# COMPOSITE SHRINKAGE ESTIMATION: REVIEW OF THEORY WITH SIMULATION STUDY AND EMPIRICAL APPLICATIONS

MICHAEL HOGUE DARIUS LI

*Bureau of Economic and Business Research*
*David Eccles School of Business*
*University of Utah*

Abstract. The increasing need for detailed data at small-area geographic levels often requires estimates from statistical surveys. While large areas often have sufficient sampling sizes for reliable measures, small areas frequently have small sampling sizes that render estimates at this granular level unusable. Conducting censuses or increasing sampling sizes might not be feasible remedies given their cost-prohibitive nature. Thus, statistical techniques are needed to improve data quality. One method involves composite shrinkage estimation, which leverages sampling strength between small areas and large areas to improve the reliability of small-area estimates. We conducted a simulation study that elucidates the usability of composite shrinkage estimators under various conditions of across-area similarity and within-area variance. Furthermore, we show a preliminary application of the statistical method, using the American Community Survey PUMS data.

## 1. Introduction to Composite Shrinkage Estimators

Composite shrinkage estimation represents one statistical method to develop efficient small-area estimators by borrowing sampling strength from the larger area that contains the small area. This design-based method contains many intuitive insights that have conceptual linkages to model-based methods and extensions of the composite estimator form, which will be reviewed in future working papers in this series. Given unreliable small-area estimates, the only recourse might be to apply the attributes of the large area to the constituent small areas. However, this would amount to assuming that the attributes of the large area exist uniformly within its boundaries.

---

*Date*: July 8, 2014.

This research was presented at the Statistics and Epidemiology Seminar hosted by the Study Design and Biostatistics Center at the University of Utah School of Medicine.

Please do not cite this working paper without the permission of the authors.

Longford shows that composite forms of small-area and larger-area estimators are in fact more efficient[1] than the dichotomous selection of either estimator [1]. The following conceptual framework of composite shrinkage estimation stems mostly from Longford. Following Longford's nomenclature and notation, we will refer to small areas as districts and larger-area estimators as "national" estimators.

1.1. **Overview.** Longford dismisses the practice of testing the null hypothesis that all district parameters $\theta_d$ are the same for all $d \in \mathcal{D}$, where $\mathcal{D}$ is the set of all districts within a particular larger area (e.g., place or county). In this framework, if the null hypothesis is rejected, then each district $d$ is assigned its district-level estimator $\hat{\theta}_d$. Otherwise, each district takes on the "national" estimator $\hat{\theta}$ [1]. Longford argues that this overly dichotomous framework does not lead to efficient district-level estimators.

Composite shrinkage estimators serve as one potential alternative. Longford discusses three variations of composite shrinkage estimators:

- composite of $\hat{\theta}_d$ and $\hat{\theta}$
- composite of $\hat{\theta}_d$ and all other district estimators $\hat{\theta}_{d'}$ $\forall d' \in \mathcal{D} \setminus \{d\}$
- composite of $\hat{\theta}_d$ and all other district estimators $\hat{\theta}_{d'}$ with spatial similarity [2]

In this paper, we will focus on the first baseline method, including its applications. The tradeoff between variance and bias underlies this baseline method. While small-area estimators are unbiased, their large sample variance could lead to unreliable estimates. On the other hand, composite shrinkage estimators incur some bias, which is offset by the reduction in variability.

1.2. **Composite Form.** The composite shrinkage estimator is formulated as the linear combination of the district estimator $\hat{\theta}_d$ and the "national" estimator $\hat{\theta}$.

$$\hat{\theta}_d^C = (1-b)\hat{\theta}_d + b\hat{\theta} \tag{1.1}$$

where $b$ is the weight assigned to the "national" estimator.

The optimal composite estimator of the form in (1.1) is obtained by minimizing $\text{MSE}\left\{\hat{\theta}_d^C; \theta_d\right\}$, which is the mean squared error of the composite estimator $\hat{\theta}_d^C$ with respect to the actual district-level paramenter $\theta_d$.

Note that $\text{MSE}\left\{\hat{\theta}_d^C; \theta_d\right\}$ can be decomposed into $\text{Var}(\hat{\theta}_d^C)$ and the squared bias $\text{B}^2(\hat{\theta}; \theta_d)$ of $\hat{\theta}$ with respect to $\theta_d$. Longford expresses the mean squared error of the composite estimator in

[1] In this context, $\hat{\theta}_1$ is more efficient than $\hat{\theta}_2$ for estimating the actual parameter $\theta$ if and only if $\text{MSE}(\hat{\theta}_1; \theta) < \text{MSE}(\hat{\theta}_2; \theta)$.

the following form:

$$\begin{aligned}\text{MSE}\left\{\hat{\theta}_d^C;\theta_d\right\} &= v_d - 2b(v_d - c_d) + b^2(v_d + v - 2c_d + B_d^2)\\ &= R_{0,d} - 2bR_{1,d} + b^2 R_{2,d}\end{aligned}$$

where

$$\begin{aligned} v_d &= \text{Var}\left(\hat{\theta}_d\right) & R_{0,d} &= v_d\\ c_d &= \text{Cov}\left(\hat{\theta}_d,\hat{\theta}\right) & R_{1,d} &= v_d - c_d\\ B_d &= \text{B}(\hat{\theta};\theta_d) = \text{E}\left[\hat{\theta}\right] - \theta_d & R_{2,d} &= v_d + v - 2c_d + B_d^2\end{aligned}$$

$\text{MSE}\left\{\hat{\theta}_d^C;\theta_d\right\}$ is a function of $b$. Let $f(b) = b^2R_{2,d} - 2bR_{1,d} + R_{0,d}$.

The minimum of $f(b)$ is attained by solving $f'(b) = 2R_{2,d}b - 2R_{1,d} = 0$.

$$b^\star = \frac{R_{1,d}}{R_{2,d}} = \frac{v_d - c_d}{v_d + v - 2c_d + B_d^2} \tag{1.2}$$

We rewrite Longford's result of (1.2) in the following way to extract some intuition behind the dynamics of the optimal weight assigned to the national estimator:

$$b^\star = \frac{\left(1 - \frac{u_d}{u_+}\right)v_d}{\left(1 - \frac{2u_d}{u_+}\right)v_d + v + (\theta - \theta_d)^2} \tag{1.3}$$

$$= \frac{\left(1 - \frac{u_d}{u_+}\right)v_d}{\left(1 - \frac{2u_d}{u_+}\right)v_d + \text{MSE}\left\{\hat{\theta};\theta_d\right\}} \tag{1.4}$$

where $\frac{u_d}{u_+}$ is the district sample as a proportion of the "national" sample. Please see Appendix A.1 for the derivation of $c_d = \frac{u_d}{u_+}v_d$, which relies on the standard assumption that the direct district-level estimators are mutually independent.

This functional form (1.4) of the optimal weight assigned to the national estimator can be thought of as a function of the district-level sampling variance $v_d$, and the mean squared error of the national estimator $\hat{\theta}$ with respect to the actual district parameter $\theta_d$. The weight assigned to the national estimator is large in the case of large district-level sample variances, given the high uncertainly of the direct estimators.

On the other hand, the optimal weight $b^\star$ is a decreasing function of the mean squared error of the national estimator $\hat{\theta}$ with respect to the actual district-level parameter $\theta_d$. This mean squared error, $\text{MSE}\left\{\hat{\theta};\theta_d\right\}$, is distinct from the one we are minimizing to obtain optimal composite estimators. As the national estimator deviates more from the actual district parameter $\theta_d$—either through large national sample variance $v$ or elevated bias between the national and district parameters—the optimal weight $b^\star$ assigned to the national estimator decreases. This makes intuitive sense, since a large mean squared error of the national estimator

$\hat{\theta}$ with respect to the actual district parameter $\theta_d$ would merit a large weight assigned to the direct district-level estimator $\hat{\theta}_d$.

