Catalogue no. 12-001-X
ISSN 1492-0921

Survey Methodology

Improved small area inference from data integration using global-local priors

by Dexter Cahoy and Joseph Sedransk

Release date: December 23, 2025

Statistics Canada Statistique Canada

Canada

How to obtain more information

For information about this product or the wide range of services and data available from Statistics Canada, visit our website, www.statcan.gc.ca.

You can also contact us by

Email at infostats@statcan.gc.ca

Telephone, from Monday to Friday, 8:30 a.m. to 4:30 p.m., at the following numbers:

- Statistical Information Service 1-800-263-1136
- National telecommunications device for the hearing impaired 1-800-363-7629
- Fax line 1-514-283-9350

Standards of service to the public

Statistics Canada is committed to serving its clients in a prompt, reliable and courteous manner. To this end, the Agency has developed standards of service which its employees observe in serving its clients. To obtain a copy of these service standards, please contact Statistics Canada toll-free at 1-800-263-1136. The service standards are also published on www.statcan.gc.ca under “Contact us” > “Standards of service to the public.”

Note of appreciation

Canada owes the success of its statistical system to a long-standing partnership between Statistics Canada, the citizens of Canada, its businesses, governments and other institutions. Accurate and timely statistical information could not be produced without their continued co-operation and goodwill.

Published by authority of the Minister responsible for Statistics Canada

© His Majesty the King in Right of Canada, as represented by the Minister of Industry, 2025

Use of this publication is governed by the Statistics Canada Open Licence Agreement.

An HTML version is also available.

Cette publication est aussi disponible en français.

Survey Methodology, December 2025
Vol. 51, No. 2, pp. 449-473
Statistics Canada, Catalogue No. 12-001-X

Improved small area inference from data integration using global-local priors

Dexter Cahoy and Joseph Sedransk[1]

Abstract

We present and apply methodology to improve inference for small area parameters by using data from several sources. This work extends Cahoy and Sedransk (2023) who showed how to integrate summary statistics from several sources. Our methodology uses hierarchical global-local prior distributions to make inferences for the proportion of individuals in Florida's counties who do not have health insurance. Results from an extensive simulation study show that this methodology will provide improved inference by using several data sources. Among the five model variants evaluated the ones using horseshoe priors for all variances have better performance than the ones using lasso priors for the local variances.

Key Words: Aberrant observations; Combining data; Health insurance; Horseshoe prior; Lasso prior; Pooling data; Shrinkage.

1. Introduction and background

Significant changes in survey sampling such as greatly reduced response rates and budgets have increased attention to developing and using new methodology to improve inferences. One promising approach is to add data from other sources to the data from a single, probability-based survey. Cahoy and Sedransk (2023) presented two methods for combining summary statistics from several sample surveys and applied them to data from the Behavioral Risk Factor Surveillance System (BRFSS) and the Small Area Health Insurance Estimates (SAHIE) Program at the US Census Bureau. These methods have more structure than those commonly used in survey sampling, and Cahoy and Sedransk (2023) showed the value of these methods in pooling such data. They also noted the potential value of this methodology in combining data from traditional sample surveys as well as non-probability surveys and, perhaps, administrative data. Cahoy and Sedransk (2023) made inference for each (Florida) county using the data only from that county, so may have lost the opportunity for further improvements in inference by combining the data across the counties. In the current paper we extend this work by providing methodology to make simultaneous inference for all counties when there are several data sources. However, the primary methodology in Cahoy and Sedransk (2023), Uncertain Pooling (Evans and Sedransk, 2001; Malec and Sedransk, 1992) cannot, at present, accommodate the computational demands of inference for many small areas together with several data sources. Thus, we use global-local (GL) models, a promising alternative with an extensive literature. There is a thorough summary in the review paper, Bhadra, Datta, Polson and Willard (2019). There are two papers in the survey sampling literature, Tang, Ghosh, Ha and Sedransk (2018) and Tang and Ghosh (2023). Results in Tang et al. (2018) motivated our choice to use GL models. Since both papers address a simpler case than we do, i.e., a single data source, we have extended the model in Tang et al. (2018).

1. Dexter Cahoy, University of Houston System, Houston, Texas, 77002. E-mail: cahoyd@uhd.edu; Joseph Sedransk, University of Maryland, College Park, Maryland, 20742. E-mail: jxs123@cwru.edu.

A simplified version of the Fay-Herriot model that Tang et al. (2018) use is

$$y_i = \theta_i + e_i, \quad \theta_i = \eta + u_i, \quad i = 1, \ldots, I \tag{1.1}$$

where y_i is the direct estimate for small area i and η is a constant. Here, $e = (e_1, \ldots, e_I)'$ and $u = (u_1, \ldots, u_I)'$ are assumed to be independent. The elements of e are independent with $e_i \sim N(0, V_i)$. Tang et al. (2018) note that V_i is assumed to be known to avoid problems with identifiability. The GL shrinkage prior for the u_i is

$$u_i \mid \lambda_i^2, \tau^2 \sim N(0, \lambda_i^2 \tau^2). \tag{1.2}$$

Instead of (1.1) and (1.2) Tang et al. (2018) replace η with a linear regression. They assume a locally uniform prior on the regression parameter, and consider several priors for the variances; see their Table 1. As seen in (1.2) there are two levels for the prior variances. The global variance, τ^2, provides shrinkage for all of the random effects while the local variances, λ_i^2, provide adjustments for the individual areas. If the priors are heavy-tailed, effective inference for both small and large random effects can be made. Tang and Ghosh (2023) extend Tang et al. (2018) by replacing (1.2) with a normal distribution with spatial structure. Another alternative to (1.2) is a spike-and-slab prior (Datta and Mandal, 2015), compared with the GL prior in Tang et al. (2018). That is, $u_i = \xi_i v_i$ where $\xi_1, \ldots, \xi_I$ are i.i.d. Bernoulli random variables with $P(\xi_i = 1) = \gamma$ and $v_i \sim \mathrm{N}(0, \zeta^2)$ independently. A recent paper, Smith and Griffin (2023), has several features that are similar to our extension of Tang et al. (2018) to inference on small areas when there are several sources. See Section 2 for details.

The rest of the paper is structured as follows. Section 2 describes our models and gives an explanation of our point estimator, the posterior mean for each FL county. Our first analysis, in Section 3, uses the data from SAHIE and BRFSS that Cahoy and Sedransk (2023) used to make inference about the proportion of adults without health insurance in the Florida counties. There are brief descriptions of the BRFSS survey and SAHIE program in Section 3.1. Credible intervals for the Florida county means are in Section 3.2 while the results of a comparison of several models using the SAHIE and BRFSS data are in Section 3.3. Section 3.4 is about the gains from using data from two sources rather than a single source. Section 4 has the results from an extensive simulation study based, in part, on the observed data set. There is also discussion of literature concerning the use of horseshoe and (Bayesian) lasso prior distributions. Section 5 has a discussion about the sampling variances, use of a larger number of data sources, model comparison and robustness, concluding with a summary.

2. Models and inference

Consideration of the two-source data for each of the counties in Florida suggested the following natural extension of the model in Tang et al. (2018). Let Y_{ij} denote the value of Y for the j-th source in the i-th county. For a fixed V_{ij},

$$
\begin{aligned}
Y_{ij} \mid \theta_{ij}, V_{ij} &\overset{\text{ind}}{\sim} N(\theta_{ij}, V_{ij}):\ j = 1,...,J;\ i = 1,...,I, \\
\theta_{ij} \mid \mu_i, \lambda_{ij}^2, \lambda_i^2, \tau_1^2 &\overset{\text{ind}}{\sim} N(\mu_i, \lambda_{ij}^2 \lambda_i^2 \tau_1^2), \\
\mu_i \mid \eta, \lambda_i^2, \tau_2^2 &\overset{\text{ind}}{\sim} N(\eta, \lambda_i^2 \tau_2^2), \\
f(\eta) &= \text{constant.}
\end{aligned} \tag{2.1}
$$

As in Cahoy and Sedransk (2023), Y is an estimate of the population proportion without health insurance. Note that the data available from BRFSS and SAHIE are only Y and the estimated standard error of Y. Here, inference for the county means, μ_i, is of primary interest.

Defining $\lambda = (\{\lambda_{ij}^2:\ j = 1,\ldots,J, i = 1,\ldots,I;\ \lambda_i^2:\ i = 1,\ldots,I\})'$ and $\tau = (\tau_1^2, \tau_2^2)'$, the joint prior density is

$$
f(\lambda, \tau) \propto \left[\prod_{i,j} f(\lambda_{ij}^2)\right] \left[\prod_i f(\lambda_i^2)\right] f(\tau_1^2)\, f(\tau_2^2).
$$

To choose the priors for the local and two global variances we have relied on discussion in the literature. Tang et al. (2018) note that Gelman (2006) and Polson and Scott (2012) prefer horseshoe priors for top level scale parameters, and there seems to be substantial additional support for this choice: for more details see Section 4.3.

A recent paper by Smith and Griffin (2023) uses a somewhat similar model in a different context, estimating demand for large assortments of differentiated goods. Their model is structured around a graphical representation of a product classification tree. In this tree, the lowest level represents the most detailed definition of product groups, while the highest level corresponds to the broadest categories. Using this tree as a framework, the authors develop hierarchical models that directly shrink product-level elasticities towards higher-level price elasticities. The structure in our problem is quite different, i.e., nesting of sources (j) within counties (i). Another assumption indicates that the objectives of the two papers are different. In the context of small area inference taking η to have a locally uniform distribution is the natural choice while Smith and Griffin (2023) assume a standard normal distribution at the highest level. Another difference is that Smith and Griffin (2023) only consider hierarchical structure for the local variances, i.e., $\lambda_{ij}^2 \lambda_i^2$ in (2.1), a limitation since we have found better performance for our problem using the non-hierarchical λ_{ij}^2 in the conditional variance of θ_{ij}. However, we have used ideas in Smith and Griffin (2023) in designing our simulation study, and clarifying some features of our approach.

