ROBUST BAYESIAN HIERARCHICAL ESTIMATION OF THE FINITE POPULATION MEAN UNDER UNEQUAL CLUSTER SAMPLING

Doris Nkan Iso-Osene[1]*, Thomas Adidaumbe Ugbe[1], Ekaette Inyang Enang[1], Christian Elendu Onwukwe[1], Joy Ogum Uket[1] and Raphael Asu Owan[1]
[1] Department of Statistics, University of Calabar, Calabar, Nigeria
*Corresponding author: doris.iso-osene@autuni.ac.nz

Abstract

Cluster sampling is widely used for studying populations that are naturally grouped, but classical estimators can perform poorly when observations contain outliers. In this study, we propose a robust Bayesian hierarchical estimator for estimating the finite population mean under unequal cluster sampling. We show that the proposed estimator has a bounded influence function and stable asymptotic mean-squared error (MSE) under ε-contamination, and we evaluated the performance of the proposed estimator using both simulated and real datasets. Simulation results show that the proposed estimator retains efficiency under correct model specification while significantly improving robustness in contaminated settings, reducing point-estimation MSE by up to 40% and posterior predictive error by up to 50%. An application to ecological parasite-load data further demonstrates improved predictive stability and moderated mean estimates relative to the Gaussian hierarchical model.

Key words: Cluster sampling; Bayesian hierarchical models; contamination; outliers; Student-t distribution

1. Introduction

Cluster sampling is a probability sampling method in which the population is divided into smaller groups, referred to as clusters. This approach is widely employed in research across education, health, marketing, environmental studies, and agriculture due to its efficiency in dealing with large, geographically dispersed populations (Ly *et al.,* 2018). In this design, clusters are selected first, followed by sampling of units within those clusters, which often results in unequal cluster sizes and varying within-cluster variance (Hansen and Lee, 2019). Consequently, estimating the population mean requires integrating sampling information with between-cluster heterogeneity (Särndal, Swensson, and Wretman, 1992). Although classical estimators remain design-consistent under ideal conditions, their performance deteriorates in the presence of outliers or distributional departures, particularly when cluster sizes are unequal (Chambers, 1986; Beaumont et al., 2013). These problems stress the requirement for robust estimation methods that remain reliable in the presence of heterogeneity and contamination in clustered survey data.

Bayesian hierarchical models estimate finite population mean under cluster sampling by embedding the sampling design within a superpopulation framework and drawing on information across clusters (Ghosh and Rao, 1994; Gelman et al., 2013). Under normal hierarchical assumptions, posterior mean estimators exhibit shrinkage toward the overall mean and provide nearly unbiased estimates when the model is well specified. However, normal models are sensitive to contamination and heavy-tailed distributions. In the presence of ε-contamination (Huber, 1964; Hampel et al., 1986), the influence function of the posterior mean becomes unbounded. Recent work compares different point estimators for variance-structure parameters in Bayesian hierarchical models, focusing on how they behave under various conditions, including contamination. Simulation experiments show that contamination can increase mean square error and predictive dispersion, and that variance components can become inflated to account for outliers (Verdebout et al., 2005).

Robust probabilistic theory emphasises bounded influence and redescending behaviour to stabilise estimates in the presence of contamination (Huber, 1964; Hampel et al., 1986). Heavy-tailed probability distributions such as the Student-t, which can be expressed as scale mixtures of normals, provide a Bayesian approach that adaptively downweigh outliers (Andrews & Mallows, 1974; Lange et al., 1989). In hierarchical models, these distributions yield observation-specific weights while preserving multilevel shrinkage (West, 1984; Gelman et al., 2013). Heavy-tailed Bayesian methods have been applied in regression and small-area estimation (Ghosh, Maiti and Roy, 2008; Chakraborty, Datta and Mandal, 2017). However, their direct application to finite population mean estimation under unequal cluster sampling, explicit contamination analysis, and practical inference remains limited. (Abuhasel, 2026).

This paper develops a robust Bayesian hierarchical estimator for the finite population mean under unequal cluster sampling. The proposed approach employs a Student-t observation level model within a two level superpopulation framework. Conditional distributions are derived to enable efficient Gibbs sampling, and several key properties are established: (i) bounded influence of the posterior mean under ε contamination, (ii) stable asymptotic mean squared error relative to normal based estimators, and (iii) robust posterior inference for variance components. By integrating contamination theory with finite population inference, the estimator is both computationally practical and decision theoretically justified. It preserves efficiency under correct model specification while reducing risk under model misspecification. The performance of the estimator is further evaluated using simulated and real ecological data, demonstrating that it maintains efficiency when the model is correctly specified and improves the accuracy of parameter estimates and the reliability of predictions in the presence of contamination.

2. Finite Population Framework

Let

$$U = \{1,2,\dots,N\} \tag{1}$$

denote a finite population of N elements, partitioned into K mutually exclusive and exhaustive clusters. Let cluster k contain N_k units, such that

$$\sum_{k=1}^{K} N_k = N. \tag{2}$$

Let Y_{ik} denote the response for unit i in cluster k, where $i = 1,\dots,N_k$. The cluster total and cluster mean are defined as

$$T_k = \sum_{i=1}^{N_k} Y_{ik}, \quad \bar{Y}_k = \frac{1}{N_k}\sum_{i=1}^{N_k} Y_{ik}. \tag{3}$$

The finite population mean is

$$\bar{Y} = \frac{1}{N}\sum_{k=1}^{K} N_k \bar{Y}_k. \tag{4}$$

Suppose that s clusters are selected using simple random sampling without replacement (SRSWOR). Let $S \subset \{1, \dots, K\}$ denote the set of sampled clusters. The classical design-unbiased estimator of the finite population mean is

$$\hat{\bar{Y}} = \frac{1}{N}\sum_{k \in S} \frac{K}{s} N_k \bar{Y}_k. \tag{5}$$

Although unbiased, this estimator may exhibit high variance when cluster sizes are unequal or when extreme within-cluster observations are present. These limitations motivate the use of hierarchical modelling approaches that borrow strength across clusters while stabilizing inference under heterogeneous and potentially contaminated data.

