Modeling Disability in Small Areas: An Area-Level Approach of Combining Two Surveys

Jiashen You*[1], Gauri S. Datta[1,2], and Jerry J. Maples[1]

[1]Center for Statistical Research and Methodology, U.S. Census Bureau
[2]Department of Statistics, University of Georgia

Abstract
Conventional small area estimation methods combine generalized linear model synthetic estimates made using covariates with direct survey estimates. Since "borrowing strength" from covariates to make quality synthetic estimates is a key motivation in small area modeling while almost all such information is collected through surveys, we recognize the need for building models that combine survey data and incorporate uncertainties in both surveys. In this study, we use the American Community Survey (ACS) to improve the disability estimates from the Survey of Income and Program Participation (SIPP). In particular, we discuss the estimation results from a bivariate Fay-Herriot model and a measurement error model as well as a comparison of estimated mean square errors.

Keywords: American Community Survey (ACS), bivariate Fay-Herriot, measurement error, mixed linear model, Survey of Income and Program Participation (SIPP)

1. Introduction

Since 1960, the National Center for Health Statistics (NCHS) has been conducting the National Health Interview Survey (NHIS), collecting information related to health topics. The first official release on nationwide disability statistics was given in NCHS (1968), where they pioneered using synthetic estimation to report the percentages of persons with activity limitation at the state level. Gonzalez (1973) points out that synthetic estimates are unbiased estimates obtained from a sample survey for a large area, under the assumption that the small areas to be estimated have the same characteristics as large areas. Even at the state level, many sample sizes were too small to provide reliable direct estimates in the 1968 release. One simple solution to avoid small area estimation altogether is to merge data collected over a spatial or temporal domain.

For many federal agencies, such challenges often arise when requests are made for them to release statistics at a level below which the primary statistical inference was targeted in the survey design (see for example, Schaible (1992)). Realizing that such issues would remain even in many nationwide surveys, the community in small area estimation research has since flourished with numerous new methods. Since indirect domain estimation techniques such as synthetic estimators cannot account for area-specific variation, explicit model-based methods have gained popularity in recent years. Ghosh and Rao (1994) provides a comprehensive review on practical examples as well as major breakthroughs, including the widely-used Fay-Herriot (FH) model by Fay and Herriot (1979). Rao (2003) notes in his book that the success of any model-based methods depends on the availability of good auxiliary data as one of the key ideas is to "borrow strength" from relatable and reliably measured covariates. However, when auxiliary variables are measured with error, one needs to exercise caution with the weights between the model-predicted synthetic values and direct estimates.

In this paper, we propose two small area estimation models that both aim at improving the SIPP disability estimates with ACS data. There have been a few dozen recent national, federally-sponsored surveys related to this topic (see Livermore et al. (2011)), partially

*Corresponding author: jiashen.you@census.gov, U.S. Census Bureau, Washington, D.C. 20233

due to the enactment of Americans with Disabilities Act (ADA) by U.S. Congress in 1990 that has prompted policymakers besides academic researchers in public health policy to search for more reliable statistics on disability. The linkability of SIPP to administrative data such as Internal Revenue Service (IRS) information makes it an appealing alternative to other similar national surveys to obtain estimates on disability statistics without having to combine several years of survey data in small areas to achieve desired reliability. We also study the effectiveness of using the American Community Survey (ACS), another survey conducted by the U.S. Census Bureau and the largest ongoing household survey of its kind to improve the model estimates from SIPP.

The remainder of this paper is structured as follows. In Section 2, we review the concept of empirical best linear unbiased predictor (EBLUP) and discuss two proposed models. In Section 3, we provide a numerical method for estimating the mean squared error for model predictions in both models. A simulation study is presented in Section 4. Details of parametrization for modeling disability data and selected small area SIPP disability estimates using our proposed models are included in Section 5. Some remarks and discussion on current and future work related to these models can be found in the last section.

2. Two Models

Let θ_i denote the population characteristic of interest in small area i whose observed value is Y_i and assume that $\theta_i = \boldsymbol{x}_i'\boldsymbol{\beta} + v_i$ where $\boldsymbol{x}_i$ are area-specific covariates of length p. The Fay and Herriot (1979) model can be expressed as a generalized linear mixed-effects model:

$$Y_i = \theta_i + e_i = \boldsymbol{x}_i'\boldsymbol{\beta} + v_i + e_i, \quad i = 1, \cdots, m, \tag{1}$$

where the model errors v_i and sampling errors e_i are assumed independent. It is further assumed that $v_i \sim N(0, \sigma_v^2)$ and $e_i \sim N(0, D_i)$, with known sampling variance D_i.

If the parameters $\boldsymbol{\beta}$ and σ_v^2 are both known, the best predictor (BP) of θ_i is

$$\tilde{\theta}_i = w_i Y_i + (1 - w_i)\boldsymbol{x}_i'\boldsymbol{\beta}, \tag{2}$$

where $w_i = \sigma_v^2/(\sigma_v^2 + D_i)$. This follows from minimizing the mean squared error (MSE) of the prediction in the form of a linear combination of Y_i and $\boldsymbol{x}_i'\beta$ where weights add up to one, i.e. $E(\alpha_i Y_i + (1 - \alpha_i)\boldsymbol{x}_i'\boldsymbol{\beta} - \theta_i)^2$.

In reality, $\boldsymbol{\beta}$ and σ_v^2 are unknown and must be estimated from the data. Fay and Herriot (1979) proposed to estimate $\boldsymbol{\beta}$ with its maximum likelihood estimator (MLE) $\tilde{\boldsymbol{\beta}} = (\boldsymbol{X}'\Sigma^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}'\Sigma^{-1}\boldsymbol{Y}$ and to estimate $\sigma_v^2 \geq 0$ by solving the estimating equation:

$$\sum_{i=1}^{m} \frac{(Y_i - \boldsymbol{x}_i'\tilde{\boldsymbol{\beta}})^2}{\sigma_v^2 + D_i} = m - p, \tag{3}$$

where $\boldsymbol{X} = (\boldsymbol{x}_1, \cdots, \boldsymbol{x}_m)'$, $\boldsymbol{Y} = (Y_1, \cdots, Y_m)'$ and $\Sigma = diag(D_i + \sigma_v^2)$.

The solution to Equation (3) gives rise to the empirical best linear unbiased predictor (EBLUP)

$$\hat{\theta}_{iFH} = \hat{w}_i Y_i + (1 - \hat{w}_i)\boldsymbol{x}_i'\tilde{\boldsymbol{\beta}}, \quad \hat{w}_i = \hat{\sigma}_v^2/(\hat{\sigma}_v^2 + D_i).$$

It is worth noting that Equation (1) assumes exact measurement on auxiliary information $\boldsymbol{x}_i$. This is unlikely to hold in many applications. We now propose two models that focus on improving the empirical best predictor by accounting for the variability in the measurement of $\boldsymbol{X}$.

2.1 Bivariate Fay-Herriot (BiFH) Model

An intuitive remedy to this problem is a bivariate Fay-Herriot (BiFH) model. We consider the case where auxiliary information is available in the form of survey data that is measured with uncertainty. Rather than treating it as a covariate $\boldsymbol{X}$, we may bring it to the other side of the formula as observed values for a related variable.

Let $\boldsymbol{Y}_i = (Y_{1i}, Y_{2i})'$ be a pair of survey estimates for area $i = 1, \cdots, m$. At the area level, the first layer of the model assumes $\boldsymbol{Y}_i = \boldsymbol{\theta}_i + \boldsymbol{e}_i$, where $\boldsymbol{e}_i \sim N((0,0)', \boldsymbol{\Sigma}_{\boldsymbol{e}_i})$. The covariance matrix $\boldsymbol{\Sigma}_{\boldsymbol{e}_i}$ is assumed to be known and its diagonal values represent sampling variances in corresponding surveys. In principle, Y_{1i} and Y_{2i} do not have to be estimates of the same population quantity, but they should be survey estimates of population quantities that are correlated to each other.

The mixed-effects model is specified by $\theta_{\xi i} = \boldsymbol{x}'_{\xi i}\boldsymbol{\beta}_\xi + v_{\xi i}$, $\xi = 1, 2$ in two levels:

$$\boldsymbol{\theta}_i = \begin{bmatrix} \theta_{1i} \\ \theta_{2i} \end{bmatrix} = \begin{pmatrix} \sum_{l=1}^{p} x_{1il}\beta_{1l} \\ \sum_{l=1}^{q} x_{2il}\beta_{2l} \end{pmatrix} + \boldsymbol{v}_i,$$

and

$$\boldsymbol{v}_i \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \boldsymbol{\Sigma}_{\boldsymbol{v}} = \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{bmatrix} \right).$$

We assume that the error terms in the level-1 models are independent of those in the level-2 model just as in the univariate FH model.

