The clp-PR and sam-RM prediction intervals

Let $X(s) = [X^T(s)_1, \ldots, X^T(s)_g]^T$, where $X(s)_i = [x_{ij}, j \in s_i]^T$, and $y(s) = [y^T(s)_1, \ldots, y^T(s)_g]^T$, where $y(s)_i = [y_{ij}, j \in s_i]^T$. Following Prasad & Rao (1990), Datta & Lahiri (2000) showed that $E(\hat{M}_i^{clp} - \eta_i)^2$, the unconditional MSE of $\hat{M}_i^{clp}$ for predicting $\eta_i$, can be approximated to second order by

$$\text{MSE}_{PR,i} = g_1i(\theta) + g_2i(\theta) + g_3i(\theta) + o_p(g^{-1}).$$

where

$$g_1i(\theta) = (1 - \gamma_i)\sigma_a^2,$$

$$g_2i(\theta) = (\bar{x}_i - \gamma_i\bar{x}_{i(s)})^T (X^T(s)_i V^{-1}(s)_i X(s)_i)^{-1} (\bar{x}_i - \gamma_i\bar{x}_{i(s)})$$

$$g_3i(\theta) = (1 - \gamma_i)^2\gamma_i\sigma_e^4\sigma_a^{-2}M(\theta),$$

with $M(\theta) = \sigma_a^4 \text{Var}(\hat{\sigma}_a^2) + \sigma_e^4 \text{Var}(\hat{\sigma}_e^2) - 2\sigma_a^2\sigma_e^2 \text{Cov}(\hat{\sigma}_a^2, \hat{\sigma}_e^2)$. The asymptotic variances and covariance of the REML estimators $\hat{\sigma}_a^2$ and $\hat{\sigma}_e^2$ needed in $M(\theta)$ (i.e. under fixed or bounded small area size
asymptotics) are given by

\[
\text{Var}(\hat{\sigma}_a^2) = \frac{2}{a} \sum_{i=1}^{g} \left\{ \frac{s_i - 1}{\sigma_e^4 + (\sigma_e^2 + s_i \sigma_a^2)^2} \right\},
\]

\[
\text{Var}(\hat{\sigma}_e^2) = \frac{2}{a} \sum_{i=1}^{g} \frac{s_i^2}{(\sigma_e^2 + s_i \sigma_a^2)^2},
\]

\[
\text{Cov}(\hat{\sigma}_a^2, \hat{\sigma}_e^2) = -\frac{2}{a} \sum_{i=1}^{g} \frac{s_i}{(\sigma_e^2 + s_i \sigma_a^2)^2},
\]

where

\[
a = \left\{ \sum_{i=1}^{g} \frac{s_i^2}{(\sigma_e^2 + s_i \sigma_a^2)^2} \right\} \left[ \sum_{i=1}^{g} \left\{ \frac{s_i - 1}{\sigma_e^4 + (\sigma_e^2 + s_i \sigma_a^2)^2} \right\} - \sum_{i=1}^{g} \frac{s_i}{(\sigma_e^2 + s_i \sigma_a^2)^2} \right].
\]

Rao & Molina (2015) showed that \(E(\hat{\mathcal{M}}_{i}^{\text{sam}} - \bar{y}_i)^2\), the unconditional MSE of \(\hat{\mathcal{M}}_{i}^{\text{sam}}\) for prediction \(\bar{y}_i\) can be approximated to second order by

\[
\text{MSE}_{\text{RM},i} = k_i^2 \left[ g_{1i}(\theta) + \tilde{g}_{2i}(\theta) + g_{3i}(\theta) \right] + (N_i - n_i) \sigma_e^2 / N_i^2,
\]

where \(\tilde{g}_{2i}(\theta)\) is obtained from \(g_{2i}(\theta)\) by replacing \(\bar{x}_i\) with \(\bar{x}_{i(r)}\).

