RESEARCH REPORT SERIES
(Statistics #2019-09)

Using American Community Survey Data to Improve Estimates from Smaller Surveys through Bivariate Small Area Estimation Models

Carolina Franco,
William R. Bell

Center for Statistical Research & Methodology
Research and Methodology Directorate
U.S. Census Bureau
Washington, D.C. 20233

Report Issued: October 25, 2019 (Updated: September 10, 2020)

Using American Community Survey Data to Improve Estimates from Smaller Surveys through Bivariate Small Area Estimation Models

Carolina Franco and William R. Bell, U.S. Census Bureau*

September 10, 2020

Abstract

We demonstrate use of bivariate area-level models to improve small area estimates from one survey by borrowing strength from related estimates from a larger survey. Specifically, we demonstrate the potential for borrowing strength from estimates from the American Community Survey (ACS), the largest U.S. household survey, to improve estimates from smaller U.S. surveys, without using regression covariates obtained from auxiliary sources. Applications presented show substantial variance reductions for state estimates of health insurance coverage from the National Health Interview Survey, and for state estimates of disability from the Survey of Income and Program Participation, when modeling these jointly with corresponding ACS estimates. A third application shows substantial variance reductions in ACS one-year county estimates of poverty of school-aged children from modeling these jointly with previous ACS five-year county estimates of school-age poverty. Simple theoretical calculations show

*The views expressed in this report are those of the authors and not the U.S. Census Bureau. All U.S. Census Bureau disclosure avoidance guidelines have been followed and estimates have been approved for release by the U.S. Census Bureau Disclosure Review Board. DRB approval number: CBDRB-FY19-357

how the amount of variance reduction depends on characteristics of the underlying data. For our applications, we examine three alternative bivariate models, starting with a simple bivariate Gaussian model. Since our applications involve modeling proportions, we also examine a bivariate binomial logit normal model, and an unmatched model that combines the Gaussian sampling model with the bivariate logit normal model for the population proportions.

KEYWORDS: combining surveys, bivariate model, health insurance coverage estimate, poverty estimate

1 Introduction

Small area estimation models can improve upon direct survey estimates by borrowing strength from auxiliary data sources, and from data for other small areas, to provide predictors with substantially reduced error variances. See Rao and Molina (2015) or Pfeffermann (2013) for reviews of small area estimation. Small area models often use regression covariates derived from auxiliary data sources such as administrative records, an approach with significant potential for variance reduction. Use of models with such covariates has some limitations, however. First, suitable covariates may simply be unavailable for some applications. Second, covariates found useful at one point in time may lose their effectiveness if the underlying data source materially changes, which can result from such things as new legislation that alters an administrative data source. For an illustration of this, see Bell et al. (2016, pp. 361-362). Finally, use of covariates increases the modeling effort. While this may be a minor obstacle when only one or a few population characteristics are of interest, it can pose a significant challenge when the goal is to improve estimates of many different population characteristics for which many different covariate data sources could be needed. In light of these limitations, small area modelers may turn to another potentially powerful source of auxiliary information, namely, to related estimates from other surveys. A related option is to borrow information from estimates of the same population characteristic from the same survey, but for an earlier time period.

We consider here the case where the smaller of two surveys defines the target population characteristic, and the goal of the modeling is to improve estimates of this target characteristic. This includes the case where the two surveys estimate ostensibly the same population characteristic, though their estimates may differ systematically due to different nonsampling errors (i.e., differences in response errors, nonignorable nonresponse errors, mode effects, concept definitions, timing, etc.)

To be specific, we focus here on borrowing information from the Census Bureau's American Community Survey (ACS) to improve estimates from smaller U.S. surveys. The ACS is a complex sample survey that encompasses many topics, asking questions about demographics, income, employment, occupation, housing, health insurance, education, veteran status, etc. Since 2012, ACS has sampled approximately 3.5 million addresses per year, though the final interviewed sample has ranged from about 2.1-2.3 million housing units due to nonresponse and to some addresses later determined not to be housing units (commercial or nonexistent units). Estimates are produced annually based on either the sample data collected the previous year or, for geographic areas with population less than 65,000, on pooled samples from the previous five years of data collection. More information about the ACS is available at www.census.gov/programs-surveys/acs/.

The ACS sample is substantially larger than the samples of other U.S. household surveys. For example, the Current Population Survey's Annual Social and Economic Supplement (CPS ASEC) samples about 100,000 addresses each year, with some of these ultimately lost to vacant housing units, non-housing units, and nonresponse. Thus, the ACS annual sample is about 25 times larger than the CPS ASEC sample. With its large sample size and the breadth of estimates produced, the ACS presents a potentially very valuable data source for small area models to improve estimates from smaller surveys.

The approach we pursue here borrows information from ACS by using simple bivariate area level models for the estimate of the target population characteristic from a smaller survey and a related ACS estimate. The models include intercept terms but no other regression covariates. Two of our specific applications include modeling relevant ACS state one-year estimates jointly with state estimates from other surveys: first, with estimates of health insurance noncoverage from the National Health Interview Survey (NHIS), and second, with estimates of total disability from the Survey of Income and Program Participation (SIPP). A third application illustrates bivariate modeling that uses ACS estimates to improve other ACS estimates. Specifically, we model previous ACS five-year county poverty estimates jointly with current ACS one-year county poverty estimates to improve the latter. Results for these three applications show that borrowing information from the ACS via these simple bivariate models can produce posterior variances significantly lower than the sampling variances of the smaller survey, or than the sampling variances of the ACS one-year estimates in the third application.

We obtain and compare results from three types of bivariate models here. One is the standard bivariate normal model, which is the particular case of the bivariate extension of the model of Fay and Herriot (1979) that uses no covariates apart from the intercepts. The multivariate FH model was first suggested by Fay (1987) and was then studied by Datta et al. (1991, 1996) and by Ghosh et al. (1996), among others. The second model we consider, used for proportions, is a no-covariates version of the bivariate binomial logit normal (BLN) model previously studied in Franco and Bell (2013, 2015). The BLN model may be more appropriate for proportions as it can naturally handle skewed data and observations (direct survey estimates) equal to zero. This model also keeps the predictions in their restricted support of $[0, 1]$. The third model we consider, also for proportions, is the bivariate extension of a particular case of the unmatched sampling and linking model of You and Rao (2002). This model combines the Gaussian sampling error model with, as in the BLN, a bivariate logit normal model for the population proportions.

Theoretical calculations done under simplifying assumptions help explain our empirical results and reveal under what conditions and to what extent one might expect to benefit from jointly modeling two survey estimates in this way. We decompose the variance reductions to show how much of the improvement from bivariate model predictions is due to pure shrinkage as from a univariate model, and how much is due to incorporating the second survey estimates into the model. These results also shed some light on how the relative sizes of the two surveys affect the variance reductions that can be achieved, suggesting that while a smaller survey can benefit from borrowing strength from a larger survey, little benefit should be expected if a larger survey attempts to borrow strength from a substantially smaller one. We give a measure reflecting the effective relative sizes of the two surveys, to quantify the difference between "smaller" and "larger".

Other authors have pursued area level modeling approaches to combining survey estimates, though typically incorporating available covariates appropriate for their particular applications. For example, Ragunathan et al. (2007) used a hierarchical Bayes trivariate model with many covariates to combine estimates of the same characteristic from two health surveys. Manzi et al. (2011) focused on combining biased small area estimates of different types (e.g, synthetic) based on survey data sources, rather than on combining direct estimates from different surveys. Wang et. al (2012) used a hierarchical Bayes model to combine estimates from multiple repeated surveys with different temporal supports.

A related alternative to jointly modeling two survey estimates is to incorporate the second survey estimates as a regression covariate while accounting for their sampling error using a measurement error model (e.g., Fuller 1987, Ybarra and Lohr 2008, Arima et. al 2017). The simple bivariate Gaussian model we consider in Section 2 is actually equivalent to a structural measurement error model with no additional covariates. Bell et al. (2019) provided some cautions about using the other main type of measurement error model, the functional measurement error model, for small area estimation. Kim et al. (2015) used a different type of measurement error model that treats the true population characteristic as a latent variable whose values for each area are fixed, unknown quantities. Another related alternative that is sometimes proposed is to simply include in a model the estimates from the second survey as a regression covariate ignoring their sampling error. We strongly warn against this approach, as such models are misspecified, which Bell et al. (2019) observed can result in suboptimal prediction and misstatement of the prediction error variances.

The bivariate models we propose here are simple to implement and can be fit with readily available software. In our applications we use a Bayesian approach with diffuse priors implemented via JAGS (Plummer, 2010). The models could alternatively be implemented from a frequentist approach by coding likelihood evaluation and maximization, as well as prediction results, in a statistical package such as R (R core team, 2017) or SAS (2010). We expect the general trends regarding variance reduction from use of a bivariate model would be similar under either a Bayesian or frequentist approach.

Section 2 defines the bivariate models and discusses model comparison. It also presents some theoretical results showing the potential for variance reduction from the bivariate model versus univariate shrinkage. Section 3 presents the three applications, demonstrating the very large variance reductions possible from bivariate modeling. Finally, Section 4 provides discussion and suggestions for future research.

2 Models, Model Comparison, and Bivariate Model Prediction Error Variance Decomposition

Suppose we have direct estimates for $i = 1, \ldots, m$ small areas of population characteristics θ_{1i} estimated by the smaller survey and θ_{2i} estimated by the ACS. θ_{1i} and θ_{2i} are assumed to be related but not identical characteristics. Since our applications involve modeling proportions, we denote the direct survey estimates appearing in our model below by p_{1i} for the smaller survey and p_{2i} for the ACS. For applications that do not involve proportions one may think of revising the notation for the observations in the model below to a more generic notation such as (y_{1i}, y_{2i}). Our focus here is on using bivariate models to borrow information from the ACS estimates p_{2i} to improve predictions of θ_{1i} relative to use of the direct estimates p_{1i}.