If we further assume that the two surveys are independent (and therefore, $\boldsymbol{\Sigma}_{\boldsymbol{e}_i} = diag(D_{1i}, D_{2i})$), we can write the joint distribution as:

$$\begin{bmatrix} Y_{1i} \\ Y_{2i} \\ \theta_{1i} \\ \theta_{2i} \end{bmatrix} \sim N\left(\begin{bmatrix} \boldsymbol{x}'_{1i}\boldsymbol{\beta}_1 \\ \boldsymbol{x}'_{2i}\boldsymbol{\beta}_2 \\ \boldsymbol{x}'_{1i}\boldsymbol{\beta}_1 \\ \boldsymbol{x}'_{2i}\boldsymbol{\beta}_2 \end{bmatrix}, \left[\begin{array}{c|c} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \hline \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{array} \right] = \left[\begin{array}{cc|cc} D_{1i} + \sigma_{11} & \sigma_{12} & \sigma_{11} & \sigma_{12} \\ \sigma_{12} & D_{2i} + \sigma_{22} & \sigma_{12} & \sigma_{22} \\ \hline \sigma_{11} & \sigma_{12} & \sigma_{11} & \sigma_{12} \\ \sigma_{12} & \sigma_{22} & \sigma_{12} & \sigma_{22} \end{array} \right] \right).$$

If one is only interested in θ_{1i} for each small area, the BLUP is obtained by taking the conditional mean (see Appendix A for derivation):

$$\tilde{\theta}_{1i} = E(\theta_{1i}|\boldsymbol{Y}_i) = Y_{1i} - \frac{D_{1i}[(D_{2i} + \sigma_{22})(Y_{1i} - \boldsymbol{x}'_{1i}\boldsymbol{\beta}_1) - \sigma_{12}(Y_{2i} - \boldsymbol{x}'_{2i}\boldsymbol{\beta}_2)]}{(D_{1i} + \sigma_{11})(D_{2i} + \sigma_{22}) - \sigma_{12}^2}.$$

2.2 Best Predictor in a Measurement Error Model

An alternative to the BiFH model where measurement error on auxiliary information is taken into consideration was studied by Ybarra and Lohr (2008). While the magnitude of variability of measurement error is assumed known, the corresponding term is not modeled as a joint population variable on the left-hand side of the equation, but remains as a covariate on the right. The resulting predictor pulls more weight toward the direct survey estimates where large measurement errors are found. Although this can be considered as a generalization of the FH model, it tends to produce predictions with less precision. It has been shown when the measurement error (ME) model in Ybarra and Lohr (2008) is applied to administrative data where very little sampling errors are observed, its EBLUPs may have larger empirical mean squared errors than those from the FH model. This is not surprising as one additional source of variation has been incorporated in the model. However, one may argue that the resulting estimates are more honest predictions of the uncertainty.

Separately, Jiang et al. (2011) proposed a new approach, called observed best prediction (OBP) that emphasizes minimizing the total mean squared prediction error (MSPE). Knowing that the best predictor assumes the form in Equation (2), the authors conjectured that model estimates are more influential in areas where sampling variances are large. Rather than using the standard EBLUP, their approach compensates by assigning larger weights to those areas where D_i's are large in estimating the regression coefficients. The corresponding $\hat{\boldsymbol{\beta}}$ is no longer the best linear unbiased estimator (BLUE), but the minimizer of $E(|\tilde{\boldsymbol{\theta}} - \boldsymbol{\theta}|^2) = \sum_{i=1}^{m} E(\tilde{\theta}_i(\boldsymbol{\beta}, \sigma_v^2) - \theta_i)^2$, where $\boldsymbol{\theta} = (\theta_i)_{1 \le i \le m}$ is the vector of all small area means and $\tilde{\boldsymbol{\theta}} = (\tilde{\theta}_i)_{1 \le i \le m}$ its best predictor as in Equation (2). The resulting estimator is different from $\hat{\theta}_{iFH}$.

The measurement error model does not emphasize the minimization of MSPE like the OBP approach does. Since OBP ignores variability in measuring auxiliary information, we now extend their work by combining the characteristics from both methods.

In addition to the usual univariate FH model assumptions, we further assume that a single covariate x_i may be observed with error:

$$X_i = x_i + f_i, \quad f_i \sim N(0, C_i),$$

where $x_1, , \cdots, x_m$ are unknown true covariate values which are measured as $X_1, \cdots, X_m$ in a survey. All error terms e_i, v_i and f_i are assumed to be mutually independent.

Given observed data $(Y_i, X_i)_{i=1,\cdots,m}$ and assuming that $(D_i, C_i)_{i=1,\cdots,m}$ are known or reliably estimated from survey data, we may write the joint density as

$$f(\boldsymbol{Y}, \boldsymbol{\theta}, \boldsymbol{X}|\boldsymbol{x}, \beta_0, \beta_1, \sigma_v^2) \propto \prod_{i=1}^{m} \frac{1}{\sqrt{\sigma_v^2}} \exp\Big\{ -\frac{(Y_i-\theta_i)^2}{2D_i} - \frac{(\theta_i - \beta_0 - \beta_1 x_i)^2}{2\sigma_v^2} - \frac{(x_i - X_i)^2}{2C_i} \Big\}.$$

Under normality, we would treat the unobserved x_i as random variables, namely, $x_i|X_i \sim N(X_i, C_i)$. Assuming a uniform prior on x_i, we can integrate it out in the likelihood function and quickly verify that $E[\theta_i|X_i, \beta_0, \beta_1, \sigma_v^2] = \beta_0 + \beta_1 X_i$ and $Var[\theta_i|X_i, \beta_0, \beta_1, \sigma_v^2] = \beta_1^2 C_i + \sigma_v^2$. This allows us to rewrite the model as:

$$\begin{cases} Y_i|\theta_i \sim N(\theta_i, D_i), \\ \theta_i|X_i, \beta_0, \beta_1, \sigma_v^2 \sim N(\beta_0 + \beta_1 X_i, \sigma_v^2 + \beta_1^2 C_i). \end{cases}$$

As in the Fay-Herriot model, assuming the parameter values for β_0, β_1 and σ_v^2 are known, we can derive the best predictor of θ_i:

$$\tilde{\theta}_{i,BP} = Y_i - \frac{D_i}{D_i + (\sigma_v^2 + \beta_1^2 C_i)}(Y_i - \beta_0 - \beta_1 X_i). \tag{4}$$

Let $\tilde{\boldsymbol{\theta}}_{BP} = (\tilde{\theta}_{1,BP}, \cdots, \tilde{\theta}_{m,BP})'$ and write

$$\boldsymbol{\Gamma} = diag\left(\frac{D_1}{D_1 + \sigma_v^2 + \beta_1^2 C_1}, \cdots, \frac{D_m}{D_m + \sigma_v^2 + \beta_1^2 C_m} \right).$$

We now apply the OBP approach to estimate the parameters β_0, β_1 and σ_v^2 by minimizing $E[(\tilde{\boldsymbol{\theta}}_{BP} - \boldsymbol{\theta})'(\tilde{\boldsymbol{\theta}}_{BP} - \boldsymbol{\theta})]$. This is equivalent to estimating $(\hat{\boldsymbol{\beta}}; \hat{\sigma}_v^2) = \arg\min Q$, where

$$Q(\boldsymbol{\beta}; \sigma_v^2) = \sum_{i=1}^{m} \frac{(Y_i - \beta_0 - \beta_1 X_i)^2 D_i^2}{(D_i + \sigma_v^2 + \beta_1^2 C_i)^2} - 2\sum_{i=1}^{m} \frac{D_i^2}{D_i + \sigma_v^2 + \beta_1^2 C_i}. \tag{5}$$

Details of the derivation of the objective function $Q(\boldsymbol{\beta}; \sigma_v^2)$ are given in Appendix C. The alternative model predictor is obtained by placing the resulting parameters into Equation (4), or $\hat{\theta}_{iOM} = \tilde{\theta}_{i,BP}(\hat{\boldsymbol{\beta}}; \hat{\sigma}_v^2)$.

3. Mean Squared Error Estimation

There are several options when it comes to estimating the mean squared error for small area model predictions. The most widely used analytic approach was first introduced in Prasad and Rao (1990). Their approach for the FH model is based on a Taylor series expansion of the true mean squared error. This was later extended in Lahiri and Rao (1995) and Datta and Lahiri (2000). Datta et al. (2005) derived an unbiased estimator for MSE of order $o(1/m)$. Just as the positive solution to Equation (3) does not always exist, the Prasad-Rao moment estimator is not guaranteed to be positive. A few computational approaches have been introduced around the same time such as the Jackknife method in Jiang et al. (2002) and Ybarra and Lohr (2008). Jiang et al. (2011) proposed their own analytic MSE estimator with numerical thresholds for the OBP.