Using the second-order approximations, unbiased estimators of \(\text{MSE}_{\text{PR}}\) (Datta & Lahiri 2000) and of \(\text{MSE}_{\text{RM}}\) (Rao & Molina 2015) are respectively given by

\[
\overline{\text{MSE}}_{\text{PR},i} = g_{1i}(\hat{\theta}) + g_{2i}(\hat{\theta}) + 2g_{3i}(\hat{\theta}).
\]

and

\[
\text{MSE}_{\text{RM},i} = k_i^2 \left[ g_{1i}(\hat{\theta}) + \tilde{g}_{2i}(\hat{\theta}) + 2g_{3i}(\hat{\theta}) \right] + (N_i - n_i) \sigma_e^2 / N_i^2.
\]

Based on the above approximations, 100(1 - \epsilon)% prediction intervals for \(\hat{\eta}_i\) and \(\bar{y}_i\) are

\[
[\hat{\mathcal{M}}_{i}^{\text{clp}} - \Phi^{-1}(1 - \epsilon/2)\overline{\text{MSE}}_{\text{PR},i}^{1/2}, \hat{\mathcal{M}}_{i}^{\text{clp}} + \Phi^{-1}(1 - \epsilon/2)\overline{\text{MSE}}_{\text{PR},i}^{1/2}]
\]

\[2\]
and

$$[\hat{M}_i^{\text{sam}} - \Phi^{-1}(1 - \varepsilon/2)\overline{\text{MSE}_{\text{RM},i}}, \hat{M}_i^{\text{sam}} + \Phi^{-1}(1 - \varepsilon/2)\overline{\text{MSE}_{\text{RM},i}}],$$

respectively.

A simulation result for the Chatterjee et al (2008) prediction interval

Table S1: Simulated coverage and length of prediction intervals when \(\alpha_i\) has a mixture distribution and \(e_{ij}\) has a normal distribution with variances \(\hat{\sigma}_\alpha^2 = 64\) and \(\hat{\sigma}_e^2 = 100\), respectively.

| Method     | sam-LW | clp-LW | clp-PR | clp-Chatterjee |
|------------|--------|--------|--------|---------------|
|            | Area   |        |        |               |
|            |  N_i   | n_i    | Cvge   | Rlen          | Cvge   | Rlen          |
|            |        |        |        |               |        |               |
|            | 1      | 40     | 20     | 0.956 0.03   | 0.948 0.03 | 0.974 0.23 | 0.739 1.65 |
|            | 2      | 86     | 43     | 0.945 0.03   | 0.931 0.06 | 0.982 0.26 | 0.846 2.75 |
|            | 3      | 166    | 42     | 0.962 0.06   | 0.966 0.05 | 0.972 0.15 | 0.765 2.38 |
|            | 4      | 105    | 26     | 0.965 0.07   | 0.956 0.06 | 0.966 0.12 | 0.701 1.67 |
|            | 5      | 181    | 45     | 0.967 0.03   | 0.966 0.03 | 0.977 0.13 | 0.767 2.42 |
|            | 6      | 190    | 48     | 0.951 0.03   | 0.954 0.02 | 0.966 0.13 | 0.772 2.5   |
|            | 7      | 47     | 20     | 0.95 0.03    | 0.947 0.01 | 0.966 0.17 | 0.724 1.52 |
|            | 8      | 124    | 31     | 0.965 0.05   | 0.954 0.03 | 0.966 0.1  | 0.716 1.84 |
|            | 9      | 183    | 46     | 0.953 0.04   | 0.954 0.03 | 0.973 0.14 | 0.782 2.47 |
|            | 10     | 128    | 32     | 0.963 0.05   | 0.957 0.05 | 0.971 0.13 | 0.753 1.94 |
|            | 11     | 113    | 28     | 0.961 0.04   | 0.956 0.02 | 0.962 0.09 | 0.703 1.69 |
|            | 12     | 193    | 48     | 0.96 0.05    | 0.953 0.04 | 0.974 0.15 | 0.767 2.56 |
|            | 13     | 113    | 28     | 0.966 0.08   | 0.963 0.06 | 0.966 0.13 | 0.715 1.78 |
|            | 14     | 148    | 37     | 0.966 0.09   | 0.964 0.08 | 0.973 0.17 | 0.788 2.25 |
|            | 15     | 132    | 33     | 0.954 0.02   | 0.948 0.01 | 0.97 0.08  | 0.721 1.87 |
Additional results for the consumer expenditure population

