

TWO-TIERED DOMAIN COMPOSITE ESTIMATION MODEL WITH PROPOSED MOMENT-MATCHING ESTIMATORS

DARIUS LI

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

ABSTRACT. The baseline method of composite shrinkage estimation of a small-area domain involves borrowing sampling strength from a larger geography. Natural extensions to this method could incorporate sampling strength from multiple domains. From an applied perspective, survey data for neighboring small areas may be sparse. Thus, we propose a two-tiered composite shrinkage estimation method that leverages sampling strength from two areas—one large area for which disaggregated data is not available and another area aggregated from smaller constituent districts with available survey data. The model also allows for the geographic flexibility in which the portions of the two areas overlap. This extension requires three moment-matching estimators to calculate the composite weights that minimize the estimated expected mean-squared error of the two-tiered domain composite estimator.

1. PROPOSED TWO-TIERED COMPOSITE ESTIMATION MODEL

Composite shrinkage estimators leverage sampling strength from large domains to improve the reliability of small-area estimates. The baseline method involves the composite form of a small area and a larger domain in which the former is situated within the latter [3]. We reviewed this method and provided additional insights via a simulation study and preliminary applications using the American Community Survey PUMS data [2]. This method can be extended to incorporate neighboring small areas [4]. However, from a practical standpoint, survey estimates for neighboring small areas may not be available on an annual basis. Furthermore, larger areas for which survey data is consistently produced may not align with the geographic boundaries of the small areas in question. Thus, we propose a practical extension to the baseline shrinkage estimation method that involves the use of two larger geographies and incorporates the flexibility of imperfect geographic alignment.

Date: August 18, 2014.

The author would like to thank Michael T. Hogue, Senior Research Statistician at the Bureau of Economic and Business Research, for his helpful review of and comments on the statistical content of this working paper.

Please do not cite this working paper without the permission of the author.

For the sake of simplicity, we will refer to small areas as districts and the two larger geographies as place and county. Despite this nomenclature, our proposed method could apply to other geographic hierarchies as long as the place includes the small area in question. The model does allow for the district to be outside the boundaries of the county. However, in practice, the county might overlap parts of the place or exist in proximity to the district in order to warrant the inclusion of the county in the model. Furthermore, place and county do not necessarily have to be corresponding geographic levels classified in the census. The selected “place” can be any geographical entity for which we have survey estimates for the district in question as well as for the other districts that comprise the place. For instance, “place” could be a collection of census tracts with available survey data. The county, on the other hand, does not necessarily have disaggregated data. A prime example of a large domain without data disaggregated to small geographic units would be the American Community Survey Public Use Microdata Areas (PUMAs). Thus, the nomenclature of place and county in this model is focused more on the availability of disaggregated district-level survey data rather than geographical hierarchy or size.

The geographical hierarchy can be formulated in two ways. In the first case, the place and county are successively larger geographies. More specifically, $\mathcal{D}_p \subset \mathcal{D}$, where \mathcal{D}_p is the set of districts within place and \mathcal{D} is the set of districts within the county. We will show that the two-tiered composite estimation model under the geographical construct of $\mathcal{D}_p \subset \mathcal{D}$ yields only marginal improvement over the baseline composite method. In the second case, $\mathcal{D}_p \not\subset \mathcal{D}$ allows for more flexibility in data applications, since the place boundaries do not have to be fully encapsulated within the county. This geographical structure lends itself to substantially improved efficiency over the baseline composite method even in cases such that the district $d \in \mathcal{D}_p \setminus \mathcal{D} \doteq \mathcal{D}_p \cap \mathcal{D}^c$.

This proposed two-tiered estimation model is perhaps most useful in cases of sparse data. For instance, we might have survey estimates for only a few small areas, which we can aggregate as a “place” for the purposes of the model. This sparsity of data could render the baseline composite method less useful if the collective sample size from these few small areas is still insufficient. Thus, this two-tiered model allows for the additional leveraging of the county survey data without having any disaggregated data for all the constituent components that comprise the county.

1.1. Composite Form. The proposed two-tiered domain composite estimation method is of the following form:

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

$$\hat{\theta}_p^C = (1 - k)\hat{\theta}_p + k\hat{\theta} \quad (1.2)$$

where $\hat{\theta}_d$ is the district-level estimator, $\hat{\theta}_p$ is the place-level estimator, and $\hat{\theta}$ is the county-level estimator. Thus, $\hat{\theta}_p^C$ is the composite estimator leveraging the sampling strength of the place and county. The district is in turn estimated by the estimator $\hat{\theta}_d^C$ —a composite of $\hat{\theta}_d$ and $\hat{\theta}_p^C$.

This two-tiered model can degenerate to the baseline composite method. In the case of $k = 1$, the model above is reduced to a baseline composite form between the district and county. In the case of $k = 0$, the model degenerates to a composite between the district and place.

Our goal is to find the optimal weights b and k assigned to the place-county composite estimator $\hat{\theta}_p^C$ and the county-level direct estimator $\hat{\theta}$. Optimality will be based on minimizing the mean squared error of $\hat{\theta}_d^C$ with respect to the unknown district parameter θ_d as shown below:

$$\begin{aligned} \text{MSE} \left\{ \hat{\theta}_d^C(b, k); \theta_d \right\} &= \text{Var} \left\{ (1-b)\hat{\theta}_d + b(1-k)\hat{\theta}_p + bk\hat{\theta} \right\} \\ &\quad + \left\{ \text{E} \left[(1-b)\hat{\theta}_d + b(1-k)\hat{\theta}_p + bk\hat{\theta} \right] - \theta_d \right\}^2 \end{aligned} \quad (1.3)$$

$$\begin{aligned} &= (1-b)^2v_d + b^2(1-k)^2v_p + b^2k^2v \\ &\quad + 2b(1-b)(1-k)c_{d,p} + 2(1-b)bkc_d + 2b^2k(1-k)c_p \\ &\quad + (-b\theta_d + b(1-k)\theta_p + bk\theta)^2 \end{aligned} \quad (1.4)$$

$$\begin{aligned} &= (1-b)^2v_d + b^2(1-k)^2v_p + b^2k^2v \\ &\quad + 2b(1-b)(1-k)w_dv_d + 2b(1-b)ku_dv_d \\ &\quad + 2b^2k(1-k)v_p \sum_{d \in \mathcal{D}_p} u_d + b^2(-\theta_d + (1-k)\theta_p + k\theta)^2 \end{aligned} \quad (1.5)$$

where

$$\begin{aligned} v_d &= \text{Var}(\hat{\theta}_d) & c_d &= \text{Cov}(\hat{\theta}_d, \hat{\theta}) \\ v_p &= \text{Var}(\hat{\theta}_p) & c_p &= \text{Cov}(\hat{\theta}_p, \hat{\theta}) \\ v &= \text{Var}(\hat{\theta}) & c_{d,p} &= \text{Cov}(\hat{\theta}_d, \hat{\theta}_p) \end{aligned}$$

and u_d and w_d are the sampling weights of district d as a proportion of the county and place sampling pool, respectively. The summation in (1.5) is over the set \mathcal{D}_p , which represents all the districts within place p . This summation is the sampling weight of all districts at the intersection of the place and county as a share of the total county sample size. Since disaggregated data may not be available for the county survey data, this sampling weight may have to be estimated.

The simplification from (1.4) to (1.5) involves the standard assumption that the estimates $\hat{\theta}_d$ are mutually independent for all $d \in \mathcal{D}$, where \mathcal{D} is the set representing all districts in the county.

1.2. Optimal Weights. Since $\text{MSE} \left\{ \hat{\theta}_d^C(b, k); \theta_d \right\}$ is a function of both b and k , the optimal two-tiered composite estimator $\hat{\theta}_d^C(b^*, k^*)$ is attained by minimizing $\text{MSE} \left\{ \hat{\theta}_d^C(b, k); \theta_d \right\}$ with respect to both b and k . Hence, the solution to the following nonlinear system of equations

resulting from taking the partial derivatives of $\text{MSE} \left\{ \hat{\theta}_d^C(b, k); \theta_d \right\}$ with respect to b and k would lead to optimal weights:

$$\begin{aligned} \frac{\partial \text{MSE} \left\{ \hat{\theta}_d^C(b, k); \theta_d \right\}}{\partial b} &= -2(1-b)v_d + 2b(1-k)^2v_p + 2bk^2v \\ &\quad + 2(1-2b)(1-k)w_dv_d + 2(1-2b)ku_dv_d \\ &\quad + 4bk(1-k)v_p \sum_{d \in \mathcal{D}_p} u_d + 2b \left\{ k(\theta - \theta_p) + (\theta_p - \theta_d) \right\}^2 = 0 \end{aligned} \quad (1.6)$$

$$\begin{aligned} \frac{\partial \text{MSE} \left\{ \hat{\theta}_d^C(b, k); \theta_d \right\}}{\partial k} &= -2b^2(1-k)v_p + 2b^2kv - 2b(1-b)w_dv_d + 2b(1-b)u_dv_d \\ &\quad + 2b^2(1-2k)v_p \sum_{d \in \mathcal{D}_p} u_d + 2b^2 \left\{ k(\theta - \theta_p)^2 + (\theta - \theta_p)(\theta_p - \theta_d) \right\} = 0 \end{aligned} \quad (1.7)$$

In the case that the place is a conglomerate of districts with no microdata to produce a place-level sample variance v_p , then the factor v_p in (1.6) and (1.7) can be replaced with $\sum_{d \in \mathcal{D}_p} w_d v_d$. One exception is that $v_p \sum_{d \in \mathcal{D}_p} u_d$ in (1.6) and (1.7) should be replaced with $\sum_{d \in \mathcal{D}_p} w_d u_d v_d$. These requirements retain the assumption that $\hat{\theta}_d$ are mutually independent for all $d \in \mathcal{D}_p$.