3. Robust Hierarchical Model

To stabilize finite population inference under unequal cluster sizes and potential contamination, we adopt a two-level hierarchical superpopulation model with a heavy-tailed observation-level specification.

3.1 Observation-Level Model

Conditional on the cluster-specific mean μ_k, we assume

$$Y_{ik} \mid \mu_k, \sigma^2, \lambda_{ik} \sim N\left(\mu_k, \frac{\sigma^2}{\lambda_{ik}}\right), \tag{6}$$

where λ_{ik} are latent scale parameters satisfying

$$\lambda_{ik} \sim \text{Gamma}\left(\frac{\nu}{2}, \frac{\nu}{2}\right). \tag{7}$$

Integrating out λ_{ik} yields the marginal distribution

$$Y_{ik} \mid \mu_k, \sigma^2 \sim t_\nu(\mu_k, \sigma^2), \tag{8}$$

where $t_\nu(\mu_k, \sigma^2)$ represents a Student-t distribution with location μ_k, scale σ^2, and degrees of freedom ν. The degrees-of-freedom parameter ν governs tail behaviour. Smaller values of ν produce heavier tails and stronger resistance to outliers, while the Gaussian model is recovered in the limiting case $\nu \to \infty$.

3.2 Cluster-Level Model

Cluster-specific means are modelled as

$$\mu_k \mid \mu, \tau^2 \sim N(\mu, \tau^2), \tag{9}$$

where μ is the overall (grand) mean, σ^2 represents within-cluster variability and τ^2 represents between-cluster variability. This structure induces shrinkage of cluster means toward the overall mean, with the degree of shrinkage governed by the relative magnitudes of σ^2 and τ^2. Combined with the heavy-tailed observation-level specification, the model simultaneously performs between-cluster borrowing of strength and unit-level robust down-weighting.

4. Prior Specification

We adopt conjugate prior distributions to preserve conditional tractability and enable efficient Gibbs sampling.

The overall mean parameter is assigned a normal prior:

$$\mu \sim N(m_0, V_0), \tag{10}$$

where m_0 and V_0 represent prior mean and variance, respectively.

The variance components are assigned inverse-gamma priors:

$$\sigma^2 \sim \mathrm{IG}(a_\sigma, b_\sigma), \tau^2 \sim \mathrm{IG}(a_\tau, b_\tau), \tag{11}$$

where $\mathrm{IG}(a, b)$ represents the inverse-gamma distribution with shape parameter a and scale parameter b. These choices preserve conjugacy under the scale-mixture representation of the Student-t model, yielding closed-form full conditional distributions for all unknown parameters.

5. Posterior Computation

Let $S \subset \{1, \dots, K\}$ denote the set of sampled clusters and let n_k represent the number of observed units in cluster $k \in S$. Posterior inference is based on the joint posterior distribution of all unknown parameters, $\{\mu_k\}_{k\in S}, \mu, \sigma^2, \tau^2, \{\lambda_{ik}\}$, which is proportional to the product of the likelihood and prior distributions.

Under the conjugate structure of Sections 3 and 4, all conditional distributions are available in closed form, enabling efficient Gibbs sampling.

5.1 Conditional Distribution of $\boldsymbol{\lambda_{ik}}$

For each observed unit i in cluster $k \in S$,

$$\lambda_{ik}|rest{\sim}Gamma\left(\frac{v+1}{2},\frac{v+(Y_{ik}-\mu_k)^2/\sigma^2}{2}\right). \tag{12}$$

These latent weights λ_{ik} adaptively down-weight extreme observations. Large residuals yield small values of λ_{ik}, thereby reducing influence.

5.2 Conditional Distribution of $\boldsymbol{\mu_k}$

For each sampled cluster $k \in S$,

$$\mu_k \mid \text{rest} \sim N(m_k, V_k), \tag{13}$$

where

$$V_k = \left(\frac{1}{\tau^2}+\frac{1}{\sigma^2}\sum_{i=1}^{n_k}\lambda_{ik}\right)^{-1},$$

$$m_k = V_k\left(\frac{\mu}{\tau^2}+\frac{1}{\sigma^2}\sum_{i=1}^{n_k}\lambda_{ik}\,Y_{ik}\right).$$

Thus, m_k is a precision-weighted average of the grand mean μ and the robust weighted cluster mean.

5.3 Conditional Distribution of $\boldsymbol{\mu}$

$$\mu \mid \text{rest} \sim N(m_\mu, V_\mu), \tag{14}$$

where

$$V_\mu = \left(\frac{1}{V_0}+\frac{|S|}{\tau^2}\right)^{-1},$$

$$m_\mu = V_\mu\left(\frac{m_0}{V_0}+\frac{1}{\tau^2}\sum_{k\in S}\mu_k\right).$$

5.4 Conditional Distribution of $\boldsymbol{\sigma^2}$

$$\sigma^2 \mid rest \sim IG\left(a_\sigma+\frac{n}{2}, b_\sigma+\frac{1}{2}\sum_{k=1}^{K}\sum_{i=1}^{n_k}\lambda_{ik}(Y_{ik}-\mu_k)^2\right). \tag{15}$$

where $n = \sum_{k\in S} n_k$is the total number of observed units

5.5 Conditional Distribution of τ^2

$$\tau^2 \mid rest \sim IG\left(a_\tau + \frac{|S|}{2}, b_\tau + \frac{1}{2}\sum_{k\in S}(\mu_k - \mu)^2\right). \tag{16}$$

5.6 Gibbs Sampling Algorithm

The posterior inference proceeds by iteratively sampling from the above conditional distributions described in sections 5.1-5.5.