1.3. **Second-Order Conditions.** Since $f$ is a polynomial in $b$, in order for a minimum to exist on $f(b) = \text{MSE}\left\{\hat{\theta}_d^C(b); \theta_d\right\}$ at some $b^\star$, $f$ must be convex on an interval containing $b^\star$, meaning that

$$f''(b) = 2R_{2,d} = 2(v_d + v - 2c_d + B_d^2) > 0 \tag{1.5}$$

Thus, the necessary and sufficient condition for the convexity of $f$ is $v_d + v + B_d^2 > 2c_d$. In order to derive a sufficient condition for the convexity of $f$, the second-order condition (1.5) is further manipulated with $c_d = \frac{u_d}{u_+} v_d$ and $\text{MSE}\left\{\hat{\theta}; \theta_d\right\} = v + B_d^2$ into the following form:

$$\left(1 - 2\frac{u_d}{u_+}\right) v_d + \text{MSE}\left\{\hat{\theta}; \theta_d\right\} > 0$$

$$\frac{1}{2}\left(1 + \frac{\text{MSE}\left\{\hat{\theta}; \theta_d\right\}}{v_d}\right) > \frac{u_d}{u_+}$$

Since $\inf \frac{1}{2}\left(1 + \frac{\text{MSE}\left\{\hat{\theta}; \theta_d\right\}}{v_d}\right) = \frac{1}{2}$, the existence of $b^\star$ minimizing $f(b)$ is guaranteed for districts with sampling weights less than half of the aggregated sample of the larger area. In practice, few districts will have sampling weights greater than half of the entire sample of the larger area than contains them. Thus, the existence of $b^\star$ will almost be a given under most conditions in which larger areas do not contain any overwhelmingly large components.

1.4. **Expected MSE.** Since $B_d^2 = (\theta - \theta_d)^2$ is difficult to estimate[2], $B_d^2$ is replaced with the district-level variance $\sigma_{\text{B}}^2 = \frac{1}{D}\sum_{d=1}^{D}(\theta_d - \theta)^2$, which is the district-level expectation of $(\theta_d - \theta)^2$, where $D$ is the total number of districts in the larger area associated with $\theta$. Instead of the optimal weight $b^\star$ assigned to the "national" estimator, we obtain a suboptimal

$$b_\dagger = \frac{v_d - c_d}{v_d + v - 2c_d + \sigma_{\text{B}}^2} \tag{1.6}$$

which minimizes the expected mean squared error $\text{eMSE}\left\{\hat{\theta}_d^C(b_d); \theta_d\right\} = \text{E}_{\mathcal{D}}\left[\text{MSE}\left\{\hat{\theta}_d^C; \theta_d\right\}\right]$.

1.5. **Estimating District-Level Variance.** We still cannot calculate $\sigma_{\text{B}}^2 = \frac{1}{D}\sum_{d=1}^{D}(\theta_d - \theta)^2$ given the unknown parameters $\theta_d$ and $\theta$. Through the method of moment matching (see Appendix A.2), $\sigma_{\text{B}}^2$ can be estimated by

$$\hat{\sigma}_{\text{B}}^2 = \frac{1}{w_+}\left\{S - \sum_{d=1}^{D} w_d(v_d - 2c_d)\right\} - v \tag{1.7}$$

[2]The "naive" estimator $(\hat{\theta} - \hat{\theta}_d)^2$ is biased for $B_d^2$. Please see Appendix A.3.

where $\frac{w_d}{w_+}$ is some weight given to each district $d$ and $S = \sum_{d=1}^{D} w_d(\hat{\theta}_d - \hat{\theta})^2$. A reasonable weight $\frac{w_d}{w_+}$ would be the district sampling weight $\frac{u_d}{u_+}$.

The resulting estimated suboptimal weight would be

$$\hat{b}_\dagger = \frac{v_d - c_d}{v_d + v - 2c_d + \hat{\sigma}^2_{\mathrm{B}}} \tag{1.8}$$

In practice, $\hat{\theta}^C_d(b^\star)$ is unattainable given the unknown $B^2_d$. Nonetheless, in the following simulation study, we will show that the suboptimal composite form $\hat{\theta}^C_d(\hat{b}_\dagger)$ performs fairly well relative to the optimal composite form $\hat{\theta}^C_d(b^\star)$.

## 2. Simulation Study

In order to better illustrate the statistical properties underlying Longford's composite shrinkage estimation method, we conduct a simulation study focused on varying conditions of across-district similarity and within-district variances.

2.1. **Method and Design.** Our simulation study involves five districts (A, B, C, D, and E) for four cases resulting from the combination of two conditions:

- similarity/dissimilarity of $\theta_d$ across districts
- large and small within-district population variances ($\sigma^2_d$)

Figures 1–4 show the distribution of the estimators $\hat{\theta}_d$, $\hat{\theta}^C_d(b^\star)$, and $\hat{\theta}^C_d(\hat{b}_\dagger)$ resulting from 1,000 repeated samples of the simulation population with the design of 2.5% simple random sample without replacement for each district. Across all four cases, the simulated district population sizes remain the same. Given our sampling design, the simulated district sample sizes also remain constant across the four simulation cases. For each estimator, the 5th ($P_5$) and 95th ($P_{95}$) percentiles from the simulated distributions are reported with the empirical MSE (Tables 1–4).

2.2. **Simulation Analysis.** In all simulation cases, Districts A and B have relatively low reductions in MSE, since they have larger sample sizes than the other districts. In fact, Districts A and B have sampling weights of nearly 42 and 28 percent, respectively. Districts with such large sampling weights probably do not need further improvement with composite estimators. Rather, their sampling strength is borrowed for the smaller districts.

Of the four simulation cases, Case SL (Figure 1) shows the most overall improvement in the estimates using composite shrinkage estimation. The optimal composite estimators $\hat{\theta}^C_d(b^\star)$ for Districts C, D, and E have empirical MSE reductions of 75.6, 87.0, and 84.8 percent, respectively. The suboptimal composite estimators $\hat{\theta}^C_d(\hat{b}_\dagger)$ for Districts C, D, and E also made large improvements with MSE reductions of 63.7, 71.5, and 72.3 percent, respectively.

These three districts leveraged across-district similarity (manifested in MSE $\left\{\hat{\theta};\theta_d\right\}$) and the large sample sizes of Districts A and B to achieve such huge improvements in the estimates. Furthermore, the large population variances underlying the districts in Case SL led to heavy leveraging of $\hat{\theta}$.

In Case DL (Figure 2), the composite estimators for Districts C, D, and E incur noticeable bias given across-district dissimilarity. One potential remedy of this issue would be to re-partition the districts into similar groups. For instance, Districts A and B could be re-grouped with other neighboring districts with large values of $\hat{\theta}_d$, while Districts C and D could leverage sampling strength from other districts with smaller values of $\hat{\theta}_d$. This would create conditions similar to those in Case SL, leading to more favorable improvements in the use of composite estimators.

Given the across-district similarity and small district variances in Case SS (Figure 3), composite estimation yields only minimal improvement to the already reliable direct district estimates. However, Case DS (Figure 4) shows that this method would not be effective for dissimilar districts with small within-district variances. Intuitively, we may not even need composite estimator methods in Cases SS and DS, since the underlying population variances for all districts are small.

FIGURE 1. Estimator Distributions from Similar Districts (Large Variance)

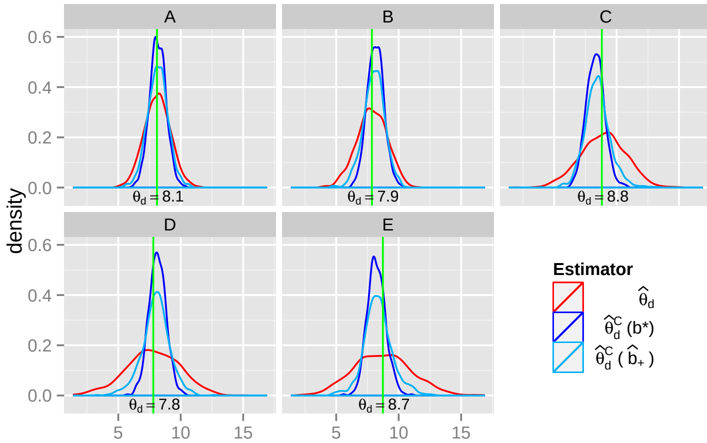



TABLE 1. Empirical MSE of Estimators from Similar Districts (Large Variance)

| | $n_d$ | $\theta_d$ | $\sigma_d$ | $\hat{\theta}_d$ | | | $\hat{\theta}_d^C(b^\star)$ | | | $\hat{\theta}_d^C(\hat{b}_\dagger)$ | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | $P_5$ | $P_{95}$ | MSE | $P_5$ | $P_{95}$ | MSE | $P_5$ | $P_{95}$ | MSE |
| A | 45 | 8.1 | 6.9 | 6.4 | 9.8 | 1.10 | 7.1 | 9.3 | 0.43 | 6.8 | 9.5 | 0.69 |
| B | 30 | 7.9 | 6.9 | 5.8 | 9.9 | 1.48 | 7.0 | 9.3 | 0.52 | 6.5 | 9.4 | 0.77 |
| C | 15 | 8.8 | 7.3 | 5.8 | 11.8 | 3.28 | 7.1 | 9.5 | 0.80 | 6.9 | 10.2 | 1.19 |
| D | 10 | 7.8 | 6.5 | 4.2 | 11.2 | 4.53 | 7.0 | 9.3 | 0.59 | 6.1 | 9.8 | 1.29 |
| E | 8 | 8.7 | 6.5 | 5.0 | 12.5 | 5.27 | 7.1 | 9.5 | 0.80 | 6.6 | 10.5 | 1.46 |

