using CPS samples aggregated across 1992-94). As described elsewhere (Slud 2003, 2004), the log-rate model for SAIPE 1993 data uses as response variable  $y_i$  the log-transformed CPS-weighted ratio estimate of county child poverty rate (among children aged 5–17 related to householders) in those counties appearing in the pooled 1992–94 CPS sample, and as predictors the logarithms of the following four variables LTAXRT, LSTMPRT, LFILRT, LCPRT:

- LTAXRT: the logarithm of the 1993 county total of Poor child IRS exemptions over total child exemptions,
- LSTMPRT: the logarithm of the ratio of the number of people receiving food stamps over the 1993 demographic estimate of county population,
- LFILRT: the logarithm of the ratio of the county number of IRS child exemptions over the 1993 estimated resident child population,
- LCPRRT: the logarithm of the child poverty rate estimated from the previous decennial census adjusted to the ‘CPS universe’ definitions of resident householders and related children.

Here  $n_i$  is the number of *households* sampled, even though the relevant sampled units would be children, because the count of CPS-sampled children was not directly available for the SAIPE data. The Fay-Herriot model (1) with this response and predictors, where  $\sigma_0^2$  is fixed (in 1993) at 0.014,  $s_i = v_e/n_i$ , and  $(\beta, v_e)$  are the unknown parameters, is the *county log-rate model* which differs from but closely approximates the model for log counts (of poor related children, by county) actually specified in SAIPE. Slud (2003, 2004) has studied the relationship of the log-rate model to the actual SAIPE log-count model, as well as the goodness of fit of small-area predictors based on the model both to the CPS data and (in decennial census years) to the census county child-poverty rates (adjusted to ‘CPS universe’ definitions). In the 1993 SAIPE data, there were 1488 counties in the 1992-94 aggregated CPS sample, of which only 1184 had a positive number of sampled poor children. Since the 304 sampled counties with no sampled poor children were ignored in fitting the model, they represent a censoring rate of  $304/1488 = 20.4\%$ . However, as indicated in Section 2, we model the dropping of counties with 0-counts by saying that the log ratios  $y_i$  of the counts of poor children divided by  $n_i$  are still normally distributed variates which are left-censored by the values  $\kappa_i = \log(1/n_i)$ , i.e. known only to take some value less than  $\kappa_i$ .

Table 5 exhibits the maximum likelihood parameter estimates of  $(\beta, v_e)$  obtained by Methods A, B, and D, with  $\sigma_0^2 = .014$  fixed. The Methods B and D yield very similar parameter estimates except for the intercept and LSTMPRT coefficients and  $v_e$ . (The LFILRT coefficients have large standard error, thus are not as different as they look.) However, as we now proceed to show, **none** of these models fits very well. As a preliminary indication, consider the predicted censoring percentage according to these models with

Table 5: Parameters fitted by Method A= FH, B= Badj, and D= Cens, to SAIPE 1993 data with predictor means subtracted and  $\sigma_0^2 = .014$  fixed.

	Int	LTAXRT	LSTMPRT	LFILRT	LCPRT	$v_e$
	$\beta_0$	$\beta_1$	$\beta_2$	$\beta_3$	$\beta_4$	
FH	-1.824	.274	.288	-.447	.379	33.1
Badj	-1.854	.235	.336	-.420	.369	35.9
Cens	-1.952	.211	.429	-.471	.365	61.8

Table 6: Parameters fitted by Method A= FH, B= Badj, and D= Cens, to SAIPE 1993 data with predictor means subtracted and  $v_e = 34.33$  fixed at value obtained by ML over  $(\beta, \sigma^2, v_e)$  within left-censored log-likelihood (5).

	Int	LTAXRT	LSTMPRT	LFILRT	LCPRT	$\sigma^2$
	$\beta_0$	$\beta_1$	$\beta_2$	$\beta_3$	$\beta_4$	
FH	-1.849	.304	.282	-.416	.338	.070
Badj	-1.889	.253	.340	-.369	.331	.077
Cens	-2.073	.293	.397	-.357	.336	.277

estimates plugged into formula (23): .091 for Method A, .096 for Method B, and .125 for Method D. Thus, although the corrections given by Methods B and D do bring the estimated value of  $c$  slightly higher, they still fall far below the observed censoring rate of .204.

One might say that the value .014 for  $\sigma^2$ , artificially fixed by a method documented in Kalton and Citro (2000) involving linear-model fitting using a linear model with the analogous predictors and the previous (in this case, the 1990) decennial census, is an obstacle to finding a close fit to the log CPS weighted ratio estimate by county that is being used as response variable. For this reason, re-fit the models by jointly maximizing the censored-data likelihood over  $(\beta, \sigma^2, v_e)$ , which can be accomplished using a function coded in R repeatedly, alternately to maximize (5) over  $(\beta, v_e)$  for fixed  $\sigma^2$  and to maximize over  $(\beta, \sigma^2)$  for fixed  $v_e$ . The resulting joint estimates are  $\sigma^2 = .277$ ,  $v_e = 34.33$ . Now fixing this  $v_e$  value and re-estimating parameters using Methods A and B gives the parameter estimates summarized in Table 6. We fixed  $v_e$  in this way because according to our model, the censored-data likelihood most fully describes the data, and we are investigating whether this fully specified left-censored model fits adequately.