The equations throughout this paper are applicable to both geographical cases of $\mathcal{D}_p \subset \mathcal{D}$ and $\mathcal{D}_p \not\subset \mathcal{D}$. The summation $\sum_{d \in \mathcal{D}_p} u_d$ could be tailored to $\sum_{d \in \mathcal{D}_p \cap \mathcal{D}} u_d$ for the case $\mathcal{D}_p \not\subset \mathcal{D}$. However, this change is not necessary, since $u_d = 0$ for all $d \notin \mathcal{D}_p \cap \mathcal{D}$.

Note that the terms $(\theta - \theta_p)^2$, $(\theta_p - \theta_d)^2$, and $(\theta - \theta_p)(\theta_p - \theta_d)$ in (1.6) and (1.7) need to be estimated, since they include the unknown district, place, and county parameters. Instead of optimizing $\text{MSE} \left\{ \hat{\theta}_d^C(b, k); \theta_d \right\}$, we have to resort to the minimization of the expected mean squared error $\text{eMSE} \left\{ \hat{\theta}_d^C(b, k); \theta_d \right\} = \text{E}_{\mathcal{D}_p} \left[\text{MSE} \left\{ \hat{\theta}_d^C(b, k); \theta_d \right\} \right]$. The resulting nonlinear system of equations¹ that solve for the optimal weights b_{\dagger} and k_{\dagger} for $\text{eMSE} \left\{ \hat{\theta}_d^C(b, k); \theta_d \right\}$ are as follows:

¹We used the `nleqslv` package in R to solve the nonlinear system of equations in (1.6) and (1.7) for the optimal weights b^* and k^* as well as the estimated variant of the nonlinear system in (1.8) and (1.9) for the suboptimal weights \hat{b}_{\dagger} and \hat{k}_{\dagger} [1].

$$\begin{aligned} \frac{\partial \text{eMSE}\{\hat{\theta}_d^C(b,k);\theta_d\}}{\partial b} &= -2(1-b)v_d + 2b(1-k)^2v_p + 2bk^2v \\ &\quad + 2(1-2b)(1-k)w_dv_d + 2(1-2b)ku_dv_d \\ &\quad + 4bk(1-k)v_p \sum_{d \in \mathcal{D}_p} u_d + 2b \left\{ k^2\sigma_p^2 + \sigma_{d,p}^2 + 2k\rho_{d,p} \right\} = 0 \end{aligned} \quad (1.8)$$

$$\begin{aligned} \frac{\partial \text{eMSE}\{\hat{\theta}_d^C(b,k);\theta_d\}}{\partial k} &= -2b^2(1-k)v_p + 2b^2kv - 2b(1-b)w_dv_d + 2b(1-b)u_dv_d \\ &\quad + 2b^2(1-2k)v_p \sum_{d \in \mathcal{D}_p} u_d + 2b^2 \left\{ k\sigma_p^2 + \rho_{d,p} \right\} = 0 \end{aligned} \quad (1.9)$$

where

$$\begin{aligned} \sigma_p^2 &= \text{E}_{\mathcal{D}_p} [(\theta - \theta_p)^2] \\ \sigma_{d,p}^2 &= \text{E}_{\mathcal{D}_p} [(\theta_p - \theta_d)^2] \\ \rho_{d,p} &= \text{E}_{\mathcal{D}_p} [(\theta - \theta_p)(\theta_p - \theta_d)] \end{aligned}$$

The terms σ_p^2 , $\sigma_{d,p}^2$, and $\rho_{d,p}$ still need to be estimated. To estimate these terms, we use the method of moment matching, employed in a similar fashion as Longford's baseline composite shrinkage method [3]. The following estimators replace their corresponding terms in (1.8) and (1.9) to minimize the estimated expected mean squared error $\widehat{\text{eMSE}}\{\hat{\theta}_d^C(b,k);\theta_d\}$ and obtain the suboptimal weights \hat{b}_\dagger and \hat{k}_\dagger . Please see Appendix A for the derivations of the following proposed estimators:

$$\hat{\sigma}_p^2 = \left(\hat{\theta} - \sum_{d \in \mathcal{D}_p} w_d \hat{\theta}_d \right)^2 - \hat{v} + 2 \sum_{d \in \mathcal{D}_p} w_d \hat{c}_d - \sum_{d \in \mathcal{D}_p} w_d^2 \hat{v}_d \quad (1.10)$$

$$\hat{\sigma}_{d,p}^2 = \sum_{d \in \mathcal{D}_p} w_d \left[(\hat{\theta}_p - \hat{\theta}_d)^2 - (\hat{v}_d - 2\hat{c}_{d,p}) \right] - \hat{v}_p \quad (1.11)$$

$$\hat{\rho}_{d,p} = \sum_{d \in \mathcal{D}_p} w_d \left[(\hat{\theta} - \hat{\theta}_p)(\hat{\theta}_p - \hat{\theta}_d) - (\hat{c}_p - \hat{c}_d + \hat{c}_{d,p}) \right] + \hat{v}_p \quad (1.12)$$

where

$$\begin{aligned} \hat{v}_d &= \widehat{\text{Var}}(\hat{\theta}_d) & \hat{c}_d &= \widehat{\text{Cov}}(\hat{\theta}_d, \hat{\theta}) \\ \hat{v}_p &= \widehat{\text{Var}}(\hat{\theta}_p) & \hat{c}_p &= \widehat{\text{Cov}}(\hat{\theta}_p, \hat{\theta}) \\ \hat{v} &= \widehat{\text{Var}}(\hat{\theta}) & \hat{c}_{d,p} &= \widehat{\text{Cov}}(\hat{\theta}_d, \hat{\theta}_p) \end{aligned}$$

With the assumption that the estimates $\hat{\theta}_d$ are mutually independent for all $d \in \mathcal{D}_p$, (1.10)–(1.12) can be further simplified to

$$\hat{\sigma}_p^2 = \left(\hat{\theta} - \sum_{d \in \mathcal{D}_p} w_d \hat{\theta}_d \right)^2 - \hat{v} + \sum_{d \in \mathcal{D}_p} (2u_d - w_d) w_d \hat{v}_d \quad (1.13)$$

$$\hat{\sigma}_{d,p}^2 = \sum_{d \in \mathcal{D}_p} w_d \left[(\hat{\theta}_p - \hat{\theta}_d)^2 - (1 - 2w_d) \hat{v}_d \right] - \hat{v}_p \quad (1.14)$$

$$\hat{\rho}_{d,p} = \sum_{d \in \mathcal{D}_p} w_d \left[(\hat{\theta} - \hat{\theta}_p)(\hat{\theta}_p - \hat{\theta}_d) + (u_d - w_d) \hat{v}_d \right] + \left(1 - \sum_{d \in \mathcal{D}_p} u_d \right) \hat{v}_p \quad (1.15)$$

2. SIMULATION STUDY OF TWO-TIERED ESTIMATION METHOD

In order to understand the robustness of the two-tiered domain composite shrinkage estimator, we will conduct a simulation study to examine the dynamics of the estimator under various conditions of district similarity and different sampling designs. While the simulation cases cannot possibly be exhaustive, they are nonetheless illustrative for understanding the conditions under which the applicability of this method is most useful.

2.1. Simulation Design. The simulated population for each case in the simulation study is comprised of five districts (A, B, C, D, and E). Districts B and C are within Place X, while Districts D and E belong to Place Y. District A is found in both places. We conduct a second set of simulation cases in Section 2.7 to explore the effects of increasing geographic overlap. In this first set of simulations, however, District A is the only overlapping district. Furthermore, the constituent districts form the entirety of the geographical extent of the corresponding places. In other words, there are no residual subdivisions within the places that do fall outside the boundaries of a district. In the estimator distributions of simulation cases (Figures 1–9), each district panel is labeled as Place, District (e.g., X,A). Note that the results of Panels X,A and Y,A are distinct in the simulation cases, since the data and sampling conditions are different, despite the overlap of District A in both places. In the case of Panel X,A disaggregated data is available for all the districts in Place X, but Place Y data is available only in aggregate form for calculating $\hat{\theta}$. Similarly, for Panel Y,A the disaggregated Place Y data is used for calculating $\hat{\theta}_p$, whereas Place X data is in aggregate form for calculating $\hat{\theta}$.

Note that the nomenclature of place and county is relative. For Place X districts, the “place” is Place X, whereas Place Y serves as the “county.” For Place Y districts, the “place” is Place Y, whereas Place X serves as the “county.” While disaggregated data to the district level is available for the purposes of the simulation study, the nonavailability of this disaggregated data is imposed for the corresponding county when conducting the simulations for each district.

2.2. Simulation Cases. For illustrative purposes, nine simulation cases are shown in this section. The simulation cases are constructed from a combination of district similarity and sampling designs.

The following are the district similarity cases:

- SS: district similarity of θ_d for both Places X and Y
- SD: district similarity within Place X and district dissimilarity within Place Y
- DD: systematic district dissimilarity for both Places X and Y

The systematic district dissimilarity in Case DD entails low and high values of θ_d for the non-overlapping districts, while District A, which overlaps both Place X and Y, takes on an intermediate value of θ_d .

Each similarity case above is paired with the following designs of simple random sampling without replacement (srswor) for a total of nine simulation cases:

- a: 2.5% srswor for each district in Places X and Y
- b: 2.5% and 5% srswor for each district in Place X and Place Y, respectively
- c: 2.5% and 10% srswor for each district in Place X and Place Y, respectively

The nomenclature for the simulation cases is a combination of the district similarity and sampling design cases. For instance, Case SDb represents the simulation in which Place X has within-district similarity of θ_d with a 2.5% srswor design, whereas Place Y has within-district dissimilarity with a 5% srswor design.