FIGURE 2. Estimator Distributions from Dissimilar Districts (Large Variance)

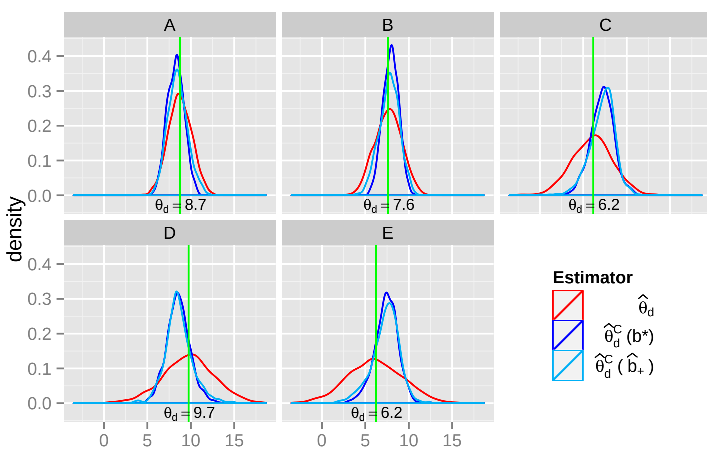



TABLE 2. Empirical MSE of Estimators from Dissimilar Districts (Large Variance)

| | $n_d$ | $\theta_d$ | $\sigma_d$ | $\hat{\theta}_d$ | | | $\hat{\theta}_d^C(b^\star)$ | | | $\hat{\theta}_d^C(\hat{b}_\dagger)$ | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | $P_5$ | $P_{95}$ | MSE | $P_5$ | $P_{95}$ | MSE | $P_5$ | $P_{95}$ | MSE |
| A | 45 | 8.7 | 9.1 | 6.6 | 11.0 | 1.79 | 6.7 | 9.8 | 1.16 | 6.7 | 10.5 | 1.37 |
| B | 30 | 7.6 | 8.5 | 5.1 | 10.2 | 2.39 | 6.4 | 9.4 | 0.90 | 5.8 | 9.7 | 1.39 |
| C | 15 | 6.2 | 9.1 | 2.1 | 10.0 | 5.62 | 4.8 | 9.2 | 2.86 | 4.6 | 9.2 | 3.35 |
| D | 10 | 9.7 | 9.5 | 4.7 | 14.6 | 8.85 | 6.4 | 10.8 | 3.27 | 6.2 | 11.5 | 3.80 |
| E | 8 | 6.2 | 9.1 | 1.1 | 11.5 | 10.00 | 5.0 | 9.5 | 3.22 | 4.3 | 9.6 | 3.78 |

FIGURE 3. Estimator Distributions from Similar Districts (Small Variance)

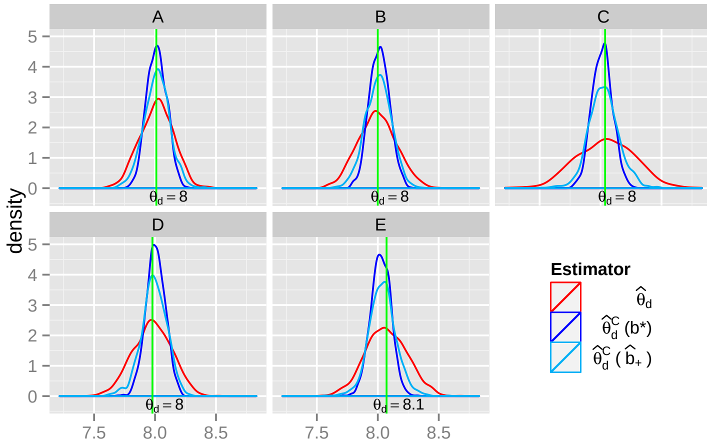



TABLE 3. Empirical MSE of Estimators from Similar Districts (Small Variance)

| | $n_d$ | $\theta_d$ | $\sigma_d$ | $\hat{\theta}_d$ | | | $\hat{\theta}^C_d(b^\star)$ | | | $\hat{\theta}^C_d(\hat{b}_\dagger)$ | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | $P_5$ | $P_{95}$ | MSE | $P_5$ | $P_{95}$ | MSE | $P_5$ | $P_{95}$ | MSE |
| A | 45 | 8.0 | 0.9 | 7.8 | 8.2 | 0.02 | 7.9 | 8.1 | 0.01 | 7.8 | 8.2 | 0.01 |
| B | 30 | 8.0 | 0.9 | 7.7 | 8.3 | 0.02 | 7.9 | 8.1 | 0.01 | 7.8 | 8.2 | 0.01 |
| C | 15 | 8.0 | 0.9 | 7.7 | 8.4 | 0.06 | 7.9 | 8.1 | 0.01 | 7.8 | 8.2 | 0.02 |
| D | 10 | 8.0 | 0.5 | 7.7 | 8.2 | 0.02 | 7.9 | 8.1 | 0.01 | 7.8 | 8.2 | 0.01 |
| E | 8 | 8.1 | 0.5 | 7.8 | 8.3 | 0.03 | 7.9 | 8.2 | 0.01 | 7.9 | 8.2 | 0.01 |

FIGURE 4. Estimator Distributions from Dissimilar Districts (Small Variance)

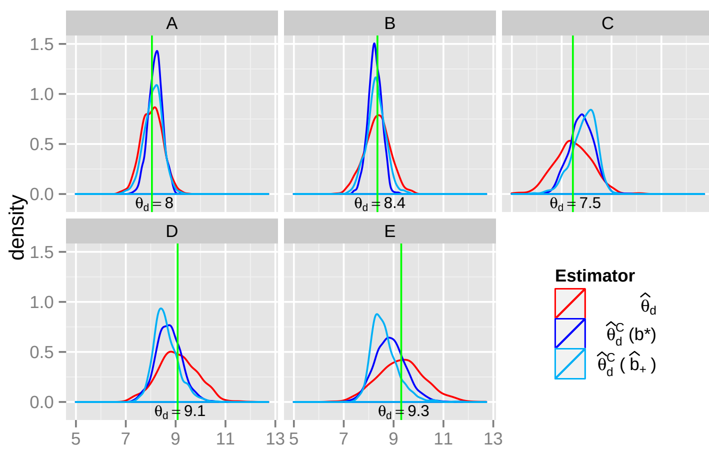



TABLE 4. Empirical MSE of Estimators from Dissimilar Districts (Small Variance)

| | $n_d$ | $\theta_d$ | $\sigma_d$ | $\hat{\theta}_d$ | | | $\hat{\theta}^C_d(b^\star)$ | | | $\hat{\theta}^C_d(\hat{b}_\dagger)$ | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | $P_5$ | $P_{95}$ | MSE | $P_5$ | $P_{95}$ | MSE | $P_5$ | $P_{95}$ | MSE |
| A | 45 | 8.0 | 2.9 | 7.4 | 8.7 | 0.19 | 7.7 | 8.7 | 0.10 | 7.5 | 8.7 | 0.14 |
| B | 30 | 8.4 | 2.8 | 7.4 | 9.2 | 0.26 | 7.8 | 8.7 | 0.09 | 7.6 | 8.9 | 0.15 |
| C | 15 | 7.5 | 2.9 | 6.2 | 8.7 | 0.57 | 6.9 | 8.6 | 0.36 | 7.0 | 8.6 | 0.45 |
| D | 10 | 9.1 | 2.4 | 7.8 | 10.4 | 0.61 | 7.9 | 9.6 | 0.39 | 7.9 | 9.5 | 0.47 |
| E | 8 | 9.3 | 2.7 | 7.8 | 10.9 | 0.87 | 7.9 | 9.9 | 0.57 | 7.9 | 9.6 | 0.75 |

**2.3. MSE Curves of Simulation Cases.** Figures 5 and 6 show the MSE curves for one sample instance from the simulated populations of Case SL and Case DL, respectively. The points on the curve are the weights $b^{\star}$, $b_{\dagger}$, and $\hat{b}_{\dagger}$ obtained from the sample instance.