Table 7: Limiting parameter estimates from Methods A=FH, B=Badj, and D=Cens, with  $v_e = 34.33$  fixed, if the left-censored model held with  $n_i$  and predictor-variable covariances as in SAIPE 1993, and if the predictors were multivariate normally distributed, identically across counties.

	Int	LTAXRT	LSTMPRT	LFILRT	LCPRT	$\sigma^2$
	$\beta_0$	$\beta_1$	$\beta_2$	$\beta_3$	$\beta_4$	
FH	-1.996	.276	.373	-.336	.316	.250
Badj	-2.074	.292	.396	-.356	.335	.281
Cens	-2.073	.293	.397	-.357	.336	.277

Again we can check whether any of the fitted models in Table 6 give adequate estimates of the censoring proportion  $c$ . The respective plug-in estimated values of  $c$  from these models are: .095 for Method A, .099 for B, and .122 for D. Thus, none of the estimation methods with the alternatively chosen value for  $v_e$  provides a close estimate of  $c$ .

Another indication of the (lack of) fit of the fully specified left-censored FH model to the data is given by the discrepancies between the Method B and Method D estimators. The formulas for the limiting values  $\beta_*$ ,  $\sigma_*^2$  of the Method A estimators and the corresponding Method-B adjusted values  $\beta_* - \delta_\beta$ ,  $\sigma_*^2 - \delta_{\sigma^2}$  can be evaluated numerically, analogously to the calculation done for Table 1, if the centered covariates  $X_i$  are treated as being multivariate normally distributed with means 0 and variances estimated from their empirical covariance matrix based on the 1488 SAIPE counties in 1993. The results of the calculation are given in Table 7.

We can see clearly that *if* the left-censored FH model held precisely with the parameters fitted by Method B in Table 6 (and if the predictor variables were *iid* multivariate normal), then the Method B and Method D parameter estimates would agree extremely closely in large- $m$  samples. This is something we saw in the  $m = 1000$  simulations in Section 6, e.g. in Table 3, but definitely do not observe in the actual SAIPE 1993 data.

The combined failure of the fitted models in Tables 5 and 6 to provide accurate  $c$  estimators or to match closely between Methods B and D strongly suggests that the left-censored FH model does not adequately fit the SAIPE 1993 data. Given that the predictor variables do linearly predict the log-rate responses very strongly (Kalton and Citro 2000, Slud 2003), how are we to understand the lack of fit? We cannot yet exclude the possibility either

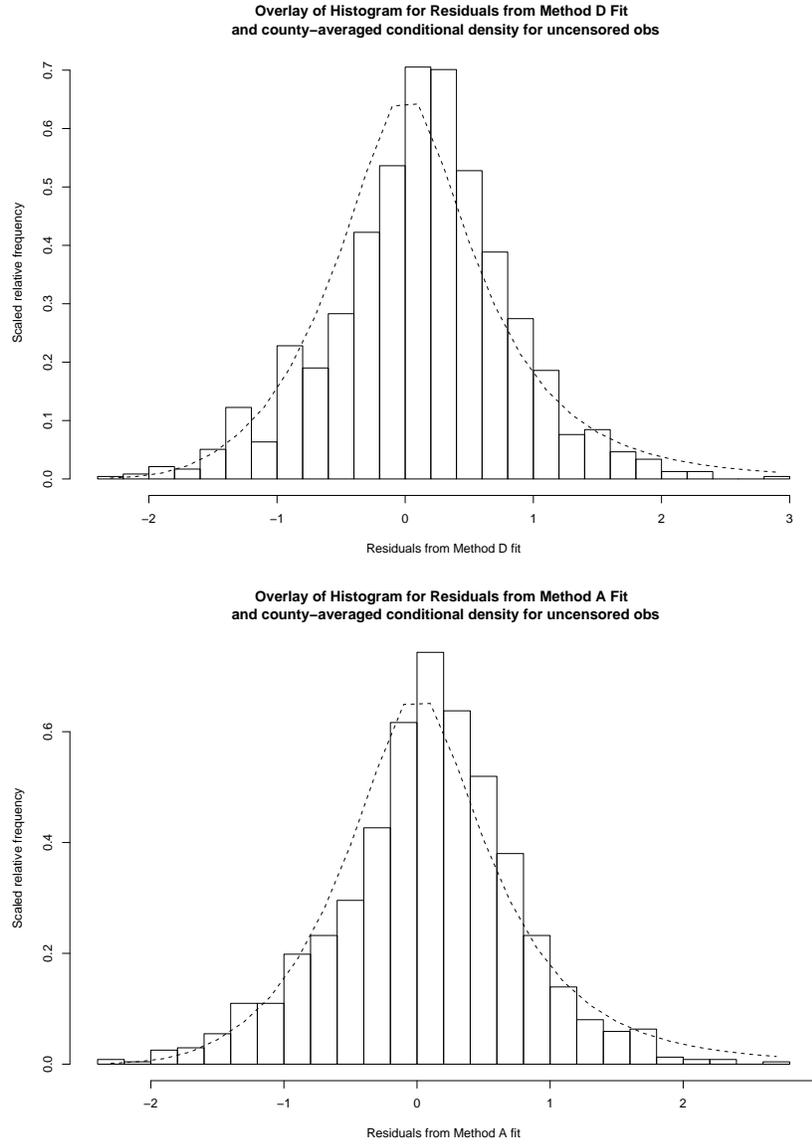


Figure 1: Histograms for residuals from Method D ( $v_e = 34.33$  fixed) and Method A ( $\sigma^2 = .014$  fixed) fits for observed SAIPE 1993 counties with non-zero counts of poor children, overlaid with estimated county-averaged conditional density for uncensored observations, as described in text.

that the underlying county random-effects  $u_i$  are non-normal or that the sample-sizes  $n_i$  figuring in the thresholds  $\kappa_i$  do not behave independently of the responses  $y_i$  conditionally given the covariates  $x_i$ . So we proceed by examining the behavior of the above-threshold residuals and comparing it to the model predictions.

