Catalogue no. 11-522-X
ISSN 1709-8211

Proceedings of Statistics Canada Symposium 2022:
Data Disaggregation: building a more representative data portrait of society

Application of sampling variance smoothing methods for small area proportion estimation

by Yong You and Mike Hidiroglou

Release date: March 25, 2024

Proceedings of Statistics Canada Symposium 2022
Data Disaggregation: Building a more-representative data portrait of society

Application of sampling variance smoothing methods for small area proportion estimation

Yong You and Mike Hidiroglou[1]

Abstract

Sampling variance smoothing is an important topic in small area estimation. In this paper, we propose sampling variance smoothing methods for small area proportion estimation. In particular, we consider the generalized variance function and design effect methods for sampling variance smoothing. We evaluate and compare the smoothed sampling variances and small area estimates based on the smoothed variance estimates through analysis of survey data from Statistics Canada. The results from real data analysis indicate that the proposed sampling variance smoothing methods work very well for small area estimation.

Key Words: Coefficient of variation, design effect, generalized variance function, log-linear model, relative error.

1. Introduction

Small area estimation has become very popular and important in both public and private agencies due to the growing demand for reliable estimates. Small area estimation is based on models that lead to reliable estimates for small areas of interest. In this paper, we focus on area level models that are based on direct survey estimates aggregated from the unit level data and area level auxiliary variables. Various area level models have been proposed in the literature to improve the precision of the direct survey estimates: a good summary of these methods is discussed in Rao and Molina (2015). The Fay-Herriot model (Fay and Herriot, 1979) is a basic area level model that is widely used in practice. The Fay-Herriot model has two components, namely, a sampling model for the direct survey estimates and a linking model for the small area parameters of interest. The sampling model assumes that there exists a direct survey estimator y_i, which is usually design unbiased, for the small area parameter θ_i such that

$$y_i = \theta_i + e_i, \ i = 1,\ldots,m, \quad (1)$$

where e_i is the sampling error associated with the direct estimator y_i and m is the number of small areas. It is customary in practice to assume that the e_i's are independently normal random variables with mean $E(e_i) = 0$ and sampling variance $Var(e_i) = \sigma_i^2$. The linking model assumes that the small area parameter of interest θ_i is related to area level auxiliary variables $x_i = (x_{i1}, \ldots, x_{ip})'$ through a linear regression model

$$\theta_i = x_i'\beta + v_i, \ i = 1,\ldots,m, \quad (2)$$

where $\beta = (\beta_1, \ldots, \beta_p)'$ is a $p \times 1$ vector of regression coefficients, and v_i's are area-specific random effects assumed to be independent and identically distributed with $E(v_i) = 0$ and $Var(v_i) = \sigma_v^2$. The assumption of normality for v_i is generally also included. The model variance σ_v^2 is unknown and needs to be estimated from the data. For the Fay-Herriot model, the sampling variance σ_i^2 is assumed to be known in model (1). As this is a very strong assumption, a smoothing or modeling approach is usually used to estimate σ_i^2. The sampling variance can be smoothed or can be modeled directly as in Wang and Fuller (2003), You and Chapman (2006), Sugasawa, Tamae and Kubokawa (2017), etc. You (2021) shows that the smoothing approach can provide more efficient and accurate model-based estimates than the modeling approach for small areas under hierarchical Bayes framework. Lesage, Beaumont and Bocci (2021) also have some discussions on the sampling variance smoothing for the Fay-Herriot model.

[1]Yong You, ICMIC, Statistics Canada, 100 Tunney's pasture driveway, Ottawa ON, K1A0T6. Contact email: yong.you@statcan.gc.ca. **Disclaimer**: The content of this paper represents the authors' opinions and not necessarily those of Statistics Canada. It describes theoretical methods that may not reflect those implemented by the Agency.