The endpoints at $b = 0$ and $b = 1$ represent the use of direct estimators $\hat{\theta}_d$ and $\hat{\theta}$, respectively. Since the district parameters are very similar in Case SL, the sampling strength from the "national" estimator $\hat{\theta}$ can be heavily leveraged. In fact, the empirical weights for all districts in Case SL are fairly close to 1. On the other hand, the weights assigned to $\hat{\theta}$ in Case DL are attenuated. While the large population variances for the districts push the empirical weights toward $b = 1$, the across-district dissimilarity pulls the weights toward $b = 0$, thereby leading to the attenuation in leveraging $\hat{\theta}$.

FIGURE 5. Case SL

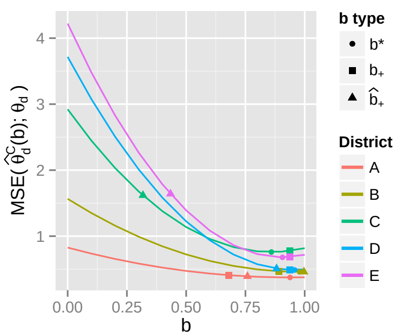
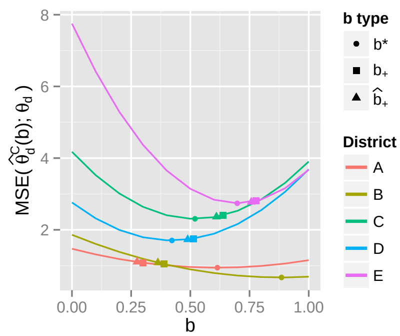



FIGURE 6. Case DL




Figures 7 and 8 show the MSE curves for one sample instance from the simulated populations of Case SS and Case DS, respectively. In Case SS, the across-district similarity with small variance can easily lead to weights closer to $b = 1$ or greater than 1. This means that the sampling strength of the "national" estimator can be heavily leveraged. On the other hand, the weights in Case DS are closer to the endpoint $b = 0$, since the district estimators are extremely dissimilar with small variance. Thus, we cannot leverage sampling strength across districts as in the other cases.

FIGURE 7. Case SS

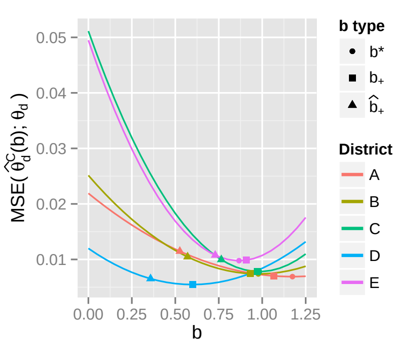
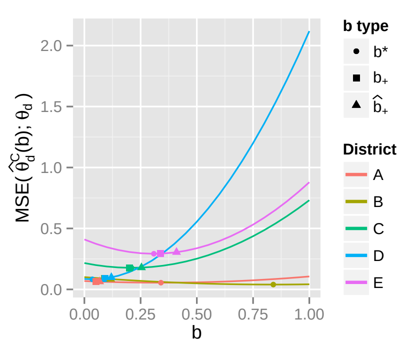



FIGURE 8. Case DS



## 3. AMERICAN COMMUNITY SURVEY PUMS

As a demonstration of the composite estimation method, we used the Public Use Microdata Sample (PUMS) from the American Community Survey (ACS) to estimate selected household attributes in Salt Lake County. Specifically, we used data from the 2008–2011 1-year ACS PUMS for the seven Public Use Microdata Areas (PUMAs) that constitute a partition of Salt Lake County (i.e. PUMAs 00501, 00502, 00503, 00504, 00505, 00506, 00507). These seven PUMAs serve as the “districts” and Salt Lake County serves as the “nation” in applying the composite estimation method. In all of the following cases, the composite estimation method was applied separately for each year.

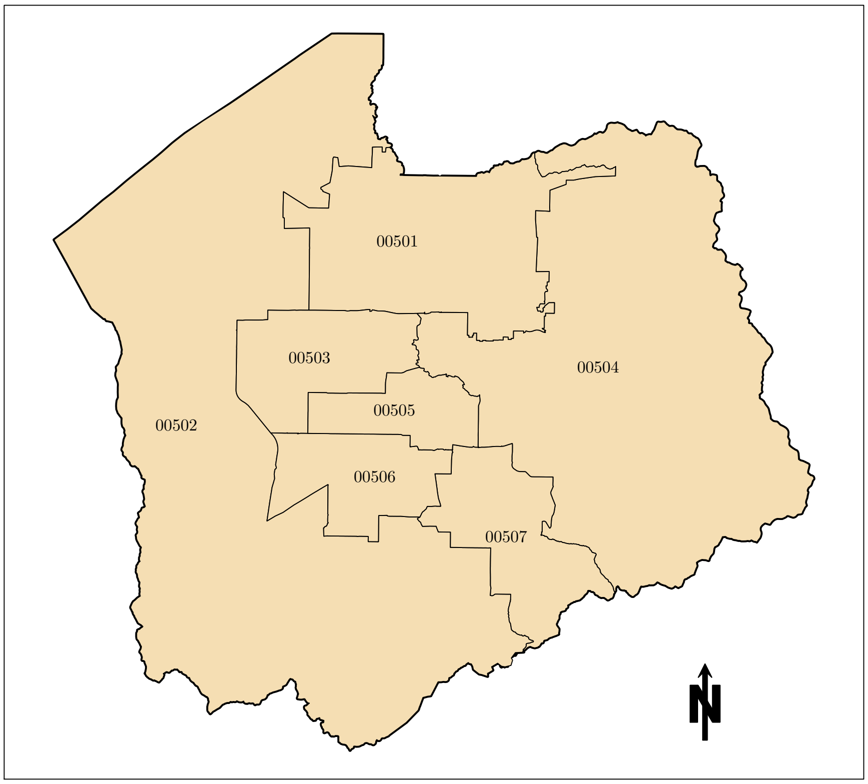



FIGURE 9. Census PUMAs within Salt Lake County

3.1. **Case 1: Persons per Household.** In this first case, we estimated the mean household size ($\mathrm{E}[\mathsf{PPH}_d]$) among all households, regardless of the household characteristics. Since no additional filter was applied to the households, the district-level sample sizes in the ACS PUMS data are very large, and thus the variance of the sampling distribution of $\overline{\mathsf{PPH}}$ is small. As a result, the empirical weights $\hat{b}_\dagger$ are very small as shown in Table 5, thereby leading to very minimal reduction in MSE. Thus, the composite estimates are closely aligned with the district estimates (Table 10). This is an expected result given the large district-level sample sizes.