Table S2: Simulated relative design-bias and design RMSEs of the point estimators sam and clp, together with the design-averages of the LW, RM and PR estimators of their RMSEs for the consumer expenditure on fresh milk products in each state in 2002. The states are in the same order as in Table 3. * identifies states in Group 3 and † identifies states in Group 2.

| STATE | ARB-sam | ARB-clp | AVE-LW | AVE-RM | AVE-PR | RMSE-sam-T | RMSE-clp-T |
|-------|---------|---------|--------|--------|--------|------------|------------|
| 16    | -0.0167 | -0.0332 | 0.4412 | 0.3701 | 0.3763 | 0.2755     | 0.2250     |
| 50†   | -0.0533 | -0.1113 | 0.4486 | 0.3733 | 0.3746 | 0.5242     | 0.5983     |
| 31    | 0.0450  | 0.0867  | 0.4790 | 0.3858 | 0.3746 | 0.2973     | 0.3749     |
| 22†   | 0.0799  | 0.1501  | 0.4840 | 0.3876 | 0.3745 | 0.3085     | 0.5257     |
| 21    | 0.0261  | 0.0497  | 0.4888 | 0.3895 | 0.3754 | 0.2457     | 0.2650     |
| 15    | -0.0139 | -0.0240 | 0.4933 | 0.3910 | 0.3756 | 0.2740     | 0.2237     |
| 32†   | -0.0768 | -0.1400 | 0.4976 | 0.3922 | 0.3750 | 0.5596     | 0.7582     |
| 37†   | 0.0847  | 0.1472  | 0.5016 | 0.3932 | 0.3745 | 0.3679     | 0.5272     |
| 1     | 0.0025  | 0.0041  | 0.4104 | 0.3430 | 0.3618 | 0.1645     | 0.1375     |
| 45    | 0.0007  | -0.0032 | 0.4143 | 0.3457 | 0.3623 | 0.3675     | 0.2918     |
| 2 *   | -0.0725 | -0.1447 | 0.3886 | 0.3282 | 0.3552 | 0.5809     | 0.8875     |
| 9 *   | 0.1437  | 0.2606  | 0.3788 | 0.3239 | 0.3543 | 0.5889     | 0.9045     |
| 41    | -0.0082 | -0.0185 | 0.3728 | 0.3168 | 0.3487 | 0.2795     | 0.2061     |
| 49    | 0.0020  | 0.0021  | 0.3562 | 0.3064 | 0.3454 | 0.2485     | 0.2101     |
| 18    | 0.0210  | 0.0403  | 0.3463 | 0.2993 | 0.3413 | 0.3341     | 0.2771     |
| 27 *  | -0.0566 | -0.1176 | 0.3395 | 0.2944 | 0.3379 | 0.5262     | 0.7639     |
| 8†    | -0.0333 | -0.0663 | 0.3268 | 0.2854 | 0.3323 | 0.4196     | 0.4474     |
| 13    | 0.0072  | 0.0114  | 0.3102 | 0.2733 | 0.3237 | 0.2245     | 0.1691     |
| 24†   | 0.0520  | 0.1032  | 0.3053 | 0.2705 | 0.3237 | 0.3205     | 0.4699     |
| 29    | 0.0145  | 0.0307  | 0.2990 | 0.2653 | 0.3191 | 0.1706     | 0.1718     |
| 53    | -0.0090 | -0.0213 | 0.2945 | 0.2622 | 0.3176 | 0.1861     | 0.1671     |
| 55    | 0.0647  | 0.0868  | 0.4673 | 0.3542 | 0.3526 | 0.3316     | 0.3692     |
| 51    | -0.0024 | -0.0033 | 0.4692 | 0.3549 | 0.3532 | 0.2490     | 0.2035     |
| 25    | -0.0063 | -0.0094 | 0.4519 | 0.3469 | 0.3487 | 0.2044     | 0.1659     |
| 4     | -0.0097 | -0.0141 | 0.4468 | 0.3442 | 0.3473 | 0.2772     | 0.2411     |
| 26    | 0.0318  | 0.0442  | 0.4327 | 0.3371 | 0.3429 | 0.2713     | 0.2599     |
| 34    | 0.0141  | 0.0187  | 0.4053 | 0.3230 | 0.3340 | 0.2577     | 0.2243     |
| 17    | 0.0271  | 0.0370  | 0.4057 | 0.3229 | 0.3337 | 0.2180     | 0.2137     |
| 39    | 0.0113  | 0.0147  | 0.3397 | 0.2868 | 0.3073 | 0.2111     | 0.1869     |
| 42    | 0.0455  | 0.0572  | 0.3181 | 0.2741 | 0.2969 | 0.2882     | 0.2954     |
| 36    | 0.0002  | -0.0003 | 0.3064 | 0.2668 | 0.2909 | 0.2584     | 0.2304     |
| 12    | -0.0286 | -0.0389 | 0.3040 | 0.2653 | 0.2897 | 0.2679     | 0.2762     |
| 48    | 0.0234  | 0.0305  | 0.2942 | 0.2591 | 0.2840 | 0.2436     | 0.2388     |
| 6     | -0.0184 | -0.0254 | 0.2577 | 0.2351 | 0.2618 | 0.2309     | 0.2338     |
Prediction intervals treating the random effects as fixed