We adopt a double bootstrap algorithm for approximating MSE in small area estimation, first introduced in Hall and Maiti (2006) and later extended in Pfeffermann and Correa (2012).The algorithm to estimate the MSE for the BiFH model is given below while the double bootstrap algorithms for other univariate models should follow from simplification:

1. Given $(\boldsymbol{Y}_1, \cdots, \boldsymbol{Y}_m)$, estimate $\hat{\boldsymbol{\beta}}$ and $\hat{\boldsymbol{\Sigma}}_{\boldsymbol{v}}$ and compute the EBLUP $\hat{\boldsymbol{\theta}}_i$.
2. Use $\hat{\boldsymbol{\beta}}$ and $\hat{\boldsymbol{\Sigma}}_{\boldsymbol{v}}$ to generate for $b_1 = 1, \cdots, B_1$:
$$\boldsymbol{Y}_i^{(b_1)} = \boldsymbol{\theta}_i^{(b_1)} + \boldsymbol{e}_i^{(b_1)}, \quad \boldsymbol{e}_i^{(b_1)} \sim N(0, \boldsymbol{\Sigma}_{\boldsymbol{e}_i}),$$
where
$$\boldsymbol{\theta}_i^{(b_1)} = \boldsymbol{X}_i\hat{\boldsymbol{\beta}} + \boldsymbol{v}_i^{(b_1)}, \quad \boldsymbol{v}_i^{(b_1)} \sim N(0, \hat{\boldsymbol{\Sigma}}_{\boldsymbol{v}}).$$
Estimate $\hat{\boldsymbol{\beta}}^{(b_1)}$ and $\hat{\boldsymbol{\Sigma}}_{\boldsymbol{v}}^{(b_1)}$. Compute
$$\hat{\boldsymbol{\theta}}_i^{EBLUP,b_1} = \boldsymbol{Y}_i^{(b_1)} - \boldsymbol{\Sigma}_{\boldsymbol{e}_i}(\hat{\boldsymbol{\Sigma}}_{\boldsymbol{v}}^{(b_1)} + \boldsymbol{\Sigma}_{\boldsymbol{e}_i})^{-1}(\boldsymbol{Y}_i^{(b_1)} - \boldsymbol{X}_i\tilde{\boldsymbol{\beta}}(\hat{\boldsymbol{\Sigma}}_{\boldsymbol{v}}^{(b_1)})).$$
3. Corresponding to each set of bootstrap data $\{(\boldsymbol{Y}_i^{(b_1)}, \boldsymbol{X}_i, \boldsymbol{\Sigma}_{\boldsymbol{e}_i})_{i=1}^m; b_1 = 1, \cdots, B_1\}$ and each pair $(\hat{\boldsymbol{\beta}}^{(b_1)}, \hat{\boldsymbol{\Sigma}}_{\boldsymbol{v}}^{(b_1)})$, generate for $b_2 = 1, \cdots, B_2$:
$$\boldsymbol{\theta}_i^{b_2(b_1)} = \boldsymbol{X}_i\hat{\boldsymbol{\beta}}^{(b_1)} + \boldsymbol{v}_i^{b_2(b_1)}, \quad \boldsymbol{v}_i^{b_2(b_1)} \sim N(0, \hat{\boldsymbol{\Sigma}}_{\boldsymbol{v}}^{(b_1)}),$$
and
$$\boldsymbol{Y}_i^{b_2(b_1)} = \boldsymbol{\theta}_i^{b_2(b_1)} + \boldsymbol{e}_i^{b_2(b_1)}, \quad \boldsymbol{e}_i^{b_2(b_1)} \sim N(0, \boldsymbol{\Sigma}_{\boldsymbol{e}_i}).$$
Estimate $\hat{\boldsymbol{\beta}}^{b_2(b_1)}$ and $\hat{\boldsymbol{\Sigma}}_{\boldsymbol{v}}^{b_2(b_1)}$ as well as compute
$$\hat{\boldsymbol{\theta}}_i^{EBLUP,b_2(b_1)} = \boldsymbol{Y}_i^{b_2(b_1)} - \boldsymbol{\Sigma}_{\boldsymbol{e}_i}(\hat{\boldsymbol{\Sigma}}_{\boldsymbol{v}}^{b_2(b_1)} + \boldsymbol{\Sigma}_{\boldsymbol{e}_i})^{-1}(\boldsymbol{Y}_i^{b_2(b_1)} - \boldsymbol{X}_i\tilde{\boldsymbol{\beta}}(\hat{\boldsymbol{\Sigma}}_{\boldsymbol{v}}^{b_2(b_1)})).$$
4. Compute
$$\boldsymbol{s}_{1i} = \frac{1}{B_1}\sum_{b_1=1}^{B_1}\{\hat{\boldsymbol{\theta}}_i^{EBLUP,b_1} - \boldsymbol{\theta}_i^{(b_1)}\}\{\hat{\boldsymbol{\theta}}_i^{EBLUP,b_1} - \boldsymbol{\theta}_i^{(b_1)}\}',$$
and
$$\boldsymbol{s}_{2i} = \frac{1}{B_1}\sum_{b_1=1}^{B_1}\frac{1}{B_2}\sum_{b_2=1}^{B_2}\{\hat{\boldsymbol{\theta}}_i^{EBLUP,b_2(b_1)} - \boldsymbol{\theta}_i^{b_2(b_1)}\}\{\hat{\boldsymbol{\theta}}_i^{EBLUP,b_2(b_1)} - \boldsymbol{\theta}_i^{b_2(b_1)}\}'.$$
Return
$$\widetilde{MSE}(\hat{\boldsymbol{\theta}}_i) = 2s_{1i} - s_{2i}.$$

In Hall and Maiti (2006), they explored different estimators of $\widetilde{MSE}(\hat{\boldsymbol{\theta}}_i) = g(s_{1i}, s_{2i})$ in terms of different choices of g for the FH model. We show that $g(s_{1i}, s_{2i}) = 2s_{1i} - s_{2i}$ is in general an unbiased estimator for MSE with order $o(m^{-1})$ in Appendix D.

4. Simulation Study

We examine simulated results using the factorial design from Ybarra and Lohr (2008) to compare the empirical mean squared errors (EMSE) of several estimators. In this example, true values of an auxiliary variable x_i are generated from a $N(5, 9)$ distribution and D_i from a gamma distribution with shape parameter 5 and scale parameter 2. For each iteration, v_i, e_i and f_i are independently generated normal variates with mean 0 and respective variance $\sigma_v^2 \in \{2, 4\}$, D_i as stated and $C_i \in \{0, c\}$ for $c = 2, 3$ or 4. Then we compute $\theta_i = 1 + 3x_i + v_i, Y_i = \theta_i + e_i$ and $X_i = x_i + f_i$ for $m \in \{20, 50, 100\}$.

The results from the simulation with $m = 50$, $\sigma_v^2 = 4$ when $k\%$ of the covariates x_i's is randomly chosen to be observed with $N(0, 3)$ errors and the remaining x_i's are observed without error, for $k \in \{20, 50, 100\}$ are shown in Table 1. The estimators that are compared side by side include:

(a) the direct estimator Y_i;

(b) the univariate FH estimator using the true covariates x_i, or $\hat{\theta}_{iF_t}$;

(c) the regular univariate FH estimator $\hat{\theta}_{iF}$ using the observed values X_i (ignoring measurement error);

(d) the OBP estimator $\hat{\theta}_{iO}$ as in Jiang et al. (2011);

(e) the ME estimator $\hat{\theta}_{iM}$ as in Ybarra and Lohr (2008);

(f) the OBPME estimator $\hat{\theta}_{iOM}$;

(g) the BiFH estimator $\hat{\theta}_{iBI}$.

$N = 1000$ repetitions, $m = 50$ small areas
$\beta_0 = 1, \beta_1 = 3, \sigma_v^2 = 4$

k	C_i	Y_i	$\hat{\theta}_{iF_t}$	$\hat{\theta}_{iF}$	$\hat{\theta}_{iO}$	$\hat{\theta}_{iM}$	$\hat{\theta}_{iOM}$	$\hat{\theta}_{iBI}$
20	0	9.97	3.17	3.65	3.70	3.52	**3.41**	3.58
	3	9.66	3.11	9.78	10.14	7.15	7.14	**7.04**
50	0	10.16	3.22	4.67	4.66	3.79	**3.72**	3.97
	3	10.07	3.22	8.04	8.26	7.46	7.43	**7.38**
100	3	10.01	3.21	7.44	7.57	7.53	7.50	**7.38**

Table 1: Simulation study on empirical mean squared errors. Within each iterations, $k\%$ of the $m = 50$ areas have $C_i = 3$ and the remaining areas have $C_i = 0$. Parameters are always estimated together, but the empirical mean squared errors are averaged separately for the areas with $C_i = 3$ and those with $C_i = 0$.

Based on the simulation setup, the estimator $\hat{\theta}_{iF_t}$ in the table above can be treated as the optimal estimator. The direct survey estimates Y_i can be seen as the worst predictor partaking no modeling effort as the EMSE is about the size of the sampling variance on average. In this comparison of the quality of estimators, we strive to obtain an EMSE that would come closest to the EMSE of $\hat{\theta}_{iF_t}$, or, simply the smallest overall EMSE. When only a portion of the covariates is measured with error, we separate the estimated values from each model and compute the individual EMSEs within both groups.