We consider next a diagnostic for lack of fit based on graphical comparison between the histogram of residuals from the Method-D or method-A model and the corresponding conditional density constructed for the observed ( $n_i > 0$ ) population of counties. Figure 1 respectively shows histograms of residuals  $y_i - x_i^{tr} \hat{\beta}^D$  (from method D with fixed  $v_e = 34.331$ ) and  $y_i - x_i^{tr} \hat{\beta}^A$  (from method A with  $\sigma^2 = .014$  fixed) for the 1184 uncensored counties, i.e. counties with  $y_i > \kappa_i$ . If the left-censored Fay-Herriot model were valid, and the parameters  $\beta_0, v_e, \sigma_0^2$  known, then in each county for which  $y_i > \kappa_i$  is known to have occurred, the conditional density of  $y_i - x_i^{tr} \beta_0$  would be

$$\frac{1}{\sqrt{\tau_i}} I_{[t \geq \kappa_i - x_i^{tr} \beta_0]} \phi(t/\sqrt{\tau_i}) / (1 - \Phi(\xi_i))$$

It follows that the displayed histogram ought to be close to the average conditional density over all such counties,

$$\frac{1}{1184} \sum_{i: y_i \geq \kappa_i} \tau_i^{-1/2} I_{[t \geq \kappa_i - x_i^{tr} \beta_0]} \phi(t/\sqrt{\tau_i}) / (1 - \Phi(\xi_i))$$

Therefore, we overlaid the histogram with this density, where the respective estimates  $(\hat{\beta}^D, (\hat{\sigma}^2)^D, 34.331)$  and  $(\hat{\beta}^A, .014, \hat{v}_e^A)$  are substituted for the parameters  $(\beta_0, \sigma^2, v_e)$ . The resulting picture is Figure 1. Although not perfect, because of a slight skewness and mean shift to the left, the fit of the ‘theoretical’ conditional densities to the histograms in these graphs is strikingly good. Both of these graphs relate only to the behavior of above-threshold response-values  $y_i$ , and they differ in that the population-wide parameters in the upper (Method D) panel are estimated based on all data, including the below-threshold observations, while in the lower panel (Method A) only the above-threshold observations were used in fitting. Partly as a result, the empirical average of the above-threshold residuals  $y_i - x_i^{tr} \hat{\beta}^D$  for the model depicted in the upper panel was 0.130, which corresponded fairly well to the estimated theoretical average

$$\frac{1}{1488} \sum_{i=1}^{1488} \frac{\phi(\hat{\xi}_i^D)}{\Phi(\hat{\xi}_i^D)} \left( (\hat{\sigma}^D)^2 + \frac{34.331}{n_i} \right)^{1/2} = .190$$

while the empirical average of 0.075 for Model A residuals from the lower panel was more sharply discrepant from its estimated theoretical average

$$\frac{1}{1488} \sum_{i=1}^{1488} \frac{\phi(\hat{\xi}_i^A)}{\Phi(\hat{\xi}_i^A)} \left( .014 + \frac{\hat{v}_e^A}{n_i} \right)^{1/2} = .160$$

Overall, the diagnostics of fit suggest that the lack of fit of the left-censored FH model to the SAIPE 1993 data are largely due to the failure of the left-censoring threshold  $\kappa_i$  in a FH model to describe the phenomenon of observing 0 poor children in a sampled county, rather than to a failure of distributional assumptions concerning random effects.

## 8 Summary and Conclusions

We have shown, theoretically and through simulations, the considerable bias and inflation of MSE that can result from ignoring left-censoring in a Fay-Herriot model by comparison with an estimation methodology (both for population and small-area parameters) based on a censored-data likelihood. We have provided a method B (estimating equations 13–15) of adjusting the parameter estimators derived by ignoring left-censoring, a method which provides answers extremely close to the censored-data ML estimators when censoring is up to about 10%. As shown in the example of Section 7, this result provides a useful check on the correctness of the model assumptions.

Theoretical results, including formulas and estimators for MSE of SAE's derived from censored-data (method D) ML estimators, are corroborated by a simulation study presented in Section 6. In the left-censored simulations, SAE's which adjust fixed-effect predictors downward in areas known to be left-censored clearly outperform SAE's which do not.

As with other Small-Area methods, the results presented here are unavoidably parametric. Yet in the motivating SAIPE example, we have seen that lack of fit is apparently due not to the failure of distributional assumptions, but to the simple model of independent sampling and left-censoring. In that example, there is in fact some clustering of samples: the CPS samples clusters of four nearby housing units, which are more similar to one another than are area-wide units. Thus, in areas with small samples, there might be a so-far unmodelled tendency for the below-threshold proportion to be larger than allowed for in model (1). In future work, we will study models with parameters for inflation of this below-threshold category.