The objective of this paper is to compare different methods to smooth the direct estimates of the sampling variances for proportions in small area estimation using the Fay-Herriot model. We proceed to do so as follows. Let $\hat{p}_{iw}$ be the direct design-based estimator for the proportion p_i for a given characteristic in the i-th area. Applying the Fay-Herriot model to $\hat{p}_{iw}$, we have

$$\hat{p}_{iw} = p_i + e_i, \tag{3}$$

with the sampling variance $Var(e_i) = \sigma_i^2$ unknown. Now let $\hat{V}_i$be the direct sampling variance estimate for σ_i^2 obtained from the survey data. Usually some of the $\hat{V}_i$'s are very unstable due to small sample sizes. We, therefore, need to obtain a smoothed estimate, $\tilde{V}_i$, for σ_i^2, and then treat the smoothed variance estimate $\tilde{V}_i$ in the sampling model (3) as known. In this paper, we compare two smoothing methods. One method is based on the generalized variance function (GVF) and the other one is based on design effects (DEFF). We then propose an average smoothed (ASM) variance estimator based on GVF and DEFF smoothed estimators. The main purpose of the paper is to promote the proposed GVF and DEFF methods. The ASM is used as an additional choice as it pools the GVF and DEFF estimates by taking their average.

There are many applications of GVF in small area estimation, see, for example, the early work of Dick (1995) and the recent application in Hidiroglou, Beaumont, and Yung (2019). DEFF can also be used in variance modeling and smoothing for small area estimation. For example, You (2008) used the smoothed design effects over time to obtain the smoothed variance and covariance matrices. Liu, Lahiri, and Kalton (2014) also applied area level models to proportions using design effects for the sampling variance smoothing and modeling. In this paper, we provide a general method to compute the design effect and propose a smoothed variance estimator based on the average design effects over areas. We will also show that the DEFF-smoothed variance estimator and the GVF-smoothed variance estimator are roughly equivalent under certain conditions. We will illustrate the sampling variance smoothing methods via application using the Canadian Labor Force Survey (LFS) survey data.

The paper is organized as follows. In section 2, we propose sampling variance smoothing methods including GVF and DEFF methods. In section 3, we compare the model-based estimates based on different smoothed sampling variance estimates using the LFS unemployment rate data. In section 4, we offer some concluding remarks.

2. Sampling Variance Smoothing Methods

2.1. Smoothing using log-linear models

In this section, we will construct a GVF model to obtain smoothed sampling variances. This procedure is widely used in practice to model the variance. We apply a log-linear regression model on the direct sampling variance $\hat{V}_i$ using the sample size n_i as the auxiliary variable in the model as follows:

$$log(\hat{V}_i) = \beta_0 + \beta_1\, log(n_i) + \varepsilon_i,\ i = 1, \ldots, m, \tag{4}$$

where the model error term is $\varepsilon_i \sim N\,(0, \tau^2)$, and the model error variance τ^2 is unknown. Note that the proposed regression model (4) is equivalent to the following model:

$$log(\hat{V}_i) = \beta_0 + \beta_1\, log(\frac{1}{n_i}) + \varepsilon_i,\ i = 1, \ldots, m, \tag{5}$$

where $log(1/n_i)$ is used as the auxiliary variable. The proposed GVF models (4) or (5) are the same models used in You (2021) for the hierarchical Bayes (HB) modeling of sampling variance. This GVF model also extends the model proposed by Souza, Moura, and Migon (2009) for sampling variances by using $log(1/n_i)$ and adding a normal random effect (ε_i) to the regression part in the model.

Let $\hat{\beta}_0$ and $\hat{\beta}_1$ denote the ordinary least square estimators of the regression coefficients β_0 and β_1. A naïve GVF-smoothed estimator of the sampling variance is obtained by taking the exponential of the fitted value:

$$\tilde{V}_i^{naive} = exp(\hat{\beta}_0 + \hat{\beta}_1\, log(n_i)). \tag{6}$$

Dick (1995) used the naïve smoothed estimator $\tilde{V}_i^{naive}$ in the application of census undercoverage small area estimation. As noted by Rivest and Belmonte (2000), the naïve smoothed estimator $\tilde{V}_i^{naive}$ underestimates the sampling variance. This can be seen as follows. If Y is a log-normal random variable with mean μ and variance σ^2, the mean of Y is $E(Y) = exp(\mu)\, exp(\tau^2/2)$. It follows that the smoothed estimator $\tilde{V}_i^{naive}$ underestimates the true values by ignoring the second term $exp(\tau^2/2)$ in the mean of the log-normal random variable. Denote as $\hat{\omega}_{RB} =$