The smallest EMSE values are consistently found in the case of OBPME and BiFH estimators. The OBPME estimator almost always outperforms both the ME estimator and the OBP by itself. This is important to us as we may attribute part of the pairwise difference as improvement upon the parent model by better handling of the data. For instance, the reduction in EMSE from $\hat{\theta}_{iO}$ may be partially caused by the fact that sampling error C_i is now incorporated in the estimator $\hat{\theta}_{iOM}$. It is also observed that when either none or 100% of the covariates is measured with error, univariate FH model performs quite well. There are even a few cases where $\hat{\theta}_{iF}$ delivers the smallest EMSE when all of C_i is 0. This is not surprising as the simulated data are built using a univariate modeling error structure and there is some small difference in numerically computing the parameters for the BiFH model versus the univariate model. In practice, only certain administrative records are treated as measured with no sampling error. Therefore, the case where all C_i's are 0 offers little insight on how to effectively incorporate different survey estimates.

In general, if the measurement error variances are all the same, then ignoring the measurement error term and applying a regular univariate FH model would work quite well since the model error term is still homoscedastic (of the size $\sigma_v^2 + \beta_1^2 C$). It is when there is variable measurement error variances that the ME, OBPME and BiFH estimator can better incorporate that information. Overall, these simulation results show a robust performance of both BiFH and OBPME estimators in terms of reducing area-specific EMSE as compared to other univariate empirical BPs.

5. Results

5.1 Disability Estimates: Data and Task

We now proceed to apply the best performing estimators to build pilot models for disability data. Our main data source comes from the disability estimates collected through the Survey of Income and Program Participation (SIPP) by the U.S. Census Bureau. The SIPP 2008 panel wave 6 topical module asked individuals aged 15 or older questions regarding adult well-being, child support agreements, support for non-household members, functional limitations and disability for adults and children as well as employer-provided health benefits. A detailed 62-item list of questions[1] pertaining to verbal, visual, physical, emotional and mental issues any person in a respondent's household may have can be found in the functional limitations and disability section.

It is worth mentioning that wave 6 of the SIPP 2008 data was collected between May and August of 2010. This is because the SIPP survey design uses a 4-month recall period, with approximately the same number of interviews being conducted in each wave of the 4-month period. The 2008 SIPP Panel started in September of 2008 centering around a "core" of labor force, program participation, and income questions while questions labeled "topical modules" are assigned to particular interviewing waves of the survey.

Our alternative data is from the American Community Survey (ACS), which is an ongoing survey sent to roughly 3.5 million housing unit addresses each year by the Census Bureau. ACS has become one of the most reliable sources for household information for current U.S. population ever since the phasing out of long forms in the decennial census in 2000. Due to its immense sample size and the fact that interviews are conducted in every county, it has become a valuable source of household data.The ACS 2010 contains 6 questions[2] related to the disability status for any person aged 5 or older in a respondent's household. Though the ACS is not meant to be a specialized health or disability survey, the majority of those questions regarding disability could be identified as similar in nature to the ones in SIPP's topical module. However, there is notable difference in the wording of survey questions. For instance, Question 17b of the personal information sheet on ACS 2010 asks whether or not the person is "blind or has serious difficulty seeing even when wearing glasses". This, in turn, can be associated with Question ADQ4 of the Functional Limitations and Disability section of SIPP 2008 wave 6 which asks if the person "has difficulty seeing the words and letters in ordinary newspaper print even when wearing glasses or contact lenses if s/he usually wear(s) them" and subsequently offers respondents three choices – "Yes", "No", and "Person is blind". SIPP also follows up with question chains such as its ADQ5 asking whether the person is "able to see the words and letters in ordinary newspaper print at all". On the other hand, the relatively small sample size in SIPP results in more direct estimates of zero that are unlikely to reflect the truth.

We choose to aggregate the SIPP disability questions into four categories: vision, hearing, mental and physical aspects of functional limitations. This allows us to identify with related ACS questions and couple the data sets in the bivariate model. "Total disability" is thus defined as having a "Yes" in any of the four main categories.

We choose to search among administrative records for covariates. In particular, the percentages of social security income (SSI) and disability income (DI) recipients are used. No sampling error is assumed for these administrative records.

5.2 Parametrization for BiFH Model

Let $\boldsymbol{Y}_{ACS} = \boldsymbol{Y}_A = (y_{11}, \cdots, y_{1m})'$, $\boldsymbol{Y}_{SIPP} = \boldsymbol{Y}_S = (y_{21}, \cdots, y_{2m})'$ and $\boldsymbol{Y} = (\boldsymbol{Y}'_A, \boldsymbol{Y}'_S)'$ be the estimated disability rates from two surveys. Exploratory analysis shows that at the "state" level (all 50 U.S. states and the District of Columbia), the correlation coefficients are

[1] http://www.census.gov/content/dam/Census/programs-surveys/sipp/questionnaires/2008/SIPP 2008 Panel Wave 06 - Topical Module Questionnaire.pdf

[2] http://www.census.gov/acs/www/Downloads/questionnaires/2010/Quest10.pdf

	$\boldsymbol{Y}_S$	pct_SSI	pct_DI
$\boldsymbol{Y}_A$	0.5906	0.6348	0.8950
$\boldsymbol{Y}_S$		0.1847	0.4317
pct_SSI			0.6806

Table 2: Correlation coefficients between $\boldsymbol{Y}_A, \boldsymbol{Y}_S$ and covariates

Let $\boldsymbol{x}_{\xi i} = (x_{\xi i1}, \cdots, x_{\xi ip})'$ denote the set of observed p covariates for small area i and $\xi \in \{A, S\}$. Since we use the same covariates for SIPP and ACS data, $\boldsymbol{X}_A = \boldsymbol{X}_S$ are both $m \times p$ matrices of covariates, and the final design matrix can be structured as

$$\boldsymbol{X} = \begin{pmatrix} \boldsymbol{X}_A & 0 \\ 0 & \boldsymbol{X}_S \end{pmatrix}.$$

Finally, let $\boldsymbol{\beta} = (\boldsymbol{\beta}'_A, \boldsymbol{\beta}'_S)'$ be the vector of regression coefficients of length $2p$. We can now simply express the regression model as

$$\boldsymbol{Y} = \begin{bmatrix} \boldsymbol{Y}_A \\ \boldsymbol{Y}_S \end{bmatrix} = \begin{bmatrix} \boldsymbol{X}_A & 0 \\ 0 & \boldsymbol{X}_S \end{bmatrix} \begin{bmatrix} \boldsymbol{\beta}_A \\ \boldsymbol{\beta}_S \end{bmatrix} + \boldsymbol{e} + \boldsymbol{v}$$

from which we can quickly derive the BLUE for regression coefficients as

$$\hat{\boldsymbol{\beta}} = (\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{Y}. \tag{6}$$

Here, $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}(\sigma_{AA}, \sigma_{SS}, \sigma_{AS})$ is a $2m \times 2m$ block diagonal covariance matrix parametrized by

$$\boldsymbol{\Sigma} = \left[\begin{array}{c|c} \boldsymbol{\Sigma}_A & \sigma_{AS}\boldsymbol{I} \\ \hline \sigma_{AS}\boldsymbol{I} & \boldsymbol{\Sigma}_S \end{array}\right], \tag{7}$$

where $\boldsymbol{\Sigma}_\xi = diag(D_{\xi 1} + \sigma_{\xi\xi}, \cdots, D_{\xi m} + \sigma_{\xi\xi})$.

We define projection matrices $\boldsymbol{P}_\xi = \boldsymbol{X}_\xi(\boldsymbol{X}'_\xi\hat{\boldsymbol{\Sigma}}_\xi^{-1}\boldsymbol{X}_\xi)^{-1}\boldsymbol{X}'_\xi\hat{\boldsymbol{\Sigma}}_\xi^{-1}$. Let $\hat{\boldsymbol{e}} = \boldsymbol{Y} - \hat{\boldsymbol{Y}} = (\hat{\boldsymbol{e}}'_A, \hat{\boldsymbol{e}}'_S)'$ where $\hat{\boldsymbol{e}}_\xi = \boldsymbol{Y}_\xi - \boldsymbol{P}_\xi\boldsymbol{Y}_\xi$. In order to guarantee the existence of a solution and a successful updates of the covariance term, we propose to estimate σ_{AA} and σ_{SS} separately prior to σ_{AS}. Justification for this method can be found in Appendix B. If $\sigma_{AA}\sigma_{SS} \neq 0$, we estimate σ_{AS} by

$$\hat{\sigma}_{AS} = \frac{\{\hat{\boldsymbol{\Sigma}}_A^{-1/2}\hat{\boldsymbol{e}}_A\}'\{\hat{\boldsymbol{\Sigma}}_S^{-1/2}\hat{\boldsymbol{e}}_S\}}{tr[\hat{\boldsymbol{\Sigma}}_A^{-1/2}\hat{\boldsymbol{\Sigma}}_S^{-1/2}(\boldsymbol{I} - \boldsymbol{P}_A)(\boldsymbol{I} - \boldsymbol{P}_S)']}. \tag{8}$$

However, the value from the above formula must be checked to ensure the validity of a corresponding correlation coefficient.