Figure 10. Estimates of PPH by PUMA, 2008–2011

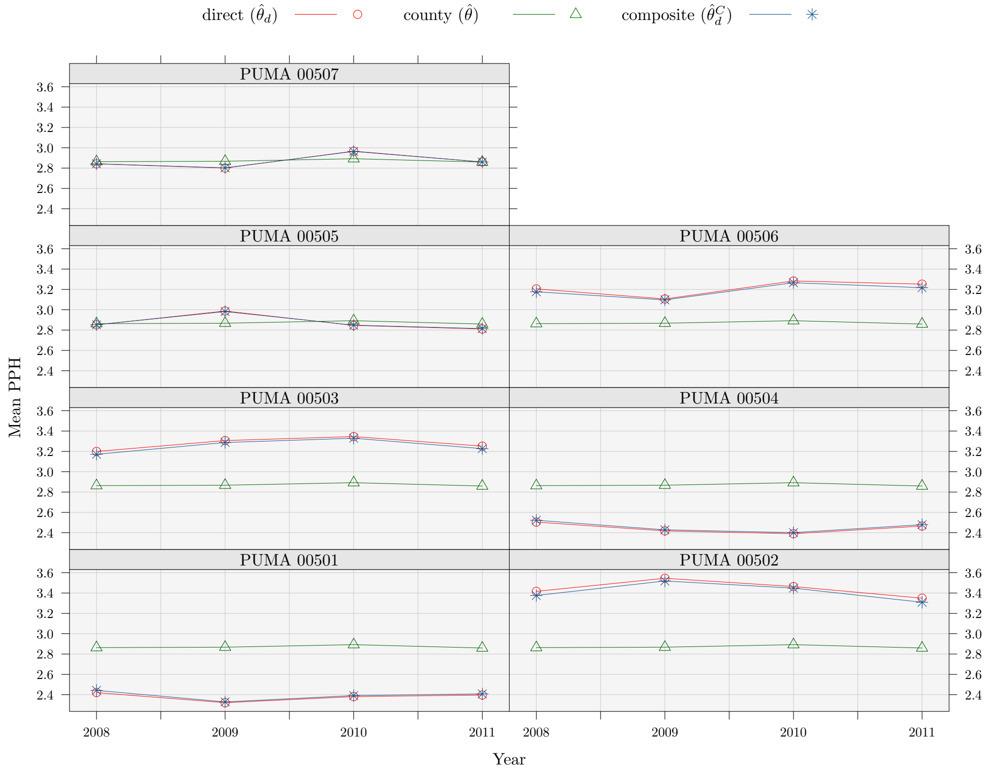



Table 5. MSE Reduction of PPH Estimates by PUMA, 2008–2011

| PUMA | year | $n_d$ | $n$ | $v_d$ | $v$ | $\hat{\theta}_d$ | $\hat{\theta}$ | $\hat{b}_\dagger$ | $\hat{\theta}_d^C$ | MSE | % ↓ MSE |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 00501 | 2008 | 653 | 3085 | 0.0083 | 0.0017 | 2.42 | 2.86 | 0.05 | 2.44 | 0.01 | 4 |
| 00502 | 2008 | 451 | 3085 | 0.0117 | 0.0017 | 3.42 | 2.86 | 0.07 | 3.38 | 0.01 | 6 |
| 00503 | 2008 | 387 | 3085 | 0.0130 | 0.0017 | 3.20 | 2.86 | 0.08 | 3.17 | 0.01 | 7 |
| 00504 | 2008 | 469 | 3085 | 0.0080 | 0.0017 | 2.50 | 2.86 | 0.05 | 2.52 | 0.01 | 4 |
| 00505 | 2008 | 348 | 3085 | 0.0120 | 0.0017 | 2.85 | 2.86 | 0.08 | 2.85 | 0.01 | 7 |
| 00506 | 2008 | 375 | 3085 | 0.0130 | 0.0017 | 3.21 | 2.86 | 0.08 | 3.18 | 0.01 | 7 |
| 00507 | 2008 | 402 | 3085 | 0.0076 | 0.0017 | 2.84 | 2.86 | 0.05 | 2.84 | 0.01 | 4 |
| 00501 | 2009 | 655 | 3102 | 0.0035 | 0.0012 | 2.32 | 2.87 | 0.01 | 2.33 | 0.00 | 1 |
| 00502 | 2009 | 468 | 3102 | 0.0091 | 0.0012 | 3.55 | 2.87 | 0.04 | 3.52 | 0.01 | 3 |
| 00503 | 2009 | 390 | 3102 | 0.0098 | 0.0012 | 3.31 | 2.87 | 0.04 | 3.29 | 0.01 | 4 |
| 00504 | 2009 | 472 | 3102 | 0.0053 | 0.0012 | 2.42 | 2.87 | 0.02 | 2.43 | 0.01 | 2 |
| 00505 | 2009 | 343 | 3102 | 0.0103 | 0.0012 | 2.99 | 2.87 | 0.05 | 2.98 | 0.01 | 4 |
| 00506 | 2009 | 364 | 3102 | 0.0088 | 0.0012 | 3.11 | 2.87 | 0.04 | 3.10 | 0.01 | 3 |
| 00507 | 2009 | 410 | 3102 | 0.0074 | 0.0012 | 2.80 | 2.87 | 0.03 | 2.80 | 0.01 | 3 |
| 00501 | 2010 | 660 | 3193 | 0.0046 | 0.0011 | 2.38 | 2.89 | 0.02 | 2.39 | 0.00 | 2 |
| 00502 | 2010 | 487 | 3193 | 0.0061 | 0.0011 | 3.46 | 2.89 | 0.03 | 3.45 | 0.01 | 2 |
| 00503 | 2010 | 404 | 3193 | 0.0079 | 0.0011 | 3.35 | 2.89 | 0.04 | 3.33 | 0.01 | 3 |
| 00504 | 2010 | 490 | 3193 | 0.0047 | 0.0011 | 2.39 | 2.89 | 0.02 | 2.40 | 0.00 | 2 |
| 00505 | 2010 | 362 | 3193 | 0.0121 | 0.0011 | 2.85 | 2.89 | 0.06 | 2.85 | 0.01 | 5 |
| 00506 | 2010 | 380 | 3193 | 0.0096 | 0.0011 | 3.28 | 2.89 | 0.04 | 3.27 | 0.01 | 4 |
| 00507 | 2010 | 410 | 3193 | 0.0081 | 0.0011 | 2.97 | 2.89 | 0.04 | 2.96 | 0.01 | 3 |
| 00501 | 2011 | 744 | 3192 | 0.0044 | 0.0008 | 2.40 | 2.86 | 0.03 | 2.41 | 0.00 | 2 |
| 00502 | 2011 | 379 | 3192 | 0.0134 | 0.0008 | 3.35 | 2.86 | 0.08 | 3.31 | 0.01 | 7 |
| 00503 | 2011 | 406 | 3192 | 0.0101 | 0.0008 | 3.25 | 2.86 | 0.06 | 3.23 | 0.01 | 5 |
| 00504 | 2011 | 472 | 3192 | 0.0055 | 0.0008 | 2.47 | 2.86 | 0.03 | 2.48 | 0.01 | 3 |
| 00505 | 2011 | 368 | 3192 | 0.0129 | 0.0008 | 2.81 | 2.86 | 0.08 | 2.81 | 0.01 | 7 |
| 00506 | 2011 | 380 | 3192 | 0.0147 | 0.0008 | 3.25 | 2.86 | 0.09 | 3.22 | 0.01 | 8 |
| 00507 | 2011 | 443 | 3192 | 0.0083 | 0.0008 | 2.86 | 2.86 | 0.05 | 2.86 | 0.01 | 4 |

3.2. **Case 2: Persons per Household among Multigenerational Households.** In this case, we applied the composite estimation method to mean household size among multigenerational households ($\mathrm{E}[\mathsf{MGPPH}_d]$). The district-level sample sizes are much smaller, and the variance of the sampling distribution of $\overline{\mathsf{MGPPH}}_d$ is much larger. Given the large variability associated with the direct district estimates, the composite estimation method can leverage $\hat{\theta}$ more heavily, as evidenced by the large empirical weights $\hat{b}_\dagger$ shown in Table 6. As a result, the composite estimates are pushed in the direction of the county estimates (Figure 11). Unlike in the first case, the composite estimator in Case 2 yields substantial MSE reduction, since the relatively large $v_d$ across the PUMAs leads to sizeable leveraging of $\hat{\theta}$.

3.3. **Extension of Case 1: PPH Using Auxiliary Data from Census 2010.** In Case 1, the method of moments estimator of $\sigma^2_{\mathrm{B}}$ produced a negative estimate for the year 2010. The simplest remedy was to use the naive and biased estimator $(\hat{\theta}_d - \hat{\theta})^2$ to ensure a positive value for this squared bias term (see Appendix A.3). This solution has the added benefit of tracking between-PUMA heterogeneity in $\theta_d - \theta$, as opposed to the method of moments estimator, which assigns one estimate of the squared bias to all districts.