At iteration $t = 1, \dots, T$, the algorithm proceeds as follows:

Step 1: Update latent weights

$$\lambda_{ik}^{(t)} \sim Gamma\left(\frac{v+1}{2}, \frac{v + \left(Y_{ik} - \mu_k^{(t-1)}\right)^2 / \sigma^{2(t-1)}}{2}\right).$$

Step 2: Update cluster-specific mean

$$\mu_k^{(t)} \sim N(m_k^{(t)}, V_k^{(t)}).$$

Step 3: Update grand mean

$$\mu^{(t)} \sim N(m_\mu^{(t)}, V_\mu^{(t)}).$$

Step 4: Update within-cluster variance

$$\sigma^{2(t)} \sim IG\left(a_\sigma + \frac{n}{2}, b_\sigma + \frac{1}{2}\sum_{k=1}^{K}\sum_{i=1}^{n_k} \lambda_{ik}^{(t)} (Y_{ik} - \mu_k^{(t)})^2\right).$$

Step 5: Update between-cluster variance

$$\tau^2 \sim IG\left(a_\tau + \frac{|S|}{2}, b_\tau + \frac{1}{2}\sum_{k\in S}\left(\mu_k^{(t)} - \mu^{(t)}\right)^2\right)$$

At each iteration, the finite population mean is computed as

$$\hat{\bar{Y}}^{(t)} = \sum_{k=1}^{K} \frac{N_k}{N} \tilde{\mu}_k^{(t)},$$

where sampled cluster means are taken from their posterior draws $\mu_k^{(t)}$, and unsampled cluster means are drawn from

$$\tilde{\mu}_k^{(t)} \sim N(\mu^{(t)}, \tau^{2(t)}).$$

Because all updates involve standard distributions and preserve conditional conjugacy, the algorithm requires no Metropolis-Hastings steps and is computationally stable.

6. Theoretical Properties

In this section, we summarize the key theoretical properties of the proposed robust Bayesian hierarchical estimator of the finite population mean. Detailed derivations are provided in Appendix A.

6.1 Shrinkage Structure

Let μ_k denote the cluster-level mean. Under the Student-t representation with latent weights λ_{ik}, the posterior mean of μ_k takes the form

$$E(\mu_k \mid \text{data}) = B_k \bar{y}_{k,w} + (1 - B_k)\mu, \tag{17}$$

where

$$\bar{y}_{k,w} = \frac{\sum_i \lambda_{ik} y_{ik}}{\sum_i \lambda_{ik}}, B_k = \frac{\tau^2}{\tau^2 + \sigma^2 / \sum_i \lambda_{ik}}.$$

Thus, the estimator is a convex combination of a robust weighted cluster mean and the grand mean. The latent weights λ_{ik} decrease for extreme observations, inducing adaptive down-weighting. Shrinkage increases when between-cluster variance τ^2 is small or cluster size is small. The estimator therefore performs adaptive, cluster-specific shrinkage while simultaneously mitigating outlier influence.

6.2 Bounded Influence

Robustness can be characterized through influence behaviour under ε-contamination.

Theorem 1 (Bounded Influence)

Under fixed degrees of freedom $\nu < \infty$ and proper priors on variance components, the influence of any single observation on the posterior mean of the finite population mean is bounded.

6.3 Redescending Property

The Student-t specification further satisfies a redescending property.

Theorem 2 (Redescending Behaviour)

For finite ν, the influence of an observation satisfies

$$\lim_{|y_{ik}|\to\infty} \frac{\partial \hat{\hat{Y}}}{\partial y_{ik}} = 0.$$

Hence, sufficiently extreme observations receive asymptotically negligible weight.

This property is stronger than bounded influence and ensures stability under arbitrarily large contamination magnitudes.

6.4 Asymptotic MSE Stability

We now examine the large-sample behaviour of the estimator.

Let $\hat{\hat{Y}}$ denote the robust finite population estimator and consider ε-contamination of the form

$$F_\varepsilon = (1-\varepsilon)F + \varepsilon G,$$

where G may have arbitrarily large variance.

Theorem 3 (Asymptotic MSE Stability)

Let the data follow an ε-contamination model

$$Y_{ik} \sim (1-\varepsilon)F_0 + \varepsilon G,$$

where F_0 has finite variance.

Let $\hat{\hat{Y}}_{RBH}$ and $\hat{\hat{Y}}_{GHB}$ denote the proposed robust and Gaussian hierarchical estimators of the finite population mean, respectively.

Then, as $n \to \infty$,

$$\text{MSE}(\hat{\hat{Y}}_{RBH}) = O\left(\frac{1}{n}\right) + O(\varepsilon),$$

and

$$\text{MSE}(\hat{\hat{Y}}_{GHB}) = O\left(\frac{1}{n}\right) + O(\varepsilon M^2),$$

with M denoting contamination magnitude.

For fixed ε and sufficiently large contamination magnitude,

$$\mathrm{MSE}(\hat{\bar{Y}}_{RBH}) < \mathrm{MSE}(\hat{\bar{Y}}_{GHB}).$$

7. Simulation Study

In this section, we evaluate the finite-sample performance of the proposed robust Bayesian hierarchical estimator (RBH) and provides numerical confirmation of the theoretical properties established in section 6. We assess efficiency under correct specification and stability under increasing contamination. The RBH estimator is compared with the Gaussian hierarchical Bayes (GHB) estimator.

7.1 Design

A finite population consisting of $K = 50$ clusters was generated under the hierarchical superpopulation model

$$\mu_k \sim N(\mu_0, \tau^2), \quad Y_{ik} \sim t_\nu(\mu_k, \sigma^2),$$

with $\mu_0 = 10$, $\tau^2 = 4$, and $\sigma^2 = 9$. Cluster sizes were unequal and drawn from Uniform (20,100) for the moderate and large-cluster scenario. From each population, $s = 20$ clusters were selected by simple random sampling without replacement.