## 9 Appendix

### 9.1 Regularity Conditions

It is assumed throughout the paper that the response-variables  $y_i$ ,  $i = 1, \dots, m$ , satisfy the model (1), and also that:

- (a) The random vectors  $x_i$  and sample sizes  $n_i$  are either uniformly bounded or are realizations of independent identically distributed variates with finite fourth moments.
- (b) Whether random or not, the vectors  $x_i \in \mathbf{R}^p$  are such that as  $m$  gets large,  $E(\frac{1}{m} \sum_{i=1}^m \|x_i\|^4)$  are uniformly bounded and with probability approaching 1 as  $m \rightarrow \infty$ ,  $\sum_{i=1}^m x_i^{\otimes 2}$  is a positive definite matrix, where for a column vector  $\mathbf{v}$ , we denote  $\mathbf{v}^{\otimes 2} = \mathbf{v} \mathbf{v}^{tr}$ .
- (c) The following large-sample (almost-sure) limits exist as  $m \rightarrow \infty$ , with error-terms  $O_P(m^{-1/2})$  :

$$\lim_m m^{-1} \sum_{i=1}^m \frac{x_i^{\otimes 2}}{\sigma^2 + s_i} \quad , \quad \lim_m m^{-1} \sum_{i=1}^m \frac{1}{2(\sigma^2 + s_i)^2}$$

- (d) For each compact subinterval  $J$  of  $(0, \infty)$ , the following large-sample limits exist uniformly over  $t \in J$ :

$$\begin{aligned} & \lim_m \frac{1}{m} \sum_{i=1}^m \frac{x_i^{\otimes 2}}{(t + s_i)^k} (1 - \Phi(\frac{\kappa_i - x_i^{tr} \beta_0}{\sqrt{\sigma_0^2 + s_i}})) \quad , \quad k = 1, 2, 3 \\ & \lim_m \frac{1}{m} \sum_{i=1}^m \frac{1}{(t + s_i)^k} (1 - \Phi(\frac{\kappa_i - x_i^{tr} \beta_0}{\sqrt{\sigma_0^2 + s_i}})) \quad , \quad k = 0, 1, 2, 3 \\ & \lim_m \frac{1}{m} \sum_{i=1}^m \frac{x_i^j \kappa_i^{1-j}}{(t + s_i)^k} \phi(\frac{\kappa_i - x_i^{tr} \beta_0}{\sqrt{\sigma_0^2 + s_i}}) \quad , \quad j = 0, 1, \quad k = 1, 2, 3 \end{aligned}$$

Moreover, the limiting proportion uncensored must be strictly less than 1:

$$c \equiv \lim_m \frac{1}{m} \sum_{i=1}^m \Phi(\frac{\kappa_i - x_i^{tr} \beta_0}{\sqrt{\sigma_0^2 + s_i}}) < 1 \quad (23)$$

## 9.2 Proof of Theorem 4.1

Beyond the regularity conditions, the Theorem assumes that the quantity  $c$  is close (but not necessarily equal) to 0. The regularity conditions directly imply the (uniform) large-sample convergence of both  $\tilde{\beta}(t)$  and  $\beta_*(t)$  to

$$\beta_0 + \left( \lim_m \frac{1}{m} \sum_{i=1}^m \frac{x_i^{\otimes 2}}{t + s_i} (1 - \Phi(\xi_i)) \right)^{-1} \left( \lim_m \frac{1}{m} \sum_{i=1}^m \frac{x_i \sqrt{\tau_i}}{t + s_i} \phi(\xi_i) \right) \quad (24)$$

and also the convergence of  $\frac{d}{dt} \tilde{\beta}(t) = \tilde{\beta}'(t)$  and of  $\frac{d}{dt} \beta_*(t) = \beta_*'(t)$  to

$$\begin{aligned} & \left( \lim_m \frac{1}{m} \sum_{i=1}^m \frac{x_i^{\otimes 2}}{t + s_i} (1 - \Phi(\xi_i)) \right)^{-1} \left( - \lim_m \frac{1}{m} \sum_{i=1}^m \frac{x_i \sqrt{\tau_i}}{(t + s_i)^2} \phi(\xi_i) \right. \\ & \quad \left. + \lim_m \frac{1}{m} \sum_{i=1}^m \frac{x_i^{\otimes 2}}{(t + s_i)^2} (1 - \Phi(\xi_i)) (\beta_*(t) - \beta_0) \right) \end{aligned} \quad (25)$$

and of  $\tilde{\beta}''(t)$ ,  $\beta_*''(t)$ , both to the same limit. Now, restricting attention to a small neighborhood of  $t$  values (not depending on  $m$ ) on which the minimum of the right-hand side of (10) is unique, we find by differentiation that the minimizer  $\sigma_*^2$  is determined as the root of the function

$$\lim_m \frac{1}{m} \sum_{i=1}^m \left[ (1 - \Phi(\xi_i)) \frac{t - \sigma_0^2}{(t + s_i)^2} - \frac{\sqrt{\tau_i}}{t + s_i} \left( x_i^{tr} \beta_*'(t) + \frac{\kappa_i - x_i^{tr} \beta_*(t)}{t + s_i} \right) \phi(\xi_i) \right]$$

whose derivative also has a uniform limit.