$exp(\hat{\tau}^2/2)$ the Rivest and Belmonte (2000) correction, where $\hat{\tau}^2$ is the estimated residual variance of the proposed log-linear regression model (4). Then a GVF-smoothed estimator, denoted as $\tilde{V}_i^{GVF.RB}$, is given by

$$\tilde{V}_i^{GVF.RB} = \tilde{V}_i^{naive} \cdot \widehat{\omega}_{RB} = \tilde{V}_i^{naive} \cdot \ exp(\hat{\tau}^2/2). \tag{7}$$

The naïve GVF estimator $\tilde{V}_i^{naive}$ in (6) underestimates the sampling variance by $exp(\hat{\tau}^2/2)$. This term is always greater than 1, and sometimes it could be large, depending on the value of $\hat{\tau}^2$.

Hidiroglou, Beaumont, and Yung (2019) proposed another correction term for the naïve estimator $\tilde{V}_i^{naive}$. Let $\tilde{V}^{naive}$ be the sum of the naïve smoothed variance estimators, that is, $\tilde{V}^{naive} = \sum_{i=1}^m \tilde{V}_i^{naive}$, and $\hat{V}^{total}$ be the sum of the direct sampling variances, that is, $\hat{V}^{total} = \sum_{i=1}^m \hat{V}_i$. Following Hidiroglou, Beaumont, and Yung (2019), we define a correction term, named as Hidiroglou, Beaumont, and Yung (HBY) correction term, as $\widehat{\omega}_{HBY} = \hat{V}^{total} / \tilde{V}^{naive}$. This leads to a second GVF-smoothed variance estimator , denoted as $\tilde{V}_i^{GVF.HBY}$. It is given by

$$\tilde{V}_i^{GVF.HBY} = \tilde{V}_i^{naive} \cdot \widehat{\omega}_{HBY} = \tilde{V}_i^{naive} \cdot \frac{\hat{V}^{total}}{\tilde{V}^{naive}}. \tag{8}$$

Note that $\widehat{\omega}_{HBY}$ is obtained as an alternative estimator of $exp(\tau^2/2)$ using the method of moments (Beaumont and Bocci, 2016). This avoids the sensitivity of the GVF model to deviations from the normality assumption of ε_i in model (4). A nice property of $\tilde{V}_i^{GVF.HBY}$ is that the sum of the smooth variance estimates is equal to the sum of the direct sampling variance estimates, that is, $\sum_{i=1}^m \tilde{V}_i^{GVF.HBY} = \sum_{i=1}^m \hat{V}_i$. This property may ensure that the smoothing procedure does not systematically overestimate or underestimate the sampling variances.

2.2. Smoothing using design effects

Let $\hat{p}_{iw}$ be the direct design-based estimate for a proportion p_i and $\hat{V}_i$ the corresponding direct sampling variance under complex design for the i-th small area. Then the estimated design effect can be approximately computed as

$$deff_i = \frac{\hat{V}_i}{\hat{p}_{iw}(1-\hat{p}_{iw})/n_i + \hat{V}_i/n_i}, \tag{9}$$

where n_i is the sample size of the i-th small area; see Gambino (2009). However, by noting that the $deff_i$ in equation (9) is not equal to 1 under simple random sampling design, we modify the $deff_i$ by multiplying it by a correction term $(n_i + 1)/n_i$:

$$deff_i = \frac{\hat{V}_i}{\hat{p}_{iw}(1-\hat{p}_{iw})/n_i + \hat{V}_i/n_i} \cdot \frac{n_i+1}{n_i}. \tag{10}$$