To treat the random effects as fixed, we rewrite the model (2.1) and (2.2) as a regression model

\[ y_{ij} = z_{ij}^* \chi + e_{ij}, \quad \text{for} \ j = 1, \ldots, N_i, \ i = 1, \ldots, g, \]

where \( z_{ij}^* = [u_i^T, x_i^{(w)}^T, v_i^T]^T \) with \( v_i \) a \( g \)-vector of zeros with a one in position \( i \), and \( \chi = [\xi^T, \beta_z^T, \alpha_1, \ldots, \alpha_g]^T \) with \( \sum_{i=1}^{g} \alpha_i = 0 \). The normal maximum likelihood estimator \( \hat{\chi} \) of \( \chi \) is the (constrained) ordinary least squares estimator which can be computed using \( \text{lm} \) in R with the sum to zero constraint on the \( \alpha_i \). The optimal model-based predictor of \( \bar{y}_i \) is the composite estimator

\[ \hat{y}_i^{\text{com-fixed}} = (1 - k_i) \bar{y}_{i(s)} + k_i \bar{z}_i^{*T} \hat{\chi}, \]

where \( \bar{z}_i^{*(r)} = [u_i^T, \bar{x}_i^{(w)(r)}^T, v_i^T]^T \) and an approximate 100(1 - \( \epsilon \))% prediction interval for \( \bar{y}_i \) is

\[ \left[ \bar{y}_i^{\text{com-fixed}} - \Phi^{-1}(1 - \epsilon / 2) k_i \left\{ \frac{\hat{\sigma}_e^2}{N_i - n_i} + \bar{z}_i^{*(r)} \hat{V}_\chi \bar{z}_i^{*} \right\}^{1/2}, \bar{y}_i^{\text{com-fixed}} + \Phi^{-1}(1 - \epsilon / 2) k_i \left\{ \frac{\hat{\sigma}_e^2}{N_i - n_i} + \bar{z}_i^{*(r)} \hat{V}_\chi \bar{z}_i^{*} \right\}^{1/2} \right], \]

where \( \hat{\sigma}_e^2 \) estimates \( \sigma_e^2 \), and \( \hat{V}_\chi \) estimates the variance of \( \hat{\chi} \). Similarly, a synthetic estimator of \( \bar{y}_i \) is

\[ \bar{y}_i^{\text{syn-fixed}} = \bar{z}_i^* \chi, \]

where \( \bar{z}_i^* = [u_i^T, x_i^{(w)}^T, v_i^T]^T \), and a second approximate 100(1 - \( \epsilon \))% prediction interval for \( \bar{y}_i \) is

\[ \left[ \bar{y}_i^{\text{syn-fixed}} - \Phi^{-1}(1 - \epsilon / 2) (\bar{z}_i^* \hat{V}_\chi \bar{z}_i^*)^{1/2}, \bar{y}_i^{\text{syn-fixed}} + \Phi^{-1}(1 - \epsilon / 2) (\bar{z}_i^* \hat{V}_\chi \bar{z}_i^*)^{1/2} \right]. \]
**Design-based simulation**