Let υ denote a generic variance and $\kappa \sim \mathcal{C}^+(0,1)$, the half-Cauchy distribution with pdf $f(y) = 2/\pi(1+y^2), y \geq 0$. Then $\upsilon = \kappa^2$ has the *horseshoe* distribution with pdf $f(\upsilon) \propto \left[(1+\upsilon^2)\sqrt{\upsilon^2}\right]^{-1}$ for $\upsilon > 0$. The *lasso* distribution has pdf $f(\upsilon) \propto \exp(-\upsilon)$ for $\upsilon \geq 0$.

Our choices are

M11a: $\lambda_{ij}^2, \lambda_i^2, \tau_1^2, \tau_2^2$ have horseshoe distributions.

M11b is the same as M11a, except that λ_{ij}^2 and λ_i^2 have lasso distributions.

M1a is the same as M11a, except that $\theta_{ij} \sim \mathrm{N}(\mu_i, \lambda_{ij}^2\tau_1^2)$.

M1b is the same as M1a, except that λ_{ij}^2 and λ_i^2 have lasso distributions.

M12 has $\lambda_{ij}^2 = \lambda_i^2 = 1$. This is the usual hierarchical model.

The first models that we chose were M1a and M1b with $\theta_{ij} \sim \mathrm{N}(\mu_i, \lambda_{ij}^2\tau_1^2)$. We added the more general M11a and M11b (with variance $\lambda_{ij}^2\lambda_i^2\tau_1^2$), after reading Smith and Griffin (2023).

To investigate the posterior mean of μ_i given the observed data and variances assume M11a and define $A_i = \lambda_i^2\tau_2^2, \xi_{ij}^2 = V_{ij} + \lambda_{ij}^2\lambda_i^2\tau_1^2, \eta_i^2 = \left(\sum_{j=1}^J \xi_{ij}^{-2}\right)^{-1}$ and $\phi_i = A_i \big/ (A_i + \eta_i^2)$.

Defining $\bar{y}_i = \sum_{j=1}^J (y_{ij} / \xi_{ij}^2) \Big/ \sum_{j=1}^J (1/\xi_{ij}^2)$, the posterior mean of μ_i is

$$E(\mu_i \mid y, \Omega) = \phi_i \bar{y}_i + (1-\phi_i)\, \bar{y}_w \tag{2.2}$$

where $\Omega = (\lambda, \tau, \{V_{ij}\colon j = 1,\ldots,J;\ i = 1,\ldots,I\})$, $y = \{y_{ij}\colon j = 1,\ldots,J;\ i = 1,\ldots,I\}$ and $\bar{y}_w = \sum_{i=1}^I (\bar{y}_i / (A_i + \eta_i^2)) \Big/ \sum_{i=1}^I (1/(A_i + \eta_i^2))$.

The *within* county weight, $\xi_{ij}^{-2} \Big/ \sum_{j=1}^J \xi_{ij}^{-2}$, reduces to $V_{ij}^{-2} \Big/ \sum_{j=1}^J V_{ij}^{-2}$, if $\lambda_{ij}^2\lambda_i^2\tau_1^2 = 0$. If not, the local and global variance components will accommodate observed values that do not conform to the standard hierarchical model M12. (Hereafter, we refer to such observations as "aberrant".) Increasing V_{ij}, being an aberrant source (large λ_{ij}^2), or an aberrant county (large λ_i^2) will reduce the weight on y_{ij}. If V_{ij} is small relative to $\lambda_{ij}^2\lambda_i^2\tau_1^2$, $\xi_{ij}^2 = \lambda_{ij}^2\lambda_i^2\tau_1^2$ which implies that $\bar{y}_i = \sum_{j=1}^J (y_{ij} / \lambda_{ij}^2) \Big/ \sum_{j=1}^J (1/\lambda_{ij}^2)$. A_i is a measure of variability *across* counties while η_i^2 is a measure of variability *within* counties (sum of sampling variance and variance across sources within counties). So, $\phi_i = A_i \big/ (A_i + \eta_i^2)$ is the standard overall shrinkage factor. Finally, $\bar{y}_w$ is a weighted average of the $\bar{y}_i$ where the weights include the adjustments based on the local variances λ_i^2 and λ_{ij}^2.

3. Analysis of data from Florida counties

3.1 Background

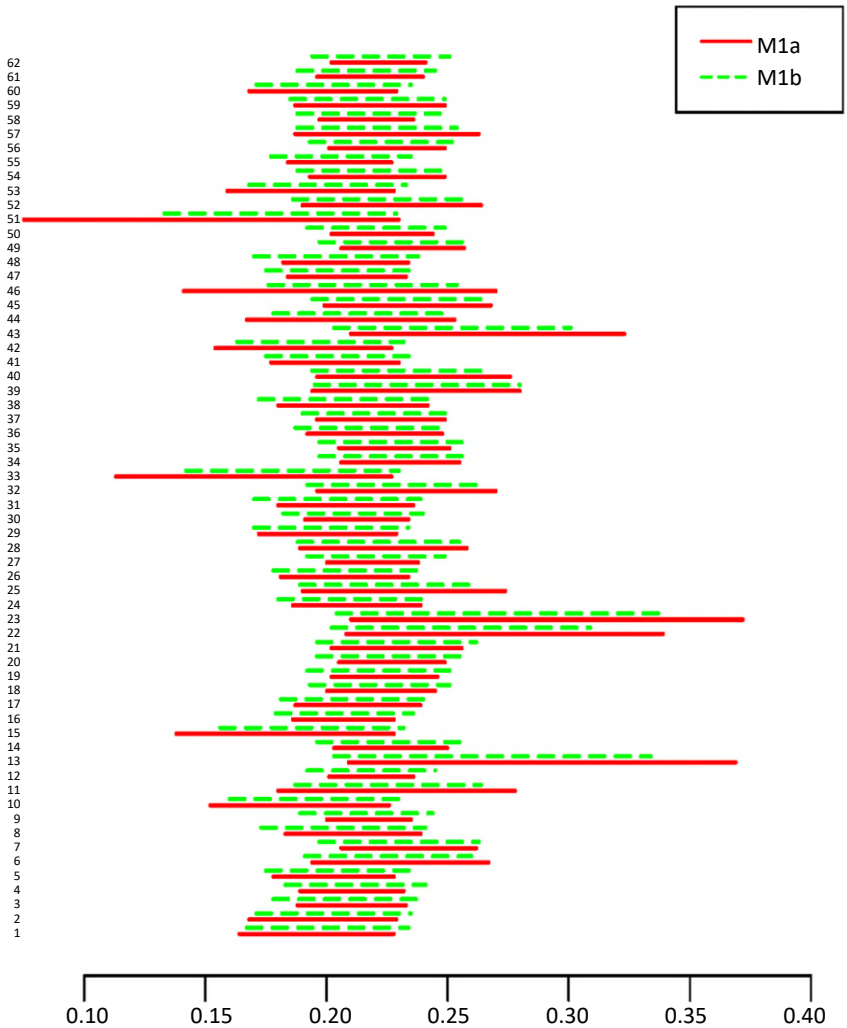
One data source is the Small Area Health Insurance Estimates Program. (The data can be obtained from https://www.census.gov/data/datasets/time-series/demo/sahie/estimates-acs.html.) The SAHIE program uses point estimates from the American Community Survey (ACS) together with administrative data such as Federal income tax returns and Medicaid/Children's Health Insurance Program (CHIP) participation rates. There is detailed area level modelling. The second source is the Behavioral Risk Factor Surveillance System, obtained through telephone interviews. (The data can be downloaded from https://www.flhealthcharts.gov/.) It uses a disproportionate stratified sample design. For additional details see Sections 3 and 7 of Cahoy and Sedransk (2023). For consistency with Cahoy and Sedransk (2023) we use 2010 data for each source. The BRFSS and SAHIE data available for each Florida county are an estimate of the population proportion uninsured (Y) and the estimated standard error of Y.

Since the analysis of the Florida county data is straightforward we present it first. As such it provides the background for the detailed simulation study in Section 4.

3.2 Inference

Figure 3.1 gives the 95% credible intervals for the Florida counties corresponding to models M1a and M1b. Overall, the pairs of intervals have similar centers: the two distributions, taken over the I counties, are almost the same (first quartiles, medians and third quartiles are within 0.002). While the median interval length is smaller for M1a (0.056) than M1b (0.064) there are eight counties where the intervals from M1a are substantially *wider*. These are situations where both the BRFSS and SAHIE point estimates are outliers or where there is a substantial difference between the two estimates. This is a first indication that M1a (with horseshoe priors for all variances) provides more appropriate inferences when there are aberrant observations. Comparing models M11a and M11b the conclusions are similar.

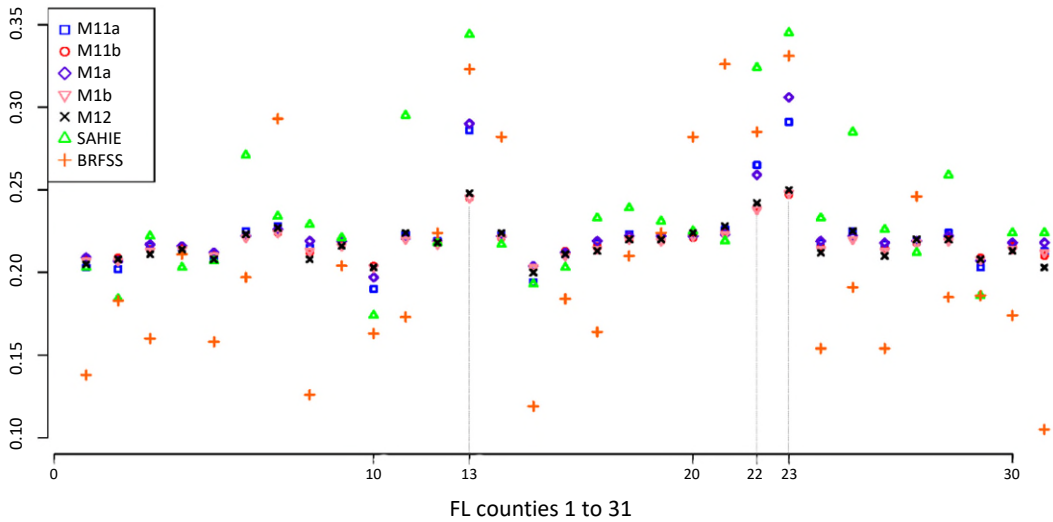
Figure 3.1 The 95% credible intervals for the 62 county means using M1a and M1b