An alternative way to estimate $B^2_d$ involves the use of data from another survey, data from the same survey but for different years, or census data. In this extension of Case 1, we used PUMA-level data from the 2010 Census as stand-ins for $\mathrm{E}[\mathsf{PPH}]$ and $\mathrm{E}[\mathsf{PPH}_d]$ and computed $B^2_d \approx (\mathrm{E}[\mathsf{PPH}] - \mathrm{E}[\mathsf{PPH}_d])^2$ for each district. This retains the benefit of tracking between-PUMA heterogeneity in $\theta_d - \theta$.

Interestingly, the composite estimates for PUMAs 00505 and 00507 yield substantial MSE reduction in using auxiliary census data to estimate $B^2_d$ (Table 7). Since the county estimates are extremely close to the estimates for PUMAs 00505 and 00507 (Figure 11), the actual $B^2_d$ for these two districts are presumably very close to 0. This could potentially be the reason that the moment-matching estimator $\hat{\sigma}^2_{\mathrm{B}}$ produced a negative value.

The increased improvement in MSE reduction using auxiliary census data suggests that the moment-matching estimator may not be an adequate method to estimate $B^2_d$, especially in cases where between-district heterogeneity is large.

Figure 11. Estimates of PPH among Multigenerational Households by PUMA, 2008–2011

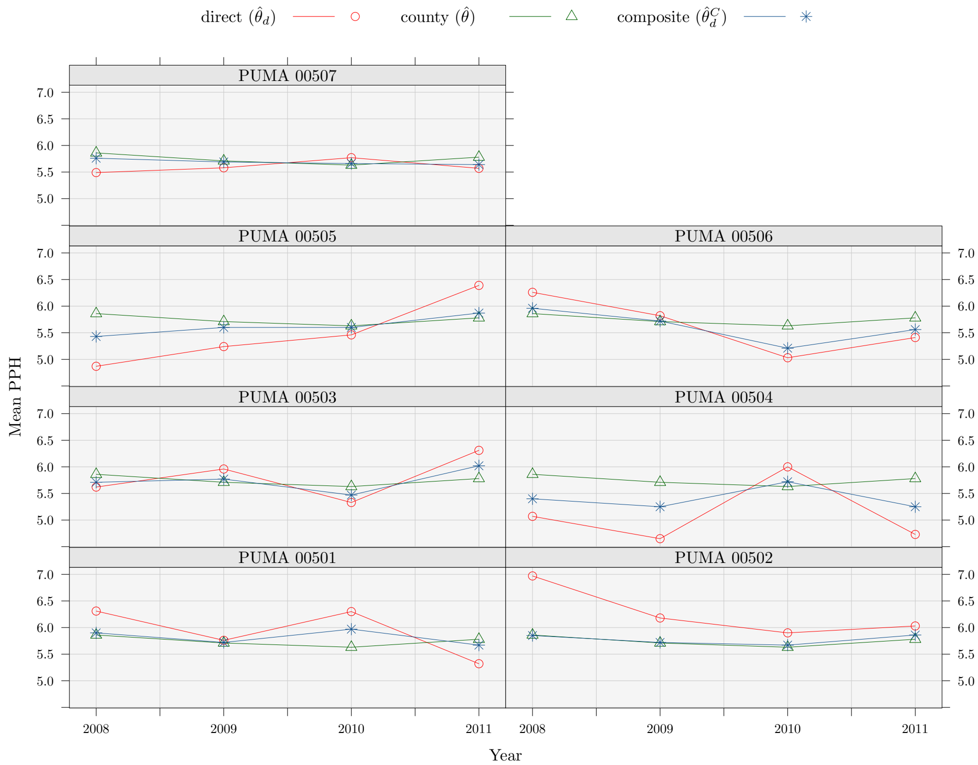



Table 6. MSE Reduction of Multigenerational PPH Estimates by PUMA, 2008–2011

| PUMA | year | $n_d$ | $n$ | $v_d$ | $v$ | $\hat{\theta}_d$ | $\hat{\theta}$ | $\hat{b}_\dagger$ | $\hat{\theta}_d^C$ | MSE | % ↓ MSE |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 00501 | 2008 | 15 | 106 | 0.7940 | 0.0637 | 6.31 | 5.86 | 0.92 | 5.90 | 0.17 | 79 |
| 00502 | 2008 | 18 | 106 | 1.0461 | 0.0637 | 6.97 | 5.86 | 1.01 | 5.85 | 0.17 | 84 |
| 00503 | 2008 | 22 | 106 | 0.1064 | 0.0637 | 5.62 | 5.86 | 0.36 | 5.71 | 0.08 | 29 |
| 00504 | 2008 | 7 | 106 | 0.1231 | 0.0637 | 5.07 | 5.86 | 0.41 | 5.40 | 0.08 | 39 |
| 00505 | 2008 | 17 | 106 | 0.2072 | 0.0637 | 4.87 | 5.86 | 0.56 | 5.43 | 0.11 | 47 |
| 00506 | 2008 | 12 | 106 | 0.4066 | 0.0637 | 6.26 | 5.86 | 0.74 | 5.96 | 0.14 | 66 |
| 00507 | 2008 | 15 | 106 | 0.3651 | 0.0637 | 5.49 | 5.86 | 0.72 | 5.76 | 0.14 | 62 |
| 00501 | 2009 | 20 | 128 | 0.1306 | 0.0293 | 5.76 | 5.71 | 0.76 | 5.72 | 0.05 | 64 |
| 00502 | 2009 | 23 | 128 | 0.2735 | 0.0293 | 6.18 | 5.71 | 0.97 | 5.72 | 0.06 | 80 |
| 00503 | 2009 | 26 | 128 | 0.1281 | 0.0293 | 5.96 | 5.71 | 0.78 | 5.77 | 0.05 | 62 |
| 00504 | 2009 | 9 | 128 | 0.0716 | 0.0293 | 4.65 | 5.71 | 0.57 | 5.25 | 0.03 | 53 |
| 00505 | 2009 | 19 | 128 | 0.1314 | 0.0293 | 5.24 | 5.71 | 0.76 | 5.60 | 0.05 | 64 |
| 00506 | 2009 | 21 | 128 | 0.2053 | 0.0293 | 5.82 | 5.71 | 0.89 | 5.72 | 0.05 | 74 |
| 00507 | 2009 | 10 | 128 | 0.1917 | 0.0293 | 5.58 | 5.71 | 0.81 | 5.69 | 0.05 | 75 |
| 00501 | 2010 | 17 | 137 | 0.4662 | 0.0234 | 6.30 | 5.63 | 0.50 | 5.97 | 0.26 | 44 |
| 00502 | 2010 | 21 | 137 | 0.3319 | 0.0234 | 5.90 | 5.63 | 0.86 | 5.67 | 0.09 | 73 |
| 00503 | 2010 | 25 | 137 | 0.1008 | 0.0234 | 5.33 | 5.63 | 0.46 | 5.47 | 0.06 | 38 |
| 00504 | 2010 | 9 | 137 | 0.4270 | 0.0234 | 6.00 | 5.63 | 0.76 | 5.72 | 0.12 | 71 |
| 00505 | 2010 | 29 | 137 | 0.1331 | 0.0234 | 5.46 | 5.63 | 0.81 | 5.60 | 0.05 | 64 |
| 00506 | 2010 | 16 | 137 | 0.1726 | 0.0234 | 5.03 | 5.63 | 0.29 | 5.21 | 0.13 | 26 |
| 00507 | 2010 | 20 | 137 | 0.1251 | 0.0234 | 5.77 | 5.63 | 0.82 | 5.66 | 0.04 | 70 |
| 00501 | 2011 | 23 | 136 | 0.2928 | 0.0373 | 5.32 | 5.78 | 0.77 | 5.67 | 0.11 | 64 |
| 00502 | 2011 | 20 | 136 | 0.2100 | 0.0373 | 6.03 | 5.78 | 0.66 | 5.86 | 0.09 | 57 |
| 00503 | 2011 | 28 | 136 | 0.1358 | 0.0373 | 6.31 | 5.78 | 0.54 | 6.02 | 0.08 | 43 |
| 00504 | 2011 | 9 | 136 | 0.1204 | 0.0373 | 4.73 | 5.78 | 0.50 | 5.25 | 0.06 | 46 |
| 00505 | 2011 | 22 | 136 | 0.4007 | 0.0373 | 6.39 | 5.78 | 0.86 | 5.87 | 0.11 | 72 |
| 00506 | 2011 | 17 | 136 | 0.0881 | 0.0373 | 5.41 | 5.78 | 0.41 | 5.56 | 0.06 | 36 |
| 00507 | 2011 | 17 | 136 | 0.0649 | 0.0373 | 5.57 | 5.78 | 0.33 | 5.64 | 0.05 | 29 |