Tables S3 and S4 show the empirical design-coverage and the design-average relative length of the intervals for the settings with variances $\sigma_a^2 = 4$ and $\sigma_e^2 = 100$ when $e_{ij}$ has a normal distribution and $\alpha_i$ has either a normal distribution or a mixture distribution; settings with non-normal $e_{ij}$ are included in the supplementary material. The areas are again presented and labeled in order of increasing size. The Monte Carlo standard errors for the design-coverage probabilities are approximately less than 0.01.

In all settings, the model holds and the within area variances are constant. Nonetheless, when $\sigma_e^2/\sigma_a^2$ is large (as in Tables S3 and S4), the kind of results we saw in Table 3 occur. In Table S3, there are 5 Group 3 areas, numbers $6$ ($\hat{\alpha}/\hat{\sigma}_a = -1.278$, $n_6 = 20$), $7$ ($\hat{\alpha}/\hat{\sigma}_a = -1.115$, $n_7 = 27$), $21$ ($\hat{\alpha}_{21}/\hat{\sigma}_a = -1.359$, $n_{21} = 41$), $24$ ($\hat{\alpha}_{24}/\hat{\sigma}_a = 0.991$, $n_{24} = 46$) and $28$ ($\hat{\alpha}_{28}/\hat{\sigma}_a = 1.195$, $n_{28} = 52$), and 6 Group 2 areas, numbers $2$ ($\hat{\alpha}/\hat{\sigma}_a = -0.290$, $n_2 = 5$), $4$ ($\hat{\alpha}/\hat{\sigma}_a = 0.860$, $n_4 = 17$), $10$ ($\hat{\alpha}/\hat{\sigma}_a = 0.948$, $n_{10} = 32$), $15$ ($\hat{\alpha}/\hat{\sigma}_a = 0.797$, $n_{15} = 36$), $22$ ($\hat{\alpha}_{22}/\hat{\sigma}_a = -0.837$, $n_{22} = 42$) and $25$ ($\hat{\alpha}_{25}/\hat{\sigma}_a = 0.810$, $n_{25} = 48$). In Table S4, there are 4 Group 3 areas, numbers $11$ ($\hat{\alpha}/\hat{\sigma}_a = 1.073$, $n_{11} = 31$), $16$ ($\hat{\alpha}_{16}/\hat{\sigma}_a = -0.807$, $n_{16} = 34$), $25$ ($\hat{\alpha}_{25}/\hat{\sigma}_a = -1.941$, $n_{25} = 44$) and $28$ ($\hat{\alpha}_{28}/\hat{\sigma}_a = 1.257$, $n_{28} = 53$), and 2 Group 2 areas, numbers $2$ ($\hat{\alpha}/\hat{\sigma}_a = -0.613$, $n_2 = 7$) and $10$ ($\hat{\alpha}/\hat{\sigma}_a = 0.789$, $n_{10} = 31$). On the other hand, when $\sigma_e^2/\sigma_a^2$ is not large, there are no Group 2 or 3 areas. Overall, we see that when $\sigma_e^2/\sigma_a^2$ is large, areas with extreme EBLUPs and small to moderate sample sizes are more difficult than other areas to estimate well in the design-based framework.
Table S3: Simulated design-coverage and design-average length of confidence intervals when $\alpha_i$ and $e_{ij}$ have normal distributions with variances $\sigma^2_{\alpha} = 4$ and $\sigma^2_e = 100$, respectively. * identifies states in Group 3 and † identifies states in Group 2.