3.3 Model comparisons

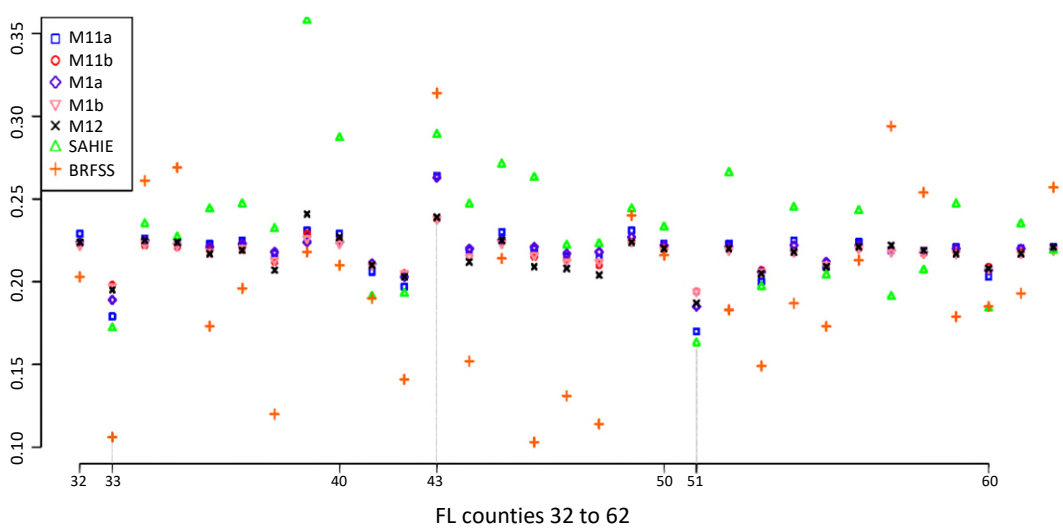
Figures 3.2 and 3.3 give the posterior mean, μ_i, for each Florida county for models M11a, M11b, M1a, M1b, the standard model M12, and the two point estimates. First, note that there is little variation in M12 across the counties. For most counties there is agreement among the five models. M1b and M11b (both lasso) have values of the posterior mean much closer to M12 than M1a and M11a (both horseshoe). This is most evident for the counties with both high and low values of the BRFSS and SAHIE point estimates such as 10, 13, 22 and 23 (see Figure 3.2). The values for M11b and M1b are similar while, typically, the values for M11a are a little more extreme than those for M1a.

Figure 3.2 Posterior means, BRFSS and SAHIE estimates for FL counties 1-31



Note: BRFSS: Behavioral Risk Factor Surveillance System; FL: Florida; SAHIE: Small Area Health Insurance Estimates.

Figure 3.3 Posterior means, BRFSS and SAHIE estimates for FL counties 32-62



Note: BRFSS: Behavioral Risk Factor Surveillance System; FL: Florida; SAHIE: Small Area Health Insurance Estimates.

For most counties there is not much difference between M1a and M1b. Generally, M1b is closer than M1a to the standard model M12. There are substantially more counties where M12 is closer to SAHIE than BRFSS. When one of SAHIE and BRFSS is small and the other is large, M1a and M1b tend to be similar and within the range defined by SAHIE and BRFSS. The most important finding is that when both SAHIE and BRFSS are large or small M1b shrinks much more than M1a. Similarly, M11b shrinks more than M11a. This is in accord with the literature. For their one stage model, i.e., without source effects, Tang et al. (2018) say that the "shrinkage factors under [lasso] priors are stochastically larger than those under [horseshoe] priors thus causing more shrinkage". Smith and Griffin (2023) add that "the tails of an exponential mixing density are lighter than the polynomial tails of the half-Cauchy, suggesting that the Bayesian lasso may tend to over-shrink large regression coefficients and under-shrink small ones relative to the horseshoe".

As seen in (2.2) the magnitude of ϕ determines the contribution of the observations from county i to the posterior mean of μ_i. Figures A.1-A.3 in the appendix show the posterior distributions of $\{\phi_i\colon i=1,\ldots,I\}$ for models M1a, M1b, M11a, M11b and M12. These figures display considerable differences among the models. For M12 the set of I distributions of ϕ show little variation which is markedly different from those for the four GL models. Comparing M1a (with horseshoe priors for all variances) with M1b (with lasso priors for the local variances and horseshoe priors for the global variances) there are notable differences, i.e., much greater variation for M1a both across and within counties. This is also seen for the counterparts of M1a and M1b, M11a and M11b, which differ from M1a and M1b by having local variances $\lambda_{ij}^2\lambda_i^2$ rather than λ_{ij}^2.

3.4 Two sources vs. one source

We use the Florida county data set to illustrate differences in inferences when there is only one data source rather than the two data sources. The one source model is

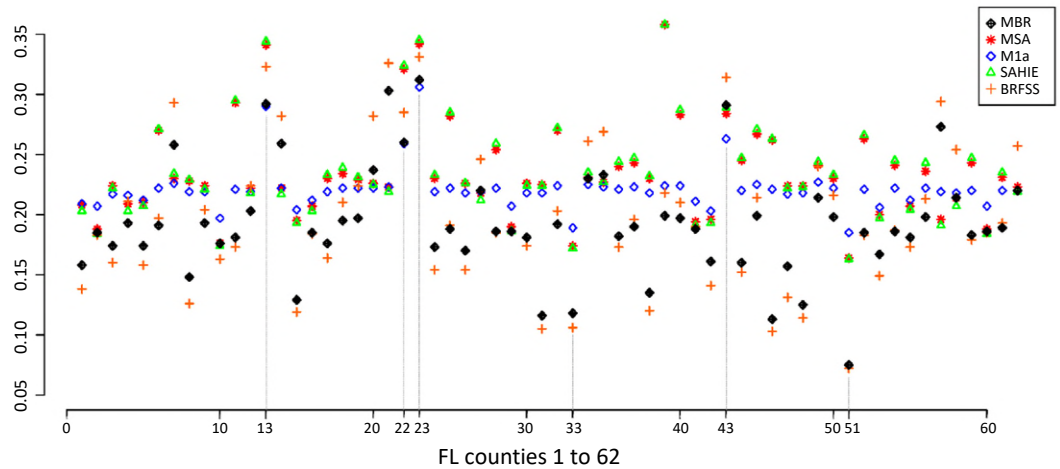
$$Y_i \mid \mu_i, V_i \overset{\text{ind}}{\sim} N(\mu_i, V_i),$$

$$\mu_i \mid \eta \overset{\text{ind}}{\sim} N(\eta, \lambda_i^2\tau^2),$$

$$f(\eta) = \text{ constant.}$$

The most meaningful comparison is with the BRFSS data set since it is the one with much larger observed standard errors, thus potentially benefiting from the addition of the SAHIE data when this is appropriate. One may consider the analysis in this section as an illustration or as a realistic choice because of (apparently) untested assumptions made in SAHIE. For each county Figure 3.4 gives the BRFSS and SAHIE point estimates and posterior means corresponding to M1a (horseshoe) and the one source horseshoe models (MBR, MSA) applied to the BRFSS and SAHIE data.

Figure 3.4 MBR, MSA, M1a, observed BRFSS and SAHIE estimates

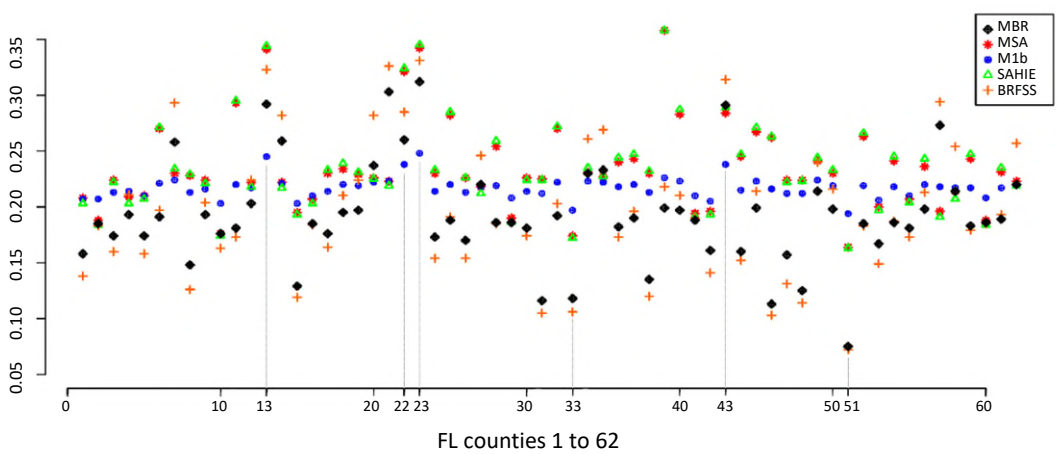


Note: BRFSS: Behavioral Risk Factor Surveillance System; FL: Florida; MBR: Model BRFSS; MSA: Model SAHIE; SAHIE: Small Area Health Insurance Estimates.

Typically, the BRFSS posterior mean (using the one source horseshoe model) adjusts the BRFSS point estimate by increasing smaller values and decreasing larger values of the latter. The M1a posterior mean adjusts the one source BRFSS posterior mean by increasing smaller values and decreasing larger values of the one source BRFSS posterior mean. With this data set the M1a posterior mean doesn't vary very much over the counties, so there are adjustments for a fairly small set of counties. But note that M1a has excellent properties, as shown in the results from the simulation study (Section 4).

The results for M11a are similar to those for M1a. This is not surprising since M11a and M1a differ only in the variance of θ_{ij} (see Section 2). With one important exception the conclusions above hold for M11b and M1b, both of which use lasso priors. Unlike for M11a and M1a, there is considerable shrinkage towards the common county value even for some counties (e.g., 13, 22, 23, 33, 43, 51) with somewhat extreme values of the BRFSS point estimate: see Figure 3.5 which has the same format as Figure 3.4.