The bivariate Gaussian model that we use here is:

$$p_{1i} = \theta_{1i} + e_{1i} \qquad \theta_{1i} = \mu_1 + u_{1i} \qquad i = 1, \ldots, m. \tag{1}$$

$$p_{2i} = \theta_{2i} + e_{2i} \qquad \theta_{2i} = \mu_2 + u_{2i} \tag{2}$$

$$\begin{bmatrix} u_{1i} \\ u_{2i} \end{bmatrix} \overset{i.i.d}{\sim} N(0, \mathbf{\Sigma}), \quad \mathbf{\Sigma} = \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{12} & \sigma_{22} \end{bmatrix} \tag{3}$$

$$\begin{bmatrix} e_{1i} \\ e_{2i} \end{bmatrix} \overset{ind.}{\sim} N(0, \mathbf{V}_i), \quad \mathbf{V}_i = \begin{bmatrix} v_{i11} & v_{i12} \\ v_{i12} & v_{i22} \end{bmatrix} \tag{4}$$

where u_{ji} for $j = 1, 2$ are the area i random effects, which are independent of the sampling errors e_{ji} in the direct estimates. In the small area estimation literature, the sampling covariance matrix $\mathbf{V}_i$ is typically assumed known, but in practice it must be estimated using survey micro data. The direct estimates of v_{i11} for areas with small to moderate sample sizes will typically be unstable, and should generally be improved. This could involve averaging variances or design effects over some areas or, for repeated surveys, over time, or fitting a generalized variance function (GVF). In Section 3 we mention how this was done for each application. For the applications to state level survey estimates in Section 3, the estimates of v_{i22} come from large ACS samples for which such improvements were not needed. Finally, note that if the samples for the two surveys are drawn independently, as will be (at least approximately) the case in all the applications in this paper, then $v_{i12} = 0$. If this condition doesn't hold, then v_{i12} must be estimated from survey microdata.

An important parameter not explicitly defined above is the correlation between the model errors, which we denote as $\rho = \text{corr}(u_{1i}, u_{2i}) = \sigma_{12}/\sqrt{\sigma_{11}\sigma_{22}}$. We expect more benefits from borrowing information from the ACS estimates p_{2i} the larger is the value of $|\rho|$.

Generalizing μ_j to $x'_{ji}\beta_j$ in (1) and (2), where β_j is a vector of regression coefficients, gives the general bivariate Fay-Herriot (1979) model. We will see from our applications, however, that large reductions in variances compared to the direct estimates can be achieved by the model given in (1)–(4) even without adding regression covariates. Equation (1) taken alone provides the corresponding univariate Gaussian model (with no covariates) for the smaller survey estimates, and similarly for equation (2) and the ACS estimates.

Since we are modeling proportions, we examine two alternative models to the bivariate Gaussian model given above. The first is a bivariate extension of a particular case of the unmatched sampling and linking model (USL) of You and Rao (2002). This model replaces equations (1) and (2) of the bivariate Gaussian model by

$$p_{1i} = \theta_{1i} + e_{1i} \qquad \text{logit}(\theta_{1i}) = \mu_1 + u_{1i} \qquad i = 1, \ldots, m. \tag{5}$$

$$p_{2i} = \theta_{2i} + e_{2i} \qquad \text{logit}(\theta_{2i}) = \mu_2 + u_{2i}. \tag{6}$$

where $\text{logit}(\theta_{1i}) = \log[\theta_{1i}/(1-\theta_{1i})]$, and similarly for $\text{logit}(\theta_{2i})$. Thus, in (5) and (6), the equations for p_{1i} and p_{2i} remain the same as in (1) and (2), but the model for the θ_{ji} is now for their logits. (You and Rao (2002) consider a general link function, not just the logit.) We retain equations (3) and (4) from the bivariate Gaussian model, though the quantities u_{1i}, u_{2i}, μ_1, μ_2, σ_{11}, σ_{22}, and ρ are not directly comparable between the Gaussian and USL models since the logit link function puts the model for $\text{logit}(\theta_{ji})$ on a different (logit) scale.

Our other alternative model for proportions is the bivariate BLN model previously studied in Franco and Bell (2013, 2015). To specify this model we define, for the smaller survey, the (estimated) effective sample sizes $\tilde{n}_{1i}$ and "effective sample numbers of successes" $\tilde{y}_{1i}$ to reproduce the direct estimated proportions p_{1i} and their estimated sampling variances v_{i11}, the latter through the usual $p(1-p)/n$ variance estimate for a proportion. Thus,

$$\begin{aligned} \tilde{n}_{1i} &= p_{1i}(1-p_{1i})/v_{i11} && (7) \\ \tilde{y}_{1i} &= p_{1i} \times \tilde{n}_{1i}. && (8) \end{aligned}$$

We make the analogous definitions of $\tilde{n}_{2i}$ and $\tilde{y}_{2i}$ for the ACS estimates. Rounding of $\tilde{n}_{1i}$, $\tilde{y}_{1i}$, $\tilde{n}_{2i}$, and $\tilde{y}_{2i}$ is needed by some statistical software. These adjustments to the observed, unweighted sample counts, and to the sample sizes, are meant to approximately capture the effects of the sampling design.

If $p_{1i} = 0$ or 1, one would probably want to replace it in (7) by some preliminary improved estimate of θ_{1i} defined to be between 0 and 1 so as to avoid obtaining an effective sample size estimate of zero. Franco and Bell (2013, 2015) illustrate this, albeit in models with regression covariates. Alternatively, if $p_{1i} = 0$ or $1, \tilde{n}_{1i}$ might be set to some fraction of the actual sample size, allowing for a design effect greater than 1. The estimates of the sampling variances v_{i11} in (7) may come from a GVF.

Following Franco and Bell (2013, 2015), we use the quantities just defined to specify the bivariate BLN model as follows:

$$\tilde{y}_{1i}|\theta_{1i}, \tilde{n}_{1i} \sim \text{Bin}(\tilde{n}_{1i}, \theta_{1i}) \qquad \text{logit}(\theta_{2i}) = \mu_1 + u_{1i} \qquad i = 1, \ldots, m. \quad (9)$$

$$\tilde{y}_{2i}|\theta_{2i}, \tilde{n}_{2i} \sim \text{Bin}(\tilde{n}_{2i}, \theta_{2i}) \qquad \text{logit}(\theta_{1i}) = \mu_2 + u_{2i} \quad (10)$$

The BLN replaces the first parts of equations (5) and (6) from the USL model with the binomial distribution assumptions while retaining the bivariate normal model for $(\text{logit}(\theta_{1i}), \text{logit}(\theta_{2i}))$. Note that the BLN model assumes conditional independence of $\tilde{y}_{1i}$ and $\tilde{y}_{2i}$ given θ_{ji} and $\tilde{n}_{ji}, j = 1, 2$. Analogous to the previous models, this holds when p_{1i} and p_{2i} are estimated from different samples drawn independently, as is the case in the applications we examine.

If the $\tilde{n}_{1i}$ are sufficiently large, and since the $\tilde{n}_{2i}$ are even larger, the binomial distributions in (9) and (10) should be approximately normal, and the BLN and USL models may then yield similar results. Some potential for differences stems from the facts that the USL model takes the sampling variances, v_{i11} and v_{i22} as known, whereas the BLN model takes the effective sample sizes, $\tilde{n}_{1i}$ and $\tilde{n}_{2i}$, as known, which allows the sampling variances $var(p_{1i}|\theta_{1i}, \tilde{n}_{1i}) = \theta_{1i}(1 - \theta_{1i})/\tilde{n}_{1i}$ to vary with θ_{1i}, and similarly for $var(p_{2i}|\theta_{2i}, \tilde{n}_{2i})$.

For Bayesian treatment of the models using JAGS, we used diffuse normal priors on the means μ_1 and μ_2 , uniform priors on σ_{11}, and σ_{22}, and a uniform$[-1, 1]$ prior on ρ. We used standard model diagnostic tools to check convergence of the MCMC chains, including Gelman-Rubin-Brooks plots (Brooks and Gelman, 1998), plots of the autocorrelation functions, and trace plots. These were computed using the *coda* package in R (Plummer et al. 2006).

2.1 Model comparisons

One would like some model selection criterion to decide between the three models described above. Rao and Wu (2001) give an overview of the various types of statistical model selection criteria. Most do not apply to our problem since our three models are non-nested (so hypothesis tests won't work) and the BLN assumes a discrete distribution for the data while the other two models assume a Gaussian distribution, so likelihood-based criteria and certain Bayesian approaches, such as DIC (van der Linde, et al. 2002), seem inapplicable. To avoid these difficulties, we look to a criterion based on mean squared prediction errors (MSPE). Rao and Wu (2001, Sec. 3) review a number of these, which differ in regard to the various overfitting penalties used. The penalties are generally functions of the number of model parameters and the number of observations. Since our three models all have the same number of parameters, and the number of observations for a given data set is a constant, we can ignore overfitting penalties and use a simple criterion based on mean squared residuals.

Since our interest lies in using p_{2i} to predict θ_{1i}, we define conditional residuals for use in model comparisons as

$$\varepsilon_i = p_{1i} - E(p_{1i}|p_{2i}). \tag{11}$$

Note that $E(\varepsilon_i|p_{2i}) = 0$ and $E(\varepsilon_i^2|p_{2i}) = \text{var}(\varepsilon_i|p_{2i}) = \text{var}(p_{1i}|p_{2i})$. The general form of criterion we use for model comparison is the weighted MSPE defined as

$$\text{WMSPE} = \sum_{i=1}^{m} w_i \varepsilon_i^2 \Bigg/ \sum_{i=1}^{m} w_i \tag{12}$$

where the $w_i \geq 0$ are suitable weights. To account for the fact that higher levels of sampling error in p_{1i} will tend to produce larger squared prediction errors, one obvious choice would be $w_i = 1/v_{i11}$. A more sophisticated choice, more fully accounting for inherent levels of predictability, would be $w_i = 1/\text{var}(p_{1i}|p_{2i})$. The latter reduces WMSPE to the mean square of the standardized conditional residuals, $[\varepsilon_i/[\text{var}(p_{1i}|p_{2i})]^{.5}]$. Still another choice would be $w_i = 1/\text{var}(\theta_{1i}|p_{2i})$. For comparing models, it seems advisable to use a single set of weights for all three models when computing their WMSPE. Our results give model comparisons using either $w_i = 1/\text{var}(p_{1i}|p_{2i})$ or $w_i^* = 1/\text{var}(\theta_{1i}|p_{2i})$, taking these for simplicity from the Gaussian model, plugging posterior means of model parameters into the calculation.