| Method | N_i | n_i | $\alpha_i/\sigma^2_\alpha$ | Direct Estimator | sam-LW | sam-RM | clp-LW | clp-PR |
|--------|-----|-----|-----------------------------|------------------|--------|--------|--------|--------|
|        |     |     |                             | Cvge   | Alen   | Cvge   | Alen   | Cvge   | Alen   |
| Area   |     |     |                             |        |        |        |        |        |        |
| 1      | 123 | 4   | 0.079                       | 0.999  | 7.418  | 1.000  | 4.915  | 0.998  | 2.243  |
| 2†     | 147 | 5   | -0.290                      | 0.994  | 7.151  | 1.000  | 4.393  | 0.941  | 2.162  |
| 3      | 114 | 7   | 0.079                       | 1.000  | 5.468  | 1.000  | 3.660  | 0.997  | 2.090  |
| 4†     | 40  | 17  | 0.860                       | 0.994  | 2.322  | 0.964  | 1.838  | 0.898  | 1.554  |
| 5      | 49  | 17  | -0.073                      | 1.000  | 3.166  | 0.997  | 1.959  | 0.989  | 1.610  |
| 6†     | 46  | 20  | -1.278                      | 0.981  | 2.192  | 0.916  | 1.680  | 0.821  | 1.450  |
| 7†     | 54  | 27  | -1.115                      | 0.995  | 2.121  | 0.937  | 1.360  | 0.903  | 1.227  |
| 8      | 57  | 28  | 0.637                       | 0.987  | 1.479  | 0.976  | 1.297  | 0.985  | 1.347  |
| 9      | 107 | 32  | -0.193                      | 1.000  | 2.144  | 0.987  | 1.497  | 0.976  | 1.479  |
| 10†    | 108 | 32  | 0.948                       | 0.998  | 2.272  | 0.966  | 1.482  | 0.919  | 1.298  |
| 11     | 111 | 33  | -0.426                      | 1.000  | 2.223  | 0.982  | 1.433  | 0.985  | 1.365  |
| 12     | 115 | 34  | -0.013                      | 1.000  | 2.202  | 0.994  | 1.439  | 0.977  | 1.269  |
| 13     | 112 | 34  | 0.304                       | 0.998  | 2.170  | 0.974  | 1.430  | 0.950  | 1.264  |
| 14     | 113 | 34  | -0.229                      | 1.000  | 2.223  | 0.982  | 1.433  | 0.994  | 1.433  |
| 15†    | 120 | 36  | 0.797                       | 0.998  | 2.094  | 0.955  | 1.394  | 0.923  | 1.239  |
| 16     | 122 | 37  | 0.015                       | 1.000  | 2.061  | 0.992  | 1.371  | 0.984  | 1.222  |
| 17     | 74  | 37  | 0.060                       | 0.999  | 1.812  | 0.970  | 1.162  | 0.953  | 1.073  |
| 18     | 128 | 38  | -0.402                      | 0.998  | 1.994  | 0.984  | 1.360  | 0.967  | 1.213  |
| 19     | 81  | 40  | -0.174                      | 1.000  | 1.555  | 0.995  | 1.124  | 0.988  | 1.042  |
| 20     | 136 | 41  | -0.096                      | 1.000  | 1.892  | 0.981  | 1.305  | 0.968  | 1.174  |
| 21†    | 82  | 41  | -1.359                      | 0.998  | 1.667  | 0.920  | 1.104  | 0.893  | 1.026  |
| 22†    | 141 | 42  | -0.837                      | 1.000  | 2.299  | 0.957  | 1.292  | 0.933  | 1.165  |
| 23     | 153 | 46  | -0.379                      | 1.000  | 1.958  | 0.979  | 1.232  | 0.966  | 1.121  |
| 24†    | 152 | 46  | 0.991                       | 0.999  | 2.000  | 0.942  | 1.231  | 0.894  | 1.119  |
| 25†    | 97  | 48  | 0.810                       | 0.999  | 1.579  | 0.962  | 1.025  | 0.949  | 0.962  |
| 26     | 171 | 51  | 0.156                       | 0.999  | 1.819  | 0.984  | 1.172  | 0.974  | 1.075  |
| 27     | 172 | 52  | -0.453                      | 1.000  | 1.720  | 0.980  | 1.158  | 0.967  | 1.064  |
| 28†    | 174 | 52  | 1.195                       | 0.996  | 1.899  | 0.931  | 1.161  | 0.898  | 1.066  |
| 29     | 183 | 55  | -0.208                      | 0.998  | 1.827  | 0.980  | 1.127  | 0.965  | 1.041  |
| 30     | 185 | 56  | 0.392                       | 1.000  | 1.796  | 0.974  | 1.115  | 0.964  | 1.033  |
**Corn data**