Figures 1–9 show the empirical distributions of $\hat{\theta}_d$, $\hat{\theta}_d^C(b^*, k^*)$, $\hat{\theta}_d^C(\hat{b}_\dagger, \hat{k}_\dagger)$, $\hat{\theta}_d^C(\hat{b}_p, 0)$, and $\hat{\theta}_d^C(\hat{b}_c, 1)$ resulting from 1,000 repeated samples of each simulation case. For consistency, the simulated population sizes are 200 for each district with very similar underlying population standard deviations σ_d . The varying sampling rates for Place Y in the simulation cases change the sampling sizes of Districts A, D, and E accordingly.

The baseline estimators $\hat{\theta}_d^C(\hat{b}_p, 0)$ and $\hat{\theta}_d^C(\hat{b}_c, 1)$ are evaluated at the estimated suboptimal weights \hat{b}_p and \hat{b}_c , which minimize the estimated expected mean-squared error of the composite of θ_d with θ_p and θ , respectively. Thus, the performance of the two-tiered estimator $\hat{\theta}_d^C(\hat{b}_\dagger, \hat{k}_\dagger)$ could be compared to the baseline estimators $\hat{\theta}_d^C(\hat{b}_p, 0)$ and $\hat{\theta}_d^C(\hat{b}_c, 1)$, given the evaluation of $\hat{\theta}_d^C(b, k)$ at suboptimal weights using the appropriate moment-matching estimators. On the other hand, the distribution of $\hat{\theta}_d^C(b^*, k^*)$ is displayed in Figures 1–9 as the optimal result of the two-tiered method. The optimal two-tiered estimator cannot be attainable in applications given the unknown parameters θ_d , θ_p , and θ . Since $\hat{\theta}_d^C(b^*, k^*)$ is evaluated at optimal weights, this estimator should not be compared to the baseline estimators presented in Figures 1–9.

2.3. **Case SS.** With district similarity across both Places X and Y, the empirical MSE reductions in using composite estimators are generally sizeable. In Case SSa (Figure 1), the two-tiered estimator $\hat{\theta}_d^C(b^*, k^*)$ yielded an additional 6 to 11 percent reduction in MSE compared to the baseline estimators $\hat{\theta}_d^C(\hat{b}_p, 0)$ and an additional 20 to 25 percent reduction in MSE compared to $\hat{\theta}_d^C(\hat{b}_c, 1)$. This sizable reduction in MSE beyond the baseline estimators stems from the small sample sizes for each district. Given such a low sampling rate, the collective sample sizes for the place and county alone provide less information than the joint leveraging of both place and county as evidenced by the empirical MSEs in Table 1.

The advantage of the two-tiered estimators begins to slowly diminish with the increased sampling rate applied to Place Y as shown in Case SSb (Figure 2) and Case SSc (Figure 3). For instance, in Figure 3, the empirical distributions of $\hat{\theta}_d^C(b^*, k^*)$ and $\hat{\theta}_d^C(\hat{b}_c, 1)$ nearly converge for Place X districts, because the large sample size of the “county” (Place Y) makes the use of Place X survey data less relevant. Nonetheless, the use of the two-tiered estimator could still yield additional reduction in MSE, albeit marginal.

FIGURE 1. Estimator Distributions for Case SSa

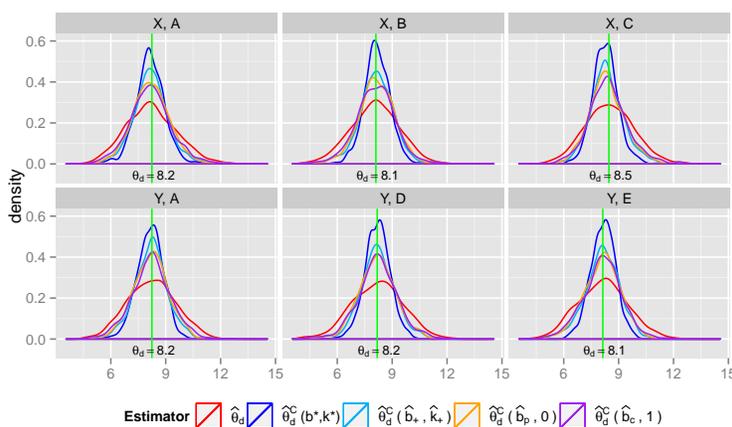


TABLE 1. Empirical MSE of Estimators for Case SSa

	n_d	σ_d	θ_d	$\hat{\theta}_d$			$\hat{\theta}_d^C(b^*, k^*)$			$\hat{\theta}_d^C(\hat{b}_p, \hat{k}_p)$			$\hat{\theta}_d^C(\hat{b}_p, 0)$			$\hat{\theta}_d^C(\hat{b}_c, 1)$			
				P_5	P_{95}	MSE	P_5	P_{95}	MSE	P_5	P_{95}	MSE	P_5	P_{95}	MSE	P_5	P_{95}	MSE	
A	X	5	3.0	8.2	6.0	10.6	1.92	6.9	9.4	0.65	6.6	9.8	1.02	6.5	9.9	1.09	6.3	10.1	1.32
A	Y	5	3.0	8.2	6.1	10.5	1.73	6.9	9.3	0.56	6.6	9.7	0.87	6.5	9.8	0.98	6.3	10.0	1.17
B	X	5	2.9	8.1	5.8	10.1	1.73	7.0	9.3	0.55	6.6	9.6	0.92	6.5	9.8	1.02	6.3	9.8	1.16
C	X	5	3.0	8.5	6.4	10.7	1.77	7.2	9.4	0.53	6.9	10.0	0.93	6.8	10.2	1.05	6.8	10.3	1.16
D	Y	5	3.2	8.2	5.9	10.4	1.95	7.0	9.4	0.60	6.7	9.8	0.99	6.5	9.9	1.10	6.4	10.0	1.23
E	Y	5	3.1	8.1	5.9	10.4	1.86	7.0	9.3	0.48	6.5	9.7	0.92	6.4	9.7	1.03	6.3	9.8	1.16

FIGURE 2. Estimator Distributions for Case SSb

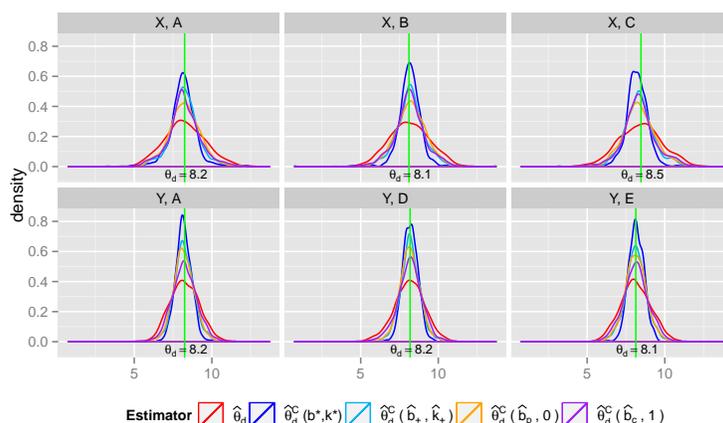


TABLE 2. Empirical MSE of Estimators for Case SSb

	n_d	σ_d	θ_d	$\hat{\theta}_d$			$\hat{\theta}_d^C(b^*, k^*)$			$\hat{\theta}_d^C(\hat{b}_\dagger, \hat{k}_\dagger)$			$\hat{\theta}_d^C(\hat{b}_p, 0)$			$\hat{\theta}_d^C(\hat{b}_c, 1)$			
				P_5	P_{95}	MSE	P_5	P_{95}	MSE	P_5	P_{95}	MSE	P_5	P_{95}	MSE	P_5	P_{95}	MSE	
A	X	5	3.0	8.2	6.1	10.5	1.77	6.8	9.3	0.66	6.7	9.8	0.90	6.5	9.9	1.04	6.4	10.0	1.12
A	Y	10	3.0	8.2	6.8	9.8	0.88	7.3	9.0	0.28	7.2	9.3	0.46	7.2	9.3	0.47	7.0	9.6	0.64
B	X	5	2.9	8.1	6.0	10.3	1.68	7.1	9.1	0.53	6.6	9.7	0.88	6.6	9.9	1.00	6.5	9.9	1.01
C	X	5	3.0	8.5	6.1	10.8	1.92	7.1	9.3	0.54	6.7	10.1	1.00	6.6	10.2	1.15	6.7	10.4	1.19
D	Y	10	3.2	8.2	6.4	9.7	0.98	7.4	9.0	0.24	7.1	9.2	0.46	7.1	9.3	0.48	6.8	9.4	0.65
E	Y	10	3.1	8.1	6.5	9.7	0.94	7.4	9.0	0.25	7.0	9.3	0.48	7.0	9.3	0.50	6.8	9.5	0.67