Using equation (10), we can re-write the design-based sampling variance $\hat{V}_i$ as

$$\hat{V}_i = deff_i \cdot \frac{\hat{p}_{iw}(1-\hat{p}_{iw})}{n_i} \cdot \left(1 + \frac{1-deff_i}{n_i}\right)^{-1}. \tag{11}$$

If the sample size n_i is large, the term $(1 + (1 - deff_i)/n_i)^{-1}$ may be negligible in (11), equation (11) reduces to

$$\hat{V}_i = deff_i \cdot \frac{\hat{p}_{iw}(1-\hat{p}_{iw})}{n_i}. \tag{12}$$

Equation (12) is used, for example, in Liu, Lahiri, and Kalton (2014) for sampling variance smoothing and modeling. However, in small area estimation, n_i can be very small, and the term $(1 + (1 - deff_i)/n_i)^{-1}$ may not be negligible.

We can compute all the design effects $deff_i$'s using (10) for all areas, and then compute the average value over all areas, thereby obtaining a smoothed design effect $\overline{deff} = \frac{1}{m}\sum_{i=1}^m deff_i$. The average proportion estimate over all areas is given by $\bar{p}_w = \frac{1}{m}\sum_{i=1}^m \hat{p}_{iw}$. Replacing the $deff_i$ by $\overline{deff}$ and $\hat{p}_{iw}$ by $\bar{p}_w$ in equation (11), a DEFF-smoothed estimator of the sampling variance for proportion estimate $\hat{p}_{iw}$ is:

$$\tilde{V}_i^{DEFF} = \overline{deff} \cdot \frac{\bar{p}_w(1-\bar{p}_w)}{n_i} \cdot \left(1 + \frac{1-\overline{deff}}{n_i}\right)^{-1}. \tag{13}$$

If the sample size n_i is large, then the term $(1 + (1 - \overline{deff})/n_i)^{-1}$ in $\tilde{V}_i^{DEFF}$ can be negligible. The smoothed variance $\tilde{V}_i^{deff}$ can then be simplified to

$$\tilde{V}_i^{DEFF} = \overline{deff} \cdot \frac{\bar{p}_w(1-\bar{p}_w)}{n_i}. \tag{14}$$

2.3. Comparison of GVF and DEFF smoothing

We now show that the GVF-estimators and the DEFF-estimator $\tilde{V}_i^{DEFF}$ can perform similarly under certain conditions. Using $\tilde{V}_i^{GVF.RB}$ as an illustration, we can express this term as:

$$\tilde{V}_i^{GVF.RB} = exp(\hat{\beta}_0 + \hat{\beta}_1 \cdot log(n_i)) \cdot exp(\frac{\hat{\tau}^2}{2}) = C_0 \cdot exp(log(n_i)^{\hat{\beta}_1}) = C_0 \cdot n_i^{\hat{\beta}_1}$$

where $C_0 = exp(\hat{\beta}_0 + \frac{\hat{\tau}^2}{2})$ is a constant. If the value of the regression coefficient $\hat{\beta}_1$ is close to -1, then the GVF-estimator $\tilde{V}_i^{GVF.RB}$ can be approximately written as $\tilde{V}_i^{GVF.RB} \approx C_0/n_i$.

The DEFF-estimator $\tilde{V}_i^{DEFF}$ can be rewritten as follows:

$$\tilde{V}_i^{DEFF} = \overline{deff} \cdot \frac{\bar{p}_w(1-\bar{p}_w)}{n_i} \cdot \left(1 + \frac{1-\overline{deff}}{n_i}\right)^{-1} = \frac{C_1}{n_i} \cdot \left(\frac{n_i + 1 - \overline{deff}}{n_i}\right)^{-1} = \frac{C_1}{n_i + 1 - \overline{deff}} \approx \frac{C_1}{n_i},$$

where $C_1 = \overline{deff} \cdot \bar{p}_w(1-\bar{p}_w)$ is a constant. Both the GVF-estimator $\tilde{V}_i^{GVF.RB}$ and the DEFF-estimator $\tilde{V}_i^{DEFF}$ are proportional to n_i^{-1} if the regression coefficient $\hat{\beta}_1$ is close to -1 in the GVF regression model. Thus under such condition, both GVF and DEFF smoothed variances should perform similarly.