An alternative derived from WMSPE is

$$\text{WMSPE*} = \sum_{i=1}^{m} w_i^* (\varepsilon_i^2 - v_{i11}) \Bigg/ \sum_{i=1}^{m} w_i^*. \tag{13}$$

For the Gaussian and USL models with parameters known, it can be shown that $E(\varepsilon_i^2|p_{2i}) = \text{var}(\theta_{1i}|p_{2i}) + v_{i11}$, so that WMSPE* estimates the weighted average MSE for prediction of the θ_{1i}. If $w_i \equiv w_i^*$, then model selections based on (12) or (13) yield the same results, though the interpretation of WMSPE* may be preferred. However, WMSPE* can be negative, the chance of which increases with increasing overall levels of sampling error. For the applications of Section 3, we use w_i for WMSPE and w_i^* for WMSPE*, so the model preferences from these two comparison criteria do not exactly agree.

Computation of the quantities $E(p_{1i}|p_{2i})$ and var$(p_{1i}|p_{2i})$ for the Gaussian model uses standard results on the multivariate normal distribution. Appendix A discusses computation of these quantities for the USL and BLN models.

2.2 Decomposing the decreases in model prediction error variances from using the bivariate Gaussian model predictors rather than the direct estimators

In this section we use simple analytical computations to help shed light on what situations could lead to large decreases in prediction error variances from using bivariate models. We assume for simplicity that the model defined by (1)–(2) is the true model with $v_{i12} = 0$, and that the model parameters are known. The latter assumption can hold approximately in cases where the number of small areas is large. Without the Gaussian assumption, the results here apply to optimal linear prediction from the model defined by (1)–(2). The results can also be taken as providing a rough indicator of the potential for variance reduction from other bivariate models, such as the bivariate BLN and USL models, as will be seen from results for the applications.

We find that the variance reductions can be expressed in terms of just three quantities: ρ, $r_{1i} = \frac{v_{i11}}{\sigma_{11}}$, and $r_{2i} = \frac{v_{i22}}{\sigma_{22}}$, the latter two being the noise to signal ratios of each of the two surveys. The relative decrease in prediction error variance from using the bivariate model predictor rather than the direct survey estimator can be expressed as:

$$\frac{v_{i11} - var(\theta_{1i}|p_{1i}, p_{2i})}{v_{i11}} = \underbrace{\left[\frac{r_{1i}}{1+r_{1i}}\right]}_{\text{var reduction, UNI vs. DIR}} \times \left[1 + \frac{1}{r_{1i}} \underbrace{\left(\frac{r_{1i}\rho^2}{(1+r_{1i})(1+r_{2i}) - \rho^2}\right)}_{\text{var reduction, BIV vs. UNI}}\right]$$

$$= \left[\frac{v_{i11} - var(\theta_{1i}|p_{1i})}{v_{i11}}\right] \quad \times \quad \left[1 + \frac{1}{r_{1i}}\left(\frac{var(\theta_{1i}|p_{1i}) - var(\theta_{1i}|p_{1i}, p_{2i})}{var(\theta_{1i}|p_{1i})}\right)\right] \tag{14}$$

The first term in brackets in the above equations is the relative decrease in variance resulting from use of the univariate model predictors compared to the direct survey estimators. The second term in brackets involves the noise to signal ratio for the first survey and the relative decrease from using the bivariate Gaussian model predictors rather than the univariate model predictors. Derivation of these results is given in Appendix B. We see there is some part of the variance reduction from bivariate modeling that can be attributed to univariate shrinkage to the mean μ_1.

We now explore how the potential for variance reduction from bivariate modeling depends on ρ, r_{1i}, and r_{2i}. To aid interpretation, for this purpose we replace r_{2i} by $k_i = r_{1i}/r_{2i}$. Notice that if $\sigma_{11} \approx \sigma_{22}$, as can happen when ACS and the smaller survey are estimating ostensibly the same, or very similar, quantities, then $k_i \approx v_{i11}/v_{i22}$. Since we would expect v_{i11} and v_{i22} to be inversely proportional to the smaller survey and ACS sample sizes, then with a little further thought, and defining effective sample sizes $\tilde{n}_{1i}$ and $\tilde{n}_{2i}$ as the sample sizes divided by the respective survey design effects (essentially what is happening in equation (7)), we can argue that, to a rough approximation, $k_i \approx \tilde{n}_{2i}/\tilde{n}_{1i}$. We can thus consider k_i a rough measure of the ratio of the effective sample size of the ACS to that of the smaller survey.

Figure 1 plots the decreases in variances from using the univariate and bivariate models, rather than the direct survey estimators, showing how the decreases depend on ρ, k_i, and r_{1i}, with r_{1i} plotted on a log scale. Panel (a) shows that the percentage decrease in variance from using the univariate model predictor rather than the direct survey estimator increases with r_{1i}. The percentage decreases are small for small values of r_{1i}, which can occur due to either large sample size (so v_{i11} is small) or substantial variation in θ_{1i} across areas (so σ_{11} is large). Substantial variance reductions exceeding 30 percent are achieved around $r_{1i} = 0.5$, and the reductions increase to quite large values as r_{1i} increases.

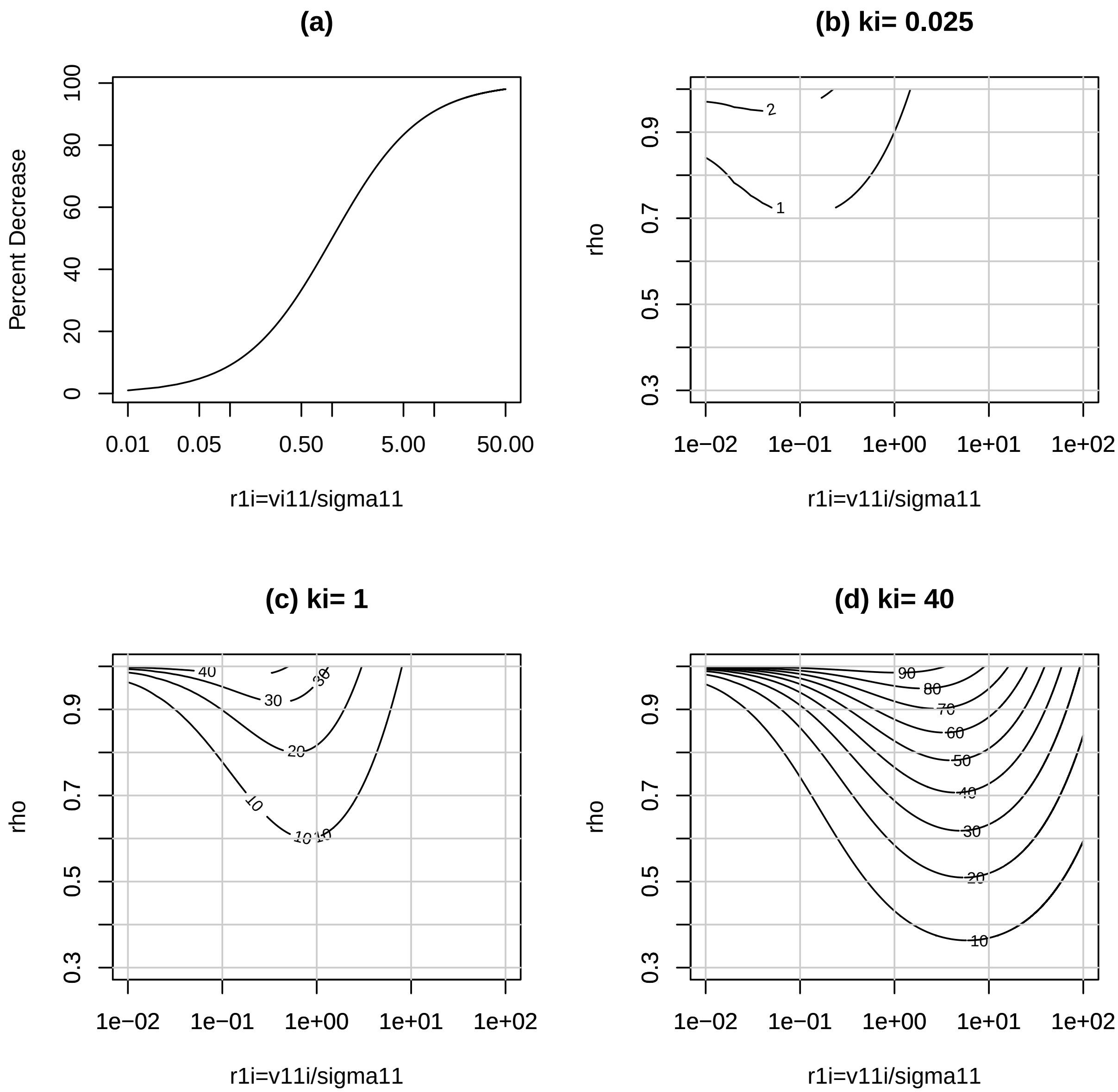

Figure 1: Plots of approximate variance decreases, based on first order approximations. (a) Plot of percentage decrease from using the univariate Gaussian model predictors rather than the direct estimates against r_{1i}. (b)-(d) Contour plots of percentage decrease from using bivariate model predictors rather than univariate Gaussian model predictors when $k_i = 0.025, 1, 40$, respectively for (b), (c), and (d), plotted over ranges of values of ρ and r_{1i}. Note that r_{1i} is plotted on the log scale.

Panels (b)–(d) of Figure 1 show contour plots of the percentage decrease in variance from using a bivariate rather than a univariate model, with contours drawn over ranges of values of ρ and r_{1i}, with plots given for the values $k_i = 0.025, 1, 40$. Panel (b) shows virtually no variance reduction from bivariate modeling versus univariate shrinkage for the case of $k_i = 0.025$, over all values of ρ and r_{1i} shown. In general, small values of $k_i < 1$ mean that the first survey is much larger than the second, the reverse of the case of interest here. The graph shows that no meaningful variance reductions should be expected from bivariate modeling for the estimates from the larger survey.