We examine the corn data presented in Battese et al (1988). This data set comprises 37 observations on the area of corn (hectares) per segment spread across 12 counties, as shown in Table S5. The objective was to predict the average area of corn (HECTARECORN) per segment in these counties. Satellite data on the number of pixels representing corn (PIXELCORN) and the number of pixels representing soybeans (PIXELSOYBEAN) in each segment were used as auxiliary variables. Additionally, we had access to the area population means for the number of pixels corresponding to corn and soybeans in each county.

We pre-processed the auxiliary variables by centering them around their respective population means, distinguishing these new variables by appending `cent` to the variable names. Further, we incorporated the population means, distinguished by appending `avg` to the variable names, as variables distinguishing between counties. This provided us with $p_b = 2$ between county variables and $p_w = 2$ within county variables.

We then fitted the model (0.1)

$$
\text{HECTARECORN}_{ij} = \beta_0 + \beta_1 \text{PIXELCORN}avg_i + \beta_2 \text{PIXELSOY}avg_i + \beta_3 \text{PIXELCORN}cent_{ij} + \beta_4 \text{PIXELSOY}cent_{ij} + \alpha_i + e_{ij}.
$$

We used the `lmer` function to estimate the parameters in the model and calculated 3 point estimates, the county sample mean, small area mean (sam) estimator and conditional linear predictor (clp), as well as the design-based variance of the sample mean (DE), the mean squared error of our proposed approximation (LW), the Prasad-Rao (PR) and Rao-Molina (RM) mean squared error estimates. The results are presented in in Table S5.

We see in Table S5, that the sam and clp predictors are very similar. The LW estimator of the MSE
proposed in this study is larger than both the RM and PR estimators when the sample size is very small, specifically in the range from 1 to 4. Nevertheless, the difference between the proposed LW estimator and both RM and PR estimators tends to decrease with increasing sample size. Notably, when the sample size reaches 6, the proposed LW estimator is smaller than both the RM and PR estimators. There is no simple pattern in the relationship between the sample mean and the sam/clp predictors or between the variance of the direct estimator, DE, and the other MSE estimators.
Table S4: Simulated design-coverage and design-average length of confidence intervals when $\alpha_i$ has a mixture distribution and $e_{ij}$ has a normal distribution with variances $\hat{\sigma}_\alpha^2 = 4$ and $\hat{\sigma}_e^2 = 100$, respectively. * identifies states in Group 3 and † identifies states in Group 2.