FIGURE 3. Estimator Distributions for Case SSc

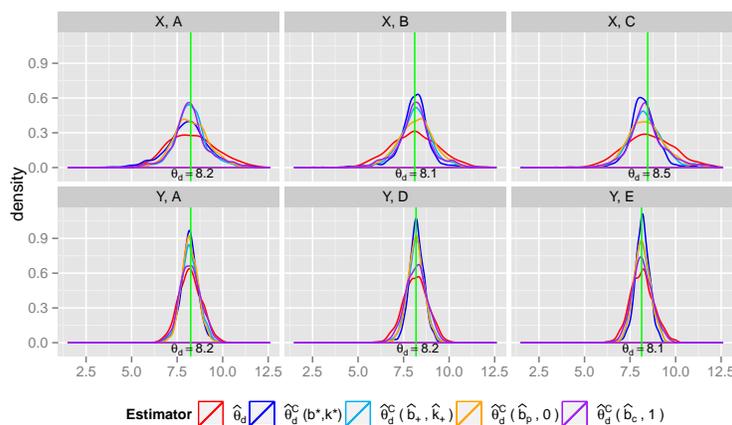


TABLE 3. Empirical MSE of Estimators for Case SSc

	n_d	σ_d	θ_d	$\hat{\theta}_d$			$\hat{\theta}_d^C(b^*, k^*)$			$\hat{\theta}_d^C(\hat{b}_\dagger, \hat{k}_\dagger)$			$\hat{\theta}_d^C(\hat{b}_p, 0)$			$\hat{\theta}_d^C(\hat{b}_c, 1)$			
				P_5	P_{95}	MSE	P_5	P_{95}	MSE	P_5	P_{95}	MSE	P_5	P_{95}	MSE	P_5	P_{95}	MSE	
A	X	5	3.0	8.2	6.3	10.6	1.75	5.9	10.0	1.44	6.9	9.8	0.77	6.8	10.0	0.94	6.7	10.1	1.00
A	Y	20	3.0	8.2	7.2	9.3	0.41	7.4	8.9	0.22	7.4	9.1	0.26	7.5	9.0	0.21	7.3	9.2	0.33
B	X	5	2.9	8.1	5.9	10.2	1.71	6.7	9.2	0.78	6.6	9.6	0.85	6.5	9.8	1.00	6.4	9.8	0.96
C	X	5	3.0	8.5	6.4	10.8	1.83	7.1	9.6	0.88	6.8	10.0	0.97	6.9	10.2	1.09	6.8	10.3	1.08
D	Y	20	3.2	8.2	7.0	9.3	0.46	7.6	8.9	0.16	7.4	9.0	0.24	7.4	8.9	0.22	7.2	9.1	0.35
E	Y	20	3.1	8.1	7.1	9.3	0.41	7.5	8.8	0.15	7.4	9.0	0.22	7.4	9.0	0.22	7.3	9.1	0.32

2.4. **Case SD.** In Cases SDa, SDb, and SDc, the actual district parameters θ_d are similar for Districts A, B, and C, whereas Place Y district parameters are dissimilar. Under this condition, the two-tiered model generally does not perform better than the baseline composite estimators for districts $d \notin \mathcal{D}_p \cap \mathcal{D}$. Even in Case SDc (Figure 6), Districts B and C do not benefit from the additional leveraging of the county survey data with four times the sampling rate of the districts in Place X. Thus, the dissimilarity of the underlying district parameters in Place Y diminishes the utility of Place Y serving as the county for Place X districts.

On the other hand, District A (within Place Y) does exhibit some increased efficiency in using $\hat{\theta}_d^C(\hat{b}_\dagger, \hat{k}_\dagger)$ rather than $\hat{\theta}_d^C(\hat{b}_c, 1)$ (Table 4). Unlike Districts B and C, District A is located at the intersection of Places X and Y. Thus, the two-tiered model can potentially counterbalance the dissimilarity within Place Y with both additional sampling strength, especially the additional District A sample data within the Place Y survey data. Note that this subtle advantage is diminished in Cases SDb and SDc in which the sampling rate of Place Y districts is increased to the point such that the overabundance of District D and E sample data introduce excessive dissimilarity and thus reduced efficiency in using the two-tiered estimator.

Despite the reduced efficiency of the two-tiered estimator under conditions of dissimilarity in Place Y, the model produces empirical distributions that generally fall in between those of the two baseline composite estimators. Thus, under certain conditions of dissimilarity, the two-tiered could in fact yield additional efficiency as shown in Case DD (Figures 7–9).

FIGURE 4. Estimator Distributions for Case SDa

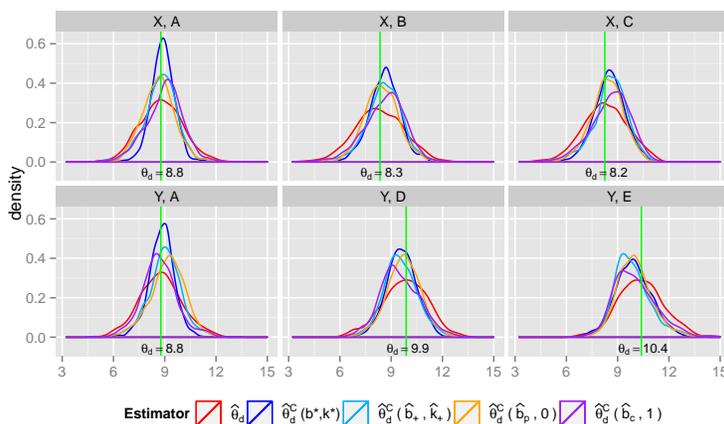


TABLE 4. Empirical MSE of Estimators for Case SDa

	n_d	σ_d	θ_d	$\hat{\theta}_d$			$\hat{\theta}_d^C(b^*, k^*)$			$\hat{\theta}_d^C(\hat{b}_\dagger, \hat{k}_\dagger)$			$\hat{\theta}_d^C(\hat{b}_p, 0)$			$\hat{\theta}_d^C(\hat{b}_c, 1)$			
				P_5	P_{95}	MSE	P_5	P_{95}	MSE	P_5	P_{95}	MSE	P_5	P_{95}	MSE	P_5	P_{95}	MSE	
A	X	5	2.8	8.8	6.8	10.9	1.55	7.8	10.0	0.48	7.1	10.3	0.97	7.1	10.3	0.98	7.1	10.7	1.23
A	Y	5	2.8	8.8	6.7	10.9	1.58	7.5	10.0	0.59	7.3	10.6	1.06	7.3	10.8	1.26	7.0	10.6	1.15
B	X	5	3.1	8.3	6.0	10.7	2.06	7.1	10.1	0.90	6.8	10.3	1.23	6.8	10.2	1.14	6.4	10.4	1.57
C	X	5	3.0	8.2	6.1	10.5	1.72	7.0	10.0	0.89	6.7	10.1	1.12	6.7	9.9	0.96	6.5	10.3	1.47
D	Y	5	3.0	9.9	7.6	11.8	1.73	8.1	11.1	0.87	8.0	11.3	1.11	8.2	11.4	0.98	7.8	11.5	1.44
E	Y	5	3.0	10.4	8.3	12.7	1.80	8.3	11.7	1.33	8.4	11.9	1.44	8.5	11.9	1.25	8.4	12.3	1.66

FIGURE 5. Estimator Distributions for Case SDb

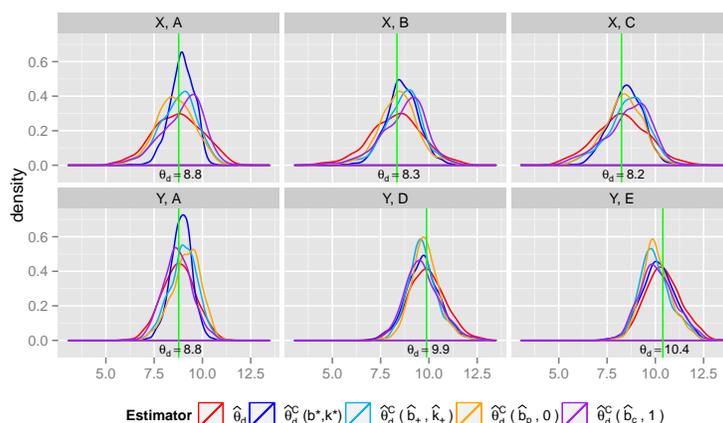


TABLE 5. Empirical MSE of Estimators for Case SDb

	n_d	σ_d	θ_d	$\hat{\theta}_d$			$\hat{\theta}_d^C(b^*, k^*)$			$\hat{\theta}_d^C(\hat{b}_\dagger, \hat{k}_\dagger)$			$\hat{\theta}_d^C(\hat{b}_p, 0)$			$\hat{\theta}_d^C(\hat{b}_c, 1)$			
				P_5	P_{95}	MSE	P_5	P_{95}	MSE	P_5	P_{95}	MSE	P_5	P_{95}	MSE	P_5	P_{95}	MSE	
A	X	5	2.8	8.8	6.7	10.8	1.58	7.8	9.9	0.42	7.1	10.2	0.97	7.0	10.2	0.98	7.0	10.5	1.22
A	Y	10	2.8	8.8	7.3	10.2	0.75	8.0	9.8	0.32	7.7	10.1	0.60	7.8	10.3	0.73	7.5	10.0	0.58
B	X	5	3.1	8.3	6.0	10.5	1.88	7.2	10.0	0.79	6.8	10.3	1.19	6.8	10.0	1.07	6.5	10.3	1.51
C	X	5	3.0	8.2	6.0	10.5	1.87	7.1	9.8	0.82	6.5	10.2	1.28	6.5	9.9	1.05	6.3	10.2	1.56
D	Y	10	3.0	9.9	8.3	11.6	0.95	8.3	11.2	0.83	8.5	11.2	0.68	8.7	11.1	0.57	8.3	11.3	0.89
E	Y	10	3.0	10.4	8.9	12.0	0.90	8.7	11.7	0.85	8.8	11.5	0.84	8.9	11.5	0.71	8.7	11.7	0.94