Considering panels (b)–(d) of Figure 1, we see that the variance reductions from using a bivariate model versus a univariate model increase with increasing k_i. Panel (c), for the case $k_i = 1$, suggests that for surveys of similar size/precision, some benefit might be obtained from using a bivariate rather than a univariate model, up to about a 40% decrease in variances, though such large variance reductions occur only for very high values of ρ and for limited values of r_{1i}.

Panel (d) plots results for $k_i = 40$, which is closer to the median k_i values for the applications presented in Section 3 that illustrate use of the ACS estimates to improve the estimates from smaller surveys. For this larger value of k_i, we see that very large reductions in variances can be achieved by using a bivariate rather than a univariate model, particularly when ρ is very high, and for "intermediate" values of r_{1i}, with r_{1i} values from about 1 to 10 yielding the biggest benefits. Note also that for moderate values of ρ, say less than 0.5, the reductions in variance are relatively low (less than 20%) regardless of the value of r_{1i}. However, in Section 3, we will see that very high values of ρ do occur in practice.

3 Three Applications

3.1 State-level health insurance coverage as measured by the NHIS

In recent years there has been significant interest in estimates of health insurance coverage for the U.S. – the proportion of the population possessing some form of health insurance – for the whole population as well as for population subgroups defined by age, race, sex, geographic area, etc. Concerns about the uninsured population have led to legislation establishing government programs designed to raise coverage, including the Affordable Care Act of 2010 and the Children's Health Insurance Program (CHIP, formerly state CHIP) of 1997, with these programs building on the earlier established Medicaid (1965) and Medicare (1966) programs. Several U.S. surveys, including the ACS, NHIS, CPS, and the SIPP, provide estimates of health insurance coverage. Estimation of insurance coverage by these surveys differs in various ways including the conceptual definitions used of insurance coverage, the methods and timing of the data collection, and the timing for release of coverage estimates. Questions about which data to use in the CHIP funding allocation formula led to a workshop in 2010 sponsored by the U.S. Department of Health and Human Services and convened by the Committee on National Statistics of the U.S. National Research Council (NRC). The workshop summary report (NRC 2010) includes a number of background papers discussing the various sources of estimates of insurance coverage. Chapter 8 of this report (Kenney and Lynch 2010) provided the assessment that "There is a general consensus that the NHIS produces the most valid coverage estimates" (p. 72) while acknowledging that "…the [NHIS] sample size is too small to produce precise annual state (and substate) estimates for most states" (p. 73).

The NHIS is a multistage probability sample of the civilian noninstitutionalized population of the U.S. that is conducted continuously by the National Center for Health Statistics through an agreement with the U.S. Census Bureau. It is a comprehensive health survey that provides estimates of three measures of health insurance noncoverage (uninsured at time of interview, uninsured for at least part of the year prior to interview, and uninsured for more than 1 year at time of interview), along with estimates of different types of insurance coverage (public and private types of coverage). The NHIS provides estimates of coverage for some states through its Early Release (ER) program, which updates its estimates quarterly. Estimates, and further information about the ER program, are available from the NHIS web page at `https://www.cdc.gov/nchs/nhis.htm`. ACS estimates of state health insurance coverage are available from American

FactFinder at `https://factfinder.census.gov/faces/nav/jsf/pages/index.xhtml`. (To be replaced soon by Data.census.gov.)

Final ER estimates based on a full calendar year's sample, and corresponding sampling standard errors, are released in May or June of the year following data collection. The estimates for 2016, based on a sample of 97,459 persons, were released on May 16, 2017. These included state estimates of noncoverage for persons of all ages for 45 states. Estimates for Alaska, District of Columbia (DC), North Dakota, South Dakota, Vermont, and Wyoming were omitted "due to considerations of sample size and precision" (Cohen et al., 2017). The number of state estimates provided varies over the years; for 2017, with a full-year sample of 78,074 persons, only 18 state estimates of coverage for persons of all ages were provided.

We mentioned in Section 2 that the direct estimators of sampling variances should be improved whenever possible prior to applying small area estimation models. The NHIS published estimates of sampling variances differ from the direct variance estimators for all but the 10 largest states. For the smaller states included in the publication, the sampling variances were calculated by assuming their design effect is the average of the design effects for the 10 largest states (Cohen et al, 2017).

Here we demonstrate how jointly modeling ACS one-year and NHIS final ER coverage estimates can substantially reduce the variances of the latter. To have a reasonable number of state estimates for the modeling, we use the NHIS 2016 estimates for 45 states of the percentage of persons of all ages lacking any health insurance coverage at time of interview. The ACS annual health insurance coverage estimates are released in September of the year following data collection. Because this is substantially later in the year than the release of the corresponding year's NHIS estimates, we model ACS estimates for 2015 rather than 2016 jointly with the 2016 NHIS estimates, for the 45 states with estimates from NHIS. This conforms to what ACS data would reasonably be available on the NHIS publication schedule. The ACS sample for 2015 included 2,305,707 household interviews nationally.

Figure 2.a plots the NHIS direct state estimates of health insurance noncoverage against the estimates from ACS. The plot suggests a strong linear relation between the two; in fact, the estimated correlation is 0.93. More of the points fall below the $y = x$ line, suggesting a possible tendency for higher estimates from the ACS. The reverse is suggested, however, by the national estimates, which are 9.0% (std. error = 0.27) for NHIS and 8.6% (std. error = 0.1) for ACS, yielding a not quite significant difference of 0.4% (std. error = 0.288).

Panels b and c of Figure 2 give histograms for the r_{1i} and k_i values. One extremely large k_i value (352) was omitted from the histogram as it distorted the plot. The r_{1i} values are all below 0.25 ($max(r_{1i}) = .246$), which puts them in the region for which Figure 1.a suggests little benefit from univariate shrinkage. The k_i values are mostly large – their first quartile is about 20 – so for most states graphs like Figure 1.d would suggest potential for significant benefits from bivariate modeling if ρ is large. In fact, panel d of Figure 2 shows that the posterior density of ρ is concentrated near 1, with a 90% credible interval of $(.94, .99)$, and a posterior mean of 0.97.

Panel e of Figure 2 shows that the NHIS direct estimates differ substantially from the bivariate Gaussian model predictions, as one would expect, although no extreme outliers appear on the plot. In contrast, Panel f shows that differences between the bivariate BLN and bivariate Gaussian predictions are rather small. The same is also true for comparisons of either to the bivariate USL predictions (results not shown).

Panels g, h, and i of Figure 2 plot the posterior variance decreases (empty circles) from the three bivariate models – Gaussian, USL, and BLN, respectively – against the sampling variances of the direct NHIS estimates. For the three states with the lowest sampling variances, the variance reductions from the Gaussian bivariate model are negligible (California) to small, about 25% (New York and Florida). For the remaining 42 states modeled, the variance improvements increase with v_{i11} from around 40% to near 80%. This assessment also applies, for the most part, to the bivariate USL and BLN results, except that their variance reductions do not show a smooth increase with v_{i11}. The large variance reductions from all three bivariate models contrast with the minimal

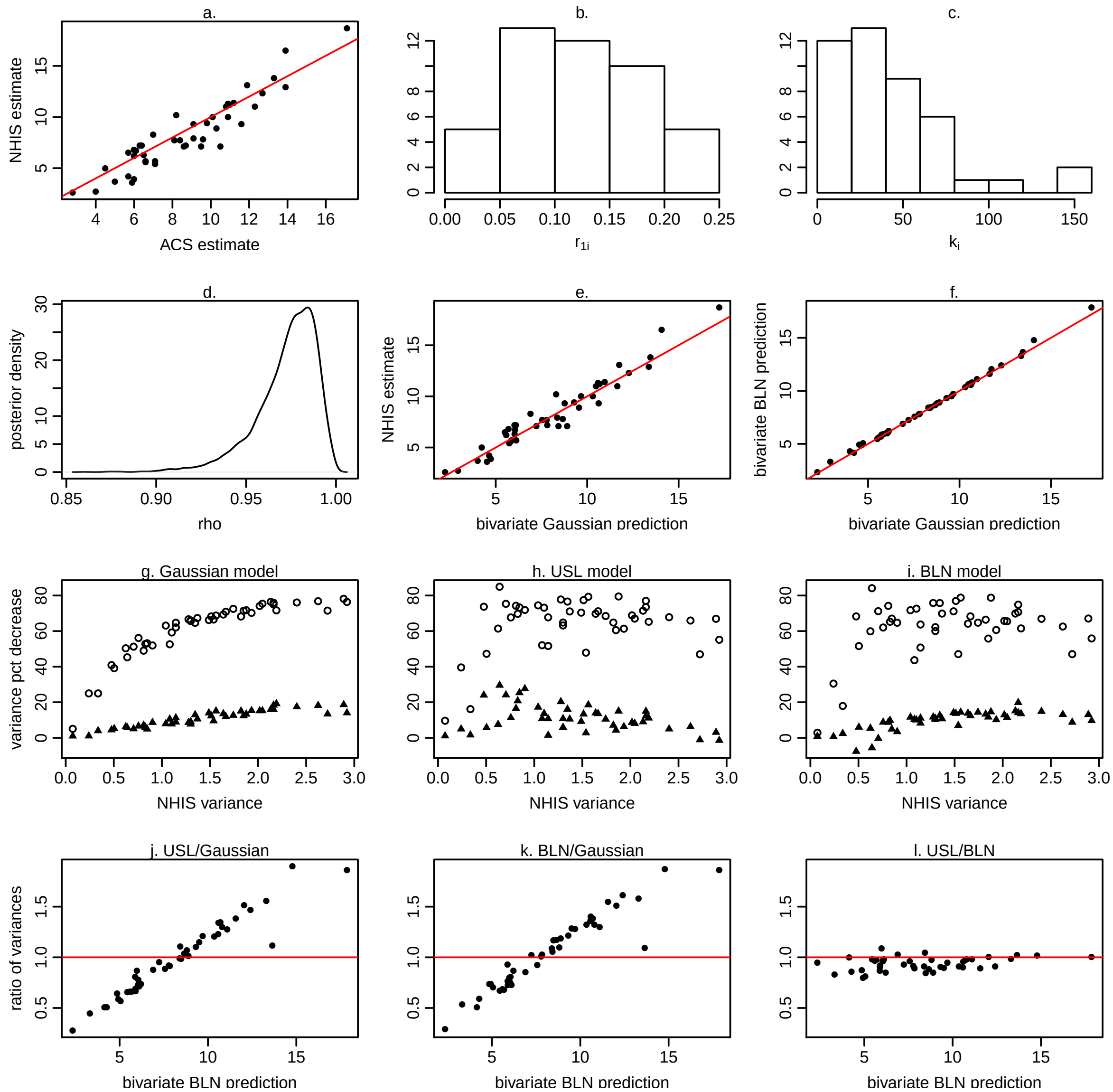

Figure 2: Bivariate modeling of 2016 NHIS and 2015 ACS estimates of health insurance noncoverage

variance reductions from univariate shrinkage shown in the plots (filled triangles).