Figure 3.5 MBR, MSA, M1b, observed BRFSS and SAHIE estimates



Note: BRFSS: Behavioral Risk Factor Surveillance System; FL: Florida; MBR: Model BRFSS; MSA: Model SAHIE; SAHIE: Small Area Health Insurance Estimates.

4. Simulation study

4.1 Introduction

Our simulation study is patterned after those in Tang et al. (2018) and Smith and Griffin (2023). In this section we compare the performance of models M1a, M1b, M11a and M11b with the base method M12. (See Section 2 for the definitions of these models.) We present the results obtained by generating data from the following models.

Let Y_{ij} denote the value of Y for the j-th source in the i-th county. For a fixed V_{ij}

$$Y_{ij} \overset{\text{ind}}{\sim} N(\theta_{ij}, V_{ij}):\ j = 1,\ldots,J;\ i = 1,\ldots,I,$$

$$\theta_{ij} \overset{\text{iid}}{\sim} N(\mu_i, \gamma_1^2),$$

$$\mu_i \overset{\text{iid}}{\sim} N(\eta, \gamma_2^2). \tag{4.1}$$

For the γ^2, we have used mixture and outlier representations. For a mixture, $\gamma^2 = \delta_1\tau_{21}^2 + (1-\delta_1)\,\tau_{22}^2$, $\delta_1 \in \{0,1\}$ with $P(\delta_1 = 1) = p_1$ and $\tau_{22}^2 =$ 0.05². For an outlier $\gamma^2 = \delta_2\tau_{11}^2$ with $\delta_2 \in \{0,1\}$ and $P(\delta_2 = 1) = p_2$. With a large number of quantities to specify we have simplified by not indexing by county or source. There are exceptions, i.e., for cases 5 and 6, described in Section 4.5.

To anchor this simulation study, we have chosen values of some quantities based on the BRFSS and SAHIE data: η is a constant, 0.25, with $I = 62$ Florida counties and $J = 2$ sources. Also, the sampling variances, V_{ij}, are the values observed in these surveys. Section 5 has further discussion about sampling variances, including the issue of identifiability noted by Tang et al. (2018).

Case 1 has both γ 's as outliers while Case 3 has both as mixtures. For Case 2 γ_1^2 is an outlier while γ_2^2 is a mixture. For Case 4 γ_1^2 is a mixture while γ_2^2 is an outlier. We considered many values of the probabilities (p_1, p_2) and variances (τ^2) and present those that are the most informative.

For each of the four cases and choice of the probabilities and the variances we generated 100 data sets. For each data set we estimated μ_i by its posterior mean, $\hat{\mu}_i$, and evaluated the fit using four common measures, i.e., average absolute relative deviation (ARB), average squared relative deviation (ASRB), average absolute deviation (AAD) and average squared deviation (ASD). We summarize for each data set as ARB $= I^{-1}\sum_{i=1}^{I}\frac{(|\hat{\mu}_i-\mu_i|)}{\mu_i}$, ASRB $= I^{-1}\sum_{i=1}^{I}\frac{(\hat{\mu}_i-\mu_i)^2}{\mu_i^2}$, AAD $= I^{-1}\sum_{i=1}^{I}|\,\hat{\mu}_i - \mu_i\,|$, and ASD $= I^{-1}\sum_{i=1}^{I}(\hat{\mu}_i - \mu_i)^2$ where I is the number of Florida counties. For each measure (ARB, ASRB, AAD, ASD) the final summary statistic is the median from the 100 data sets. To simplify, we present results only for ARB and ASRB.

4.2 Posterior computation

All of the calculations were implemented in R using the Advanced Cyberinfrastructure Coordination Ecosystem: Services & Support (ACCESS)'s Pittsburgh Supercomputing Center. We generated 100 data

samples and calculated posterior estimates using 18K MCMC runs and 3K burn-ins for each sample. To test convergence, we simulated five chains of length 7K (with 2K burn-ins) for the posterior distributions of the μ_i's, and obtained values of the split-R variant of Gelman's diagnostic close to 1.0 for all 62 counties. Section 11.4 of Gelman, Carlin, Stern, Dunson, Vehtari and Rubin (2013) has a discussion of convergence measures including the split-R variant. The full conditional distributions for models M11a and M11b are given in Appendix A. The corresponding Gibbs algorithms for M1a and M1b can be obtained by simplifying the given expressions.

4.3 Overall evaluation

Our comparisons are made relative to the standard model, M12, as discrepancy ratios. Model 12 is given in Section 2: it is the special case of the global-local model with $\lambda_{ij}^2 = \lambda_i^2 = 1$. For model Mx (x=1a, 1b, 11a, 11b) and ARB the discrepancy ratio is denoted by ARB(Mx)/ARB(M12) with small values of the ratio showing gains for Mx relative to M12. We summarize the distributions (over the set of specifications) of the discrepancy ratio for models M1a, M1b, M11a and M11b using the medians.

For Cases 1-4 the medians of ARB(M1a)/ARB(M12), ARB(M11a)/ARB(M12), ARB(M1b)/ARB(M12) and ARB(M11b)/ARB(M12) are (0.589, 0.633, 0.802, 0.853), (0.733, 0.760, 0.851, 0.874), (0.780, 0.821, 0.870, 0.915), and (0.786, 0.913, 0.835, 0.929). They show considerable gains for all of the global-local models with M1a having the smallest values (largest gain). The results for ASRB are similar.

Additional information is in Tables S1-S4 in the Supplementary Material (https://arxiv.org/pdf/2412.07824) corresponding to the four cases. In each table we present for ARB and ASRB, and each of the four discrepancy ratios, the minimum value, first quartile, median, mean, third quartile and maximum value (taken over the set of specifications).

Table 4.1 gives an overall summary. For each discrepancy ratio, case and model, Mx, the cell entry is the number of specifications where Mx has the smallest value among the set of models. Clearly, model M1a is preferred. Moreover, M1a has a smaller value of ARB(M1a)/ARB(M12) than ARB(M1b)/ARB(M12) for 99 of the 122 settings. The other horseshoe model, M11a, has a smaller value of ARB(M11a)/ARB(M12) than ARB(M11b)/ARB(M12) for 95 settings. The results for ASRB show a greater preference for these horseshoe models. See Table S21 for a breakdown by case.

These results lend additional credence to the literature showing a preference for the horseshoe prior. For example, Smith and Griffin (2023) say that "the tails of an exponential mixing density are lighter than the polynomial tails of the half-Cauchy, suggesting that the Bayesian lasso may tend to over-shrink large regression coefficients and under-shrink small ones relative to the horseshoe". Makalic and Schmidt (2016) add that the "choice of a half-Cauchy prior distribution over the global and local hyperparameters results in aggressive shrinkage of small coefficients (i.e., noise) and virtually no shrinkage of sufficiently large coefficients (i.e., signal). This is contrast to the well-known Bayesian lasso ... and Bayesian ridge hierarchies where the shrinkage effect is uniform across all coefficients".

For the simpler model in (1.1) and (1.2) Bhadra et al. (2019) show why the horseshoe prior is effective. This argument also holds for inference about the source effects in our (2.1), since the first two expressions in (2.1) are essentially the same as the model in (1.1) and (1.2) (drop the subscript i). For our preferred model, M1a, take $\lambda_i^2 = 1$. Then

$$E(\theta_{ij} \mid y_{ij}) = \left[1 - E(\kappa_{ij} \mid y_{ij})\right] y_{ij} + E(\kappa_{ij} \mid y_{ij})\, \mu_i \tag{4.2}$$

where the shrinkage weight, κ, is $\kappa_{ij} = V_{ij} \big/ (V_{ij} + \lambda_{ij}^2 \tau_1^2)$. Bhadra et al. (2019) write the marginal likelihood of y_{ij} in terms of κ_{ij}, noting that the posterior density of the shrinkage weight identifies signals and noise by letting $\kappa_{ij} \to 0$ and $\kappa_{ij} \to 1$. The marginal likelihood is zero when the shrinkage weight is zero, so it does not help identify the signals. However, as shown in Bhadra et al. (2019) using the horseshoe prior enables the posterior to approach both zero and one. This gives an indication why M1a, with horseshoe prior at the source level, has better performance than M1b, with lasso prior at the source level.

Table 4.1
For each case: number of specifications where a model has the best performance

		ARB		
Case	M11a	M11b	M1a	M1b
1	0	1	29	0
2	8	0	28	0
3	0	1	16	3
4	0	0	17	19
		ASRB		
Case	M11a	M11b	M1a	M1b
1	5	0	15	0
2	6	0	30	0
3	9	0	8	3
4	2	1	21	12

Note: ARB: average absolute relative deviation; ASRB: average squared relative deviation.

Tables 4.2-4.5 give more detailed information for M1a. Each table refers to a case, e.g., in Table 4.2 both γ^2's are outliers. Each row corresponds to a specification of the probabilities and the variances, τ^2. The first set of columns identify the probabilities and the τ^2, e.g., for Case 1 the components of γ_1^2 in (4.1) corresponding to the outlier model and the components of γ_2^2 in (4.1) corresponding to the outlier model. The last two columns display the discrepancy ratios for ARB and ASRB corresponding to model M1a, i.e., ARB(M1a)/ARB(M12) and ASRB(M1a)/ASRB(M12). Tables S5-S16 in the Supplementary Material give the same information for the other models. These tables can be used to investigate the effects of changing the values of the probabilities and variances: see Section 4.4.

4.4 Changing probabilities and variances

We next investigate the effects of changing the probabilities and values of the variances (τ^2). This is a new evaluation since these quantities are fixed in Tang et al. (2018) and Smith and Griffin (2023). Motivated by the results in Section 3.3 and the literature referenced in Section 2 we only summarize our findings about

properties of M1a. Values for the probabilities and variances were chosen, initially, based on the results in Section 3. Other values were added to try to sharpen the conclusions. In Tables 4.2-4.5 we present the results for a subset of the choices of the probabilities and variances that we considered. For the most part the gains for M1a increase as the probabilities and variances increase. There are some exceptions, not explained by further investigation.