Under the hypotheses of Theorem 4.1,

$$\beta_*(t) \stackrel{P}{\approx} \lim_m \left( \frac{1}{m} \sum_{i=1}^m \frac{x_i^{\otimes 2}}{t + s_i} \right)^{-1} \frac{1}{m} \sum_{i=1}^m \frac{x_i y_i}{t + s_i} = \beta_0$$

and  $\sigma_*^2 \approx \sigma_0^2$ , and by (24) and the boundedness away from 1 of  $c$  in (23),

$$\beta_*'(\sigma_*^2), \beta_* - \beta_0 = \mathcal{O}_P \left( \lim_m \frac{1}{m} \sum_{i=1}^m \frac{\sqrt{\tau_i}}{\sigma_*^2 + s_i} \phi(\xi_i) \right)$$

It follows by the inspection of the displayed function above with root  $\sigma_*^2$  that to top order,

$$\lim_m \frac{1}{m} \sum_{i=1}^m \left[ (1 - \Phi(\xi_i)) \frac{\sigma_*^2 - \sigma_0^2}{\tau_i^2} - \frac{\phi(\xi_i)}{\sqrt{\tau_i}} \left( \frac{\kappa_i - x_i^{tr} \beta_0}{\tau_i} \right) \right] \approx 0$$

proving the first approximate equality of Theorem 4.1. To prove the second, successively substitute  $\sigma_*^2, \sigma_0^2$  for  $t$  into (24) and compare the two results using (25) and Taylor's theorem to find

$$\beta_* - \beta_0 = \left( \lim_m \frac{1}{m} \sum_{i=1}^m \frac{x_i^{\otimes 2}}{\tau_i} (1 - \Phi(\xi_i)) \right)^{-1} \left( \lim_m \frac{1}{m} \sum_{i=1}^m \frac{x_i}{\sqrt{\tau_i}} \phi(\xi_i) \right) \cdot (1 + o_P(1))$$

The proof is complete.  $\square$

### 9.3 Formulas for $A, \Sigma$

We begin by collecting for future reference the formulas and notations

$$\xi_i = \frac{\kappa_i - x_i^{tr} \beta_0}{\sqrt{\tau_i}} \quad , \quad \Delta_i = x_i^{tr} (\beta_* - \beta_0)$$

and for  $k = 1, \dots, 4$ ,

$$\mu_{ki} \equiv E \left( I_{[y_i \geq \kappa_i]} (y_i - x_i^{tr} \beta_*)^k \right) = \int_{\xi_i}^{\infty} (z \sqrt{\tau_i} - \Delta_i)^k \phi(z) dz$$

Then

$$\mu_{1i} = \sqrt{\tau_i} \phi(\xi_i) - \Delta_i (1 - \Phi(\xi_i))$$

$$\mu_{2i} = (\tau_i + \Delta_i^2) (1 - \Phi(\xi_i)) + \sqrt{\tau_i} (\sqrt{\tau_i} \xi_i - 2\Delta_i) \phi(\xi_i)$$

$$\begin{aligned} \mu_{3i} &= \sqrt{\tau_i} \left( \tau_i (\xi_i^2 + 2) - 3\xi_i \Delta_i \sqrt{\tau_i} + 3\Delta_i^2 \right) \phi(\xi_i) \\ &\quad - (\Delta_i^3 + 3\tau_i \Delta_i) (1 - \Phi(\xi_i)) \end{aligned}$$

$$\begin{aligned} \mu_{4i} &= \sqrt{\tau_i} \left( \tau_i^{3/2} (\xi_i^3 + 3\xi_i) - 4\tau_i \Delta_i (\xi_i^2 + 2) + 6\tau_i^{1/2} \xi_i \Delta_i^2 - 4\Delta_i^3 \right) \phi(\xi_i) \\ &\quad + \left( \Delta_i^4 + 3\tau_i^2 + 6\tau_i \Delta_i^2 \right) (1 - \Phi(\xi_i)) \end{aligned}$$

In terms of these notations, we apply the formulas for  $A, \Sigma$  given below (11) to obtain explicitly:

$$A_{11} = \frac{1}{m} \sum_{i=1}^m x_i^{\otimes 2} \frac{1 - \Phi(\xi_i)}{\sigma_*^2 + s_i} \quad (26)$$

$$A_{12} = \frac{1}{m} \sum_{i=1}^m x_i \frac{\mu_{1i}}{(\sigma_*^2 + s_i)^2} \quad (27)$$

$$A_{22} = \frac{1}{m} \sum_{i=1}^m \frac{1}{(\sigma_*^2 + s_i)^3} \left[ \mu_{2i} - \frac{\sigma_*^2 + s_i}{2} (1 - \Phi(\xi_i)) \right] \quad (28)$$

$$\Sigma_{11} = \frac{1}{m} \sum_{i=1}^m \frac{x_i^{\otimes 2}}{(\sigma_*^2 + s_i)^2} (\mu_{2i} - \mu_{1i}^2) \quad (29)$$

$$\Sigma_{12} = \frac{1}{m} \sum_{i=1}^m \frac{x_i}{2(\sigma_*^2 + s_i)^2} \left( \frac{\mu_{3i}}{\sigma_*^2 + s_i} - \mu_{1i} \left( \frac{\mu_{2i}}{\sigma_*^2 + s_i} + \Phi(\xi_i) \right) \right) \quad (30)$$

$$\begin{aligned} \Sigma_{22} = \frac{1}{m} \sum_{i=1}^m \frac{1}{4(\sigma_*^2 + s_i)^2} & \left[ \frac{\mu_{4i}}{(\sigma_*^2 + s_i)^2} - \frac{2\mu_{2i}}{\sigma_*^2 + s_i} + 1 - \Phi(\xi_i) \right. \\ & \left. - \left( \frac{\mu_{2i}}{\sigma_*^2 + s_i} - 1 + \Phi(\xi_i) \right)^2 \right] \quad (31) \end{aligned}$$