To assess robustness, ε-contamination was introduced:

$$Y_{ik} \sim (1-\varepsilon)N\left(\mu_k, \frac{\sigma^2}{\lambda_{ik}}\right) + \varepsilon N(\mu_k, M^2),$$

with $\varepsilon \in \{0, 0.10, 0.15, 0.20, 0.25, 0.30\}$ and $M \in \{50, 100\}$. We compared: GHB and RBH with $\nu \in \{3,4,8\}$. Each scenario was replicated 1,000 times. Performance metrics included (i) Mean squared error (MSE) of the finite population mean, (ii) Posterior predictive MSE (PPMSE) and (iii) 95% credible interval coverage.

7.2 Results: Moderate and Large Clusters

Correct Specification (ε = 0)

In Table 1 (no contamination), the RBH and GHB exhibited nearly equal performance. The GHB had an MSE of 0.1345, while the RBH MSE ranged from 0.1345 to 0.1354, corresponding to less than 1% efficiency loss. Coverage probability for both RBH and GHB was close to nominal (93-94%). The Posterior Predictive MSE (PPMSE) performance was slightly improved under RBH (PPMSE ≈ 15) for degrees of freedom 4 and 3 and (PPMSE ≈ 18) for degree of freedom 8 while GHB (PPMSE ≈ 18). These results confirm that the heavy-tailed specification preserves efficiency under ideal conditions.

Moderate Contamination (ε = 0.10-0.15)

In Table 2 (moderate contamination, $\varepsilon = 0.10$, $M = 50$), the GHB estimator produced an MSE of 0.266, while the proposed RBH estimator ($\nu = 4$) achieved an MSE of 0.230, corresponding to a 14% reduction relative to GHB. The improvement is even more evident in predictive performance: PPMSE decreased from 521 under GHB to 272 under RBH, representing nearly a 48% reduction. Under stronger contamination (Table 3; $\varepsilon = 0.15$, $M = 100$), the MSE of the GHB estimator increased greatly to 0.860, while the RBH estimator ($\nu = 4$) achieved an MSE of 0.672, reflecting a 22% reduction. Predictive error was also nearly halved, with PPMSE decreasing from 3018 for GHB to 1532 for RBH. In addition, the GHB estimator exhibited strong over-coverage (approximately 98%), indicating inflated variance, but the RBH estimator produced tighter intervals, although with moderate under-coverage.

Severe Contamination (ε = 0.20-0.30)

In Table 4 (severe contamination, $\varepsilon = 0.20, 0.25, 0.30$, $M = 100$), the GHB estimator produced an average MSE of 1.476, while the proposed RBH estimator ($\nu = 4$) yielded an average MSE of 0.914, corresponding to a 38% reduction in MSE relative to GHB. The improvement is even more evident in predictive performance: PPMSE decreased from 6021.944 under GHB to 2691.589 under RBH, representing nearly a 50% reduction. Coverage probability also illustrates the robustness-calibration trade-off: GHB maintained persistent over-coverage (approximately 98%), but RBH coverage ranged from 57% to 91%, increasing with degree of freedom (ν).

8. Real Data Application

To assess the practical application of the proposed estimator, a secondary dataset from an ecological study of sandstone-cave bats in Enugu State, Nigeria, was used to evaluate its ability to estimate the mean parasite load. The data includes 35 caves, excluding cave 17 because no observations are recorded. Here, we used parasite load as the unit-level outcome, treating caves as naturally occurring clusters.

8.1 Results

Table 3 shows the estimated mean parasite load, along with 95% credible intervals (CrI) and posterior predictive mean squared errors (PPMSE). The GHB estimator yielded an estimated finite-population mean parasite load of 5.12, with a 95% credible interval (CrI) of (4.58, 5.67). The corresponding PPMSE was 46.47, reflecting strong predictive dispersion under the normality assumption, but the proposed RBH estimator produced consistently lower mean estimates across all degrees-of-freedom specifications. For $\nu = 4$, the estimated mean was 3.87 (95% CrI: 3.46, 4.29) with PPMSE = 33.01, representing approximately a 29% reduction in predictive error relative to the Gaussian estimator. For $\nu = 3$, the mean further decreased to 3.61 (PPMSE = 33.02), while for $\nu = 8$, the estimate was 4.33 (PPMSE = 36.43), still significantly lower than the GHB estimator.

9. Discussion

The simulation study provides clear empirical support for the theoretical results established in Theorems 1-3. When the model is correctly specified ($\varepsilon = 0$), the Gaussian hierarchical Bayes (GHB) estimator and the proposed robust Bayesian hierarchical (RBH) estimator display nearly identical mean-squared error and achieve coverage probabilities close to their nominal levels. This aligns with the asymptotic efficiency properties implied by Theorem 3 and the limiting Gaussian case as $\nu \to \infty$. Consequently, adopting a heavy-tailed specification does not impose an efficiency penalty under ideal conditions (Lange, Little & Taylor, 1989).

As contamination increases, the divergence predicted by robustness theory becomes evident. Theorem 1 shows that the RBH posterior mean has bounded influence under ε-contamination, whereas the Gaussian estimator does not. This distinction is reflected in the simulations: GHB reacts to contamination mainly through inflated variance, producing wider credible intervals and greater predictive dispersion. In contrast, RBH down-weights extreme observations, maintaining stable estimation error.

Theorem 3's MSE-stability result is also confirmed numerically. Under moderate to severe contamination, RBH reduces point-estimation MSE by 14-45% and lowers predictive MSE by roughly 40-50% relative to GHB. Although very small degrees of freedom may lead to under-coverage, moderate values achieve a practical balance between robustness and interval calibration, consistent with the redescending behaviour described in Theorem 2.