In practice, we may initialize $\sigma_{AA} = \sigma_{SS} = \sigma_{AS} = 0$. The following algorithm is implemented within each iteration until convergence:

1. Set up $\hat{\boldsymbol{\Sigma}}_\xi^{(i)}$ and $\hat{\boldsymbol{\Sigma}}^{(i)}$ as in Equation (7).
2. Compute $\hat{\boldsymbol{\beta}}^{(i+1)}$ as in Equation (6).
3. Find unique solutions $\sigma_{AA}^{(i+1)}, \sigma_{SS}^{(i+1)}$ as in Equation (9) in Appendix B.
4. Compute and check the validity of $\sigma_{AS}^{(i+1)}$. This step involves checking the Cauchy-Schwarz inequality:
   - If $\sigma_{AA}^{(i+1)}\sigma_{SS}^{(i+1)} = 0$, set $\sigma_{AS}^{(i+1)} = 0$.
   - If $[\sigma_{AS}^{(i+1)}]^2 > \sigma_{AA}^{(i+1)}\sigma_{SS}^{(i+1)}$, set $\sigma_{AS}^{(i+1)} = sign(\sigma_{AS}^{(i+1)})\sqrt{\sigma_{AA}^{(i+1)}\sigma_{SS}^{(i+1)}}$, where $sign(x) = 1_{(x>0)} - 1_{(x<0)}$.

5.3 Estimating Disability in SIPP

Following the results from the simulation study, we consider the two proposed models along with the univariate FH model to estimate the disability rates from the data. Unlike the simulated example where true values of the unknown parameters are known, we rely on the implicit assumption that direct survey estimates are unbiased regardless of the corresponding sample sizes. This, in turn, translates to having faith in the execution of sampling design and the expectation of only reasonable deviation of model estimates away from the survey estimates. A glance of estimated rates of previously defined "total" diability for all 50 states and the District of Columbia is shown in Figure 1, sorted by their SIPP sample sizes on the left margin. All three model estimates are displayed on a naive 90% confidence interval constructed using the corresponding direct estimate and survey variance for that area.

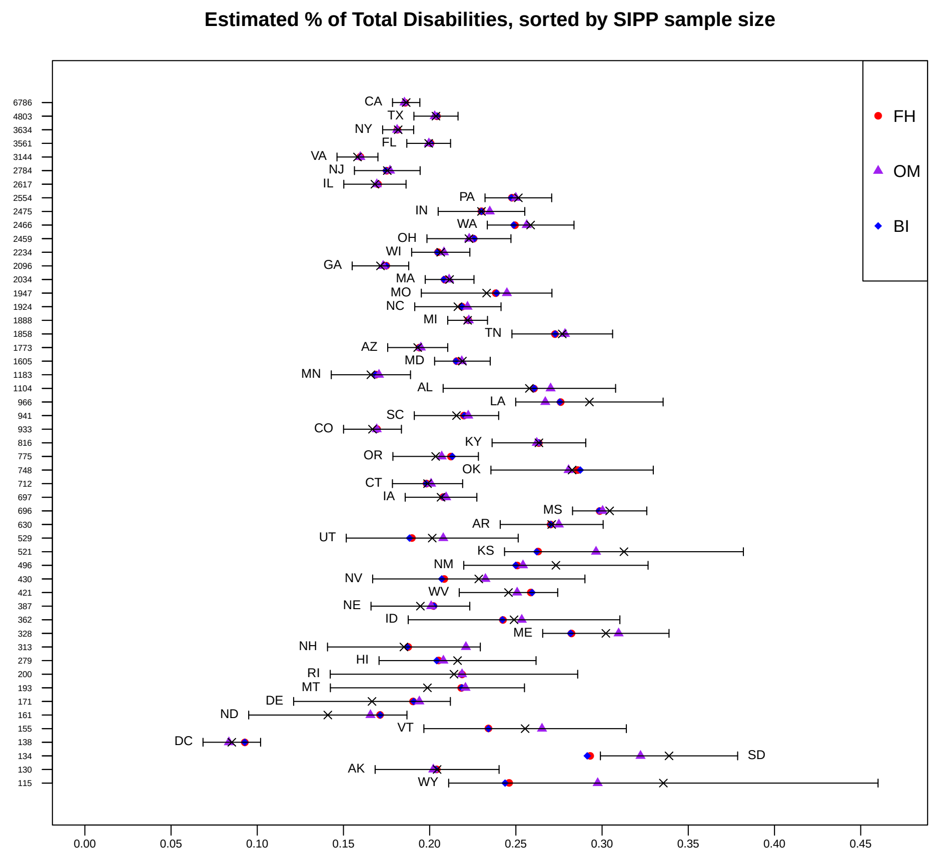

Figure 1: Comparison of FH, OM, and BiFH model estimates of state-level total disabilities on naive 90% confidence intervals centered at direct estimates (marked with "x")

As sample size increases, all three model estimates tend to converge to the direct estimates. Overall, the FH and the BiFH models produce very similar estimates. With respect to the naive confidence intervals, South Dakota is the only state whose FH and BiFH estimates do not lie within the naive 90% confidence interval. The difference plot of the model estimates in excess of direct survey estimates are shown in Figure 2 in Appendix E.

The estimated parameter values are listed in Table 3. Note that the regression parameters for the BiFH model are not directly comparable to those in univariate models where

ACS survey estimates are used as a predictor. The correlation coefficient of the residuals terms in the BiFH model is estimated to be $\hat{\rho}_{AS} = 0.6059$.

Model	$\hat{\sigma}_v^2$ (or $\hat{\sigma}_{SS}$)	$\hat{\beta}_0$	$\hat{\beta}_{ssi}$	$\hat{\beta}_{di}$	$\hat{\beta}_{ACS}$
FH	0.0009	0.0071	-0.0116	-0.0116	1.7889
OBPME	0.0030	-0.0243	-0.0286	-0.0459	2.9913
BiFH	0.0013	0.1408	-0.0089	0.0358	

Table 3: Estimated values of model parameters $\Psi = (\sigma_v^2; \beta_0, \beta_{ssi}, \beta_{di}, \beta_{ACS})$ for FH and OBPME model and $\Psi_{Bi} = (\sigma_{SS}; \beta_0^{(S)}, \beta_{ssi}^{(S)}, \beta_{di}^{(S)})$ for BiFH model

Based on the model estimates, the univariate FH model tends to give the smallest estimated MSE, ignoring measurement error in the ACS. This is not surprising as the theoretical approaches of estimating the MSE of the FH model such as Datta et al. (2005) may also underestimate the term. The OBPME model draws its estimates closer to the direct survey values in areas with large sampling uncertainty and produces overall the smallest absolute residuals from survey estimates. BiFH model produces larger estimated MSE when compared head-to-head with FH model in areas where both models give estimates that are farther away from direct estimates. This trade-off may favor the BiFH model with respect to coverage when confidence interval of desired level is computed.

6. Discussion

In this article, we are motivated to extend the basic FH model in small area estimation problems to better accommodate covariates measured with error. Rather than simply including the extra information as plain covariates while ignoring the sampling variability entirely, we present two model-assisted approaches that are built to deliver the best predictors as a linear combination of model and direct survey estimates. Our simulated examples have shown that both proposed estimators perform better than several existing models in terms of reducing empirical mean squared errors under a univariate setting while BiFH model outperforms unvariate models when bivariate modeling error is assumed.

In general, the OBPME estimator benefits from both the advantage of OBP estimator, that is, the robustness and reduced mean squared error when model is misspecified and the feature of modeling covariates with known sampling error of the ME model. Due to the fact that both parent estimators are well studied, making further improvement as well as changing means of estimation (for example, to hierarchical Bayes) may be less challenging. On the other hand, multivariate FH models such as BiFH are more general and usually deliver optimal estimates in terms of minimal associated empirical mean squared errors in simulations. However, the updating scheme of model parameters, especially those in the modeling error matrix $\boldsymbol{\Sigma_v}$ is not limited to the version provided in this paper and it may pose computational challenges as the dimension increases.

Since all model-assisted approaches in SAE depend heavily on having quality covariates, we believe the use of ACS data indeed "lends strength" to our models so that disability estimates at the state level seem valid and consistent with the survey estimates. This is a good start but more work needs to be done for incorporating the ACS data at a lower level before making any meaningful comparison of proposed models with FH estimates, or even direct survey estimates. For example, SIPP has less than 2 in 5 counties on average that records some type of disability and therefore one of our goals in building the county-level model for disability is to make estimates for counties that are either not sampled or have zero survey estimates. On the other hand, while ACS have many fewer counties among over 3100 that record no disability, one may wish to consider proper transformation and/or a different modeling approach, for instance, such as zero-inflated beta binomial model. Furthermore, although the parametric double boostrap method for computing estimated MSE in Section 3 is second-ordered unbiased in the form of $2s_{1i} - s_{2i}$, it is not guaranteed to be positive. In practice, one may use other forms suggested in Hall and Maiti (2006).We intend to conduct more extensive research that addresses some of these concerns before new estimates are made official.