FIGURE 12. Estimates of PPH by PUMA using Auxiliary Data from Census 2010, 2008–2011

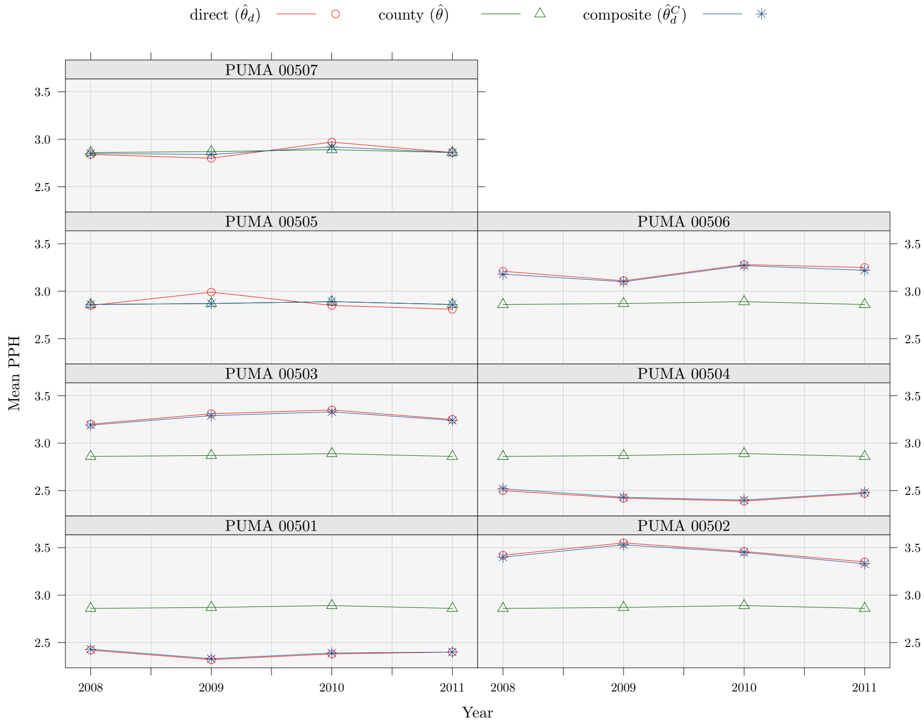



TABLE 7. MSE Reduction of PPH by PUMA using Auxiliary Census Data, 2008–2011

| PUMA | year | $n_d$ | $n$ | $v_d$ | $v$ | $\hat{\theta}_d$ | $\hat{\theta}$ | $\hat{b}_\dagger$ | $\hat{\theta}_d^C$ | MSE | % ↓ MSE |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 00501 | 2008 | 653 | 3085 | 0.0083 | 0.0017 | 2.42 | 2.86 | 0.02 | 2.43 | 0.01 | 2 |
| 00502 | 2008 | 451 | 3085 | 0.0117 | 0.0017 | 3.42 | 2.86 | 0.03 | 3.40 | 0.01 | 3 |
| 00503 | 2008 | 387 | 3085 | 0.0130 | 0.0017 | 3.20 | 2.86 | 0.04 | 3.19 | 0.01 | 3 |
| 00504 | 2008 | 469 | 3085 | 0.0080 | 0.0017 | 2.50 | 2.86 | 0.03 | 2.52 | 0.01 | 3 |
| 00505 | 2008 | 348 | 3085 | 0.0120 | 0.0017 | 2.85 | 2.86 | 0.96 | 2.86 | 0.00 | 85 |
| 00506 | 2008 | 375 | 3085 | 0.0130 | 0.0017 | 3.21 | 2.86 | 0.06 | 3.18 | 0.01 | 5 |
| 00507 | 2008 | 402 | 3085 | 0.0076 | 0.0017 | 2.84 | 2.86 | 0.61 | 2.85 | 0.00 | 53 |
| 00501 | 2009 | 655 | 3102 | 0.0035 | 0.0012 | 2.32 | 2.87 | 0.01 | 2.33 | 0.00 | 1 |
| 00502 | 2009 | 468 | 3102 | 0.0091 | 0.0012 | 3.55 | 2.87 | 0.03 | 3.53 | 0.01 | 2 |
| 00503 | 2009 | 390 | 3102 | 0.0098 | 0.0012 | 3.31 | 2.87 | 0.03 | 3.29 | 0.01 | 3 |
| 00504 | 2009 | 472 | 3102 | 0.0053 | 0.0012 | 2.42 | 2.87 | 0.02 | 2.43 | 0.01 | 2 |
| 00505 | 2009 | 343 | 3102 | 0.0103 | 0.0012 | 2.99 | 2.87 | 0.98 | 2.87 | 0.00 | 87 |
| 00506 | 2009 | 364 | 3102 | 0.0088 | 0.0012 | 3.11 | 2.87 | 0.04 | 3.10 | 0.01 | 4 |
| 00507 | 2009 | 410 | 3102 | 0.0074 | 0.0012 | 2.80 | 2.87 | 0.63 | 2.84 | 0.00 | 55 |
| 00501 | 2010 | 660 | 3193 | 0.0046 | 0.0011 | 2.38 | 2.89 | 0.01 | 2.39 | 0.00 | 1 |
| 00502 | 2010 | 487 | 3193 | 0.0061 | 0.0011 | 3.46 | 2.89 | 0.02 | 3.45 | 0.01 | 1 |
| 00503 | 2010 | 404 | 3193 | 0.0079 | 0.0011 | 3.35 | 2.89 | 0.02 | 3.33 | 0.01 | 2 |
| 00504 | 2010 | 490 | 3193 | 0.0047 | 0.0011 | 2.39 | 2.89 | 0.02 | 2.40 | 0.00 | 2 |
| 00505 | 2010 | 362 | 3193 | 0.0121 | 0.0011 | 2.85 | 2.89 | 1.02 | 2.89 | 0.00 | 90 |
| 00506 | 2010 | 380 | 3193 | 0.0096 | 0.0011 | 3.28 | 2.89 | 0.05 | 3.27 | 0.01 | 4 |
| 00507 | 2010 | 410 | 3193 | 0.0081 | 0.0011 | 2.97 | 2.89 | 0.66 | 2.92 | 0.00 | 58 |
| 00501 | 2011 | 744 | 3192 | 0.0044 | 0.0008 | 2.40 | 2.86 | 0.01 | 2.40 | 0.00 | 1 |
| 00502 | 2011 | 379 | 3192 | 0.0134 | 0.0008 | 3.35 | 2.86 | 0.04 | 3.33 | 0.01 | 3 |
| 00503 | 2011 | 406 | 3192 | 0.0101 | 0.0008 | 3.25 | 2.86 | 0.03 | 3.24 | 0.01 | 3 |
| 00504 | 2011 | 472 | 3192 | 0.0055 | 0.0008 | 2.47 | 2.86 | 0.02 | 2.48 | 0.01 | 2 |
| 00505 | 2011 | 368 | 3192 | 0.0129 | 0.0008 | 2.81 | 2.86 | 1.05 | 2.86 | 0.00 | 93 |
| 00506 | 2011 | 380 | 3192 | 0.0147 | 0.0008 | 3.25 | 2.86 | 0.07 | 3.22 | 0.01 | 6 |
| 00507 | 2011 | 443 | 3192 | 0.0083 | 0.0008 | 2.86 | 2.86 | 0.69 | 2.86 | 0.00 | 60 |

## 4. Conclusion

We provided a review of Longford's composite estimation framework, focusing on the baseline composite form of the district and national estimators. Additional intuitive elements of this framework have been expounded and demonstrated in the simulation study and initial application using the ACS PUMS data.