The real-data application reinforces these findings: in the cave parasite dataset, RBH moderates mean estimates and reduces predictive MSE by 22-29%, mirroring the contamination patterns observed in simulation. Overall, the agreement between theory, simulation, and empirical performance demonstrates that the proposed estimator preserves efficiency under correct specification while stabilizing estimates under misspecification.

Conclusion

In this study, we developed a robust Bayesian hierarchical estimator for the finite population mean under unequal cluster sampling by embedding a Student-t observation model within a two-level hierarchical framework. The resulting estimator exhibits key robustness properties, including bounded influence, redescending behaviour, and stable asymptotic mean-squared error. Both the simulation results and the real-data application show that the proposed estimator maintains high efficiency under correct model specification while delivering substantial gains under contamination by reducing both estimation error and predictive risk.

Overall, the proposed estimator offers a practical and theoretically grounded solution for finite-population inference in clustered settings. Its ability to remain efficient in ideal conditions while providing markedly improved reliability in the presence of heavy-tailed data or influential observations makes it a valuable tool for applied survey work.

Declaration of competing interest: The authors declare that they have no known competing financial interests or personal relationship that could have appeared to influence the research reported in this paper.

Data availability: Data will be made available on request.

12. References

Abuhasel, K. A. (2026). A robust methodology for finite population mean estimation based on generalized M estimation. *Scientific Reports.* 16(1). https://doi.org/10.1038/s41598-026-35592-5

Andrews, D. F., & Mallows, C. L. (1974). Scale mixtures of normal distributions. *Journal of the Royal Statistical Society: Series B (Methodological)*, 36(1), 99-102

Beaumont, J. F., Haziza, D., & Ruiz-Gazen, A. (2013). A unified approach to robust estimation in finite population sampling. *Biometrika*. 100(2), 359-375.

Chakraborty, A., Datta, G. S., & Mandal, A. (2017). Robust hierarchical Bayes small area estimation for nested error linear regression models. *Journal of Statistical Planning and Inference*. 187, 41-56. https://doi.org/10.1016/j.jspi.2016.10.004

Chambers, R. L. (1986). Outlier robust finite population estimation. *Journal of the American Statistical Association*. 81(396), 1063-1069.

Ghosh, M., Maiti, T., & Roy, A. (2008). Influence functions and robust Bayes and empirical Bayes small area estimation. *Biometrika*. 95(3), 573-585. https://doi.org/10.1093/biomet/asn02

Ghosh, M. and Rao, J. N. K. (1994). Small Area Estimation: An Appraisal. *Statistics Science*. 9, 65-93. https://doi.org/10.1214/ss/1177010647

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

Hampel, F. R. (1974). The influence curve and its role in robust estimation. *Journal of the American Statistical Association*. 69(346), 383-393.

Hansen, B. E. & Lee, S. (2010). Asymptotic theory for clustered samples. *Journal of Econometrics*. 210(2), 268-290. https://doi.org/10.1016/j.jeconom.2019.02.001.

Huber, P. J. (1964). Robust estimation of a location parameter. *Annals of Mathematical Statistics*, 35(1), 73-101.

Lange, K. L., Little, R. J., & Taylor, J. M. (1989). Robust statistical modelling using the t distribution. *Journal of the American Statistical Association*. *84*(408), 881-896.

Ly T, Cockburn M, Langholz B. (2018). Cost-efficient case-control cluster sampling designs for population-based epidemiological studies. *Spat Spatiotemporal Epidemiol*. :95-105. doi: 10.1016/j.sste.2018.05.002. Epub 2018 Jun 15. PMID: 30390934.

Särndal, C. E., Swensson, B., & Wretman, J. (1992). Model assisted survey sampling. Springer.

Verdebout, J. M., Verdebout, J. M. & Verdebout, J. M. (2005). Robust Estimation of Multivariate Covariance Components. *Biometrics.* 61(1), pp. 162-170. https://doi.org/10.1111/j.0006-341X.2005.03004.x

West, M. (1984). Outlier models and prior distributions in Bayesian linear regression. *Journal of the Royal Statistical Society: Series B (Methodological)*. 46(3), 431-439.

12. Appendix

Appendix A: Derivations

This appendix provides formal derivations omitted from the main text.

A.1 Derivation of Full Conditional for λ_{ik}

The observation-level model is

$Y_{ik}|\mu_k, \sigma^2, \lambda_{ik} \sim N\left(\mu_k, \frac{\sigma^2}{\lambda_{ik}}\right),$

with prior

$$\lambda_{ik} \sim Gamma\left(\frac{v}{2}, \frac{v}{2}\right)$$

The conditional posterior is proportional to

$$\lambda_{ik}{}^{1/2} exp\left(-\frac{\lambda_{ik}(Y_{ik}-\mu_k)^2}{2\sigma^2}\right) * \lambda_{ik}^{\frac{v}{2}-1} exp\left(-\frac{v}{2}\lambda_{ik}\right).$$

Collecting terms yields

$$\lambda_{ik} \mid \text{rest} \sim Gamma\left(\frac{\nu+1}{2}, \frac{\nu+(Y_{ik}-\mu_k)^2/\sigma^2}{2}\right).$$

A.2 Derivation of Full Conditional for $\boldsymbol{\mu_k}$

For a sampled cluster $k \in S$, the observation-level model is

$$Y_{ik} \mid \mu_k, \sigma^2, \lambda_{ik} \sim N\left(\mu_k, \frac{\sigma^2}{\lambda_{ik}}\right), i = 1, \dots, n_k.$$

Hence the likelihood contribution for cluster k is

$$p(\mathrm{Y}_k \mid \mu_k, \sigma^2, \lambda_k) \propto \prod_{i=1}^{n_k} \left(\frac{\lambda_{ik}}{\sigma^2}\right)^{1/2} \exp\left\{-\frac{\lambda_{ik}}{2\sigma^2}(Y_{ik} - \mu_k)^2\right\}.$$