A. Conditional Distribution of $\theta_{1i}|\boldsymbol{Y}_i$

We have

$$\boldsymbol{Y}_i|\boldsymbol{\theta}_i = \begin{bmatrix} Y_{1i} \\ Y_{2i} \end{bmatrix} \Big| \begin{bmatrix} \theta_{1i} \\ \theta_{2i} \end{bmatrix} \sim N\left(\begin{bmatrix} \theta_{1i} \\ \theta_{2i} \end{bmatrix}, \boldsymbol{\Sigma}_{\boldsymbol{e}_i} \right),$$

and assuming that

$$\begin{bmatrix} \theta_{1i} \\ \theta_{2i} \end{bmatrix} \sim N\left(\boldsymbol{X}_i'\boldsymbol{\beta} = \begin{bmatrix} \boldsymbol{x}_{1i}'\boldsymbol{\beta}_1 \\ \boldsymbol{x}_{2i}'\boldsymbol{\beta}_2 \end{bmatrix}, \boldsymbol{\Sigma}_{\boldsymbol{v}} = \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{12} & \sigma_{22} \end{bmatrix} \right),$$

one can show that

$$f(\boldsymbol{\theta}_i|\boldsymbol{Y}_i) \propto \exp\left\{ -\frac{1}{2}(\boldsymbol{Y}_i - \boldsymbol{\theta}_i)'\boldsymbol{\Sigma}_{\boldsymbol{e}_i}^{-1}(\boldsymbol{Y}_i - \boldsymbol{\theta}_i) - \frac{1}{2}(\boldsymbol{\theta}_i - \boldsymbol{X}_i'\boldsymbol{\beta})'\boldsymbol{\Sigma}_{\boldsymbol{v}}^{-1}(\boldsymbol{\theta}_i - \boldsymbol{X}_i'\boldsymbol{\beta}) \right\}.$$

The BLUP for $\boldsymbol{\theta}_i$ is

$$\tilde{\boldsymbol{\theta}}_i = \boldsymbol{Y}_i - (\boldsymbol{\Sigma}_{\boldsymbol{e}_i}^{-1} + \boldsymbol{\Sigma}_{\boldsymbol{v}}^{-1})^{-1}\boldsymbol{\Sigma}_{\boldsymbol{v}}^{-1}(\boldsymbol{Y}_i - \boldsymbol{X}_i'\boldsymbol{\beta}).$$

In the case where $\boldsymbol{\Sigma}_{\boldsymbol{e}_i} = diag(D_{1i}, D_{2i})$, we may further simplify

$$\begin{aligned} (\boldsymbol{\Sigma}_{\boldsymbol{e}_i}^{-1} + \boldsymbol{\Sigma}_{\boldsymbol{v}}^{-1})^{-1}\boldsymbol{\Sigma}_{\boldsymbol{v}}^{-1} &= \boldsymbol{\Sigma}_{\boldsymbol{e}_i}(\boldsymbol{\Sigma}_{\boldsymbol{e}_i} + \boldsymbol{\Sigma}_{\boldsymbol{v}})^{-1} \\ &= \begin{bmatrix} D_{1i} & 0 \\ 0 & D_{2i} \end{bmatrix} \begin{bmatrix} D_{1i} + \sigma_{11} & \sigma_{12} \\ \sigma_{12} & D_{2i} + \sigma_{22} \end{bmatrix}^{-1} \\ &= \frac{1}{(D_{1i} + \sigma_{11})(D_{2i} + \sigma_{22}) - \sigma_{12}^2} \begin{bmatrix} D_{1i}(D_{2i} + \sigma_{22}) & -D_{1i}\sigma_{12} \\ -D_{2i}\sigma_{12} & D_{2i}(D_{1i} + \sigma_{11}) \end{bmatrix}. \end{aligned}$$

Hence, if the BLUP of the first element is sought, we will have

$$\tilde{\theta}_{1i} = Y_{1i} - \frac{D_{1i}[(D_{2i} + \sigma_{22})(Y_{1i} - \boldsymbol{x}_{1i}'\boldsymbol{\beta}_1) - \sigma_{12}(Y_{2i} - \boldsymbol{x}_{2i}'\boldsymbol{\beta}_2)]}{(D_{1i} + \sigma_{11})(D_{2i} + \sigma_{22}) - \sigma_{12}^2}.$$

B. An Updating Scheme for Covariance σ_{12}

We start with deriving an estimate for $\sigma_{\xi\xi}, \xi \in \{1, 2\}$, one at a time. We have

$$\begin{aligned} E[\hat{\boldsymbol{e}}_\xi'\boldsymbol{\Sigma}_\xi^{-1}\hat{\boldsymbol{e}}_\xi] &= tr[\boldsymbol{\Sigma}_\xi^{-1}Var(\hat{\boldsymbol{e}}_\xi)] \\ &= tr[\boldsymbol{\Sigma}_\xi^{-1}(\boldsymbol{I} - \boldsymbol{P}_\xi)\boldsymbol{\Sigma}_\xi(\boldsymbol{I} - \boldsymbol{P}_\xi)'] \\ &= tr[(\boldsymbol{I} - \boldsymbol{\Sigma}_\xi^{-1}\boldsymbol{X}_\xi(\boldsymbol{X}_\xi'\boldsymbol{\Sigma}_\xi^{-1}\boldsymbol{X}_\xi)^{-1}\boldsymbol{X}_\xi')(\boldsymbol{I} - \boldsymbol{P}_\xi)'] \\ &= tr[(\boldsymbol{I} - \boldsymbol{P}_\xi)'(\boldsymbol{I} - \boldsymbol{P}_\xi)'] \\ &= tr[\boldsymbol{I} - \boldsymbol{P}_\xi' - \boldsymbol{P}_\xi' + \boldsymbol{P}_\xi'\boldsymbol{P}_\xi'] \\ &= tr[\boldsymbol{I} - \boldsymbol{P}_\xi - \boldsymbol{P}_\xi + \boldsymbol{P}_\xi] \\ &= m - tr(\boldsymbol{P}_\xi) = m - rank(\boldsymbol{X}_\xi'\boldsymbol{X}_\xi). \end{aligned} \tag{9}$$

We can estimate σ_{11} and σ_{22} by solving for the solutions in Equation (9) just like the case of the univariate FH model, or set them to 0 if a positive value cannot be found. If either of σ_{11} or σ_{22} is zero, we set σ_{12} to 0. Otherwise, we set off to estimating σ_{12} between $\hat{\boldsymbol{e}}_1$ and $\hat{\boldsymbol{e}}_2$. Note that

$$\text{Cov}(\boldsymbol{Y}_1, \boldsymbol{Y}_2) = \text{Cov}(\boldsymbol{e}_1 + \boldsymbol{v}_1, \boldsymbol{e}_2 + \boldsymbol{v}_2) = \text{Cov}(\boldsymbol{v}_1, \boldsymbol{v}_2) = \sigma_{12}\boldsymbol{I}.$$

Then

$$\begin{aligned} \text{Cov}(\hat{\boldsymbol{e}}_1, \hat{\boldsymbol{e}}_2) &= \text{Cov}((\boldsymbol{I} - \boldsymbol{P}_1)\boldsymbol{Y}_1, (\boldsymbol{I} - \boldsymbol{P}_2)\boldsymbol{P}_2) \\ &= (\boldsymbol{I} - \boldsymbol{P}_1)\text{Cov}(\boldsymbol{Y}_1, \boldsymbol{Y}_2)(\boldsymbol{I} - \boldsymbol{P}_2)' \\ &= \sigma_{12}(\boldsymbol{I} - \boldsymbol{P}_1)(\boldsymbol{I} - \boldsymbol{P}_2)' \end{aligned}$$

Denote $\boldsymbol{\Sigma}_{\xi}^{-1/2} = diag((D_{\xi 1}+\sigma_{\xi\xi})^{-1/2}, \cdots, (D_{\xi m}+\sigma_{\xi\xi})^{-1/2})$. Therefore,

$$\begin{aligned}
E[(\boldsymbol{\Sigma}_1^{-1/2}\hat{\boldsymbol{e}}_1)'(\boldsymbol{\Sigma}_2^{-1/2}\hat{\boldsymbol{e}}_2)] &= E\left\{tr[\hat{\boldsymbol{e}}_1'\boldsymbol{\Sigma}_1^{-1/2}\boldsymbol{\Sigma}_2^{-1/2}\hat{\boldsymbol{e}}_2]\right\} \\
&= E\left\{tr(\boldsymbol{\Sigma}_1^{-1/2}\boldsymbol{\Sigma}_2^{-1/2})\hat{\boldsymbol{e}}_1\hat{\boldsymbol{e}}_2'\right\} \\
&= tr[(\boldsymbol{\Sigma}_1^{-1/2}\boldsymbol{\Sigma}_2^{-1/2})\mathrm{Cov}(\hat{\boldsymbol{e}}_2, \hat{\boldsymbol{e}}_1)] \\
&= tr[(\boldsymbol{\Sigma}_1^{-1/2}\boldsymbol{\Sigma}_2^{-1/2})\sigma_{12}(\boldsymbol{I}-\boldsymbol{P}_2)(\boldsymbol{I}-\boldsymbol{P}_1)'].
\end{aligned}$$