Note that when censoring is absent, i.e. all thresholds  $\kappa_i$  tend to  $-\infty$ , then  $a_* = a_0$  and by well-known properties (skewness and kurtosis 0) of the normal distribution,  $\mu_{1i} = \mu_{3i} = 0$ ,  $\mu_{2i} = \tau_i$ ,  $\mu_{4i} = 3\tau_i^2$ . It follows in this case that  $A_{11} = \Sigma_{11} = \Sigma_\beta^{-1}$ ,  $A_{12} = \Sigma_{12} = 0$ ,  $A_{22} = \Sigma_{22} = \Sigma_{\sigma^2}^{-1}$ .

#### 9.4 Derivation of Modified MSE-D Formula

We develop here a more precise formula for MSE based on the estimator  $\hat{\vartheta}_i^D$  given in formula (21).

$$\begin{aligned} MSE_i^D &= E(\hat{\vartheta}_i^D - \vartheta_i)^2 \\ &= E \left\{ x_i^{tr} \hat{\beta} + I_{[y_i \geq \kappa_i]} \hat{\gamma}_i (y_i - x_i^{tr} \hat{\beta}) - I_{[y_i < \kappa_i]} \sqrt{\hat{\tau}_i} \hat{\gamma}_i \frac{\phi(\hat{\xi}_i)}{\Phi(\hat{\xi}_i)} - x_i^{tr} \beta_0 - u_i \right\}^2 \\ &= E \left\{ x_i^{tr} (\hat{\beta} - \beta_0) + I_{[y_i \geq \kappa_i]} (\hat{\gamma}_i (u_i + e_i) - u_i - \hat{\gamma}_i x_i^{tr} (\hat{\beta} - \beta_0)) \right. \\ &\quad \left. + I_{[y_i < \kappa_i]} (-\hat{\gamma}_i \sqrt{\hat{\tau}_i} \frac{\phi(\hat{\xi}_i)}{\Phi(\hat{\xi}_i)} - u_i) \right\}^2 \end{aligned}$$

Using the independence of  $\alpha_i \equiv \gamma_i(u_i + e_i) - u_i$  from  $\{u_j + e_j\}_{j=1}^m$ , and therefore from the estimators  $(\hat{\beta}, \hat{\sigma}^2) = (\hat{\beta}^D, (\hat{\sigma}^D)^2)$ , we evaluate the 0'th

order term in the last expression and recognize the cross-terms as 0, finding the last display equal to

$$\begin{aligned}
& \frac{s_i \sigma_0^2}{\tau_i} + E \left\{ x_i^{tr} (\hat{\beta} - \beta_0) + I_{[y_i \geq \kappa_i]} (\hat{\gamma}_i - \gamma_i) (u_i + e_i) \right. \\
& \quad \left. - I_{[y_i \geq \kappa_i]} \hat{\gamma}_i x_i^{tr} (\hat{\beta} - \beta_0) - I_{[y_i < \kappa_i]} \left( \gamma_i (u_i + e_i) + \hat{\gamma}_i \sqrt{\hat{\tau}_i} \frac{\phi(\hat{\xi}_i)}{\Phi(\hat{\xi}_i)} \right) \right\}^2 \\
= & \quad (22) - 2 E \left\{ \left( x_i^{tr} (\hat{\beta} - \beta_0) + I_{[y_i \geq \kappa_i]} [(\hat{\gamma}_i - \gamma_i) (u_i + e_i) - \hat{\gamma}_i x_i^{tr} (\hat{\beta} - \beta_0)] \right. \right. \\
& \quad \left. \left. - I_{[y_i < \kappa_i]} \left( \hat{\gamma}_i \frac{\phi}{\Phi} \Big|_{\hat{\xi}_i} \sqrt{\hat{\tau}_i} - \gamma_i \frac{\phi}{\Phi} \Big|_{\xi_i} \sqrt{\tau_i} \right) \cdot I_{[y_i < \kappa_i]} \gamma_i (u_i + e_i + \frac{\phi(\xi_i)}{\Phi(\xi_i)} \sqrt{\tau_i}) \right\} \\
& \quad + E \left\{ x_i^{tr} (\hat{\beta} - \beta_0) + I_{[y_i \geq \kappa_i]} (\hat{\gamma}_i - \gamma_i) (u_i + e_i) - I_{[y_i \geq \kappa_i]} \hat{\gamma}_i x_i^{tr} (\hat{\beta} - \beta_0) \right. \\
& \quad \left. - I_{[y_i < \kappa_i]} \left( \hat{\gamma}_i \frac{\phi}{\Phi} \Big|_{\hat{\xi}_i} \sqrt{\hat{\tau}_i} - \gamma_i \frac{\phi}{\Phi} \Big|_{\xi_i} \sqrt{\tau_i} \right) \right\}^2
\end{aligned}$$

where (22) in the last formula denotes the 0'th order formula given in equation (22).