Panels j, k, and l plot ratios of the posterior variances for the three pairs of bivariate models against posterior means for the bivariate BLN model. The two plots that involve the Gaussian model show substantial differences from the USL and BLN models, with the ratios increasing roughly linearly with the point predictions. These differences will definitely affect statistical inferences, such as prediction intervals, made from the models. The posterior variance ratios of the USL to the BLN model reflect smaller differences, with those from the USL tending to be slightly lower.

Table 2 provides some summary statistics on the variance reductions from v_{i11} for univariate shrinkage (for the Gaussian model) and for the three bivariate models. These results reflect the corresponding results of panels g, h, and i of Figure 2 discussed above. Table 2 also provides the model comparison statistics WMSPE and WMSPE* defined in Section 2.1. These statistics favor the USL and BLN models over the Gaussian model, and are somewhat neutral between the USL and BLN, whose differences are small. The univariate model statistics, which measure weighted sums of squared differences from the overall mean, are much higher than those from the bivariate models.

	percentage variance reductions					**model comparison stats**	
model	**mean**	**1st q.**	**median**	**3rd q.**	**max**	**WMSPE**	**WMSPE***
univariate Gaussian	11	7	11	15	19	11.01	10.48
bivariate Gaussian	62	53	66	72	78	1.41	0.138
bivariate USL	65	61	68	74	85	1.37	0.093
bivariate BLN	62	60	66	71	84	1.38	0.074

Table 1: 2016 NHIS state health insurance noncoverage estimates: Percent variance reductions from direct estimates for the univariate and bivariate models, and the model comparison statistics, WMSPE and WMSPE*.

The main conclusion from the results presented is that bivariate modeling of the NHIS and ACS state estimates of health insurance noncoverage can yield predictions with substantially reduced variances compared to the direct estimates from NHIS for most states. This could allow NCHS to publish estimates for more states, perhaps for all states and DC.

3.2 State-level total disability as measured by the SIPP

A number of U.S. surveys provide data on disability, including the ACS, CPS, and SIPP. Yang and Tan (2018) and Livermore, et al. (2011) provide overviews of U.S. federal disability data sources. The web page, "How Disability Data are Collected from The Survey of Income and Program Participation," on the Census Bureau web site (at `https://www.census.gov/topics/health/disability/guidance/data-collection-sipp.html`), notes that a SIPP redesign in 2014 included putting the six disability questions asked in ACS on the SIPP, along with three additional questions each on child disability and work disability. Prior to this, the 2008 SIPP panel collected more detailed disability data in three topical modules, and this data provided the basis for the "*Americans With Disabilities*" P70 report series. In reference to these topic modules it is noted that, "While the [SIPP] disability measure covers a broader spectrum of activities, a drawback to the SIPP as a data source is the relatively small sample size." Thus, here we examine bivariate modeling of ACS and SIPP state disability estimates to try to reduce the variances of estimates from the 2008 SIPP panel.

Previously, You, Datta, and Maples (2014) pursued bivariate modeling of ACS and SIPP 2008 panel state disability estimates, using models that included two covariates drawn from administrative records. The two covariates used were percentages of the population receiving benefits from the supplemental security income and disability income programs, respectively. Here, we examine the potential for improving the SIPP total disability estimates analyzed by You, et al. via bivariate models of SIPP and ACS data without using any covariates.

You, et al. defined "total disability" for the SIPP data as the estimated number of persons age 15 and over having any of four general types of disability – vision, hearing, mental functional limitations, and physical functional limitations. For comparable ACS data, they took ACS estimates of persons 15+ with a positive response to any of the six ACS disability questions. The state-level SIPP estimates of total disability were based on data collected between May and August of 2010 from wave 6 of the 2008 SIPP panel, which provided interviews from about 35,000 housing units (Sundukchi and Westra, 2015). The 2010 ACS sample obtained interviews from 1.92 million housing units (`https://www.census.gov/acs/www/methodology/sample-size-and-data-quality/sample-size/`). The 2010 ACS estimates were released in September of 2011 and so would have been available for use with data from wave 6 of the 2008 SIPP panel, whose estimates were released in July 2012. The estimates used by You, et al. were custom tabulations of SIPP data. The ACS estimates are available from American FactFinder. To smooth out the SIPP direct variance estimates, we took the national-level estimate of the design-effect, and assumed that the same design effect held for all the individual states.

Figure 3.a plots the SIPP direct state estimates of total disability against the estimates from ACS. The plot suggests a positive relation between the two, and the estimated correlation is 0.55. It is clear, however, that the two set of estimates are estimating different underlying quantities, as the SIPP estimates are higher than those of the ACS for all but one state. The dotted regression line further highlights the differences.

Panels b and c of Figure 3 give histograms for the r_{1i} and k_i values. The values of r_{1i} are rather large for some states, with a maximum of 2.35 and a third quartile of about 0.56. The aforementioned value of the third quartile suggests that for several states we should see some notable reductions in variances from using a pure shrinkage model, since Figure 1 suggests reductions of about 33% for r_{1i} of about 0.5. The k_i values are also large, with a median of 43. As in the NHIS application, these k_i values suggest potential for significant benefits from bivariate modeling if ρ is large. Panel d of Figure 3 shows the posterior density of ρ, which has a posterior mean of 0.79, with a 90% credible interval of $(0.64, 0.90)$. The density is not as highly concentrated near 1 as in the NHIS application, but still shows solid evidence of a high value of ρ.

Panel e of Figure 3 shows that the SIPP direct estimates differ substantially from the bivariate Gaussian model predictions. Though there do not appear to be large systematic differences, we see a bit of shrinkage taking place, with some of the more extreme direct estimates becoming less

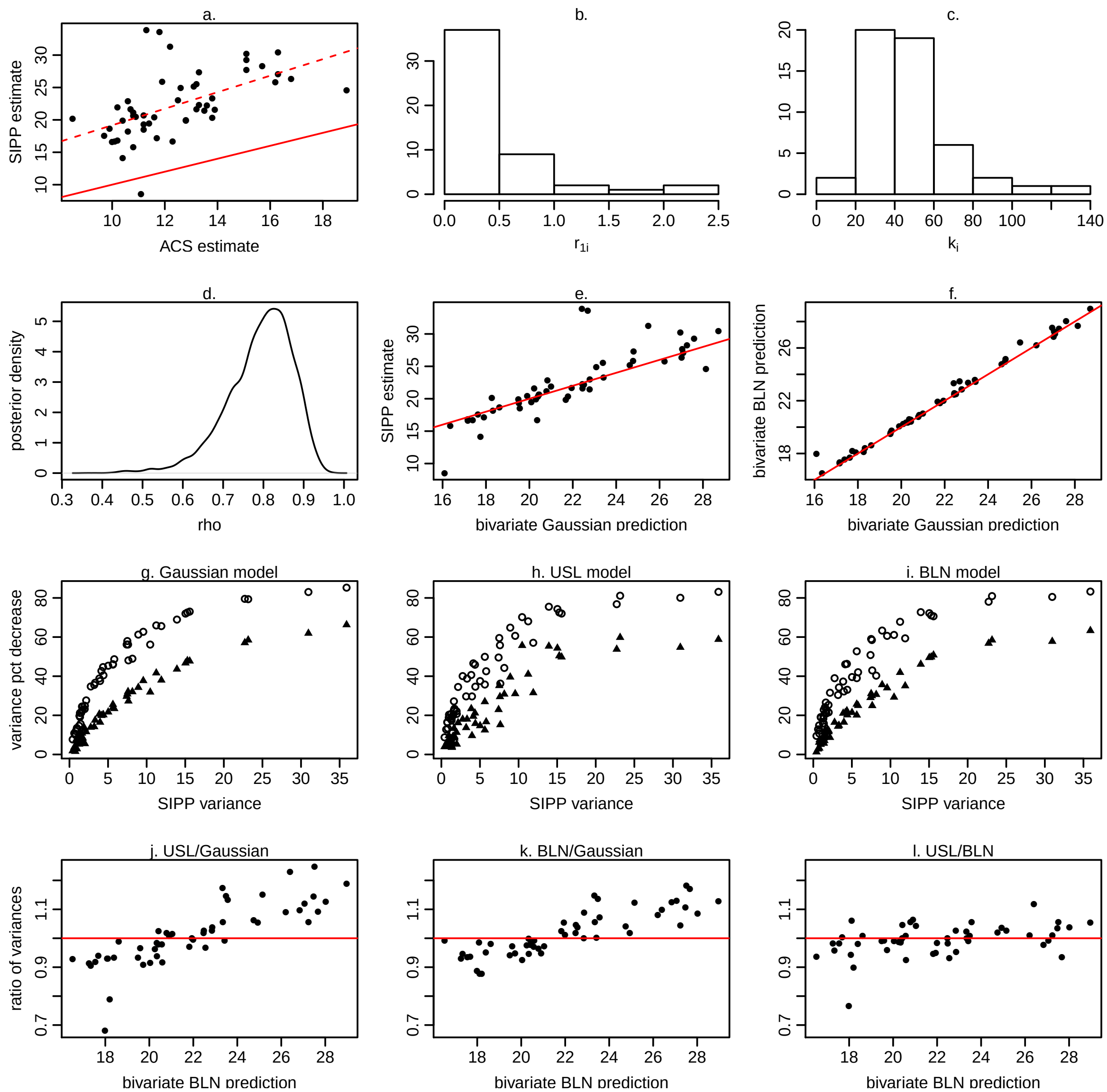

Figure 3: Bivariate modeling of 2010 SIPP and ACS estimates of total disability

extreme after modeling. Note for instance, the two points in the top center, or the point in the lower left. Panel f shows that differences between the bivariate BLN and bivariate Gaussian predictions are rather small. The same is true for comparisons that involve the bivariate USL predictions (results not shown).