Table 4.2
Simulation specifications and ARB(M1a)/ARB(M12), ASRB(M1a)/ASRB(M12) for Case 1

	p_2 for μ_i	p_2 for θ_{ij}	τ_{11} for μ_i	τ_{11} for θ_{ij}	M1a/M12(ARB)	M1a/M12(ASRB)
1	0.100	0.100	0.025	0.025	0.803	0.670
2			0.050	0.050	0.563	0.466
3			0.100	0.100	0.345	0.226
4			0.200	0.200	0.244	0.058
5			0.050	0.100	0.487	0.327
6			0.100	0.050	0.425	0.282
7	0.200	0.200	0.025	0.025	0.828	0.832
8			0.050	0.050	0.635	0.601
9			0.100	0.100	0.429	0.293
10			0.200	0.200	0.298	0.123
11			0.050	0.100	0.636	0.544
12			0.100	0.050	0.562	0.486
13	0.400	0.400	0.025	0.025	0.896	0.874
14			0.050	0.050	0.772	0.784
15			0.100	0.100	0.636	0.500
16			0.200	0.200	0.612	0.481
17			0.050	0.100	0.780	0.727
18			0.100	0.050	0.728	0.711
19	0.100	0.200	0.025	0.025	0.923	0.888
20			0.050	0.050	0.595	0.442
21			0.100	0.100	0.405	0.328
22			0.200	0.200	0.255	0.096
23			0.050	0.100	0.609	0.579
24			0.100	0.050	0.456	0.370
25	0.200	0.100	0.025	0.025	0.796	0.773
26			0.050	0.050	0.600	0.504
27			0.100	0.100	0.415	0.240
28			0.200	0.200	0.309	0.093
29			0.050	0.100	0.544	0.369
30			0.100	0.050	0.583	0.433

Note: ARB: average absolute relative deviation; ASRB: average squared relative deviation.

Consider Case 1 (both "outliers", Table 4.2) and ARB. The gains for M1a (vs. M12) increase, as expected, as both τ_{11} 's, $\tau_{11}(\mu_i)$ for μ_i and $\tau_{11}(\theta_{ij})$ for θ_{ij} increase. The gains for M1a are generally larger for $\left(\tau_{11}(\mu_i), \tau_{11}(\theta_{ij})\right) =$ (0.10, 0.05) than for (0.05, 0.10). That is, the gains are larger when the variance of μ_i is larger than the variance of θ_{ij} in the two outlier models. The gains for M1a decrease as both p_2 's, $p_2(\mu_i)$ and $p_2(\theta_{ij})$, increase from 0.1 to 0.2 to 0.4. This occurs because larger p_2 in the outlier model makes it closer to M12. There is no pattern in the gains associated with $\left(p_2(\mu_i), p_2(\theta_{ij})\right) =$ (0.1, 0.2) vs. those associated with $\left(p_2(\mu_i), p_2(\theta_{ij})\right) =$ (0.2, 0.1). The patterns for ASRB are similar to those for ARB.

Table 4.3
Simulation specifications and ARB(M1a)/ARB(M12), ASRB(M1a)/ASRB(M12) for Case 2

	p_1 for μ_i	p_2 for θ_{ij}	τ_{21} for μ_i	τ_{11} for θ_{ij}	M1a/M12(ARB)	M1a/M12(ASRB)
1	0.100	0.100	0.100	0.050	0.942	0.892
2				0.100	0.725	0.520
3				0.200	0.502	0.255
4			0.200	0.050	0.894	0.816
5				0.100	0.723	0.707
6				0.200	0.551	0.375
7		0.200	0.100	0.050	0.936	0.887
8				0.100	0.709	0.473
9				0.200	0.487	0.166
10			0.200	0.050	0.887	0.823
11				0.100	0.739	0.569
12				0.200	0.498	0.217
13		0.400	0.100	0.050	0.951	0.873
14				0.100	0.742	0.481
15				0.200	0.525	0.259
16			0.200	0.050	0.923	0.862
17				0.100	0.754	0.617
18				0.200	0.556	0.330
19	0.200	0.100	0.100	0.050	0.959	0.940
20				0.100	0.713	0.579
21				0.200	0.483	0.252
22			0.200	0.050	0.924	0.826
23				0.100	0.702	0.570
24				0.200	0.519	0.405
25		0.200	0.100	0.050	0.943	0.892
26				0.100	0.764	0.616
27				0.200	0.465	0.203
28			0.200	0.050	0.886	0.723
29				0.100	0.749	0.515
30				0.200	0.492	0.245
31		0.400	0.100	0.050	0.917	0.725
32				0.100	0.775	0.603
33				0.200	0.521	0.260
34			0.200	0.050	0.919	0.793
35				0.100	0.727	0.668
36				0.200	0.572	0.411

Note: ARB: average absolute relative deviation; ASRB: average squared relative deviation.

Next, consider Case 2 (μ_i is "mixture"; θ_{ij} is "outlier", Table 4.3) and ARB. The gains for M1a (vs. M12) increase, as expected, as $\tau_{21}(=\tau_{11})$ increases. The gains for M1a are larger for $(\tau_{21}, \tau_{11}) =$ (0.1, 0.2) than for (0.2, 0.1). That is, the gains are larger when the variance of θ_{ij} is larger than the variance of μ_i. The gains for M1a increase as $p_1(=p_2)$ increases from 0.1 to 0.2 while there are only small differences for $(p_1, p_2) =$ (0.1, 0.2) vs. (0.2, 0.1). The patterns for ASRB are similar to those for ARB for the specifications with changing variances. For the specifications with changing probabilities the patterns are inconsistent.

Case 3 has both μ_i and θ_{ij} as "mixtures" (Table 4.4). For ARB the gains increase as τ_{21} for μ_i increases from 0.1 to 0.2 to 0.4 while the changes in the gains are usually small as τ_{21} for θ_{ij} increases from 0.1 to 0.2 to 0.4. There are only small changes as p_1 for μ_i, $p_1(\mu_i)$, increases from 0.1 to 0.2, and no pattern as p_1 for θ_{ij}, $p_1(\theta_{ij})$, increases. The patterns for ASRB are similar to those for ARB.

Case 4 has an "outlier" model for μ_i and a "mixture" model for θ_{ij} (Table 4.5). For ARB the gains increase as τ_{11} increases from 0.05 to 0.10 to 0.20 and τ_{21} increases from 0.1 to 0.2 to 0.4. For almost all specifications there is an increase in the gains as p_1 and p_2 increase from 0.1 to 0.2. The patterns for ASRB are similar to those for ARB.

Table 4.4
Simulation specifications and ARB(M1a)/ARB(M12), ASRB(M1a)/ASRB(M12) for Case 3

	p_1 for μ_i	p_1 for θ_{ij}	τ_{21} for μ_i	τ_{21} for θ_{ij}	M1a/M12(ARB)	M1a/M12(ASRB)
1	0.100	0.100	0.100	0.100	1.050	1.102
2			0.200	0.200	0.844	0.452
3			0.400	0.400	0.655	0.305
4			0.200	0.400	0.680	0.214
5			0.400	0.200	0.828	0.562
6	0.200	0.200	0.100	0.100	1.001	0.960
7			0.200	0.200	0.766	0.470
8			0.400	0.400	0.636	0.427
9			0.200	0.400	0.619	0.275
10			0.400	0.200	0.783	0.763
11	0.100	0.200	0.100	0.100	1.025	0.997
12			0.200	0.200	0.786	0.527
13			0.400	0.400	0.619	0.353
14			0.200	0.400	0.671	0.429
15			0.400	0.200	0.778	0.612
16	0.200	0.100	0.100	0.100	1.006	1.005
17			0.200	0.200	0.852	0.563
18			0.400	0.400	0.609	0.301
19			0.200	0.400	0.596	0.296
20			0.400	0.200	0.904	0.893

Note: ARB: average absolute relative deviation; ASRB: average squared relative deviation.

Table 4.5
Simulation specifications and ARB(M1a)/ARB(M12), ASRB(M1a)/ASRB(M12) for Case 4

	p_2 for μ_i	τ_{11} for μ_i	p_1 for θ_{ij}	τ_{21} for θ_{ij}	M1a/M12(ARB)	M1a/M12(ASRB)
1	0.100	0.050	0.100	0.100	1.353	1.545
2				0.200	1.175	1.344
3				0.400	0.815	0.777
4			0.200	0.100	1.270	1.402
5				0.200	1.122	1.254
6				0.400	0.775	0.807
7		0.100	0.100	0.100	1.059	0.725
8				0.200	0.865	0.584
9				0.400	0.749	0.504
10			0.200	0.100	1.025	0.693
11				0.200	0.804	0.684
12				0.400	0.760	0.817
13		0.200	0.100	0.100	0.687	0.444
14				0.200	0.651	0.298
15				0.400	0.548	0.223
16			0.200	0.100	0.673	0.429
17				0.200	0.644	0.294
18				0.400	0.582	0.424

Note: ARB: average absolute relative deviation; ASRB: average squared relative deviation.

Table 4.5 (continued)
Simulation specifications and ARB(M1a)/ARB(M12), ASRB(M1a)/ASRB(M12) for Case 4

	p_2 for μ_i	τ_{11} for μ_i	p_1 for θ_{ij}	τ_{21} for θ_{ij}	M1a/M12(ARB)	M1a/M12(ASRB)
19	0.200	0.050	0.100	0.100	1.183	1.288
20				0.200	1.088	1.002
21				0.400	0.875	0.679
22			0.200	0.100	1.216	1.076
23				0.200	1.009	1.007
24				0.400	0.747	0.822
25		0.100	0.100	0.100	0.832	0.586
26				0.200	0.797	0.472
27				0.400	0.710	0.426
28			0.200	0.100	0.903	0.637
29				0.200	0.821	0.594
30				0.400	0.733	0.610
31		0.200	0.100	0.100	0.695	0.559
32				0.200	0.645	0.385
33				0.400	0.502	0.160
34			0.200	0.100	0.679	0.450
35				0.200	0.616	0.429
36				0.400	0.590	0.374

Note: ARB: average absolute relative deviation; ASRB: average squared relative deviation.