FIGURE 6. Estimator Distributions for Case SDc

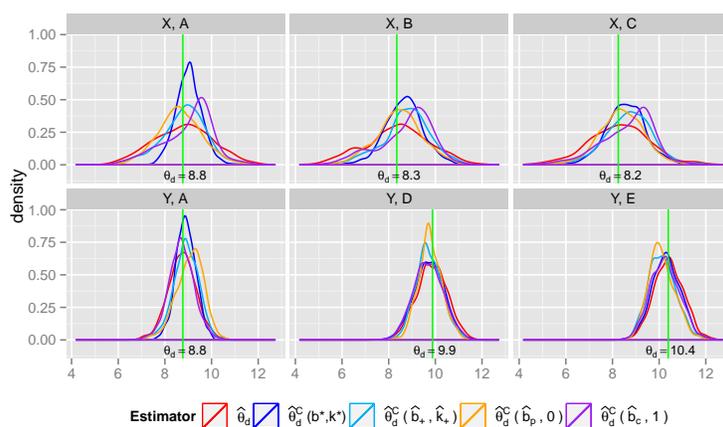


TABLE 6. Empirical MSE of Estimators for Case SDc

	n_d	σ_d	θ_d	$\hat{\theta}_d$			$\hat{\theta}_d^C(b^*, k^*)$			$\hat{\theta}_d^C(\hat{b}_\dagger, \hat{k}_\dagger)$			$\hat{\theta}_d^C(\hat{b}_p, 0)$			$\hat{\theta}_d^C(\hat{b}_c, 1)$			
				P_5	P_{95}	MSE	P_5	P_{95}	MSE	P_5	P_{95}	MSE	P_5	P_{95}	MSE	P_5	P_{95}	MSE	
A	X	5	2.8	8.8	6.7	10.8	1.58	8.1	9.9	0.38	7.0	10.3	0.95	7.1	10.3	0.93	7.0	10.4	1.21
A	Y	20	2.8	8.8	7.8	9.7	0.36	8.2	9.6	0.19	8.0	9.8	0.32	8.1	10.0	0.44	7.8	9.6	0.30
B	X	5	3.1	8.3	6.0	10.6	1.88	7.4	10.0	0.73	6.8	10.4	1.22	6.8	10.1	1.02	6.5	10.3	1.50
C	X	5	3.0	8.2	6.1	10.4	1.71	7.2	10.0	0.85	6.8	10.2	1.28	6.8	9.9	1.00	6.5	10.1	1.47
D	Y	20	3.0	9.9	8.8	10.9	0.43	8.7	10.8	0.42	8.9	10.7	0.34	9.0	10.7	0.27	8.7	10.8	0.43
E	Y	20	3.0	10.4	9.4	11.4	0.40	9.3	11.3	0.37	9.2	11.1	0.41	9.3	11.1	0.39	9.2	11.3	0.42

2.5. **Case DD.** In this set of simulations, we created a systematic dissimilarity within Places X and Y, whereby the former has lower values of θ_d and the latter has higher values of θ_d . Furthermore, the overlapping District A has a value of θ_d in between the dissimilar parameters of the non-overlapping districts. In practice, we could easily have districts with intermediate values given their intersection between two larger areas. Thus, this set of simulations test the efficiency of the two-tiered estimator under the condition that the underlying parameter is distributed spatially.

District A yields increased efficiency with the two-tiered estimator in Case DDa with additional MSE reductions of 20 percent for Panel X,A compared to the baseline composite estimator $\hat{\theta}_d^C(\hat{b}_p, 0)$ (Figure 7). This additional improvement in efficiency is retained fairly consistently for District A with the increasing sampling rate for Place Y as shown in Cases DDb and DDc (Figures 8 and 9). While $\hat{\theta}_d^C(\hat{b}_\dagger, \hat{k}_\dagger)$ yields an empirical MSE slightly larger than that of $\hat{\theta}_d^C(\hat{b}_p, 0)$ for Panel Y,A, the former is empirically less biased than the latter and thus potentially more favorable.

In fact, Panels X,A and Y,A in Figures 7–9 show the utility of the two-tiered estimator in curbing the incurred bias from the two baseline indicators. While the baseline composite estimator $\hat{\theta}_d^C(\hat{b}_p, 0)$ is slightly downward biased with the low underlying parameters in Districts B and C, $\hat{\theta}_d^C(\hat{b}_c, 1)$ is slightly upward biased given the high parameters assigned to Districts D and E. Thus, the two-tiered estimator counterbalances the incurred bias from both baseline composite estimators. For the non-overlapping districts, this effect diminishes, since the underlying parameters for these districts belong to one of the two extremities.

FIGURE 7. Estimator Distributions for Case DDa

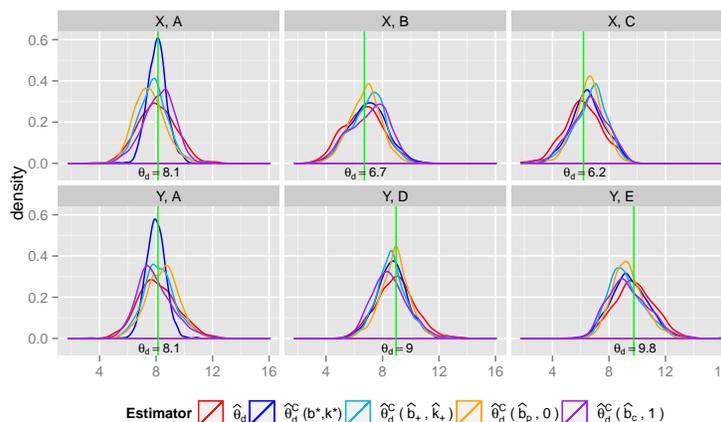


TABLE 7. Empirical MSE of Estimators for Case DDa

	n_d	σ_d	θ_d	$\hat{\theta}_d$			$\hat{\theta}_d^C(b^*, k^*)$			$\hat{\theta}_d^C(\hat{b}_\dagger, \hat{k}_\dagger)$			$\hat{\theta}_d^C(\hat{b}_p, 0)$			$\hat{\theta}_d^C(\hat{b}_c, 1)$			
				P_5	P_{95}	MSE	P_5	P_{95}	MSE	P_5	P_{95}	MSE	P_5	P_{95}	MSE	P_5	P_{95}	MSE	
A	X	5	3.1	8.1	5.9	10.2	1.78	7.0	9.3	0.49	6.1	9.5	1.24	5.9	9.4	1.54	6.2	9.9	1.29
A	Y	5	3.1	8.1	5.9	10.6	2.00	6.8	9.1	0.51	6.4	10.4	1.41	6.4	10.2	1.39	6.0	10.2	1.69
B	X	5	3.2	6.7	4.4	9.1	2.01	4.9	9.2	1.79	4.9	8.9	1.53	4.9	8.6	1.22	4.7	9.1	2.02
C	X	5	3.0	6.2	4.0	8.5	1.83	4.5	8.6	1.54	4.6	8.4	1.59	4.9	8.2	1.24	4.4	8.6	1.85
D	Y	5	3.1	9.0	6.7	11.2	1.92	6.8	10.6	1.34	7.0	10.6	1.39	7.2	10.7	1.19	6.7	10.8	1.85
E	Y	5	3.2	9.8	7.4	12.1	2.07	7.3	11.6	1.85	7.4	11.5	1.87	7.6	11.4	1.51	7.3	11.8	2.20

FIGURE 8. Estimator Distributions for Case DDb

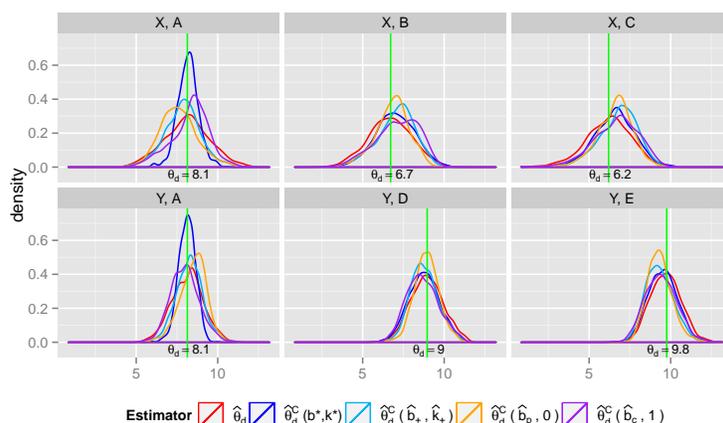


TABLE 8. Empirical MSE of Estimators for Case DDb

	n_d	σ_d	θ_d	$\hat{\theta}_d$			$\hat{\theta}_d^C(b^*, k^*)$			$\hat{\theta}_d^C(\hat{b}_\dagger, \hat{k}_\dagger)$			$\hat{\theta}_d^C(\hat{b}_p, 0)$			$\hat{\theta}_d^C(\hat{b}_c, 1)$			
				P_5	P_{95}	MSE	P_5	P_{95}	MSE	P_5	P_{95}	MSE	P_5	P_{95}	MSE	P_5	P_{95}	MSE	
A	X	5	3.1	8.1	5.7	10.4	1.94	7.2	9.2	0.41	5.9	9.7	1.34	5.9	9.7	1.59	6.1	10.0	1.41
A	Y	10	3.1	8.1	6.5	9.8	0.93	7.3	8.9	0.28	6.9	9.7	0.72	7.1	9.7	0.78	6.6	9.5	0.80
B	X	5	3.2	6.7	4.4	8.9	1.77	4.9	8.9	1.54	4.9	8.8	1.51	4.8	8.5	1.14	4.7	9.0	1.94
C	X	5	3.0	6.2	3.8	8.4	1.98	4.5	8.6	1.69	4.6	8.5	1.83	4.6	8.2	1.39	4.2	8.7	2.13
D	Y	10	3.1	9.0	7.4	10.6	0.98	7.2	10.4	0.92	7.5	10.3	0.75	7.7	10.2	0.58	7.2	10.4	1.00
E	Y	10	3.2	9.8	8.2	11.2	0.88	8.1	11.0	0.86	8.0	10.8	0.93	8.1	10.6	0.76	7.9	11.0	1.02