Now we check that the middle term (the cross-term) in the last formula is  $o_P(1/m)$  for large  $m$ . To do this, we represent

$$\hat{\beta} - \beta_0 = (mI)^{-1} \nabla \log \text{Lik}(\beta_0, \sigma_0^2) + O_P(m^{-1})$$

which results in a calculation

$$\begin{aligned}
& \text{Cov} \left( \hat{\beta} - \beta_0, I_{[y_i < \kappa_i]} \gamma_i \left( \frac{u_i + e_i}{\sqrt{\tau_i}} + \frac{\phi(\xi_i)}{\Phi(\xi_i)} \right) \right) = o\left(\frac{1}{m}\right) + \\
& \quad \text{Cov} \left( ((mI)^{-1})_{11} \frac{x_i}{\sqrt{\tau_i}} \left( I_{[y_i \geq \kappa_i]} \frac{u_i + e_i}{\sqrt{\tau_i}} - I_{[y_i < \kappa_i]} \frac{\phi(\xi_i)}{\Phi(\xi_i)} \right) \right. \\
& \quad \left. - ((mI)^{-1})_{12} \frac{1}{2\tau_i} \left( I_{[y_i \geq \kappa_i]} \left( 1 - \frac{(u_i + e_i)^2}{\tau_i} \right) + I_{[y_i < \kappa_i]} \xi_i \frac{\phi(\xi_i)}{\Phi(\xi_i)} \right), \right. \\
& \quad \left. I_{[y_i < \kappa_i]} \gamma_i \left( \frac{u_i + e_i}{\sqrt{\tau_i}} + \frac{\phi(\xi_i)}{\Phi(\xi_i)} \right) \right)
\end{aligned}$$

where the first  $p$  rows of the inverted per-observation information matrix  $I^{-1}$  have been block-decomposed into an upper-left  $p \times p$  block  $(I^{-1})_{11}$  and a  $p \times 1$  block  $(I^{-1})_{12}$ . Then the last covariance-expression is also  $o(1/m)$ . A similar calculation gives the same  $o(1/m)$  result for the cross-term involving  $\hat{\gamma}_i - \gamma_i$  in  $MSE_i^D$ . Thus, apart from  $o(1/m)$  remainders,

$$MSE_i^D = (22) + E \left\{ x_i^{tr} (\hat{\beta} - \beta_0) + I_{[y_i \geq \kappa_i]} (\hat{\gamma}_i - \gamma_i) (u_i + e_i) \right\}$$

$$- I_{[y_i \geq \kappa_i]} \hat{\gamma}_i x_i^{tr} (\hat{\beta} - \beta_0) - I_{[y_i < \kappa_i]} \left( \hat{\gamma}_i \frac{\phi}{\Phi} \Big|_{\hat{\xi}_i} \sqrt{\hat{\tau}_i} - \gamma_i \frac{\phi}{\Phi} \Big|_{\xi_i} \sqrt{\tau_i} \right) \Big\}^2$$

After a further application of the delta method to the terms in the last expected square, we find

$$MSE_i^D = (21) + E \left\{ a_i x_i^{tr} (\hat{\beta} - \beta_0) + b_i (\hat{\gamma}_i - \gamma_i) \right\}^2 \quad (32)$$

where

$$a_i \equiv 1 - \gamma_i I_{[y_i \geq \kappa_i]} + \gamma_i I_{[y_i < \kappa_i]} \frac{\phi(\xi_i)}{\Phi(\xi_i)} \left( -\xi_i - \frac{\phi(\xi_i)}{\Phi(\xi_i)} \right)$$

$$b_i \equiv I_{[y_i \geq \kappa_i]} (u_i + e_i) - I_{[y_i < \kappa_i]} \frac{\sqrt{\tau_i} \phi(\xi_i)}{2(1 - \gamma_i) \Phi(\xi_i)} \left( 2 - \gamma_i + \left( \xi_i + \frac{\phi(\xi_i)}{\Phi(\xi_i)} \right) \xi_i \gamma_i \right)$$

Using the same representation of  $\hat{\beta} - \beta_0$  given above, we can recognize by ML theory that the asymptotic covariance matrix of  $(\hat{\beta} - \beta_0, \hat{\sigma}^2 - \sigma_0^2)$  is  $(mI)^{-1}$ . The identity  $\gamma_i = 1 - s_i/(s_i + \sigma^2)$ , together with the delta-method, then gives the asymptotic covariance matrix for  $(x_i^{tr}(\hat{\beta} - \beta_0), \hat{\gamma}_i - \gamma_i)$  as

$$M_i \equiv \begin{pmatrix} x_i^{tr} (mI)^{-1} {}_{11} x_i & \frac{(1-\gamma_i)^2}{s_i} x_i^{tr} (mI)^{-1} {}_{12} \\ \frac{(1-\gamma_i)^2}{s_i} x_i^{tr} (mI)^{-1} {}_{12} & \frac{(1-\gamma_i)^4}{s_i^2} (mI)^{-1} {}_{22} \end{pmatrix}$$