4.5 Two sources vs. one source

With data from two sources one could make inference using the data from only one source. Our investigation started with data generated from Cases 1-4. For each case and specification we calculated ARB and ASRB using models M1a, M1b, M11a, M11b and M12. We also used MSA and MBR, one source models (Section 3.4) using only the SAHIE and BRFSS data. As in Section 3.2 we summarize for each case using discrepancy ratios, here A(M12)/A(M1a), A(MSA)/A(M1a) and A(MBR)/A(M1a) where A denotes either ARB or ASRB. Tables 4.6-4.9 summarize the results for Cases 1-4. For each ratio we give the minimum value, first quartile, median, mean, third quartile and maximum value among the set of specifications for the case. Except for a few minimum values all discrepancy ratios exceed 1, meaning that there are losses by using M12, MSA and MBR relative to M1a. For ARB the median values of the ratio ARB(MBR)/ARB(M1a) are very large, i.e., (2.530, 2.255, 1.782, 2.703) for Cases 1-4. This shows the benefit of using a GL model, here M1a. The median values for ASRB(MBR)/ASRB(M1a) (4.593, 5.319, 3.636, 6.021) are even larger and provide additional evidence about using M1a.

Similarly, the four distributions of A(MSA)/A(M1a) have large values, albeit usually smaller than for A(MBR)/A(M1a). For example, for ARB the medians of {ARB(MSA)/ARB(M1a), ARB(MBR)/ARB(M1a)} are {(1.875, 2.530), (1.255, 2.255), (1.724, 1.782), (2.879, 2.703)}.

We next refined the specifications in Cases 1-4 by assuming different distributions for the θ 's. This is more realistic in that, typically, the two sources would be expected to have different characteristics. We also simplified by assuming no aberrant values at the county level. Throughout we generate data with Source 1 as BRFSS and Source 2 as SAHIE. For Case 5 the model used to generate the data is

$$y_{ij} \overset{\text{ind}}{\sim} N(\theta_{ij}, V_{ij})$$

$$\theta_{i1} \overset{\text{ind}}{\sim} N(\mu_i, \tau_1^2)$$

$$\theta_{i2} \overset{\text{ind}}{\sim} N(\mu_i, \delta\tau_2^2)$$

where $\delta \in \{0,1\}$ and $P(\delta = 1) = p$. Finally, $\mu_i \overset{\text{iid}}{\sim} N(0.25, \tau^2)$.

Table 4.6
Summary statistics for ARB and ASRB for Case 1: One vs. two sources

	ARB		
	M12/M1a	MSA/M1a	MBR/M1a
Min.	1.141	1.089	1.458
1st.Qu.	1.494	1.516	2.019
Median	1.722	1.875	2.530
Mean	1.996	2.044	2.659
3rd.Qu.	2.219	2.316	3.242
Max	4.737	3.950	4.969
	ASRB		
	M12/M1a	MSA/M1a	MBR/M1a
Min	1.176	1.319	1.826
1st.Qu	1.786	2.153	2.790
Median	2.235	3.934	4.593
Mean	3.329	5.741	6.526
3rd.Qu	3.553	7.445	8.436
Max	18.581	32.331	30.281

Note: ARB: average absolute relative deviation; ASRB: average squared relative deviation; BRFSS: Behavioral Risk Factor Surveillance System; MBR: Model BRFSS; MSA: Model SAHIE; SAHIE: Small Area Health Insurance Estimates.

Table 4.7
Summary statistics for ARB and ASRB for Case 2: One vs. two sources

	ARB		
	M12/M1a	MSA/M1a	MBR/M1a
Min	1.012	0.965	1.871
1st.Qu	1.123	1.107	2.044
Median	1.345	1.255	2.255
Mean	1.482	1.364	2.235
3rd.Qu	1.947	1.620	2.419
Max	2.113	2.054	2.608
	ASRB		
	M12/M1a	MSA/M1a	MBR/M1a
1st.Qu	1.247	1.639	3.568
Median	1.650	2.520	4.716
Mean	2.279	3.546	5.319
3rd.Qu	3.147	5.855	6.963
Max	5.246	8.317	10.078

Note: ARB: average absolute relative deviation; ASRB: average squared relative deviation; BRFSS: Behavioral Risk Factor Surveillance System; MBR: Model BRFSS; MSA: Model SAHIE; SAHIE: Small Area Health Insurance Estimates.

Table 4.8
Summary statistics for ARB and ASRB for Case 3: One vs. two sources

	ARB		
	M12/M1a	MSA/M1a	MBR/M1a
Min	0.943	1.370	1.445
1st.Qu	1.155	1.551	1.594
Median	1.254	1.724	1.782
Mean	1.304	1.813	1.851
3rd.Qu	1.510	1.964	1.985
Max	1.733	2.643	2.653
	ASRB		
	M12/M1a	MSA/M1a	MBR/M1a
Min	0.935	1.787	2.139
1st.Qu	1.081	2.198	2.395
Median	1.881	3.509	3.636
Mean	2.096	3.846	4.127
3rd.Qu	2.321	4.307	4.556
Max	4.867	8.459	12.665

Note: ARB: average absolute relative deviation; ASRB: average squared relative deviation; BRFSS: Behavioral Risk Factor Surveillance System; MBR: Model BRFSS; MSA: Model SAHIE; SAHIE: Small Area Health Insurance Estimates.

Table 4.9
Summary statistics for ARB and ASRB for Case 4: One vs. two sources

	ARB		
	M12/M1a	MSA/M1a	MBR/M1a
Min	0.726	1.681	1.590
1st.Qu	1.019	2.400	2.317
Median	1.260	2.879	2.703
Mean	1.269	3.150	2.904
3rd.Qu	1.382	3.564	3.316
Max	2.116	6.413	5.740
	ASRB		
	M12/M1a	MSA/M1a	MBR/M1a
Min	0.530	1.326	1.154
1st.Qu	1.002	3.203	3.393
Median	1.407	6.507	6.021
Mean	1.817	11.151	10.253
3rd.Qu	2.040	11.854	12.164
Max	5.935	56.195	46.653

Note: ARB: average absolute relative deviation; ASRB: average squared relative deviation; BRFSS: Behavioral Risk Factor Surveillance System; MBR: Model BRFSS; MSA: Model SAHIE; SAHIE: Small Area Health Insurance Estimates.

We used 18 specifications which included all arrangements of $\tau = 0.05$, $\tau_1 = (0.005, 0.010)$, $p = (0.10, 0.20, 0.40)$ and $\tau_2 = (0.05, 0.10, 0.20)$: see Table 4.10. The results, summarized in Table 4.11, show almost no specifications where using the BRFSS data as a single source is preferable to using M1a. For ARB and ASRB the median values of A(MBR)/A(M1a) (across the 18 specifications) are quite large, i.e., 1.663 and 2.070. (Note that another study, using larger values of τ, produced very similar results.) Looking at the 18 individual ratios, A(MBR)/A(M1a), the values of the ratios decrease as p and τ_2 increase, as expected. The value of τ_1 has only a minimal effect. See Tables 4.10, S17 and S18.

Table 4.10
Simulation specifications for Case 5: $\mu_i \sim N(0.25, \tau^2), \theta_{i1} \sim N(\mu_i, \tau_1^2), \theta_{i2} \sim N(\mu_i, \delta\tau_2^2), \delta \sim \text{Ber}(p)$

	τ	τ_1	p	τ_2
1	0.05	0.005	0.1	0.05
2				0.1
3				0.2
4			0.2	0.05
5				0.1
6				0.2
7			0.4	0.05
8				0.1
9				0.2
10		0.01	0.1	0.05
11				0.1
12				0.2
13			0.2	0.05
14				0.1
15				0.2
16			0.4	0.05
17				0.1
18				0.2

Table 4.11
Summary statistics for ARB and ASRB for Case 5

	ARB		
	M12/M1a	MSA/M1a	MBR/M1a
Min	1.000	1.080	1.088
1st.Qu	1.065	1.346	1.443
Median	1.162	1.485	1.663
Mean	1.285	1.786	1.694
3rd.Qu	1.326	2.002	2.025
Max	2.030	3.435	2.342
	ASRB		
	M12/M1a	MSA/M1a	MBR/M1a
Min	1.000	1.468	0.962
1st.Qu	1.133	2.233	1.712
Median	1.348	4.369	2.070
Mean	1.681	7.033	2.377
3rd.Qu	1.851	11.494	3.112
Max	3.899	20.498	4.465

Note: ARB: average absolute relative deviation; ASRB: average squared relative deviation; BRFSS: Behavioral Risk Factor Surveillance System; MBR: Model BRFSS; MSA: Model SAHIE; SAHIE: Small Area Health Insurance Estimates.

For Case 6 the model used to generate the data is the same as that for Case 5 except that $p = 1$. Since Case 6 assumes no aberrations a global-local model is unnecessary. Thus, it is not surprising that there are specifications where MBR is preferred. For example, the medians of A(MBR)/A(M1a) over the 12 specifications are 0.854 and 0.731 for ARB and ASRB (see Table 4.13). The 12 specifications are listed in Table 4.12. As seen in Tables S19 and S20 the value of τ_1 has little effect. However, as τ_2 increases A(MBR)/A(M1a) decreases while A(MSA)/A(M1a) increases. Thus, for large values of τ_2 there are benefits to using MBR rather than the GL model M1a. But, as seen in Case 5, it seems unlikely that using only one source of data will be preferable when (a) there are significant aberrant observations, and (b) the single source is the one with small bias and large sampling variance.

Table 4.12
Simulation specifications for Case 6: $\mu_i \sim N\ (0.25,\ \tau^2), \theta_{ij} \sim N(\mu_i, \tau_j^2)$

	τ	τ_1	τ_2
1	0.05	0.005	0.01
2			0.02
3			0.05
4			0.1
5			0.2
6			0.4
7		0.01	0.01
8			0.02
9			0.05
10			0.1
11			0.2
12			0.4

Table 4.13
Summary statistics for ARB and ASRB for Case 6

	ARB		
	M12/M1a	MSA/M1a	MBR/M1a
Min	0.954	1.065	0.546
1st.Qu	0.967	1.208	0.634
Median	1.000	2.293	0.854
Mean	1.027	3.283	1.097
3rd.Qu	1.099	4.596	1.495
Max	1.170	8.230	2.168
	ASRB		
	M12/M1a	MSA/M1a	MBR/M1a
Min	0.882	1.083	0.293
1st.Qu	0.911	1.452	0.380
Median	0.975	5.067	0.731
Mean	1.030	15.203	1.547
3rd.Qu	1.182	18.322	2.420
Max	1.289	60.459	4.745

Note: ARB: average absolute relative deviation; ASRB: average squared relative deviation; BRFSS: Behavioral Risk Factor Surveillance System; MBR: Model BRFSS; MSA: Model SAHIE; SAHIE: Small Area Health Insurance Estimates.