Ignoring terms not involving μ_k, the kernel is

$$p(\mathrm{Y}_k \mid \mu_k) \propto \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n_k} \lambda_{ik}\,(Y_{ik} - \mu_k)^2\right\}.$$

From the cluster-level model,

$$\mu_k \mid \mu, \tau^2 \sim N(\mu, \tau^2),$$

so, the prior is

$$p(\mu_k \mid \mu) \propto \exp\left\{-\frac{1}{2\tau^2}(\mu_k - \mu)^2\right\}.$$

The full conditional is proportional to

$$p(\mu_k \mid \text{rest}) \propto \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n_k} \lambda_{ik}\,(Y_{ik} - \mu_k)^2 - \frac{1}{2\tau^2}(\mu_k - \mu)^2\right\}.$$

Expanding the quadratic term

$$(Y_{ik} - \mu_k)^2 = Y_{ik}^2 - 2Y_{ik}\mu_k + \mu_k^2.$$

Thus,

$$\sum_{i=1}^{n_k} \lambda_{ik}\,(Y_{ik} - \mu_k)^2 = \sum_{i=1}^{n_k} \lambda_{ik}\,Y_{ik}^2 - 2\mu_k \sum_{i=1}^{n_k} \lambda_{ik}\,Y_{ik} + \mu_k^2 \sum_{i=1}^{n_k} \lambda_{ik}\,.$$

Collecting only terms involving μ_k, the exponent becomes

$$-\frac{1}{2}\left[\left(\frac{1}{\tau^2} + \frac{1}{\sigma^2}\textstyle\sum_{i=1}^{n_k} \lambda_{ik}\right)\mu_k^2 - 2\left(\frac{\mu}{\tau^2} + \frac{1}{\sigma^2}\textstyle\sum_{i=1}^{n_k} \lambda_{ik}\,Y_{ik}\right)\mu_k\right].$$

Completing the square

$$\frac{\left(\frac{1}{\tau^2}+\frac{1}{\sigma^2}\sum_{i=1}^{n_k}\lambda_{ik}\right)}{2}\left[\left(\mu_k-\left(\frac{\mu}{\tau^2}+\frac{1}{\sigma^2}\sum_{i=1}^{n_k}\lambda_{ik}\,Y_{ik}\right)\right)^2\right]$$

This quadratic form implies a Normal distribution with:

$$V_k=\left(\frac{1}{\tau^2}+\frac{1}{\sigma^2}\sum_{i=1}^{n_k}\lambda_{ik}\right)^{-1},$$

$$m_k=V_k\left(\frac{\mu}{\tau^2}+\frac{1}{\sigma^2}\sum_{i=1}^{n_k}\lambda_{ik}\,Y_{ik}\right).$$

Hence

$$\mu_k \mid \text{rest} \sim N(m_k, V_k)$$

A.3 Derivation of Full Conditional for μ

Recall:

Prior:

$$\mu \sim N(m_0, V_0)$$

Cluster-level model:

$$\mu_k \mid \mu, \tau^2 \sim N(\mu, \tau^2), k \in S.$$

The conditional posterior of μ is proportional to:

$$p(\mu \mid \text{rest}) \propto p(\mu)\prod_{k\in S} p(\,\mu_k \mid \mu, \tau^2).$$

Substituting the densities:

$$p(\mu) \propto \exp\left\{-\frac{1}{2V_0}(\mu-m_0)^2\right\},$$

$$p(\mu_k \mid \mu) \propto \exp\left\{-\frac{1}{2\tau^2}(\mu_k-\mu)^2\right\}.$$

Thus,

$$p(\mu \mid \text{rest}) \propto \exp\left\{-\frac{1}{2V_0}(\mu-m_0)^2-\frac{1}{2\tau^2}\sum_{k\in S}(\,\mu_k-\mu)^2\right\}.$$

Expanding the quadratic terms yields

$$(\mu - m_0)^2 = \mu^2 - 2m_0\mu + m_0^2,$$

$$(\mu_k - \mu)^2 = \mu_k^2 - 2\mu_k\mu + \mu^2.$$

Then,

$$\sum_{k\in S}(\mu_k - \mu)^2 = \sum_{k\in S}\mu_k^2 - 2\mu\sum_{k\in S}\mu_k + \mid S \mid \mu^2.$$

Substituting into the log-posterior and collecting terms involving μ:

$$-\frac{1}{2}\left[\left(\frac{1}{V_0} + \frac{\mid S \mid}{\tau^2}\right)\mu^2 - 2\left(\frac{m_0}{V_0} + \frac{1}{\tau^2}\textstyle\sum_{k\in S}\mu_k\right)\mu\right].$$

Completing the square gives

$$-\frac{\left(\frac{1}{V_0} + \frac{\mid S \mid}{\tau^2}\right)}{2}\left[\left(\mu - \left(\frac{m_0}{V_0} + \frac{1}{\tau^2}\textstyle\sum_{k\in S}\mu_k\right)\right)^2\right]$$

This is quadratic in μ, so the posterior is Normal with:

$$V_\mu = \left(\frac{1}{V_0} + \frac{\mid S \mid}{\tau^2}\right)^{-1},$$

$$m_\mu = V_\mu\left(\frac{m_0}{V_0} + \frac{1}{\tau^2}\textstyle\sum_{k\in S}\mu_k\right).$$

Therefore,

$$\mu | rest \sim N\left(m_\mu, V_\mu\right).$$

A.4 Derivation of Full Conditional for σ^2

From the observation-level model:

$Y_{ik}|\mu_k, \sigma^2, \lambda_{ik} \sim N\left(\mu_k, \frac{\sigma^2}{\lambda_{ik}}\right).$

Thus, the likelihood contribution is:

$$p(Y \mid \sigma^2) \propto (\sigma^2)^{-n/2}\exp\left\{-\frac{1}{2\sigma^2}\sum_{k\in S}\sum_{i=1}^{n_k}\lambda_{ik}(Y_{ik} - \mu_k)^2\right\}.$$

The prior is:

$$\sigma^2 \sim \mathrm{IG}(a_\sigma, b_\sigma),$$

with density:

$$p(\sigma^2) \propto (\sigma^2)^{-(a_\sigma+1)}\exp\left\{-\frac{b_\sigma}{\sigma^2}\right\}.$$

Multiplying likelihood and prior gives

$$p(\sigma^2 \mid \text{rest}) \propto (\sigma^2)^{-(a_\sigma+1+n/2)}\exp\left\{-\frac{1}{2\sigma^2}\sum_{k\in S}\sum_{i=1}^{n_k}\lambda_{ik}(Y_{ik}-\mu_k)^2\right\}.$$

This is the kernel of an inverse-gamma distribution.