We proceed to substitute into formula (32). Since  $(a_i, b_i)$  is asymptotically independent of  $(\hat{\beta} - \beta_0, \hat{\sigma}^2 - \sigma_0^2)$  to top order, we find (still ignoring  $o(1/m)$  remainders)

$$MSE_i^D = (21) + tr \left( E \begin{pmatrix} a_i^2 & a_i b_i \\ a_i b_i & b_i^2 \end{pmatrix} M_i \right) \quad (33)$$

If we denote

$$C_i \equiv E \begin{pmatrix} a_i^2 & a_i b_i \\ a_i b_i & b_i^2 \end{pmatrix}$$

then

$$C_{i,11} = E(a_i^2) = (1 - \gamma_i)^2 (1 - \Phi(\xi_i)) + \Phi(\xi_i) \left\{ 1 - \gamma_i \frac{\phi(\xi_i)}{\Phi(\xi_i)} \left( \xi_i + \frac{\phi(\xi_i)}{\Phi(\xi_i)} \right) \right\}^2$$

$$C_{i,12} = C_{i,21} = E(a_i b_i) = \sqrt{\tau_i} \phi(\xi_i) (1 - \gamma_i) - \frac{\sqrt{\tau_i} \phi(\xi_i)}{2(1 - \gamma_i)}.$$

$$C_{i,22} = \tau_i (1 - \Phi(\xi_i) + \xi_i \phi(\xi_i)) + \frac{\tau_i \phi^2(\xi_i)}{4(1 - \gamma_i)^2 \Phi(\xi_i)} \left( 2 - \gamma_i + \left( \xi_i + \frac{\phi(\xi_i)}{\Phi(\xi_i)} \right) \xi_i \gamma_i \right) \left( 1 - \gamma_i \frac{\phi(\xi_i)}{\Phi(\xi_i)} \left( \xi_i + \frac{\phi(\xi_i)}{\Phi(\xi_i)} \right) \right)^2$$

Note that the formula for the per-observation  $(p+1) \times (p+1)$  information matrix  $I$  is the same as  $I(\beta_0, \sigma_0^2)$  in (19). Using that formula together with the formulas given above for the  $2 \times 2$  matrices  $M_i$  and  $C_i$ , we obtain  $MSE_i^D$  from formula (21) in the paper and formula (33) above.

**Acknowledgments.** The first author's research was conducted at the Census Bureau, Statistical Research Division. The second author's research was supported at the Census Bureau by an ASA/Census fellowship, and also by NSF grant No. SES-0318184. We thank Nancy Kirkendall of EIA for bringing the DOE left-censored surveys to our attention.

## 10 REFERENCES

- Buckley, J. and James, I. (1979), Linear regression with censored data, *Biometrika* **66**, 429-436.
- Citro, C. and Kalton, G., eds. (2000) **Small-Area Estimates of School-Age Children in Poverty: Evaluation of Current Methodology**, (National Research Council), Washington DC: Nat. Acad. Press.
- Datta, G. and Lahiri, P. (2001) A unified measure of uncertainty of estimated best linear unbiased predictors in small area estimation problems. *Statistica Sinica* **10**, 613-627.
- Fay, R. and Herriot, R. (1979) Estimates of income for small places: an application of James-Stein procedures to census data. *JASA* **74**, 341-53.
- Ghosh, M., and Rao, J.N.K. (1994) Small Area Estimation: an appraisal. *Statist. Sci.* **9**, 55-93.
- Klein, J. and Moeschberger, M. (2003) **Survival Analysis: Techniques for Censored and Truncated Data**, 2nd ed. New York: Springer.

- Koul, H., Susarla, V. and van Ryzin, J. (1981) Regression analysis with randomly right-censored data. *Ann. Statist.* **9**, 1276-88.
- Lahiri, P., and Rao, J.N.K. (1995) Robust estimation of mean squared error of small area estimators *Jour. Amer. Statist. Assoc.* **90**, 758-766.
- Nusser, S.M., and Goebel, J.J. (1997) The national resource inventory: a long-term multi-resource monitoring program. *Environmental and Ecological Statistics*, **4**, 181-204.
- Prasad, N. and Rao, J.N.K. (1990) The estimation of mean-squared errors of small-area estimators. *Jour. Amer. Statist. Assoc.* **85**, 163-171.
- R Development Core Team (2004). **R: A language and environment for statistical computing**. R Foundation for Statistical Computing, Vienna, Austria. <http://www.R-project.org>.
- Rao, J.N.K. (2003) **Small Area Estimation**. New York: John Wiley.
- Ritov, Y. (1990) Estimation in a linear regression model with censored data, *Ann. Statist.* **18**, 303-328.
- Slud, E. (2003) Assessing fit of SAIPE models to Census and CPS county child-poverty rates. Federal Committee on Statistical Methodology Research Conference Proceedings, Nov. 2003, Arlington, VA.
- Slud, E. (2004) Small area estimation errors in SAIPE using GLM versus FH models. Survey Research Methods Proc. Amer. Statist. Assoc, Joint Statistical Meetings, Toronto.
- Slud, E. and Maiti, T. (2004) MSE estimation in transformed Fay-Herriot models, with application to SAIPE. Preprint, submitted.
- Tsiatis, A. (1990), Estimating regression parameters using linear rank tests for censored data, *Ann. Statist.* **18**, 354-372.
- White, H. (1982) Maximum likelihood estimation of misspecified models. *Econometrica* **50**, 1-25.
- Ying, Z. (1993), A large sample study of rank estimation for censored regression data, *Ann. Statist.* **21**, 76-99.