5. Discussion and summary

In the context of small area estimation from a single survey, Datta and Lahiri (1995) developed a hierarchical Bayes method using a general class of scale mixture of normal distributions (see Model 1 on page 314). For certain choices of prior their hierarchical Bayes methodology ensures robustness against outliers (see their Theorem 5). In the context of a single survey the horseshoe prior is included in the class of priors proposed in Datta and Lahiri (1995), so protects against outliers in the sense of their Theorem 5. Extension to the several survey case, as in this paper, would be valuable.

A referee asked about formal model comparisons. For the model in (1.1) and (1.2) (with η as a linear regression) Tang et al. (2018) consider choosing a single model using deviance information criterion (DIC), and find in their simulation study that "the selected model produces deviation measures close to the smallest among all candidate models and better coverage rate than the DM model". However, there has been extensive criticism of DIC, and one should be cautious about using it for models such as (1.1) and (1.2) and

for the more complex models that we have used. Plummer (2008) is a thorough investigation of DIC that details several significant problems. Carlin and Louis (2009) identify additional problems with DIC, and also describe (Section 4.6.2) another general approach, i.e., predictive model selection.

It is of interest to investigate the effect of having a larger number of sources of information. We did a preliminary study using the four cases that are the basis of our simulation study. To illustrate, consider the 30 specifications for Case 1 in Table 4.2. We generated data as in Section 3 for $J = 2$ and $J = 4$ sources. The V_{ij} were selected as a bootstrap sample from the $\hat{V}_{ij}$. There are, as expected, substantial gains as measured using the ratios $(\mathrm{A}(\mathrm{Mx}(J=4))/\mathrm{A}(\mathrm{Mx}(J=2))$ where A denotes ARB or ASRB. But, they vary considerably over the choices of the probabilities and variances. For example, for M1a the smallest and largest values of the ratio are 0.142 and 0.767. Consider M1a and ARB: as both τ_{11} 's increase the gains from using $J = 4$ sources decrease. As both p_2 's increase from 0.10 to 0.20 to 0.40 the dominant pattern is for the ratio to decrease, then increase. Further study is needed to get more definitive results.

Making inference about the sampling variances, the V_{ij}, is a challenging problem. The most common method is to assume a model for the $\hat{V}_{ij}$, the estimates of the V_{ij}. However, with only a single sample, $\{\hat{V}_{ij}\colon j=1,\ldots,J;\ i=1,\ldots,I\}$, one cannot verify the model. Methodology proposed in the literature is reviewed by Cahoy and Sedransk (2023). In the GL setting, but with only one data set, Tang et al. (2018) assume fixed V_i to avoid problems with identifiability. Our situation is more complex, i.e., with small areas and several data sources.

As future research we envision modelling the $\{\hat{V}_{ij}\colon j=1,\ldots,J;\ i=1,\ldots,I\}$ using data sets from prior years. An additional extension is to include covariates in our models.

The objective of this paper is to present methodology to improve inference for "small area" parameters by using data from several sources. While there is an extensive literature about global-local models there appears to be only one other paper, Smith and Griffin (2023), that uses a several stage model. Moreover, the application in Smith and Griffin (2023) is completely different from ours. Consequently, it was necessary to supplement our analysis of the Florida county data with an extensive simulation study to show properties of this methodology.

We introduce a set of global-local models and explain how these models combine information across counties and sources. Analyzing the BRFSS and SAHIE data using these models provides an opportunity to show how these models treat aberrant observations. This also includes a comparison of inferences using all of the data with inferences based on only one data source. In the extensive simulation study the data are selected from both outlier and mixture models. These results show that the proposed methodology can be used to provide improved inferences for small area parameters by using several data sources.

Acknowledgements

The authors are grateful to Advanced Cyberinfrastructure Coordination Ecosystem: Services & Support (ACCESS) for the computing support, and to several reviewers who posed interesting questions, valuable for future research.

Appendix

A. The ϕ_i's

Figure A.1 The ϕ_i 's for M1a (top) and M1b (bottom)

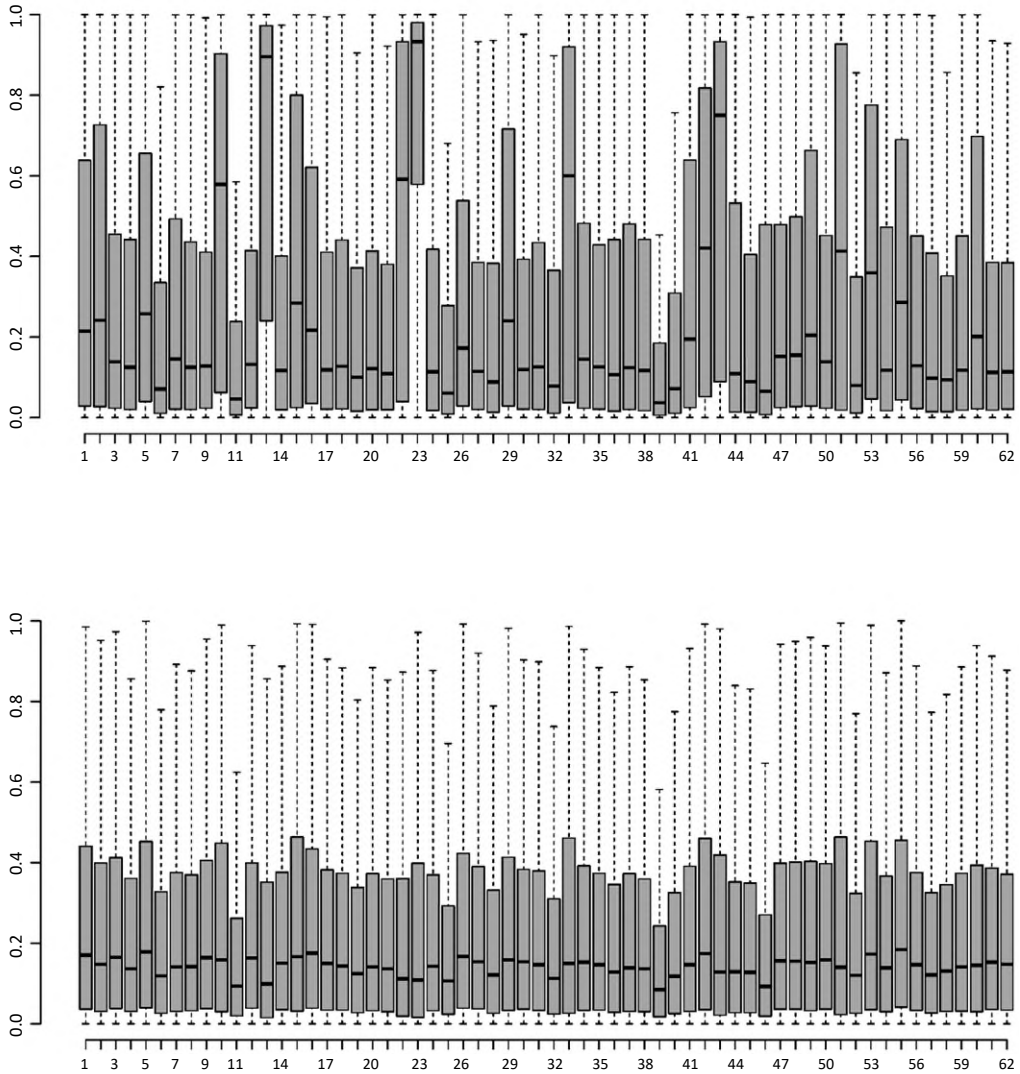
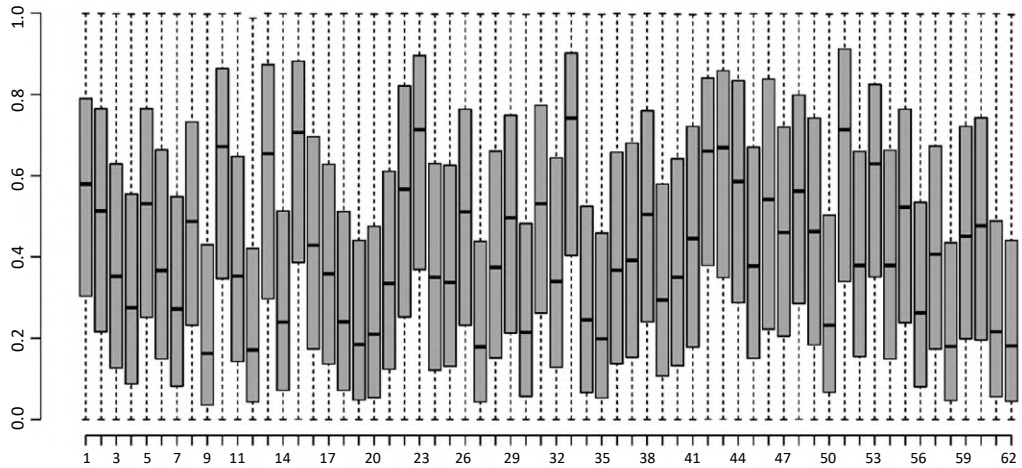


Figure A.2 The ϕ_i 's for M11a (top) and M11b (bottom)