In practical applications, to make use of both the GVF and DEFF smoothed estimates, we can define an average smoothed (ASM) estimator $\tilde{V}_i^{ASM}=(\tilde{V}_i^{GVF.RB}+\tilde{V}_i^{GVF.HBY}+\tilde{V}_i^{DEFF})/3$ as a simple data pooling method to obtain the final smoothed variance estimate. As we will see in the LFS small area application in Section 3, the average smoothed estimator $\tilde{V}_i^{ASM}$ can perform very well and lead to large bias and CV reduction for small area estimates.

3. LFS Small Area Estimation Using Smoothed Sampling Variances

In this section, we apply the variance smoothing methods to the Canadian Labour Force Survey (LFS) data and compare the small area estimates based on the smoothed sampling variances. The LFS produces monthly estimates of the unemployment rate at national and provincial levels. The LFS also releases unemployment estimates for sub-provincial areas such as Census Metropolitan Areas (CMAs) and Census Agglomerations (CAs) across Canada. However, the direct estimates are not reliable for sub-provincial areas because the sample sizes in some areas are quite small. The various small area estimation models for LFS are discussed, for example, in Hidiroglou, Beaumont, and Yung (2019), Lesage, Beaumont and Bocci (2021), You, Rao, and Gambino (2003), and You (2008, 2021). We apply the Fay-Herriot model given by (1) and (2) to the May 2016 unemployment rate estimates at the CMA/CA level. We consider using four smoothed variance estimators in the LFS application, namely, $\tilde{V}_i^{GVF.RB}$, $\tilde{V}_i^{GVF.HBY}$, $\tilde{V}_i^{DEFF}$ and the average smoothed estimator $\tilde{V}_i^{ASM}=(\tilde{V}_i^{GVF.RB}+\tilde{V}_i^{GVF.HBY}+\tilde{V}_i^{DEFF})/3$. We consider the EBLUP approach in the application. The details of the EBLUP estimator and related MSE estimation based on the Fay-Herriot model can be found, for example, in Rao and Molina (2015) and You (2021). Local area employment insurance monthly beneficiary rate is used as an auxiliary variable in the Fay-Herriot model as in Hidiroglou, Beaumont, and Yung (2019) and You (2008, 2021). The model-based estimates and the direct estimates are compared with the census estimates to evaluate the effects of sampling variance smoothing.

We first obtain the smoothed sampling variances for all the 128 CMA/CAs using the proposed $\tilde{V}_i^{GVF.RB}$, $\tilde{V}_i^{GVF.HBY}$, $\tilde{V}_i^{DEFF}$ and $\tilde{V}_i^{ASM}$. For the GVF model (4), the regression estimates are $\hat{\beta}_0 = -3.194$ and $\hat{\beta}_1 = -0.901$. The RB residual correction term $exp(\hat{\tau}^2/2)$ is equal to 1.467 and the HBY correction term is $\hat{\omega}_{HBY} = \hat{V}^{total}/\tilde{V}^{naive}$ =1.786. As the regression coefficient $\hat{\beta}_1 = -0.901$ is close to -1, and the difference between two correction terms is not large, we should expect a similar smoothed sampling variances for the LFS data. We applied the Fay-Herriot model to the 128 CMA/CA LFS unemployment rate data with the four different smoothed sampling variances and obtained the corresponding EBLUP estimates. The details of the EBLUP estimator with REML method to estimate the variance component can be found, for example, in You (2021) and Rao and Molina (2015). The small area EBLUP estimates are compared via the absolute relative error (ARE) of the direct and EBLUP estimates with respect to the census estimates for each CMA/CA as follows: $\text{ARE}_i = \left|(\theta_i^{Census} - \theta_i^{Est})/\theta_i^{Census}\right|$, where θ_i^{Est} is the direct or the EBLUP estimate and θ_i^{Census} is the corresponding census value of the LFS unemployment rate. It is a common practice to evaluate the model-based estimates with the census values, for example, as in Hidiroglou, Beaumont and Yung (2019) and You (2021). We then take the average of AREs over CMA/CAs by different subgroups with respect to sample size, same as in Hidiroglou, Beaumont and Yung (2019). Table 3.1 presents the average ARE for the direct LFS and EBLUP estimators based on different input sampling variance estimates. For comparison, we also used the direct sampling variance as input sampling variance in the Fay-Herriot model. For example, EBLUP(DIR) represents that the direct (DIR) sampling variance estimate is used in the Fay-Herriot model, EBLUP(GVF.RB) represents the smoothed sampling variance estimate $\tilde{V}_i^{GVF.RB}$ (GVF.RB) is used, etc.