Finally, we can estimate σ_{12} by

$$\hat{\sigma}_{12} = \frac{(\hat{\boldsymbol{\Sigma}}_1^{-1/2}\hat{\boldsymbol{e}}_1)'(\hat{\boldsymbol{\Sigma}}_2^{-1/2}\hat{\boldsymbol{e}}_2)}{tr[(\hat{\boldsymbol{\Sigma}}_1^{-1/2}\hat{\boldsymbol{\Sigma}}_2^{-1/2})(\boldsymbol{I}-\boldsymbol{P}_2)(\boldsymbol{I}-\boldsymbol{P}_1)']} = \frac{(\hat{\boldsymbol{\Sigma}}_1^{-1/2}\hat{\boldsymbol{e}}_1)'(\hat{\boldsymbol{\Sigma}}_2^{-1/2}\hat{\boldsymbol{e}}_2)}{\hat{M}_1-\hat{M}_2-\hat{M}_3+\hat{M}_4},$$

where

$$\begin{aligned}
\hat{M}_1 &= \sum_{i=1}^{m}(D_{1i}+\hat{\sigma}_{11})^{-1/2}(D_{2i}+\hat{\sigma}_{22})^{-1/2}, \\
\hat{M}_2 &= \sum_{i=1}^{m}(D_{1i}+\hat{\sigma}_{11})^{-1/2}(D_{2i}+\hat{\sigma}_{22})^{-1/2}\boldsymbol{P}_{1ii}, \\
\hat{M}_3 &= \sum_{i=1}^{m}(D_{1i}+\hat{\sigma}_{11})^{-1/2}(D_{2i}+\hat{\sigma}_{22})^{-1/2}\boldsymbol{P}_{2ii}, \\
\hat{M}_4 &= \sum_{i=1}^{m}(D_{1i}+\hat{\sigma}_{11})^{-1/2}(D_{2i}+\hat{\sigma}_{22})^{-1/2}\boldsymbol{R}_{ii},
\end{aligned}$$

where

$$\boldsymbol{R} = \boldsymbol{P}_2\boldsymbol{P}_1' = \boldsymbol{X}_2(\boldsymbol{X}_2'\hat{\boldsymbol{\Sigma}}_2^{-1}\boldsymbol{X}_2)^{-1}\boldsymbol{X}_2'\hat{\boldsymbol{\Sigma}}_2^{-1}\hat{\boldsymbol{\Sigma}}_1^{-1}\boldsymbol{X}_1(\boldsymbol{X}_1'\hat{\boldsymbol{\Sigma}}_1^{-1}\boldsymbol{X}_1)^{-1}\boldsymbol{X}_1'.$$

C. Derivation of the Objective Function for OBPME Estimator

We show below the steps for deriving the function $Q(\boldsymbol{\beta};\sigma_v^2)$ in Equation (5).

$$\begin{aligned}
& E[(\tilde{\boldsymbol{\theta}}_{BP}-\boldsymbol{\theta})'(\tilde{\boldsymbol{\theta}}_{BP}-\boldsymbol{\theta})] \\
= \; & E[\{(\boldsymbol{Y}-\boldsymbol{\theta})-\boldsymbol{\Gamma}(\boldsymbol{Y}-\boldsymbol{X\beta})\}'\{(\boldsymbol{Y}-\boldsymbol{\theta})-\boldsymbol{\Gamma}(\boldsymbol{Y}-\boldsymbol{X\beta})\}] \\
= \; & E[(\boldsymbol{Y}-\boldsymbol{X\beta})'\boldsymbol{\Gamma}^2(\boldsymbol{Y}-\boldsymbol{X\beta})+(\boldsymbol{Y}-\boldsymbol{\theta})'(\boldsymbol{Y}-\boldsymbol{\theta})-2(\boldsymbol{Y}-\boldsymbol{\theta})'\boldsymbol{\Gamma}(\boldsymbol{Y}-\boldsymbol{X\theta})] \\
= \; & E[(\boldsymbol{Y}-\boldsymbol{X\beta})'\boldsymbol{\Gamma}^2(\boldsymbol{Y}-\boldsymbol{X\beta})]+tr(\boldsymbol{D})-2E[\boldsymbol{e}'\boldsymbol{\Gamma}(\boldsymbol{Y}-\boldsymbol{X\beta})] \\
= \; & E[(\boldsymbol{Y}-\boldsymbol{X\beta})'\boldsymbol{\Gamma}^2(\boldsymbol{Y}-\boldsymbol{X\beta})]+tr(\boldsymbol{D})-2E[\boldsymbol{e}'\boldsymbol{\Gamma}\boldsymbol{Y}] \\
= \; & E[(\boldsymbol{Y}-\boldsymbol{X\beta})'\boldsymbol{\Gamma}^2(\boldsymbol{Y}-\boldsymbol{X\beta})]+tr(\boldsymbol{D})-2E[\boldsymbol{e}'\boldsymbol{\Gamma}(\boldsymbol{\theta}+\boldsymbol{e})] \\
= \; & E[(\boldsymbol{Y}-\boldsymbol{X\beta})'\boldsymbol{\Gamma}^2(\boldsymbol{Y}-\boldsymbol{X\beta})]+tr(\boldsymbol{D})-2E[\boldsymbol{e}'\boldsymbol{\Gamma}\boldsymbol{e}] \\
= \; & E[(\boldsymbol{Y}-\boldsymbol{X\beta})'\boldsymbol{\Gamma}^2(\boldsymbol{Y}-\boldsymbol{X\beta})]+tr(\boldsymbol{D})-2tr(\boldsymbol{\Gamma}\boldsymbol{D}) \\
= \; & E\left[\sum_{i=1}^{m}\frac{(Y_i-\beta_0-\beta_1X_i)^2D_i^2}{(D_i+\sigma_v^2+\beta_1^2C_i)^2}\right]+\sum_{i=1}^{m}D_i-2\sum_{i=1}^{m}\frac{D_i^2}{D_i+\sigma_v^2+\beta_1^2C_i}.
\end{aligned}$$

D. A Second-order Unbiased Estimator for MSE

Following the proof in Hall and Maiti (2006) that shows $2s_{1i}-s_{2i}$ is an unbiased estimator for $MSE(\hat{\theta}_{iFH})$, we sketch an outline to show that this estimator is second-order unbiased, i.e. of order $o(m^{-1})$ for any estimator $\hat{\theta}_i$ in this paper. Given data $\boldsymbol{Y}=(Y_1,\cdots,Y_m)'$ and parameter Ψ as in Table 3, write the naive MSE estimator as

$$M_i(\Psi) = MSE(\hat{\theta}_i) = E_{\Psi}[\hat{\theta}_i-\theta_i]^2 = k_i(\Psi)+t_i(\Psi)+o(m^{-1}),$$

where $k_i(\Psi) = E_{\Psi}(\tilde{\theta}_{i_{BP}}-\theta_i)^2 = O(1)$ is the dominating term. For instance, $k_i(\Psi) = k_i(\sigma_v^2) = D_iw_i = \frac{D_i\sigma_v^2}{\sigma_v^2+D_i}$ for the FH model; and $t_i(\Psi)=O(m^{-1})$ collects other terms in the expansion.

We estimate θ_i by $\hat{\theta}_i$ and Ψ by $\hat{\Psi}$. Using $\hat{\Psi}$ and the model, generate $\boldsymbol{\theta}^{(b_1)}$ and $\boldsymbol{Y}^{(b_1)}$, $b_1 = 1, \cdots, B_1$. Then using $\boldsymbol{Y}^{(b_1)}$, we estimate $\theta_i^{(b_1)}$ by $\hat{\theta}_i^{(b_1)}$ and $\Psi^{(b_1)}$ by $\hat{\Psi}^{(b_1)}$. Let

$$M_i(\hat{\Psi}) = E_{\hat{\Psi}}[\hat{\theta}_i^{(b_1)} - \theta_i^{(b_1)}]^2 = k_i(\hat{\Psi}) + t_i(\hat{\Psi}) + o(m^{-1}).$$

Note that $E_\Psi[k_i(\hat{\Psi})] = k_i(\Psi) + b_i(\Psi) + o(m^{-1})$ since Hall and Maiti (2006) has shown that the bias of estimating $k_i(\hat{\Psi})$ is expected to be of the same order of t_i.

Define

$$s_{1i} = \frac{1}{B_1} \sum_{b_1=1}^{B_1} (\hat{\theta}_i^{(b_1)} - \theta_i^{(b_1)})^2.$$

This gives us

$$E_{\hat{\Psi}}[s_{1i}] = E_{\hat{\Psi}}(\hat{\theta}_i^{(b_1)} - \theta_i^{(b_1)})^2 = M_i(\hat{\Psi}) = k_i(\hat{\Psi}) + t_i(\hat{\Psi}) + o(m^{-1}).$$

Now, generate $\boldsymbol{\theta}^{b_2(b_1)}$ and $\boldsymbol{Y}^{b_2(b_1)}$ using the model and $\hat{\Psi}^{(b_1)}$, $b_2 = 1, \cdots, B_2$ and get estimates $\hat{\theta}_i^{b_2(b_1)}$ and $\hat{\Psi}^{b_2(b_1)}$.