Hence,

$\sigma^2|rest \sim IG\left(a_\sigma+\frac{n}{2}, b_\sigma+\frac{1}{2}\sum_{k\in S}\sum_{i=1}^{n_k}\lambda_{ik}(Y_{ik}-\mu_k)^2\right).$

A.5 Derivation of Full Conditional for τ^2

From the cluster-level model:

$$\mu_k \mid \mu, \tau^2 \sim N(\mu, \tau^2).$$

Thus, the likelihood term is:

$$p(\{\mu_k\} \mid \tau^2) \propto (\tau^2)^{-|S|/2}\exp\left\{-\frac{1}{2\tau^2}\sum_{k\in S}(\mu_k-\mu)^2\right\}.$$

The prior is:

$$\tau^2 \sim \text{IG}(a_\tau, b_\tau),$$

with density:

$$p(\tau^2) \propto (\tau^2)^{-(a_\tau+1)}\exp\left\{-\frac{b_\tau}{\tau^2}\right\}.$$

Multiplying the likelihood and the prior:

$p(\tau^2 \mid \text{rest}) \propto (\tau^2)^{-(a_\tau+1+|S|/2)}exp\left\{-\frac{1}{\tau^2}\left[b_\tau+\frac{1}{2}\sum_{k\in S}(\mu_k-\mu)^2\right]\right\}.$

This is inverse-gamma.

Therefore,

$\tau^2 \mid \text{rest} \sim \text{IG}\left(a_\tau + \frac{|S|}{2}, b_\tau + \frac{1}{2}\sum_{k\in S}(\mu_k - \mu)^2\right)$.

A.6 Proof of Bounded Influence (Theorem 1)

For extreme observations $|Y_{ik}| \to \infty$,

$$\lambda_{ik} = \frac{\nu + 1}{\nu + (Y_{ik} - \mu_k)^2/\sigma^2} = O(|Y_{ik}|^{-2}).$$

Thus, the contribution of Y_{ik} to the weighted cluster mean is

$$\lambda_{ik} Y_{ik} = O(|Y_{ik}|^{-1}),$$

which converges to zero. Therefore, the derivative of the posterior mean with respect to Y_{ik} remains bounded.

A.7 Proof of Redescending Behaviour (Theorem 2)

We prove the result in three steps.

Step 1: Student-t as a Scale Mixture

Under the robust hierarchical specification,

$$Y_{ik} \mid \mu_k, \sigma^2, \lambda_{ik} \sim N\left(\mu_k, \frac{\sigma^2}{\lambda_{ik}}\right),$$

with latent weights

$$\lambda_{ik} \sim \text{Gamma}\left(\frac{\nu}{2}, \frac{\nu}{2}\right).$$

The conditional posterior expectation of λ_{ik} is

$$E(\lambda_{ik} \mid y_{ik}) = \frac{\nu + 1}{\nu + \frac{(y_{ik} - \mu_k)^2}{\sigma^2}}.$$

For large residuals,

$$|y_{ik}| \to \infty,$$

we have

$$\frac{(y_{ik} - \mu_k)^2}{\sigma^2} \to \infty,$$

and therefore

$E(\lambda_{ik} \mid y_{ik}) \sim \frac{\nu+1}{(y_{ik}-\mu_k)^2/\sigma^2} = O\left(\frac{1}{y_{ik}^2}\right).$

Thus,

$$\lambda_{ik} \to 0 \text{as} \mid y_{ik} \mid \to \infty.$$

Step 2: Effect on the Posterior Mean of μ_k

From Appendix A.2, the posterior mean of μ_k is

$$m_k = V_k \left(\frac{\mu}{\tau^2} + \frac{1}{\sigma^2} \sum_{i=1}^{n_k} \lambda_{ik}\, y_{ik}\right),$$

where

$$V_k = \left(\frac{1}{\tau^2} + \frac{1}{\sigma^2} \sum_{i=1}^{n_k} \lambda_{ik}\right)^{-1}.$$

Let observation y_{ik} diverge while all others remain fixed.

Since $\lambda_{ik} = O(1/y_{ik}^2)$,

$$\lambda_{ik} y_{ik} = O\left(\frac{1}{y_{ik}}\right) \to 0.$$

Similarly,

$$\sum_{i=1}^{n_k} \lambda_{ik} = O\left(\frac{1}{y_{ik}^2}\right) \to 0.$$

Therefore,

$$V_k \to \tau^2,$$

and

$$m_k \to \mu.$$

Thus,

$$\frac{\partial m_k}{\partial y_{ik}} \to 0.$$

The cluster mean becomes insensitive to extreme observations.