Table 3.1: Comparison of ARE for EBLUP estimates based on the different input sampling variances

CMA/CAs	Direct LFS	EBLUP (DIR)	EBLUP (GVF.RB)	EBLUP (GVF.HBY)	EBLUP (DEFF)	EBLUP (ASM)
25 smallest areas	0.489	0.279	0.181	0.184	0.180	0.182
Next 25 smallest areas	0.338	0.214	0.146	0.147	0.146	0.146
Next 25 smallest areas	0.276	0.198	0.138	0.143	0.134	0.138
Next 25 smallest areas	0.198	0.161	0.134	0.141	0.130	0.135
28 largest areas	0.132	0.125	0.099	0.108	0.091	0.099
Overall areas	0.283	0.194	0.138	0.144	0.135	0.139

It is clear from Table 3.1 that the EBLUP estimates substantially improve the direct estimates by reducing ARE. Even with the use of the direct sampling variance estimates, EBLUP(DIR) results in much smaller ARE than the direct survey estimator. However, by using the smoothed sampling variance estimates, EBLUP performs substantially much better than the direct estimator. The AREs are reduced by each area group and overall areas. In general, all the EBLUPs with the four smoothed sampling variances perform very similarly. Among the EBLUP estimators using smoothed sampling variances, EBLUP(GVF.HBY) has slightly larger ARE than others, and the EBLUP(DEFF) has slightly smaller ARE. For example, over all the 128 CMA/CAs, the respective ARE's of EBLUP(GVF.RB), EBLUP(GVF.HBY) and EBLUP(DEFF) are 0.138, 0.144, and 0.135. So EBLUP(DEFF) performs the best in terms of relative error. For the average smoothed sampling variance $\tilde{V}_i^{ASM}$, the EBLUP(ASM) has overall ARE value 0.139, which is between the ARE values of EBLUPs using GVF and DEFF. The EBLUP(ASM) performs very well.

For average CV, EBLUP also reduces the CV substantially over the direct estimator. The direct LFS estimator has average CV 39.4%, EBLUP(DIR) has average CV 24.5%, whereas EBLUP(GVF.RB) has average CV 10.3%, EBLUP(GVF.HBY) has a slightly smaller average CV 8.2%, and EBLUP(DEFF) has average CV value 11.8%. The EBLUP(ASM) has average CV 10.2%. Thus, using smoothed sampling variances substantially reduces the CV for EBLUPs, and again the CV for EBLUP(ASM) is between the CV values of EBLUP using GVF and DEFF variances. EBLUP(ASM) has smaller ARE than EBLUP(GVF.RB) and EBLUP(GVF.HBY) and has smaller CV than EBLUP(GVF.RB) and EBLUP(DEFF). The use of averaged smoothed sampling variances $\tilde{V}_i^{ASM}$ leads to a balanced reduction for both ARE and CV. It is clear that the average smoothed estimator $\tilde{V}_i^{ASM}$ performs very well.