FIGURE 9. Estimator Distributions for Case DDc

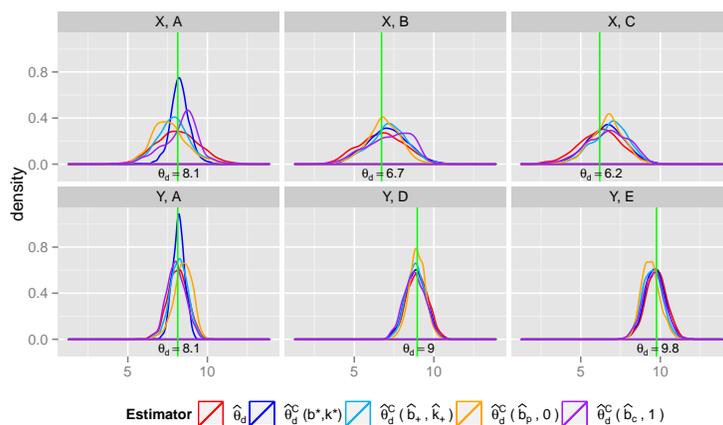


TABLE 9. Empirical MSE of Estimators for Case DDc

	n_d	σ_d	θ_d	$\hat{\theta}_d$			$\hat{\theta}_d^C(b^*, k^*)$			$\hat{\theta}_d^C(\hat{b}_\dagger, \hat{k}_\dagger)$			$\hat{\theta}_d^C(\hat{b}_p, 0)$			$\hat{\theta}_d^C(\hat{b}_c, 1)$			
				P_5	P_{95}	MSE	P_5	P_{95}	MSE	P_5	P_{95}	MSE	P_5	P_{95}	MSE	P_5	P_{95}	MSE	
A	X	5	3.1	8.1	5.9	10.3	1.88	7.3	9.3	0.38	6.1	9.7	1.31	5.9	9.7	1.57	6.2	9.9	1.36
A	Y	20	3.1	8.1	7.1	9.2	0.40	7.6	8.8	0.15	7.3	9.1	0.33	7.5	9.3	0.46	7.1	9.0	0.37
B	X	5	3.2	6.7	4.5	9.1	2.03	5.0	9.1	1.70	4.9	9.1	1.71	5.0	8.5	1.19	4.7	9.1	2.13
C	X	5	3.0	6.2	3.9	8.4	1.80	4.6	8.5	1.58	4.8	8.6	1.75	4.9	8.2	1.30	4.4	8.7	1.98
D	Y	20	3.1	9.0	7.9	10.0	0.43	7.8	9.9	0.42	7.9	9.9	0.36	8.1	9.8	0.26	7.7	9.9	0.47
E	Y	20	3.2	9.8	8.7	10.8	0.42	8.6	10.7	0.41	8.5	10.5	0.45	8.5	10.4	0.46	8.5	10.7	0.47

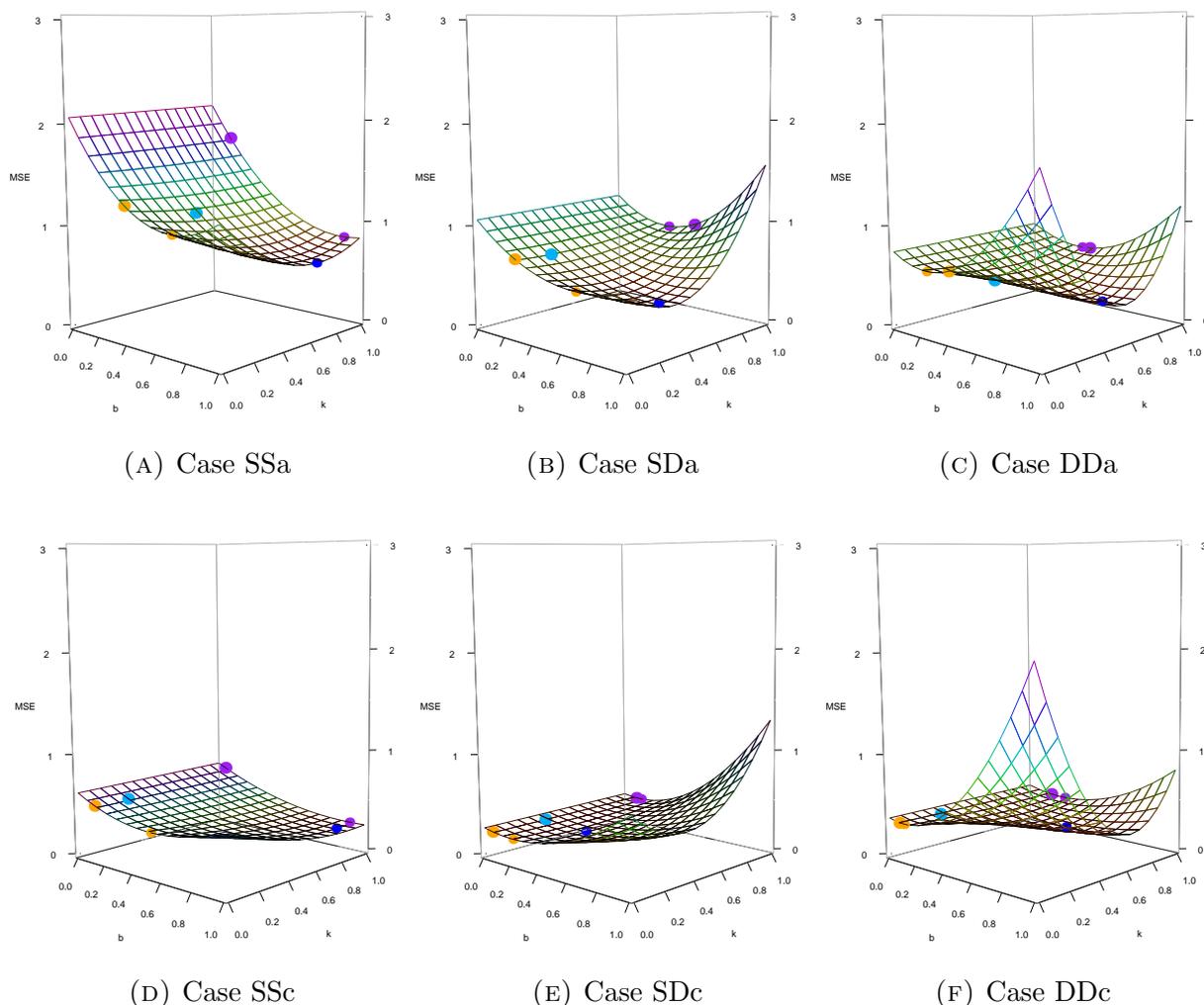


FIGURE 10. MSE Surface Plots for Panel X,A Sample Instances

The surface plots above show the resulting MSEs for different combinations of the weights b and k for sample instances of Panel X,A in six of the nine simulation cases. The large and small points are the MSEs corresponding to the suboptimal and optimal estimators, respectively. The dark and sky blue points are the MSEs for the optimal and suboptimal two-tiered estimators. Notice that the boundary curves at $k = 0$ and $k = 1$ represent the MSE curves for the baseline methods that leverage $\hat{\theta}_d$ with $\hat{\theta}_p$ and $\hat{\theta}$, respectively.

2.6. MSE Surface Plot of Sample Instances. Figure 10 shows the mean-squared error surface plots of a sample instance of Panel X,A for each of the selected simulation cases (Cases SS, SD, and DD under sampling designs a and c). The top panel (Figures 10a–10c) show sample instances for Cases SSa, SDa, and DDa, while the bottom panel (Figures 10d–10f) display sample instances for Cases SSa, SDc, and DDc.

For each panel in Figure 10, the x - and y -axes are the weights b and k in the two-tiered composite estimation model. While b represents the weight assigned to $\hat{\theta}_p^C$ (the composite of $\hat{\theta}_p$ and $\hat{\theta}$), k allocates additional weight toward $\hat{\theta}$ and thus away from $\hat{\theta}_p$.

The color scheme of the points on the MSE surface plots is consistent with that of the estimator distributions of the simulation cases in Figures 1–9. The large and small orange points on the surface plot boundary at $k = 0$ represent the actual mean-squared error $\text{MSE} \left\{ \hat{\theta}_d^C(b, 0); \theta_d \right\}$ evaluated at b_p^* and \hat{b}_p , which minimize $\text{MSE} \left\{ \hat{\theta}_d^C(b, 0); \theta_d \right\}$ and $\widehat{\text{eMSE}} \left\{ \hat{\theta}_d^C(b, 0); \theta_d \right\}$, respectively. Similarly, the large and small purple points on the boundary at $k = 1$ represent the actual mean-squared error $\text{MSE} \left\{ \hat{\theta}_d^C(b, 1); \theta_d \right\}$ evaluated at b_c^* and \hat{b}_c , which minimize $\text{MSE} \left\{ \hat{\theta}_d^C(b, 1); \theta_d \right\}$ and $\widehat{\text{eMSE}} \left\{ \hat{\theta}_d^C(b, 1); \theta_d \right\}$. More simply put, these boundary points indicate the actual mean-squared error evaluated at the optimal and estimated suboptimal weights for the two baseline composite estimators. The actual mean-squared error is not calculable in practice, but for the purposes of the simulation study, the underlying district, place, and county parameters are available to make these calculations.