Define

$$s_{2i} = \frac{1}{B_1} \sum_{b_1=1}^{B_1} \frac{1}{B_2} \sum_{b_2=1}^{B_2} (\hat{\theta}_i^{b_2(b_1)} - \theta_i^{b_2(b_1)})^2.$$

$$\begin{aligned} E_{\hat{\Psi}}[s_{2i}] &= E_{\hat{\Psi}}\left[\frac{1}{B_1}\sum_{b_1=1}^{B_1} E_{\hat{\Psi}^{(b_1)}}(\hat{\theta}_i^{b_2(b_1)} - \theta_i^{b_2(b_1)})^2\right] \\ &= E_{\hat{\Psi}}\left[\frac{1}{B_1}\sum_{b_1=1}^{B_1} M_i(\hat{\Psi}^{(b_1)})\right] \\ &= E_{\hat{\Psi}}[M_i(\hat{\Psi}^{(b_1)})] \\ &= E_{\hat{\Psi}}[k_i(\hat{\Psi}^{(b_1)}) + t_i(\hat{\Psi}^{(b_1)})] + o(m^{-1}) \\ &= \{k_i(\hat{\Psi}) + b_i(\hat{\Psi})\} + t_i(\hat{\Psi}) + o(m^{-1}). \end{aligned}$$

Hence,

$$\begin{aligned} E_\Psi[2s_{1i}] &= 2\{k_i(\Psi) + b_i(\Psi)\} + 2t_i(\Psi) + o(m^{-1}). \\ E_\Psi[s_{2i}] &= \{[k_i(\Psi) + b_i(\Psi)] + b_i(\Psi)\} + t_i(\Psi) + o(m^{-1}). \end{aligned}$$

Therefore,

$$E_\Psi[2s_{1i} - s_{2i}] = MSE(\hat{\theta}_i) + o(m^{-1}). \tag{10}$$

$g(s_{1i}, s_{2i}) = 2s_{1i} - s_{2i}$ is a second-order unbiased estimator for $MSE(\hat{\theta}_i)$.

E. Difference Plot of Model Estimates, Less Direct Estimates

The plot listed shows the difference between all respective model estimates and direct estimates.

References

Datta, G.S., Fay, R.E. and Ghosh, M., Hierarchical and Empirical Multivariate Bayes Analysis in Small Area Estimation, *Proceedings of the Seventh Annual Research Conference of the Bureau of the Census*, U.S. Department of Commerce, 1991, pp. 63-79.

Datta, G.S. and Lahiri, P., A Unified Measure of Uncertainty of Estimated Best Linear Unbiased Predictiors in Small Area Estimation Problems, *Statistica Sinica*, **10**, 2000, pp. 613-627.

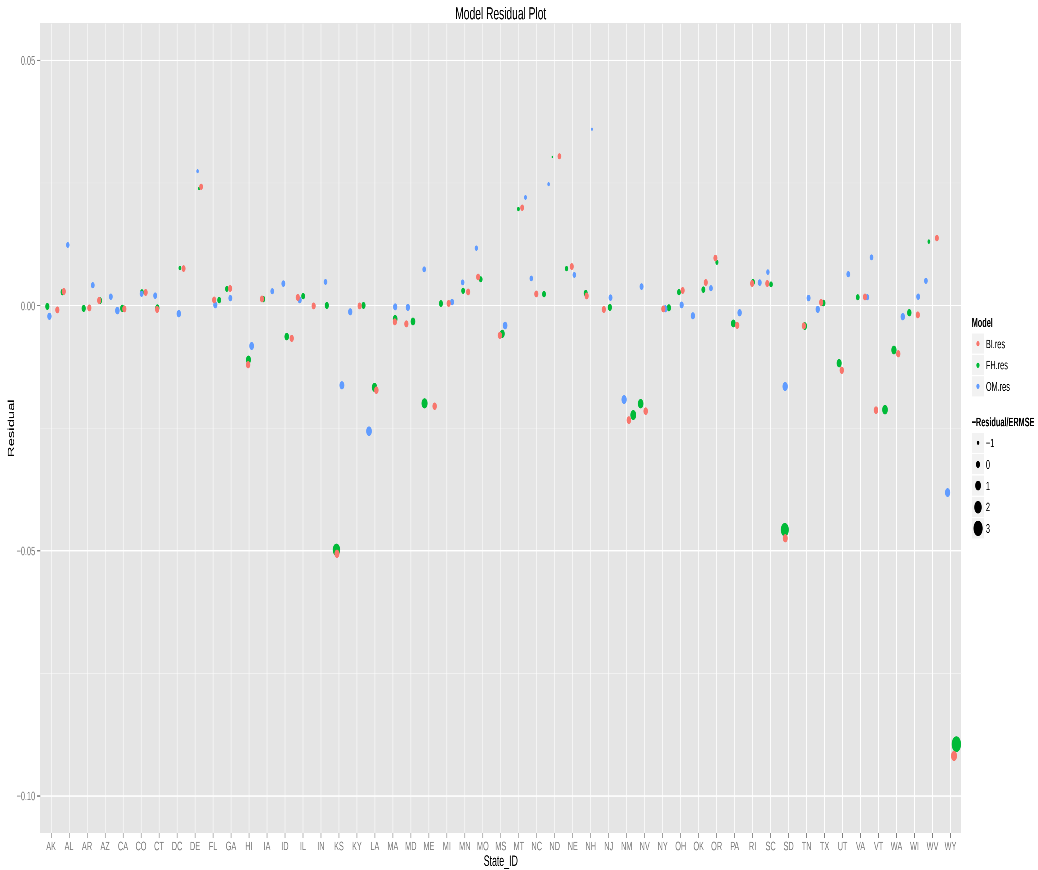

Figure 2: Plot of FH, OM, and BiFH model residulas in excess of direct estimates for state-level total disabilities. Size of the bubble indicates the magnitude of the difference standardized by the estimated RMSE from the bootstrap method.

Datta, G.S., Rao, J.N.K. and Smith, D.D., On Measuring the Variability of Small Area Estimators under a Basic Area Level Model, *Biometrika*, **92** (1), 2005, pp. 183 - 196.

Fay, R.E. and Herriot, R.A., Estimates of Income for Small Places: An Application of James-Stein Procedures to Census Data, *Journal of the American Statistical Association*, **74** (366), 1979, pp. 269-277.

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

Gonzalez, M.E., Use and Evaluation of Synthetic Estimates, *Proceedings of the Social Statistics Section, American Statistical Association*, 1973, pp. 33-36.

Hall, P. and Maiti, T., On Parametric Bootstrap Methods for Small Area Prediction, *Journal of Royal Statistical Society, Series B*, **68** (2), 2006, pp. 221 - 238.

Hall, P. and Maiti, T., Nonparametric Estimation of Mean-Squared Prediction Error in Nested-Error Regression Models, *The Annals of Statistics*, **34** (4), 2006, pp. 1733 - 1750.

Jiang, J., Lahiri, P. and Wan, S-M., A Unified Jackknife Theory, *Annals of Statistics*, **30** (6), 2002, pp. 1782-1810.

Jiang, J., Nguyen, T. and Rao, J.S., Best Predictive Small Area Estimation, *Journal of the American Statistical Association*, **106** (494), 2011, pp. 732-745.

Lahiri, P. and Rao, J.N.K., Robust Estimation of Mean Squared Error of Small Area Estimators, *Journal of the American Statistical Association*, **90** (430), 1995, pp. 758-766.

Livermore, G., Whalen, D., Prenovitz, S., Aggarwal, R. and Bardos, M., Disability Data in National Survey, *Mathematica Policy Research*, 2011.

National Center for Health Statistics, Synthetic State Estimates of Disability, *P.H.S. Publication No. 1759, U.S. Government Printing Office, Washington, DC*, 1968.

Pfeffermann, D. and Correa, S., Empirical Bootstrap Bias Correction and Estimation of Prediction Mean Square Error in Small Area Estimation, *Biometrika*, **99** (2), 2012, pp. 457-472.

Prasad, N.G.N. and Rao, J.N.K., The Estimation of the Mean Squared Error of Small-Area Estimators, *Journal of the American Statistical Association*, **85** (409), 1990, pp. 163-171.

Rao, J.N.K., Small Area Estimation, *Wiley Series in Survey Methodology*, Wiley, 2003.

Schaible, W.L., Use of Small Area Statistics in U.S. Federal Programs, *Small Area Statistics and Survey Designs* (G. Kalton, J. Kordos and R. Plateck eds.), Central Statistical Office, Warsaw, 1992, pp. 95-114.

Ybarra, L.M.R. and Lohr, S.L., Small Area Estimation When Auxiliary Information Is Measured With Error, *Biometrika*, **95** (4), 2008, pp. 919 - 931.