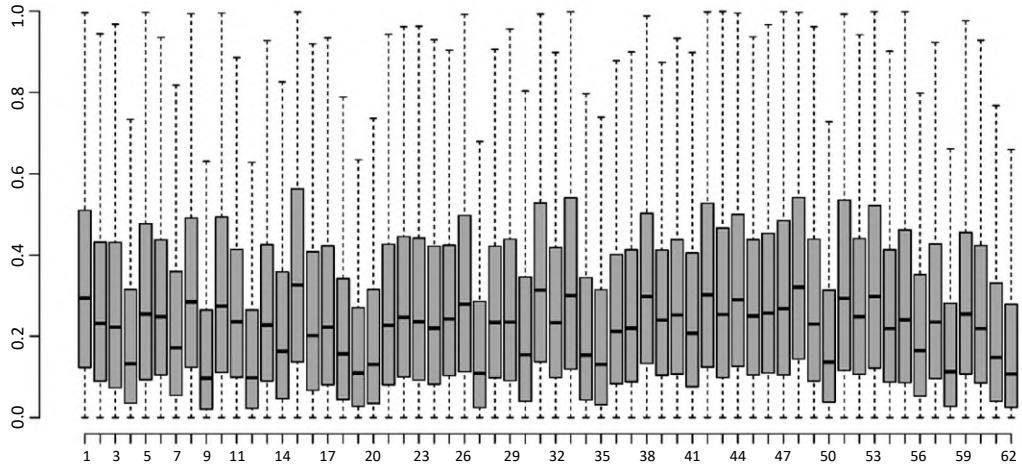
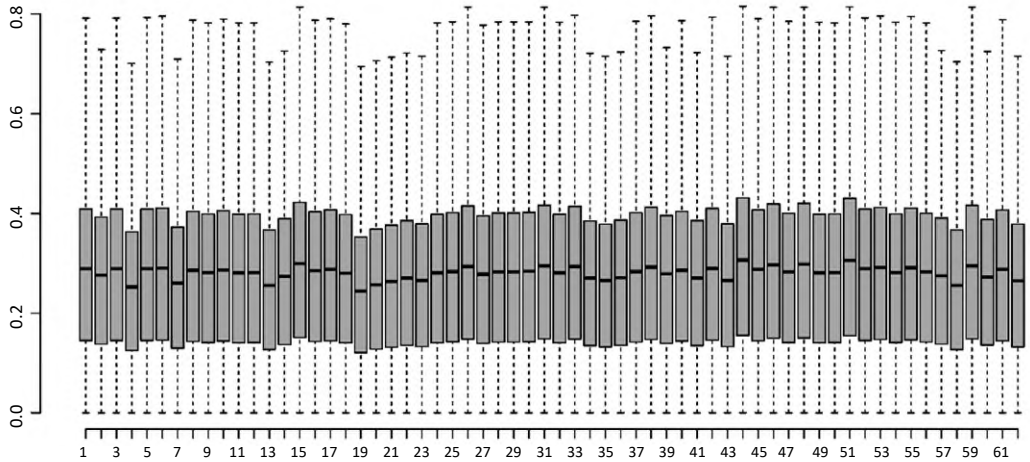


Figure A.3 The ϕ_i 's for M12



B. Full conditional distributions for M11a and M11b

We present below the full conditional distributions for the interaction effects, θ_{ij}, the small area expected values, μ_i, and the local and global variances.

Interaction effects: $\theta_{ij} \mid \text{else} \sim N(\bar{\theta}_{ij}, V_{\theta_{ij}})$ where

$$\bar{\theta}_{ij} = \frac{Y_{ij}/V_{ij} + \mu_i/\lambda_{ij}^2\lambda_i^2\tau_1^2}{1/V_{ij} + 1/\lambda_{ij}^2\lambda_i^2\tau_1^2} \text{ and } V_{\theta_{ij}} = \frac{1}{1/V_{ij} + 1/\lambda_{ij}^2\lambda_i^2\tau_1^2}.$$

Small area means: $\mu_i \mid \text{else} \sim N(\bar{\mu}_i, V_{\mu_i})$ where

$$\bar{\mu}_i, = \frac{\bar{\theta}_i c_i + \eta d_i}{c_i + d_i}, \qquad V_{\mu_i} = \frac{1}{c_i + d_i},$$

$a_{ij} = \lambda_{ij}^2\lambda_i^2\tau_1^2, c_i = \sum_j (1/a_{ij}), d_i = 1/(\lambda_i^2\tau_2^2)$, and $\bar{\theta}_i = \frac{\sum_j (\theta_{ij}/a_{ij})}{\sum_j (1/a_{ij})}$.

Local variances: Following Makalic and Schmidt (2016), we use the following Inverse Gamma (IG (shape, rate)) scale mixture representation: If $\kappa \mid \xi \sim \text{IG}(1/2, 1/\xi)$ and $\xi \sim \text{IG}(1/2, 1)$ then $\kappa \sim \mathcal{C}^+(0,1)$ and $\nu = \kappa^2$ (see Section 2) has the horseshoe distribution. Note that the IG (shape, rate) density is proportional to $x^{(-1-\text{shape})}e^{(-\text{rate}/x)}$.

Using the horseshoe prior on the local variances,

$$\lambda_{ij}^2 \mid \text{else} \sim \text{IG}\left(1, \frac{(\theta_{ij} - \mu_i)^2}{2\lambda_i^2\tau_1^2} + \frac{1}{\xi}\right); \qquad \xi \mid \lambda_{ij}^2 \sim \text{IG}\left(1, 1 + \frac{1}{\lambda_{ij}^2}\right).$$

$$\lambda_i^2 \mid \text{else} \sim \text{IG}\left(\frac{(J+4)}{2} - 1, \sum_j \frac{(\theta_{ij} - \mu_i)^2}{2\lambda_{ij}^2\tau_1^2} + \frac{(\mu_i - \eta)^2}{2\tau_2^2} + \frac{1}{\xi}\right); \; \xi \mid \lambda_i^2 \sim \text{IG}\left(1, 1 + \frac{1}{\lambda_i^2}\right).$$

Using the *lasso* prior $(\propto \exp(-\kappa^2))$ on the local variances, and the Generalized Inverse Gaussian distribution GIG (shape, $\chi, \psi) \propto x^{(\text{shape}-1)}e^{-\frac{1}{2}(\chi/x + \psi x)}$,

$$\lambda_{ij}^2 \mid \text{ else} \sim \text{GIG}\left(1/2, \frac{(\theta_{ij} - \mu_i)^2}{\lambda_i^2\tau_1^2}, 2\right)$$

and

$$\lambda_i^2 \mid \text{ else} \sim \text{GIG}\left(\frac{(-J+1)}{2}, \sum_j \frac{(\theta_{ij} - \mu_i)^2}{\lambda_{ij}^2\tau_1^2} + \frac{(\mu_i - \eta)^2}{\tau_2^2}, 2\right).$$

Global variances: Using the horseshoe prior on the global variances, we have the following scale mixture representation:

$$\tau_1^2 \mid \text{else} \sim \text{IG}\left(\frac{(IJ+3)}{2} - 1, \frac{(\theta_{ij} - \mu_i)^2}{2\lambda_{ij}^2\lambda_i^2} + \frac{1}{\xi}\right); \qquad \xi \mid \tau_1^2 \sim \text{IG}\left(1, 1 + \frac{1}{\tau_1^2}\right).$$

$$\tau_2^2 \mid \text{else} \sim \text{IG}\left(\frac{(I+3)}{2} - 1, \sum_j \frac{(\theta_{ij} - \mu_i)^2}{2\lambda_i^2} + \frac{1}{\xi}\right); \qquad \xi \mid \tau_2^2 \sim \text{IG}\left(1, 1 + \frac{1}{\tau_2^2}\right).$$

References

Bhadra, A., Datta, J., Polson, N. and Willard, B. (2019). Lasso meets horseshoe: A survey. *Statistical Science*, 34(3), 405-427. https://www.jstor.org/stable/26874188.

Cahoy, D. and Sedransk, J. (2023). Combining data from surveys and related sources. *Survey Methodology*, 49(1), 69-87. Paper available at https://www150.statcan.gc.ca/n1/en/pub/12-001-x/2023001/article/00003-eng.pdf.

Carlin, B. and Louis, T. (2009). *Bayesian Methods for Data Analysis* (3rd ed.). Chapman and Hall/CRC.

Datta, G. and Lahiri, P. (1995). Robust hierarchical Bayes estimation of small area characteristics in the presence of covariates and outliers. *Journal of Multivariate Analysis,* 54(2), 310-328.

Datta, G. and Mandal, A. (2015). Small area estimation with uncertain random effects. *Journal of the American Statistical Association*, 110, 1735-1744.

Evans, R. and Sedransk, J. (2001). Combining data from experiments that may be similar. *Biometrika*, 88(3), 643-656.

Gelman, A. (2006). Prior distributions for variance parameters in hierarchical models. *Bayesian Analysis*, 1, 515-533.

Gelman, A., Carlin, J., Stern, H., Dunson, D., Vehtari, A. and Rubin, D. (2013). *Bayesian Data Analysis* (3rd ed.). Chapman and Hall/CRC. https://doi.org/10.1201/b16018.

Makalic, E. and Schmidt, D. (2016). A simple sampler for the horseshoe estimator. *IEEE Signal Processing Letters*, 23(1), 179-182, Jan. 2016, DOI: 10.1109/LSP.2015.2503725.

Malec, D. and Sedransk, J. (1992). Bayesian methodology for combining the results from different experiments when the specifications for pooling are uncertain. *Biometrika*, 79(3), 593-601.

Plummer, M. (2008). Penalized loss functions for Bayesian model comparisons. *Biostatistics*, 9(3), 523-539.

Polson, N. and Scott, J. (2012). On the half-Cauchy prior for a global scale parameter. *Bayesian Analysis*, 7, 887-902.

Smith, A. and Griffin, J. (2023). Shrinkage priors for high-dimensional demand estimation. *Quant Mark Econ*, 21, 95-146, DOI: https://doi.org/10.1007/s11129-022-09260-7.

Tang, X. and Ghosh, M. (2023). Global-local priors for spatial small area estimates. *Calcutta Statistical Association Bulletin*, 75(2), 141-154.

Tang, X., Ghosh, M., Ha, N. and Sedransk, J. (2018). Modeling random effects using global-local shrinkage priors in small area estimation. *Journal of the American Statistical Association*, 113(524), 1476-1489, DOI: 10.1080/01621459.2017.1419135.