The effect of sampling design on the optimal weights for the estimator model is most visible in Case SSa (Figures 10a). In this simulation case, each district in both places has a 2.5% srswor design. This leads to large optimal weights b_p^* and b_c^* , weighing $\hat{\theta}_p$ or $\hat{\theta}$, respectively, very heavily in calculating the baseline composite estimators. The two-tiered estimator yields even further reductions in MSE by allocating less weight to $\hat{\theta}_d$ (manifested in the large value of b^*) than in the baseline estimators in order to incorporate survey data from both the place and county as shown in Figure 10a. Similarly, the estimated suboptimal two-tiered estimator $\hat{\theta}_d^C(\hat{b}_\dagger, \hat{k}_\dagger)$ yields a considerably lower MSE than its corresponding estimated suboptimal baseline estimators $\hat{\theta}_d^C(\hat{b}_p, 0)$ and $\hat{\theta}_d^C(\hat{b}_c, 1)$.

On the other hand, with the sampling rate increased to 10% for Place Y districts in Case SSc (Figure 10d), the further reductions in MSE using the two-tiered estimator becomes marginal. As evidenced in Figure 10d, the MSE evaluated at the optimal weights for the two-tiered estimator is only marginally lower than that of the baseline composite estimator $\hat{\theta}_d^C(b_p^*, 1)$ leveraging $\hat{\theta}_d$ and $\hat{\theta}$. Since Panel X,A utilizes Place Y survey data to calculate the county estimator, the baseline composite estimator leveraging sampling strength between $\hat{\theta}_d$ and $\hat{\theta}$ could easily reduce MSE dramatically with across-district similarity and the high sampling rate of Place Y in Case SSc. This means that the baseline estimator $\hat{\theta}_d^C(b_p^*, 1)$ has fully incorporated the county survey data with its large sampling weight, rendering the further leveraging of place survey data extraneous. Introducing Place X survey data in the two-tiered estimator model does not yield enough new information for a large reduction in MSE beyond that achieved using the baseline estimator leveraging $\hat{\theta}_d$ and $\hat{\theta}$.

On the other hand, with district dissimilarity within the county in Case SD, the baseline estimator $\hat{\theta}_d^C(b_c^*, 1)$ becomes an unattractive composite estimator. While the other baseline estimator $\hat{\theta}_d^C(b_p^*, 0)$ would be more appropriate given the homogeneity in Place X, the optimal two-tiered estimator could potentially yield further reductions in MSE by leveraging the

additional survey data for District A within the county as shown in Figure 10b. Despite this additional reduction in MSE for the optimal two-tiered estimator, the additional error incurred by the moment-matching estimators used to calculate the suboptimal two-tier estimator $\hat{\theta}_d^C(\hat{b}_\dagger, \hat{k}_\dagger)$ reduces the efficiency obtained with the optimal two-tiered estimator considerably. In Case SDc (Figure 10e), the optimal two-tiered estimator yields only marginal improvement, since the large county sampling rate introduces excessive data for districts dissimilar to District A. While District A is sampled at a larger rate within the county, the inability to disaggregate this subset from the dissimilar districts reduces the benefit of a larger sampling rate.

The panels for Cases DDa and DDc (Figures 10c and 10f) show that systematic dissimilarity could potentially yield additional MSE reduction using the two-tiered estimator. In Case DDa, the large MSE at the two corners intersecting $b = 1$ at $k = 0$ and $k = 1$ represent the large error associated with weighing the place and county estimators too heavily given district dissimilarity within both domains. While the MSE at the corner of $b = 1$ and $k = 1$ is lessened in Case DDc with the large county sampling rate, the MSE remains elevated in the other corner of $b = 1$ and $k = 0$, since the place sampling design remains at 2.5% srswor. Interestingly, in Case DDa, the optimal weight k^* is about 0.57, meaning that $\hat{\theta}_p$ and $\hat{\theta}$ are weighted similarly for the place-county estimator $\hat{\theta}_p^C$. In Case DDc, the optimal weight k^* in the selected sample instance increases to 0.6, since the higher county sampling rate increases the sampling strength of $\hat{\theta}$. However, the two-tiered estimator does not achieve a considerable reduction in MSE in Case DDc, since the increased county sampling rate makes the baseline estimator $\hat{\theta}_d^C(b_c^*, 1)$ a more suitable alternative.

2.7. Effect of Geographic Alignment. For all simulation cases presented in the previous section, District A was the only district overlapping Place X and Place Y. We will now examine the effect of geographic alignment on the performance of the two-tiered composite estimator model by increasing the overlap between Places X and Y. As with the previous set of simulations, Place X comprises of Districts A, B, and C, while Districts D and E are within Place Y. Unlike in the first set of simulations, we will focus on Panel X,A. From the perspective of Panel X,A, Place X is the place, while Place Y is the county. We will retain this geographic classification for this set of simulations.

The following are three different cases of geographic overlap that we will examine:

- Case A: $\mathcal{D}_p \cap \mathcal{D} = \{A\}$
- Case AB: $\mathcal{D}_p \cap \mathcal{D} = \{A, B\}$
- Case ABC: $\mathcal{D}_p \subset \mathcal{D}$, where $\mathcal{D}_p = \{A, B, C\}$

Each geographic alignment case is then paired with sampling designs a and c used in the main simulation study. For simplicity, we will focus on the conditions of Case SS in simulating the geographic alignment cases.

Figure 11 shows the empirical distributions of $\hat{\theta}_d$, $\hat{\theta}_d^C(\hat{b}_\dagger, \hat{k}_\dagger)$, $\hat{\theta}_d^C(\hat{b}_p, 0)$ and $\hat{\theta}_d^C(\hat{b}_c, 1)$ for Panel X,A in Cases A, AB, and ABC under both sampling designs a and c. Case A under sampling

design a shows the most visible reduction in MSE using the two-tiered estimator. The empirical distribution of the two-tiered estimator slowly converges to that of the baseline estimator $\hat{\theta}_d^C(\hat{b}_c, 1)$ with increasing geographic overlap as seen in the top three panels in Figure 11. The increasing geographic overlap between Places X and Y reduces the utility of leveraging Place X survey data as the place estimator, since Place Y encapsulates districts that also belong to Place X.

The bottom panel in Figure 11 shows that the increasing county sampling rate compounds the issue of increasing geographic alignment, causing a further reduction in the utility of the two-tiered estimator. In fact, the empirical MSE for the two-tiered estimator $\hat{\theta}_d^C(\hat{b}_\dagger, \hat{k}_\dagger)$ is slightly higher than that of the baseline estimator $\hat{\theta}_d^C(\hat{b}_c, 1)$. This reversal in the empirical MSEs could be attributed to the tradeoff between the additional leveraging of estimators in the two-tiered model and the incurred error from using three moment-matching estimators. In other words, the increasing county sampling rate makes the composite form of $\hat{\theta}_d$ and $\hat{\theta}$ such an efficient estimator that the cost of incurring additional error from three moment-matching estimators outweighs the benefits of leveraging place survey data already encapsulated in the county estimator.

This second simulation study demonstrates that geographic overlap and sampling designs can play a role in determining the tradeoffs between using the two-tiered estimator and the baseline composite estimators.

FIGURE 11. Estimator Distributions for Panel X,A in Case SS

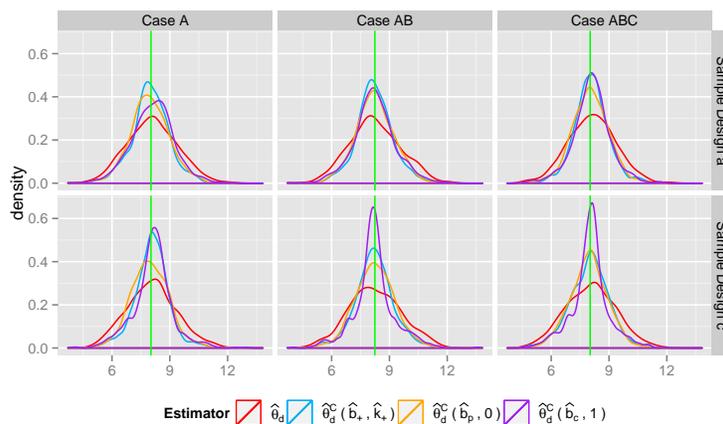


TABLE 10. Empirical MSE of Estimators for Panel X,A in Case SS

		$\hat{\theta}_d$			$\hat{\theta}_d^C(\hat{b}_\dagger, \hat{k}_\dagger)$			$\hat{\theta}_d^C(\hat{b}_p, 0)$			$\hat{\theta}_d^C(\hat{b}_c, 1)$		
		P_5	P_{95}	MSE	P_5	P_{95}	MSE	P_5	P_{95}	MSE	P_5	P_{95}	MSE
A	a	6.0	10.3	1.74	6.3	9.6	0.96	6.3	9.7	1.06	6.3	9.8	1.22
AB	a	6.2	10.5	1.79	6.7	9.9	0.91	6.6	10.0	1.02	6.5	10.0	1.16
ABC	a	6.1	10.1	1.55	6.6	9.5	0.79	6.5	9.6	0.89	6.5	9.7	0.93
A	c	5.9	10.3	1.73	6.5	9.4	0.83	6.4	9.6	0.99	6.3	9.7	1.05
AB	c	6.0	10.6	1.92	6.6	9.8	0.98	6.6	9.8	1.07	6.4	10.1	1.13
ABC	c	6.0	10.2	1.72	6.2	9.6	1.04	6.3	9.6	0.96	6.2	9.7	0.95

3. CONCLUSION

We extended Longford's baseline composite estimation method with a two-tiered composite estimation model consisting of place and county parameters. Given the added complexity to the model, we proposed three moment-matching estimators required for minimizing the estimated expected mean squared error of the two-tiered model with respect to the actual district parameter.