Panels g, h, and i of Figure 3 plot the posterior variance decreases (empty circles) from the three bivariate models – Gaussian, USL, and BLN, respectively – against the sampling variances of the direct SIPP estimates. The variance improvements increase with v_{i11} to over 80% for all three models. The variance improvements from the univariate shrinkage model (triangles) are also notable, and can be as high as $60-66\%$, depending on the model, with a third quartile of $31-32\%$ for the BSL, USL, and Gaussian models. The patterns of increase as the SIPP sampling variance increases are similar for all three models, with the plot smoother for the Gaussian model.

Panels j, k, and l plot ratios of the posterior variances for the three pairs of bivariate models against posterior means for the bivariate BLN model. The two plots that involve the Gaussian model show systematic differences from the USL and BLN models, again with the ratios increasing roughly linearly with the point predictions. The posterior variance ratios of the USL to the BLN model reflect slightly smaller differences, with neither model having higher variances than the other overall, and no clear pattern, with one possible outlier with a USL/BLN ratio of about 0.76.

Table 2 provides some summary statistics on the variance reductions from v_{i11} for univariate shrinkage (for the Gaussian model) and for the three bivariate models. These results reflect the corresponding results of panels g, h, and i of Figure 3 discussed above. Table 2 also provides the model comparison statistics WMSPE and WMSPE* defined in Section 2.1. These statistics do not distinguish well among the three bivariate models, but do favor all bivariate models over the univariate ones.

	percentage variance reductions					**model comparison stats**	
model	**mean**	**1st q.**	**median**	**3rd q.**	**max**	**WMSPE**	**WMSPE***
univariate Gaussian	22	8	20	32	66	22.1	19.6
bivariate Gaussian	41	21	39	57	85	13.4	11.8
bivariate USL	40	21	36	58	83	13.5	11.6
bivariate BLN	40	21	37	59	83	13.4	11.7

Table 2: Percent variance reductions from direct estimates for the univariate and bivariate models, and the model comparison statistics, WMSPE and WMSPE*.

We conclude that bivariate modeling of the SIPP and ACS state estimates of total disability yields predictions with substantially reduced variances for most states compared to the direct estimates from SIPP. Some of this improvement is due to pure (univariate) shrinkage, and some is due to borrowing strength from ACS.

3.3 ACS One Year Estimates Borrowing Strength from the Five-Year Estimates

In this section, we illustrate using ACS estimates to improve ACS estimates. Specifically, we use 2007-2011 ACS 5-year county estimates of rates of school-aged children in poverty to improve the corresponding ACS one-year estimates for 2012. The specific variable of rates of school-aged children in poverty is chosen as an illustration, but we expect similar benefits could be achieved for other ACS one-year estimates. Despite its very large national sample size, ACS publishes one-year estimates only for counties of 65,000 or more populations, and even for these counties some estimates will be suppressed due to high sampling variances. Here we model 1-year and 5-year estimates for 3,137 counties, omitting 5 U.S. counties which were not consistently defined across the different years of data. Many of the 1-year estimates are unpublished. These 1-year and 5-year estimates have an estimated correlation of 0.58.

To smooth the direct variances, for both the ACS 5-year estimates and the ACS 1-year estimates, we use the GVF described in Franco and Bell (2013). The GVF uses preliminary estimates of the county proportions derived from a nonlinear regression using county-level covariates tabulated from SNAP and IRS data. Those model-based preliminary estimates are then used to fit a model for the direct variances of the county-level ACS estimates. These GVF estimates were also used in Franco and Bell (2015), Arima et al. (2017), and Bell et. al (2019), for different sets of years.

Our graphical displays in this section differ from those of Sections 3.1 and 3.2 for two reasons. First, because of disclosure limitation restrictions on the unpublished ACS 1-year county estimates, and second, because some of the original plots were unclear when produced for more than $3,000$ data points. Hence, we exclude the analogous plots to some of the panels in Figures 2 and 3, but remark that the general trends are similar for these plots as we saw in the NHIS and SIPP applications. For instance, point predictions from the different bivariate models are similar, while, as before, the corresponding posterior variances can be quite different, with the posterior variance ratios between the pairs of models increasing with the point predictions for the USL/Gaussian and BLN/Gaussian ratios, as we saw in panels j and k of Figures 2 and 3.

Panels a and b of Figure 4 show histograms for the values of r_{1i} and k_i. The median of r_{1i} is 0.47 with a mean of 0.66, and a 95th percentile of 1.90. This implies some of the values of r_{1i} are quite high, indicating potential for variance decreases from a pure shrinkage model. The values of k_i are more moderate than in the NHIS and SIPP applications, with the first and third quartiles between 3 and 4, though this is still large enough that one might expect to benefit from borrowing strength from the 5-year estimates provided that ρ is very high. In fact, panel c of Figure 4 shows that the posterior density of ρ for the bivariate Gaussian model is quite concentrated at values near one. The posterior mean of ρ is 0.94 with a 90% credible interval of $(0.93, 0.95)$.

Figure 4, panels d-f show histograms of the percentage reductions of the posterior variances relative to the sampling variances of the direct estimates, of the univariate and bivariate Gaussian, USL, and BLN models, respectively. Each panel overlays the histogram of percentage variance reductions from using a bivariate model, shown in gray, with that from using the corresponding univariate shrinkage model, shown in white. Overall, we see large benefits for a considerable share of the counties from the bivariate modeling with the 5-year estimates. Note that for each model the variance reductions from the bivariate models are considerably larger than those from univariate shrinkage – the bulk of the distributions of the bivariate models' percent differences are concentrated around higher values than is the case for univariate shrinkage.

Table 3 provides some summary statistics on the variance reductions from v_{i11} for univariate

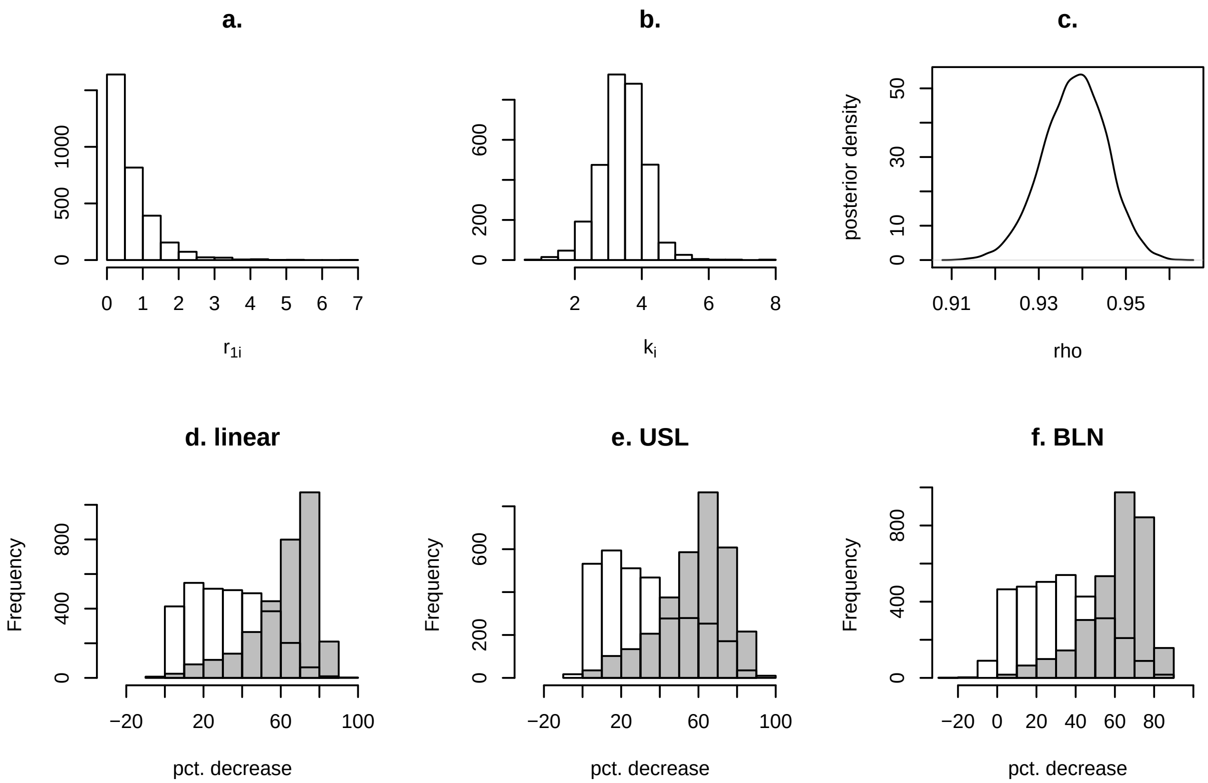

Figure 4: Bivariate modeling of 2012 ACS and 2007-2011 ACS estimates of school-aged children in poverty

shrinkage (for the Gaussian model) and for the three bivariate models. These results reflect the corresponding results of panels d-f of Figure 4 discussed above. Table 3 also provides the model comparison statistics WMSPE and WMSPE* defined in Section 2.1. Again, these statistics do not distinguish well among the three bivariate models, but do favor all bivariate models over the univariate ones.