| Method       | N_i | n_i | $\alpha_i/\hat{\sigma}_\alpha^2$ | sam-LW | sam-RM | clp-LW | clp-PR |
|--------------|-----|-----|----------------------------------|-------|-------|-------|-------|
|              |     |     |                                  | Cvge  | Alen  | Cvge  | Alen  |
| Baseline     | 1   | 146 | 0.117                            | 0.998 | 6.302 | 1.000 | 4.867 |
|              | 2†  | 113 | -0.613                           | 0.967 | 5.017 | 0.998 | 3.613 |
|              | 3   | 77  | 0.098                            | 1.000 | 4.783 | 0.997 | 1.863 |
|              | 4   | 40  | -0.010                           | 1.000 | 3.251 | 0.994 | 1.911 |
|              | 5   | 47  | -0.035                           | 1.000 | 2.748 | 0.994 | 1.504 |
|              | 6   | 42  | 0.367                            | 1.000 | 3.438 | 0.990 | 1.542 |
|              | 7   | 49  | -0.245                           | 1.000 | 3.372 | 0.995 | 1.507 |
|              | 8   | 60  | 0.506                            | 1.000 | 2.257 | 0.981 | 1.274 |
|              | 9   | 104 | -0.004                           | 1.000 | 2.123 | 0.996 | 1.485 |
|              | 10† | 103 | 0.789                            | 0.999 | 2.355 | 0.961 | 1.236 |
|              | 11† | 102 | 1.073                            | 0.996 | 2.056 | 0.933 | 1.235 |
|              | 12  | 109 | 0.084                            | 1.000 | 2.220 | 0.991 | 1.435 |
|              | 13  | 112 | -0.115                           | 1.000 | 2.261 | 0.999 | 1.413 |
|              | 14  | 114 | 0.006                            | 1.000 | 2.105 | 0.995 | 1.418 |
|              | 15  | 113 | -0.031                           | 0.999 | 2.314 | 0.996 | 1.415 |
|              | 16† | 114 | -0.807                           | 1.000 | 2.452 | 0.929 | 1.418 |
|              | 17  | 74  | 0.143                            | 1.000 | 1.512 | 0.997 | 1.417 |
|              | 18  | 75  | -0.375                           | 0.999 | 1.869 | 0.960 | 1.417 |
|              | 19  | 125 | -0.197                           | 0.998 | 2.068 | 0.980 | 1.616 |
|              | 20  | 127 | -0.049                           | 0.999 | 1.945 | 0.992 | 1.340 |
|              | 21  | 134 | -0.299                           | 1.000 | 2.169 | 0.981 | 1.307 |
|              | 22  | 82  | 0.487                            | 0.998 | 1.752 | 0.969 | 1.090 |
|              | 23  | 86  | 0.171                            | 0.999 | 1.667 | 0.965 | 1.064 |
|              | 24  | 147 | -0.265                           | 0.997 | 1.831 | 0.987 | 1.246 |
|              | 25† | 148 | -1.941                           | 0.961 | 1.962 | 0.704 | 1.247 |
|              | 26  | 88  | 0.267                            | 1.000 | 1.609 | 0.977 | 1.052 |
|              | 27  | 91  | 0.019                            | 1.000 | 1.631 | 0.994 | 1.023 |
|              | 28† | 176 | 1.257                            | 0.998 | 1.772 | 0.924 | 1.133 |
|              | 29  | 181 | -0.127                           | 0.999 | 1.930 | 0.984 | 1.125 |
|              | 30  | 200 | 0.063                            | 0.991 | 1.604 | 0.987 | 1.066 |

Table S5: Small area estimation for reported hectares of corn in 12 Iowa Counties

| County         | Area size | sample size | sample mean | DE    | sam-LW | sam-RM | clp-LW | clp-PR |
|----------------|-----------|-------------|-------------|-------|-------|-------|-------|-------|
| Cerro Gordo    | 545       | 1           | 0.199       | 165.760| —     | 104.929| 100.367| 105.800|
| Hamilton       | 566       | 1           | -0.159      | 96.320 | —     | 144.880| 100.512| 100.716|
| Worth          | 395       | 1           | -0.147      | 76.080 | —     | 120.919| 95.523 | 93.796 |
| Humboldt       | 424       | 2           | 0.329       | 150.875| 1180.862| 100.694| 100.456| 105.204|
| Franklin       | 564       | 3           | 0.212       | 158.623| 10.786 | 137.060| 100.514| 77.220 |
| Pocahontas     | 570       | 3           | -0.086     | 102.523| 624.721| 111.231| 101.547| 78.461 |
| Winnebago      | 402       | 3           | -0.042     | 112.773| 308.713| 116.989| 101.322| 77.190 |
| Wright         | 567       | 3           | 0.242       | 144.297| 966.822| 119.754| 101.544| 78.962 |
| Webster        | 687       | 4           | 0.075       | 117.595| 112.743| 111.870| 101.836| 76.117 |
| Hancock        | 569       | 5           | -0.205     | 109.382| 48.621 | 121.837| 101.947| 59.873 |
| Kossuth        | 965       | 5           | -0.033     | 110.252| 29.218 | 112.278| 60.933 | 58.861 |
| Hardin         | 556       | 6           | -0.251     | 114.810| 205.886| 127.496| 50.491 | 54.477 |

RM is intended to be used with sam; PR is intended to be used with clp.