Step 3: Effect on the Finite Population Mean

The estimator of the finite population mean is

$$\hat{\hat{Y}} = \sum_{k=1}^{K} \frac{N_k}{N} \tilde{\mu}_k,$$

where sampled clusters use posterior draws of μ_k and unsampled clusters are drawn from $N(\mu, \tau^2)$. Since

$$\frac{\partial \tilde{\mu}_k}{\partial y_{ik}} \to 0,$$

and the estimator is linear in $\tilde{\mu}_k$,

$$\frac{\partial \hat{\hat{Y}}}{\partial y_{ik}} = \frac{N_k}{N} \frac{\partial \tilde{\mu}_k}{\partial y_{ik}} \to 0.$$

Hence,

$$\lim_{|y_{ik}| \to \infty} \frac{\partial \hat{\hat{Y}}}{\partial y_{ik}} = 0.$$

A 8. Proof of Asymptotic MSE Stability Under Contamination (Theorem 3)

Step 1: MSE Decomposition

For any estimator $\hat{\theta}$,

$$\mathrm{MSE}(\hat{\theta}) = \mathrm{Var}(\hat{\theta}) + \mathrm{Bias}(\hat{\theta})^2.$$

Thus, we analyse variance and squared bias separately.

Step 2: Asymptotic Variance

Under regularity conditions for hierarchical Bayesian estimators with proper priors and identifiable parameters (Ghosh & Rao, 1994), posterior means satisfy

$$\hat{\hat{Y}} - \bar{Y} = O_p\left(\frac{1}{\sqrt{n}}\right),$$

implying

$$\mathrm{Var}(\hat{\hat{Y}}) = O\left(\frac{1}{n}\right).$$

This rate holds for both Gaussian and Student-t specifications, since the hierarchical structure is identical at the cluster level and the Student-t model reduces to Gaussian as $\nu \to \infty$.

Thus,

$$\text{Var}_{RBH} = O\left(\frac{1}{n}\right), \text{Var}_{GHB} = O\left(\frac{1}{n}\right).$$

Step 3: Bias Under ε-Contamination

Under contamination,

$$Y_{ik} = Z_{ik} + \delta_{ik},$$

where

$Z_{ik} \sim F_0$,

$\delta_{ik} \sim G$ with probability ε.

The Gaussian hierarchical estimator has unbounded influence (Theorem 1). Thus, the bias induced by contamination scales linearly with contamination magnitude:

$$\text{Bias}_{GHB} = O(\varepsilon M).$$

Therefore,

$$\text{Bias}^2_{GHB} = O(\varepsilon^2 M^2).$$

For moderate ε,

$$\text{Bias}^2_{GHB} = O(\varepsilon M^2).$$

From Theorem 2, the influence function of the Student-t estimator is redescending. The effective weight for extreme observations satisfies

$$\lambda_{ik} = O\left(\frac{1}{(y_{ik} - \mu_k)^2}\right).$$

Hence contamination magnitude does not enter linearly into the estimator.

The induced bias is bounded:

$$\text{Bias}_{RBH} = O(\varepsilon).$$

Therefore,

$$\text{Bias}^2_{RBH} = O(\varepsilon^2).$$

Step 4: Combining Terms

Thus,

$$\mathrm{MSE}_{RBH} = O\left(\frac{1}{n}\right) + O(\varepsilon^2),$$

$$\mathrm{MSE}_{GHB} = O\left(\frac{1}{n}\right) + O(\varepsilon M^2).$$

For fixed ε and increasing contamination magnitude M,

$$\mathrm{MSE}_{GHB} \to \infty, \mathrm{MSE}_{RBH} \text{ remains bounded.}$$

Hence,

$$\mathrm{MSE}_{RBH} < \mathrm{MSE}_{GHB} \text{ for sufficiently large } M.$$

Step 5: Efficiency Under Correct Specification

When $\varepsilon = 0$,

$$\mathrm{Bias}_{RBH} = 0, \mathrm{Bias}_{GHB} = 0.$$

Both estimators share the same asymptotic variance order $O(1/n)$. Hence,

$$\mathrm{MSE}_{RBH} = \mathrm{MSE}_{GHB} + o(1).$$

Thus, no asymptotic efficiency loss occurs under correct specification.

Appendix B: Results

Table 1

Simulation summary of estimator performance for the finite population mean under correct model specification (ε = 0).

Model	MSE	PPMSE	Coverage
RBH $(v = 3)$	0.1354181	14.66271	0.933
RBH $(v = 4)$	0.1351272	15.21522	0.930
RBH $(v = 8)$	0.1347677	16.31511	0.931
GHB	0.1345436	18.01968	0.936

Table 2

Simulation summary of estimator performance for the finite population mean under moderate contamination (ε = 0.10, M = 50).

Model	MSE	PPMSE	Coverage
RBH $(v = 3)$	0.23132	269.8860	0.843
RBH $(v = 4)$	0.22957	272.2289	0.848
RBH $(v = 8)$	0.22312	283.8958	0.860
GHB	0.26628	521.1379	0.971

Table 3

Simulation summary of estimator performance for the finite population mean under moderate contamination (ε = 0.15, M = 100).

Model	MSE	PPMSE	Coverage
RBH ($v = 3$)	0.67833	1524.293	0.625
RBH ($v = 4$)	0.67222	1532.205	0.627
RBH ($v = 8$)	0.61674	1641.386	0.720
GHB	0.86000	3018.138	0.982

Table 4

Simulation summary of estimator performance for the finite population mean under severe contamination (ε = 0.20, 0.25, 0.30 M = 100).

Model	MSE	PPMSE	Coverage
RBH ($v = 3$)	0.96040	2585.775	0.567
RBH ($v = 4$)	0.91398	2691.589	0.661
RBH ($v = 8$)	0.81250	3222.999	0.910
GHB	1.47578	6021.944	0.977

Table 5

Posterior performance summary for the finite population mean parasite load under Gaussian and robust hierarchical models

Model	**Mean estimate**	**Credible interval (Crl) 2.5%**	**Credible interval (Crl) 97.5%**	**PPMSE**
RBH ($v = 3$)	3.605832	3.2351284	4.033470	33.01687
RBH ($v = 4$)	3.871797	3.459431	4.287054	33.00856
RBH ($v = 8$)	4.327445	3.894394	4.763547	34.43002
GHB	5.123672	4.581576	5.668019	46.46945