Panel f of Figure 4 shows some negative percent differences for posterior variances with the univariate BLN model (univariate shrinkage) compared to sampling variances of the direct estimates. These arise from the fact, noted in Section 2, that the BLN model holds the effective sample sizes, $\tilde{n}_{1i}$, constant, but allows the sampling variances, $var(p_{1i}|\theta_{1i}, \tilde{n}_{1i})$, to vary with θ_{1i}. If, for a given area, posterior information about θ_{1i} suggests that it differs substantially from the pro-

model	percentage variance reductions					model comparison stats	
	mean	**1st q.**	**median**	**3rd q.**	**95th percentile**	**WMSPE**	**WMSPE***
univariate Gaussian	33	17	32	47	65	0.0119	0.0103
bivariate Gaussian	62	54	67	74	81	0.0058	0.0045
bivariate USL	58	48	62	71	82	0.0057	0.0044
bivariate BLN	60	52	64	73	83	0.0058	0.0044

Table 3: Percent variance reductions from direct estimates for the univariate and bivariate models, and the model comparison statistics, WMSPE and WMSPE*.

portion effectively reflected in v_{i11}, then the sampling variance implicitly used in the calculation of the posterior variance of θ_{1i} can be substantially different from v_{i11}. If this implicit sampling variance is substantially larger than v_{i11}, then the resulting posterior variance of θ_{1i} can be larger than v_{i11}, especially for the univariate shrinkage model. With over $3,000$ county estimates calculated, this occurred for univariate shrinkage with the BLN a number of times. To address this issue, for the purpose of estimating percent variance reductions for the BLN we replaced the original sampling variances, v_{i11}, with $\hat{\theta}_{1i}(1-\hat{\theta}_{1i})/\tilde{n}_{1i}$, where $\hat{\theta}_{1i}$ is the posterior mean of θ_{1i}. This reduced the number of negative variance improvements, though some still remain in panel f. We calculated this same correction to the BLN results for the NHIS and SIPP applications, but the effects there were minor, and so are not shown in Sections 3.1 and 3.2.

It is natural for this application to ask if a time series model would be more suitable than the bivariate model, as we have multiple years of 1-year estimates available. This question was studied by Franco and Bell (2015) using a BLN model with a first-order autoregressive (AR(1)) structure for the random effects, with application to ACS county estimates of poverty rates for school-aged children. We also studied analogous linear Gaussian time series models, and models of different orders up through AR(5) models. The results showed empirically and analytically that even when the AR(1) model is true, non-trivial variance reductions from using an AR(1) model rather than a bivariate model occur only in very limited circumstances, and did not occur for the application studied, even when excluding available regression covariates from the models. The AR(5) model did not fare much better. Hence, we focused on the bivariate model here.

4 Discussion and Conclusions

The theoretical calculations in Section 2 shed light on the conditions that lead to large decreases in variances from bivariate modeling of two survey estimates, and the three applications in Section 3 illustrate that these conditions do occur for modeling ACS estimates jointly with estimates from a smaller survey. Though univariate shrinkage produced some of the decreases in variances we saw in the three applications, substantial additional benefits resulted from borrowing strength from the ACS estimates. Given the breadth of estimates provided by the ACS, and their geographic detail, the results in this paper suggest a large potential for bivariate modeling to improve small area estimates from the many smaller U.S. surveys.

A key factor in the amount of variance reduction achieved from bivariate modeling is the correlation, ρ, between the population characteristics estimated by the ACS and by the smaller survey. When ACS estimates ostensibly the same characteristic as the smaller survey, as for the NHIS and SIPP applications shown, one would expect ρ to be high. But even if ACS does not estimate ostensibly the same quantity as the target population characteristic estimated by the smaller survey, beneficial variance reductions can be possible if ACS estimates a related characteristic strongly correlated with the target population characteristic.

Though the potential benefits of bivariate modeling seem clear, some issues of methodology deserve further study. One such problem is that of improving direct estimates of sampling variances from small samples. Direct sampling variance estimates can be expected to be more unstable than the corresponding direct survey point estimates that the small area models aim to improve. In our examples, we used simple methods to smooth the direct sampling variances, but further study of this problem is warranted.

Another issue that deserves more attention is model selection/model diagnostics. For each application the point predictions from the three models tended to look rather similar, leading to similar looking diagnostics. This also explains why the WMSPE criteria we explored did not do a very good job distinguishing between the three alternative bivariate models considered, though they did show that the bivariate models performed substantially better than univariate shrinkage for all three applications. Posterior variances did differ substantially, especially between the Gaussian and the other two models, recommending further investigation to determine which model best reflects prediction uncertainty for a given application.

References

Arima, S., Bell, W.R., Datta, G.S., Franco, C., and Liseo, B. (2017). Multivariate Fay-Herriot Bayesian Estimation of Small Area Means Under Functional Measurement Error. *Journal of the Royal Statistical Society A*, **180**, 1191–1209, DOI:10.1111/rssa.12321.

Bell, William R., Chung, Hee-Cheol, Datta, Gauri S., and Franco, Carolina (2019), "Measurement Error in Small Area Estimation: Functional Versus Structural Versus Naive Models," *Survey Methodology*, **45**, 61–80.

Bell, W. R., Basel, W. W., and Maples, J. J. (2016). An Overview of the U.S. Census Bureau's Small Area Income and Poverty Estimates (SAIPE) Program. In *Analysis of Poverty Data by Small Area Methods*, Monica Pratesi (ed.), London: Wiley, 349-378.

Brooks, S P. and Gelman, A. (1998). General Methods for Monitoring Convergence of Iterative Simulations. *Journal of Computational and Graphical Statistics*, **7**, 434-455.

Cohen, Robin A., Zammitti, Emily P., and Martinez, Michael E. (2017). *Health Insurance Coverage: Early Release of Estimates from the National Health Interview Survey, 2016*. Division of Health Interview Statistics, National Center for Health Statistics, May, 2017. `https://www.cdc.gov/nchs/data/nhis/earlyrelease/insur201705.pdf`.

Last accessed July 11, 2019.

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

Datta, G.S., Ghosh, M., Nangia, N. and Natarajan, K. (1996). Estimation of Median Income of Four-Person Families: A Bayesian Approach, in *Bayesian Analysis in Statistics and Econometrics,* eds.: D.A. Berry, K.M. Chaloner and J.K. Geweke. New York: Wiley, 129-140.

Fay, R. E. (1987). Application of Multivariate Regression to Small Domain Estimation, in *Small Area Statistics*, eds. R. Platek, J.N.K. Rao, C.E. Sarndal, and M.P. Singh, New York: Wiley, 91–102.

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

Franco, C. and Bell, W. R. (2013). Applying Bivariate Binomial/Logit Normal Models to Small Area Estimation. *Proceedings of the American Statistical Association, Section on Survey Research Methods*, 690–702, URL http://ww2.amstat.org/sections/srms/Proceedings/.

Franco, C. and Bell, W. R. (2015). Borrowing Information Over Time in Binomial/Logit Normal Models for Small Area Estimation. *Statistics in Transition (new series)* and *Survey Methodology*, joint issue on Small Area Estimation, **16**, 563–584, available at http://stat.gov.pl/en/sit-en/issues-and-articles-sit/previous-issues/volume-16-number-4-december-2015/.

Fuller, W. A. (1987). *Measurement Error Models*. New York: Wiley.

Ghosh, M., Nangia, N. and Kim, D. (1996). Estimation of Median Income of Four-Person Families: A Bayesian Time Series Approach. *Journal of the American Statistical Association*, **91**, 1423-1431.

Kenney, Genevieve and Lynch, Victoria (2010). Monitoring Children's Health Insurance Coverage Under CHIPRA Using Federal Surveys, chapter 8 in *Databases for Estimating Health Insurance Coverage for Children: A Workshop Summary*. Thomas J. Plewes, rapporteur. Committee on National Statistics, Division of Behavioral and Social Sciences and Education. Washington, DC: The National Academies Press.

Kim, J. K., Park, S., and Kim, S. (2015). Small Area Estimation Combining Information from Several Sources. *Survey Methodology*, **41**, 1, 21-36.

Livermore, Gina, Whalen, Denise, Prenovitz, Sarah, Aggarwal, Raina, and Bardos, Maura (2011). *Disability Data in National Surveys*. Mathematica Policy Research, report submitted to the Office of the Assistant Secretary for Planning and Evaluation, U.S. Department of Health and Human Services. http://cire.mathematica-mpr.com/media/publications/pdfs/disability/data_national_surveys.pdf.

Lohr, S. L. and Prasad, N. G. N. (2013). Small Area Estimation with Auxiliary Survey Data. *The Canadian Journal of Statistics*, **31**, 383-396.

Manzi, G., Spiegelhalter, D. J., Turner, R. M., Flowers, J., and Thompson, S. G. (2011). Modeling bias in combining small area prevalence estimates from multiple surveys. *Journal of the Royal Statistical Society A*, **174**, 31–50, https://doi.org/10.1111/j.1467-985X.2010.00648.x.

National Research Council (2010). *Databases for Estimating Health Insurance Coverage for Children: A Workshop Summary*. Thomas J. Plewes, rapporteur. Committee on National Statistics, Division of Behavioral and Social Sciences and Education. Washington, DC: The National Academies Press.

Pfeffermann, Danny (2013). New Important Developments in Small Area Estimation. *Statistical Science*, **28**, 40–68.

Plummer, M. (2010). JAGS - Just Another Gibbs Sampler (JAGS 2.1.0., May 12, 2010). Available at https://sourceforge.net/projects/mcmc-jags.

Plummer M., Best, N., Cowles, K. and Vines, K. (2006). CODA: Convergence Diagnosis and Output Analysis for MCMC *R News*, **6**, 1, 1-11. `http://CRAN.R-project.org/doc/Rnews/Rnews_2006-1.pdf`.

R Core Team (2017). *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing. URL https://www.R-project.org/.

Raghunathan, T.E, Xie, D., Schenker, N., Parsons, V. L., Davis W. W., Dodd, K. W., and Feuer, E. J. (2007). Combining Information From Two Surveys to Estimate County-Level Prevalence Rates of Cancer Risk Factors and Screening. *Journal of the American Statistical Association*, **102**, 474–486.

Rao, J.N.K. and Molina, Isabel (2015). *Small Area Estimation*, second edition, Hoboken, New Jersey: John Wiley.

Rao, C. R. and Wu, Y. (2001). On Model Selection (with discussion). Chapter 2 in *Model Selection*, IMS Lecture Notes–Monograph Series, Volume 38, Beachwood, Ohio: Institute of Mathematical Statistics, 1–64.

SAS Institute Inc. (2010). The NLMIXED Procedure. SAS version 9.2 Help and Documentation.