While this baseline composite form leads to huge improvements in the reliability of small-area estimates, especially in the case of across-district similarity, further gains could be made with additional extensions. For instance, the initial testing of the ACS PUMS data suggest that temporal similarity should be incorporated. Furthermore, the simulation study shows that across-district dissimilarity could hinder improvements in estimates. Thus, the composite form might need to be reformulated to leverage similarity across neighboring districts rather than between each individual district and the larger area.

The estimation of the squared bias $B_d^2$ will require further research, since this term could easily affect the extent to which the composite estimator improves small-area estimates. Further simulation studies might be needed in order to understand the conditions under which the moment-matching estimator $\hat{\sigma}_{\mathrm{B}}^2$ breaks down. The application of auxiliary census data in this initial application of the ACS PUMS data yielded some improvement. Potentially, the composite estimation method could be employed on various estimators of $B_d^2$ using different auxiliary data to produce more reliable estimates of the squared bias that take into account between-district heterogeneity.

This baseline composite estimation method provides a solid foundation from which to make sensible extensions that improve the reliability of small-area estimates.

## APPENDIX A. DERIVATIONS

A.1. **Covariance.** In many of the equational forms in this paper, the equality $c_d = \frac{u_d}{u_+} v_d$ is used. This rests upon the standard assumption that $\hat{\theta}_d$ are mutually independent for all $d \in \mathcal{D}$, making $\text{Cov}(\hat{\theta}_d, \hat{\theta}_j) = 0$ for all $j \neq d$.

*Proof.*

$$c_d = \text{Cov}\left(\hat{\theta}_d, \hat{\theta}\right) \tag{A.1}$$

$$= \text{Cov}\left(\hat{\theta}_d, \frac{\sum_{i=1}^{D} u_i \hat{\theta}_i}{u_+}\right) \tag{A.2}$$

$$= \text{Cov}\left(\hat{\theta}_d, \frac{u_d \hat{\theta}_d}{u_+}\right) + \sum_{j \neq d} \frac{u_j}{u_+} \text{Cov}\left(\hat{\theta}_d, \hat{\theta}_j\right) \tag{A.3}$$

$$= \frac{u_d}{u_+} v_d \tag{A.4}$$

□

A.2. **Method of Moment Matching.** The following shows brief derivations to obtain $\hat{\sigma}^2_{\text{B}}$ by the method of moment matching. First, solve for $\text{E}[S]$.

$$\begin{aligned}
\text{E}[S] &= \text{E}\left[\sum_{d=1}^{D} w_d (\hat{\theta}_d - \hat{\theta})^2\right] \\
&= \sum_{d=1}^{D} w_d \, \text{E}\left[(\hat{\theta}_d - \hat{\theta})^2\right] \\
&= \sum_{d=1}^{D} w_d \left\{v_d + v - 2c_d + (\theta_d - \theta)^2\right\}
\end{aligned}$$

Second, take the district-level expectation of $\text{E}[S]$ in order to recover the term $\sigma^2_B$.

$$\begin{aligned}
\text{E}_{\mathcal{D}}\left\{\text{E}[S]\right\} &= \sum_{d=1}^{D} w_d (v_d + v - 2c_d) + \sum_{d=1}^{D} w_d \text{E}_{\mathcal{D}}\left[(\theta_d - \theta)^2\right] \\
&= \sum_{d=1}^{D} w_d (v_d + v - 2c_d) + \sum_{d=1}^{D} w_d \sigma^2_{\text{B}} \\
&= \sum_{d=1}^{D} w_d (v_d - 2c_d) + w_+ (\sigma^2_{\text{B}} + v)
\end{aligned}$$

where $w_+ = \sum_{d=1}^{D} w_d$.

Third, apply the method of moment matching in which $S$ is equated with $\mathrm{E}_{\mathcal{D}}\left\{\mathrm{E}\left[S\right]\right\}$ rather than $\mathrm{E}\left[S\right]$.

$$S = \mathrm{E}_{\mathcal{D}}\left\{\mathrm{E}\left[S\right]\right\}$$
$$S = \sum_{d=1}^{D} w_d(v_d - 2c_d) + w_+(\sigma_{\mathrm{B}}^2 + v)$$

By solving the equation above for $\sigma_{\mathrm{B}}^2$, we obtain (1.7).

A.3. **Estimating $B_d^2$.** The term $B_d^2(\hat{\theta};\theta_d)$, defined as $\left(\mathrm{E}\left[\hat{\theta}\right] - \theta_d\right)^2$, appears in the function $b^\star$. Therefore, in order to make $b^\star$ operational, we need an appropriate estimator for $B_d^2(\hat{\theta};\theta_d)$. Note that since $\hat{\theta}$ is unbiased for $\theta$, $B_d^2(\hat{\theta};\theta_d) = (\theta - \theta_d)^2$.

A "naive" estimator for $B_d^2$ is $\widetilde{B}_d^2 := \left(\hat{\theta} - \hat{\theta}_d\right)^2$.

**Result 1 (bias of the "naive" estimator).** *The following shows that $\widetilde{B}_d^2$ is biased for $B_d^2$ and that the bias equals $v - 2c_d + v_d$.*

*Proof.*

$$\begin{aligned}
\mathrm{E}\left[\widetilde{B}_d^2\right] &= \mathrm{E}\left[(\hat{\theta} - \hat{\theta}_d)^2\right] \\
&= \mathrm{E}\left[\hat{\theta}^2 - 2\hat{\theta}\hat{\theta}_d + \hat{\theta}_d^2\right] \\
&= \mathrm{E}\left[\hat{\theta}^2\right] - 2\,\mathrm{E}\left[\hat{\theta}\hat{\theta}_d\right] + \mathrm{E}\left[\hat{\theta}_d^2\right] \\
&= \left[\mathrm{Var}\,(\hat{\theta}) + \left(\mathrm{E}\left[\hat{\theta}\right]\right)^2\right] - 2\left[\mathrm{Cov}\,(\hat{\theta},\hat{\theta}_d) + \mathrm{E}\left[\hat{\theta}\right]\mathrm{E}\left[\hat{\theta}_d\right]\right] + \left[\mathrm{Var}\,(\hat{\theta}_d) + \left(\mathrm{E}\left[\hat{\theta}_d\right]\right)^2\right] \\
&= \mathrm{Var}\,(\hat{\theta}) - 2\,\mathrm{Cov}\,(\hat{\theta},\hat{\theta}_d) + \mathrm{Var}\,(\hat{\theta}_d) + \theta^2 - 2\theta\theta_d + \theta_d^2 \\
&= \mathrm{Var}\,(\hat{\theta}) - 2\,\mathrm{Cov}\,(\hat{\theta},\hat{\theta}_d) + \mathrm{Var}\,(\hat{\theta}_d) + (\theta - \theta_d)^2 \\
&= v - 2c_d + v_d + B_d^2
\end{aligned}$$

□

*Note that in the case where the districts form a partition of the larger region that contains them, we make the substitution $c_d = \frac{u_d}{u_+}v_d$ (see Appendix A.1), which reduces the final line in the proof above to $v + v_d\left(1 - 2\frac{u_d}{u+}\right) + B_d^2$.*

Result 1 shows that $\widetilde{B}_d^2$ is upward biased by $v - 2c_d + v_d = v + v_d\left(1 - 2\frac{u_d}{u+}\right)$. Thus, the estimator $\breve{B}_d^2 := \widetilde{B}_d^2 - \hat{v} - \hat{v}_d\left(1 - 2\frac{u_d}{u+}\right) = \left(\hat{\theta} - \hat{\theta}_d\right)^2 - \hat{v} - \hat{v}_d\left(1 - 2\frac{u_d}{u+}\right)$ is unbiased for $B_d^2$.

The estimator $\breve{B}_d^2$ has two disadvantages compared to $\widetilde{B}_d^2$: (1) It can be negative and (2) it will generally have a larger sampling variance. $\breve{B}_d^2$ will be negative when $\hat{\theta}$ and $\hat{\theta}_d$ are close, $\hat{v}$ is large, $\hat{v}_d$ is large, or $u_d$ is small.

## REFERENCES

[1] LONGFORD, N. T. *Missing Data and Small-Area Estimation: Modern Analytical Equipment for the Survey Statistician.* Springer, New York, 2005.

[2] LONGFORD, N. T. Small area estimation with spatial similarity. *Computational Statistics and Data Analysis 54* (2010), 1151–1166.