The simulation studies show that the two-tiered model yields the most additional efficiency under conditions of across-district similarity for both the place and county. Systematic dissimilarity with the overlapping district taking on an intermediate value also leads to increased efficiency with the two-tiered estimator, which counterbalances the incurred bias of the two baseline composite estimators. For nearly all the simulation cases, increasing the sampling rate reduces the utility of the two-tiered model, since one of the baseline composite estimators could potentially capture most of the reduction in MSE. Even in such cases, the empirical distributions of the two-tiered estimator often converges to those of the baseline composite estimator with the largest empirical efficiency. In addition to increased county sampling rates, increasing geographic overlap between the place and county diminishes the utility of the two-tiered county, since the county survey data incorporates more information about districts overlapping the place and county, rendering the additional leveraging of the place survey data less relevant.

The proposed moment-matching estimator does lead to the additional requirement of disaggregated data for the districts within the place boundaries. However, the nonlinear system used to solve the optimal weights in the model do not necessarily have this additional data requirement if the moment-matching estimators were replaced with auxiliary data. Nonetheless, this composite estimation extension expands the data flexibility in applications, since the place boundaries do not need to be completely encapsulated within those of the county. With this increased flexibility, the two-tiered composite estimation model could more readily combine various survey data with differing geographical specifications.

APPENDIX A. PROPOSED MOMENT-MATCHING ESTIMATORS

In the baseline composite shrinkage estimation, Longford employed the method of moment matching to estimate the between-area variance term [3]. We adopt this method in a similar fashion to estimate the terms σ_p^2 , $\sigma_{d,p}^2$, and $\rho_{d,p}$ in (1.8) and (1.9) with proposed moment-matching estimators.

A.1. **Estimating σ_p^2 .** In order to estimate $\sigma_p^2 = \mathbb{E}_{\mathcal{D}_p} [(\theta - \theta_p)^2]$, we will set a simple estimator to a particular moment that will recover the term σ_p^2 . First, let

$$S_p = (\hat{\theta} - \hat{\theta}_p)^2 = \left(\hat{\theta} - \sum_{d \in \mathcal{D}_p} w_d \hat{\theta}_d \right)^2 \quad (\text{A.1})$$

Second, take the expectation of S_p :

$$\mathbb{E}[S_p] = \mathbb{E}[\hat{\theta}^2] - 2 \sum_{d \in \mathcal{D}_p} w_d \mathbb{E}[\hat{\theta} \hat{\theta}_d] + \mathbb{E} \left[\sum_{d \in \mathcal{D}_p} w_d^2 \hat{\theta}_d^2 + \sum_{i \in \mathcal{D}_p} \sum_{j \neq i} w_i w_j \hat{\theta}_i \hat{\theta}_j \right] \quad (\text{A.2})$$

$$= v - 2 \sum_{d \in \mathcal{D}_p} w_d c_d + \sum_{d \in \mathcal{D}_p} w_d^2 v_d + \sum_{i \in \mathcal{D}_p} \sum_{j \neq i} w_i w_j \text{Cov}(\hat{\theta}_i, \hat{\theta}_j) + \theta^2 - 2 \sum_{d \in \mathcal{D}_p} w_d \theta \theta_d + \sum_{i \in \mathcal{D}_p} \sum_{j \neq i} w_i w_j \theta_i \theta_j + \sum_{d \in \mathcal{D}_p} w_d^2 \theta_d^2 \quad (\text{A.3})$$

$$= v - 2 \sum_{d \in \mathcal{D}_p} w_d c_d + \sum_{d \in \mathcal{D}_p} w_d^2 v_d + \left(\theta - \sum_{d \in \mathcal{D}_p} w_d \theta_d \right)^2 \quad (\text{A.4})$$

Note that the between-district covariance term in (A.3) disappears from (A.4) with the assumption that the estimates $\hat{\theta}_d$ are mutually independent for all $d \in \mathcal{D}_p$.

Third, take the expectation of $\mathbb{E}[S_p]$ over \mathcal{D}_p , the set of districts within the place boundaries.

$$\mathbb{E}_{\mathcal{D}_p} \left\{ \mathbb{E}[S_p] \right\} = v - 2 \sum_{d \in \mathcal{D}_p} w_d c_d + \sum_{d \in \mathcal{D}_p} w_d^2 v_d + \mathbb{E}_{\mathcal{D}_p} \left[\left(\theta - \sum_{d \in \mathcal{D}_p} w_d \theta_d \right)^2 \right] \quad (\text{A.5})$$

$$= v - 2 \sum_{d \in \mathcal{D}_p} w_d c_d + \sum_{d \in \mathcal{D}_p} w_d^2 v_d + \sigma_p^2 \quad (\text{A.6})$$

Lastly, rather than setting the simple estimator S equal to $E[S_p]$, we will match S with $E_{\mathcal{D}_p} \{E[S_p]\}$ in order to solve for $\hat{\sigma}_p^2$.

$$\begin{aligned} S_p &= E_{\mathcal{D}_p} \{E[S_p]\} \\ S_p &= v - 2 \sum_{d \in \mathcal{D}_p} w_d c_d + \sum_{d \in \mathcal{D}_p} w_d^2 v_d + \sigma_p^2 \\ \hat{\sigma}_p^2 &= S_p - \hat{v} + 2 \sum_{d \in \mathcal{D}_p} w_d \hat{c}_d - \sum_{d \in \mathcal{D}_p} w_d^2 \hat{v}_d \end{aligned} \quad (\text{A.7})$$

The result in (A.7) is equivalent to the estimator in (1.10).

A.2. Estimating $\sigma_{d,p}^2$. To estimate $\sigma_{d,p}^2 = E_{\mathcal{D}_p} [(\theta_p - \theta_d)^2]$, we use a weighted sum of squares $S_{d,p}$ as a simple estimator for moment matching, shown below:

$$S_{d,p} = \sum_{d \in \mathcal{D}_p} w_d (\hat{\theta}_p - \hat{\theta}_d)^2 \quad (\text{A.8})$$

Then, we take the expectation of $S_{d,p}$:

$$E[S_{d,p}] = \sum_{d \in \mathcal{D}_p} w_d \{v_p + v_d - 2c_{d,p} + (\theta_p - \theta_d)^2\} \quad (\text{A.9})$$

In preparation for moment matching, we take the expectation for $E[S_{d,p}]$ over \mathcal{D}_p :

$$E_{\mathcal{D}_p} \{E[S]\} = \sum_{d \in \mathcal{D}_p} w_d (v_d + v_p - 2c_{d,p}) + \sum_{d \in \mathcal{D}_p} w_d E_{\mathcal{D}_p} [(\theta_p - \theta_d)^2] \quad (\text{A.10})$$

$$= \sum_{d \in \mathcal{D}_p} w_d (v_d - 2c_{d,p}) + v_p + \sigma_{d,p}^2 \quad (\text{A.11})$$

By matching the simple estimator in (A.8) with (A.11),

$$\hat{\sigma}_{d,p}^2 = S_{d,p} - \sum_{d \in \mathcal{D}_p} w_d (\hat{v}_d - 2\hat{c}_{d,p}) - \hat{v}_p \quad (\text{A.12})$$

which is equivalent to the estimator in (1.11).

A.3. **Estimating** $\rho_{d,p}$. To estimate $\rho_{d,p} = E_{\mathcal{D}_p} [(\theta - \theta_p)(\theta_p - \theta_d)]$, we select the following simple estimator for moment matching:

$$R_{d,p} = \sum_{d \in \mathcal{D}_p} w_d (\hat{\theta} - \hat{\theta}_p) (\hat{\theta}_p - \hat{\theta}_d) \quad (\text{A.13})$$

We then take the expectation of the simple estimator $R_{d,p}$:

$$E [R_{d,p}] = \sum_{d \in \mathcal{D}_p} w_d \{c_p - c_d - v_p + c_{d,p} + (\theta - \theta_p)(\theta_p - \theta_d)\} \quad (\text{A.14})$$

In order to recover the term $\rho_{d,p}$, we must then take the expectation of $E [R_{d,p}]$ over \mathcal{D}_p :

$$E_{\mathcal{D}_p} \left\{ E [R_{d,p}] \right\} = \sum_{d \in \mathcal{D}_p} w_d (c_p - c_d + c_{d,p}) + \sum_{d \in \mathcal{D}_p} w_d E_{\mathcal{D}_p} [(\theta - \theta_p)(\theta_p - \theta_d)] - v_p \quad (\text{A.15})$$

$$= \sum_{d \in \mathcal{D}_p} w_d (c_p - c_d + c_{d,p}) + \rho_{d,p} - v_p \quad (\text{A.16})$$

By matching the simple estimator $R_{d,p}$ with the moment $E_{\mathcal{D}_p} \left\{ E [R_{d,p}] \right\}$ and solving for $\rho_{d,p}$, we obtain

$$\hat{\rho}_{d,p} = R_{d,p} - \sum_{d \in \mathcal{D}_p} w_d (\hat{c}_p - \hat{c}_d + \hat{c}_{d,p}) + \hat{v}_p \quad (\text{A.17})$$

which is equivalent to the estimator in (1.12).

REFERENCES

- [1] HASSELMAN, B. *nleqslv: Solve systems of non linear equations*, 2014. R package version 2.1.1.
- [2] HOGUE, M., AND LI, D. Composite Shrinkage Estimation: Review of Theory with Simulation Study and Empirical Applications. Economic and Statistical Working Paper Series. Bureau of Economic and Business Research, University of Utah, 2014.
- [3] LONGFORD, N. T. *Missing Data and Small-Area Estimation: Modern Analytical Equipment for the Survey Statistician*. Springer, New York, 2005.
- [4] LONGFORD, N. T. Small area estimation with spatial similarity. *Computational Statistics and Data Analysis* 54 (2010), 1151–1166.