Sundukchi, Mahdi and Westra, Ashley (2015). Survey of Income and Program Participation (SIPP) 2008 Panel: Source and Accuracy Statement for Longitudinal Analysis of Waves 1 to 16 Public Use Files (S&A-19). URL https://www.census.gov/programs-surveys/sipp/techdocumentation/source-accuracy-statements/source-accuracy-statements-2008.html.

van der Linde, Angelika, Carlin, Bradley P., Best, Nicola G., and Spiegelhalter, David J. (2002). Bayesian Measures of Model Complexity and Fit (with discussion). *Journal of the Royal Statistical Society, Series B*, **64**, 583–639.

Wang, J.C., Holan, S.H., Nandram, B., Barboza, W., Toto, C., and Anderson, E. (2012). A Bayesian Approach to Estimating Agricultural Yield Based on Multiple Repeated Surveys. *Journal of Agricultural, Biological, and Environmental Statistics*, **17**, 84–106, `https://doi.org/`
10.1007/s13253-011-0067-5.

Yang, K. Lisa and Tan, Hock E. (2018). "Disability Statistics: Online Resource for U.S. Disability Statistics," Institute on Employment and Disability, Cornell University, available at http://www.disabilitystatistics.org/sources.cfm.

Ybarra, L. M. R. and Lohr, S. L. (2008). Small Area Estimation when Auxiliary Information is Measured with Error. *Biometrika*, **95**, 919–931.

You, Jiashen, Datta, Gauri S., Maples, Jerry J. (2014). Modeling Disability in Small Areas: An Area-Level Approach of Combining Two Surveys. Research Report RRS2010-08, Center for Statistical Research and Methodology, U.S. Census Bureau, http://www.census.gov/srd/papers/pdf/rrs2014-11.pdf, last accessed 9/19/2018.

You, Yong and Rao, J. N. K. (2002). Small Area Estimation Using Unmatched Sampling and Linking Models. *The Canadian Journal of Statistics*, **30**, 3–15.

Appendix A: Computations needed for the WMSPE statistics

We show how to compute the quantities $E(p_1|p_2)$ and var$(p_1|p_2)$ needed to compute the WMSPE measures of Section 2.1 for the USL and BLN models. (Those for the Gaussian model follow from standard results.) To simplify notation, we drop the i subscript in the following, understanding that the results below would be applied for each area $i = 1, \dots, m$.

We start with the USL model given by (5) and (6). Let $z_1 = \text{logit}(\theta_1) = \log[\theta_1/(1-\theta_1)]$ and $z_2 = \text{logit}(\theta_2) = \log[\theta_2/(1-\theta_2)]$, and note that the inverse logit function is $\theta_1 = [1+\exp(-z_1)]^{-1}$. From standard results for the bivariate Gaussian model for $(z_1, z_2)'$, $z_1|z_2 \sim N(\mu_1+(\sigma_{12}/\sigma_{22})(z_2-\mu_2), \sigma_{11} - \sigma_{12}^2/\sigma_{22})$. Let $\phi(z_1|z_2)$ denote the corresponding density. For any integrable function $f(z_1)$ defined on $(-\infty, \infty)$, we have

$$E[f(z_1)|z_2] = E[f(z_1)|\theta_2] = \int_{-\infty}^{\infty} f(z_1)\phi(z_1|z_2)dz_1.$$

Functions $f(z_1)$ of interest are $f(z_1) = \theta_1 = [1 + \exp(-z_1)]^{-1}$ and $f(z_1) = \theta_1^2$, for which we calculate

$$E(\theta_1|\theta_2) = \int_{-\infty}^{\infty} [1 + \exp(-z_1)]^{-1}\phi(z_1|z_2)dz_1 \tag{15}$$

$$E(\theta_1^2|\theta_2) = \int_{-\infty}^{\infty} [1 + \exp(-z_1)]^{-2}\phi(z_1|z_2)dz_1 \tag{16}$$

$$\text{var}(\theta_1|\theta_2) = E(\theta_1^2|\theta_2) - [E(\theta_1|\theta_2)]^2. \tag{17}$$

Given values for the model parameters (in our applications, we use their posterior means from the fitted bivariate model), and given a value for $z_2 = \text{logit}(\theta_2)$, we can compute (15) and (16) by numerical integration, which thus also gives (17). We used the integrate function of R (R Core Team 2017) for this purpose.

Now

$$\begin{aligned} E(p_1|p_2) &= E_{\theta_2|p_2}[E(p_1|p_2,\theta_2)] \\ &= E_{\theta_2|p_2}\{E_{\theta_1|\theta_2}[E(p_1|\theta_1,p_2,\theta_2)]\} \\ &= E_{\theta_2|p_2}[E_{\theta_1|\theta_2}(\theta_1)] \\ &= E_{\theta_2|p_2}[E(\theta_1|\theta_2)]. \end{aligned} \tag{18}$$

To compute $E_{\theta_2|p_2}[E(\theta_1|\theta_2)]$, we took posterior simulations of $\theta_2|p_2$ obtained from JAGS using the univariate USL model for p_2 given by (6). One way to do this is to run the bivariate USL model fixing $\rho = 0$. For each such simulated value of θ_2, we computed $E(\theta_1|\theta_2)$ from (15), and then averaged these results over the simulations. The averaging over simulations was done within R using functions from CODA (Plummer, et al. 2006).

We also have

$$\begin{aligned} \text{var}(p_1|p_2) &= E_{\theta_2|p_2}[\text{var}(p_1|p_2,\theta_2)] + \text{var}_{\theta_2|p_2}[E(p_1|p_2,\theta_2)] \\ &= E_{\theta_2|p_2}[\text{var}(p_1|\theta_2)] + \text{var}_{\theta_2|p_2}[E(p_1|\theta_2)]. \end{aligned} \tag{19}$$

The second term in (19) equals $\text{var}_{\theta_2|p_2}[E(\theta_1|\theta_2)]$ since $E(p_1|\theta_2) = E_{\theta_1|\theta_2}[E(p_1|\theta_1,\theta_2)] = E_{\theta_1|\theta_2}(\theta_1)$. This can be computed analogously to the computation of (18) by taking the variance of the values of $E(\theta_1|\theta_2)$ computed for the simulations of $\theta_2|p_2$, again using CODA. For the first term in (19), we note that

$$\begin{aligned} \text{var}(p_1|\theta_2) &= E_{\theta_1|\theta_2}[\text{var}(p_1|\theta_1,\theta_2)] + \text{var}_{\theta_1|\theta_2}[E(p_1|\theta_1,\theta_2)] \quad (20) \\ &= E_{\theta_1|\theta_2}(v_{11}) + \text{var}_{\theta_1|\theta_2}(\theta_1) \\ &= v_{11} + \text{var}(\theta_1|\theta_2). \end{aligned}$$

The first term in (19), $E_{\theta_2|p_2}[\text{var}(p_1|\theta_2)]$, can thus be computed by adding v_{11} to the average of the values of var$(\theta_1|\theta_2)$ computed over the simulations of $\theta_2|p_2$, once again using CODA.

The derivation for the BLN model is the same as above through equation (20). The second term in (20) then becomes var$(\theta_1|\theta_2)$, as above, but for computing the first term in (20) we have

$$\begin{aligned} E_{\theta_1|\theta_2}[\text{var}(p_1|\theta_1,\theta_2)] &= E_{\theta_1|\theta_2}[\text{var}(p_1|\theta_1)] \\ &= E_{\theta_1|\theta_2}[\theta_1(1-\theta_1)/\tilde{n}_1] \\ &= \tilde{n}_1^{-1}[E(\theta_1|\theta_2) - E(\theta_1^2|\theta_2)]. \end{aligned}$$

Appendix B: Sketch of proof of the decomposition formula (14)

As in Appendix A, we simplify notation by dropping the i subscript in the following. Since predicting θ_1 is equivalent to predicting e_1, by the properties of the multivariate normal distribution, we have

$$\text{var}(\theta_1|p_1) = \text{var}(e_1|p_1) = v_{11} - v_{11}^2(\sigma_{11}+v_{11})^{-1} = \frac{v_{11}\sigma_{11}}{\sigma_{11}+v_{11}}. \tag{21}$$

Similarly, we have

$$\text{var}(\boldsymbol{\theta}|\mathbf{p}) = \text{var}(\mathbf{e}|\mathbf{p}) = \mathbf{V} - \mathbf{V}(\boldsymbol{\Sigma}+\mathbf{V})^{-1}\mathbf{V},$$

where $\boldsymbol{\theta} = (\theta_1,\theta_2)$, $\mathbf{e} = (e_1,e_2)$, and $\mathbf{p} = (p_1,p_2)$. Simplifying, and assuming $v_{12}=0$, we obtain, for the $(1,1)$ element of the resulting conditional variance matrix

$$\text{var}(\theta_1|\mathbf{p}) = v_{11} - \frac{v_{11}^2(v_{22}+\sigma_{22})}{(v_{11}+\sigma_{11})(v_{22}+\sigma_{22}) - \rho^2\sigma_{11}\sigma_{22}}. \tag{22}$$

To compute the percentage decrease in posterior variances from using the univariate model versus the direct estimators, we have

$$\frac{v_{11} - \text{var}(\theta_1|p_1)}{v_{11}} = \frac{v_{11} - v_{11}\sigma_{11}/(\sigma_{11} + v_{11})}{v_{11}} = \frac{v_{11}}{\sigma_{11} + v_{11}} = \frac{r_1}{1 + r_1}$$

Similarly, using (21) and (22) and simplifying we obtain

$$\frac{\text{var}(\theta_1|p_1) - \text{var}(\theta_1|\mathbf{p})}{\text{var}(\theta_1|p_1)} = \frac{\rho^2 v_{11}\sigma_{22}}{(v_{11} + \sigma_{11})(v_{22} + \sigma_{22}) - \rho^2\sigma_{11}\sigma_{22}} = \frac{r_1\rho^2}{(r_1 + 1)(r_2 + 1) - \rho^2} \tag{23}$$

Computing the percentage difference between the bivariate model posterior variances and the direct estimators, and expressing it in terms of r_1, r_2, and ρ gives

$$\frac{v_{11} - \text{var}(\theta_1|\mathbf{p})}{v_{11}} = \frac{r_1(r_2 + 1)}{(r_1 + 1)(r_2 + 1) - \rho^2}$$

which, after some algebra, can be shown to be equivalent to expression (14).