Lesage, Beaumont, and Bocci (2021) considered the following smoothing model, denoted as LBB model, for sampling variance smoothing:

$$log(\hat{V}_i) = \beta_0 + \beta_1\, log(z_i) + \beta_2\, log(1 - z_i) + \beta_3\, log(n_i) + \varepsilon_i,\ i = 1, \ldots, m, \tag{15}$$

where z_i is the employment insurance beneficiary rate used in the Fay-Herriot model as auxiliary variable to obtain the EBLUP estimators. By applying the LBB smoothing model (15) to the 128 area sampling variance data, we have the following regression estimates $\hat{\beta}_0 = -4.443$, $\hat{\beta}_1 = -0.486$, $\hat{\beta}_2 = -29.139$ and $\hat{\beta}_3 = -0.886$. The residual correction term $\widehat{\omega}_{RB} = exp(\hat{\tau}^2/2)$ is equal to 1.461 and the HBY correction term is $\widehat{\omega}_{HBY} = \hat{V}^{total}/\tilde{V}^{naive}$ =1.782. We denote as $\tilde{V}_i^{LBB.RB}$ the smoothed variance estimator based on the LBB model (15) using formula (7) with a correction term $\widehat{\omega}_{RB}$=1.461. Similarly, let $\tilde{V}_i^{LBB.HBY}$ be the smoothed variance estimator based on the LBB model (15) using formula (8) with a correction term $\widehat{\omega}_{HBY}$ =1.782. We now compare the EBLUP estimates based on the LBB smoothing model and the proposed smoothing method. In particular, we compare the proposed EBLUP(ASM) to EBLUP estimates using $\tilde{V}_i^{LBB.RB}$ and $\tilde{V}_i^{LBB.HBY}$, e.g., EBLUP(LBB.RB) and EBLUP(LBB.HBY).

Table 3.2: Comparison of ARE based on different GVF models and smoothed sampling variances

CMA/CAs	Direct LFS	EBLUP(ASM)	EBLUP(LBB.RB)	EBLUP(LBB.HBY)
25 smallest areas	0.489	0.182	0.181	0.183
Next 25 smallest areas	0.338	0.146	0.144	0.145
Next 25 smallest areas	0.276	0.138	0.137	0.142
Next 25 smallest areas	0.198	0.135	0.135	0.141
28 largest areas	0.132	0.099	0.099	0.108
Overall areas	0.283	0.139	0.138	0.143

Table 3.2 presents the average ARE to compare the effects of variance smoothing using ASM and LBB model. It is clear from Table 3.2 that all EBLUP estimates perform very well and improve the direct survey estimates by substantially reducing the ARE with respect to the census values. EBLUP(ASM) and EBLUP(LBB.RB) perform almost the same, and EBLUP(LBB.HBY) has slightly larger ARE, same as EBLUP(GVF.HBY) in Table 3.1. EBLUP(LBB.HBY) and EBLUP(GVF.HBY) perform almost identically by comparing the results in Table 3.1 and Table 3.2. In terms of CV, EBLUP(LBB.RB) and EBLUP(ASM) have the same average CV 10.2%, and EBLUP(LBB.HBY) has the same average CV 8.2% as EBLUP(GVF.HBY). The LFS small area application shows that the proposed GVF model (4) and the proposed sampling variance smoothing methods GVF, DEFF and ASM perform very well by comparing the EBLUP estimates with the census values and other GVF smoothing model for LFS application, e.g., Lesage, Beaumont and Bocci (2021).

4. Conclusion

In this paper, we have proposed sampling variance smoothing estimators using the generalized variance function method and smoothed design effect method for small area estimation. The proposed smoothing models and methods only require the use of the sample size in the model and computation of design effects. The proposed estimators $\tilde{V}_i^{GVF.RB}$, $\tilde{V}_i^{GVF.HBY}$ and $\tilde{V}_i^{DEFF}$ usually result in similar smoothed variance estimates. In practical applications, we may use the average smoothed estimator $\tilde{V}_i^{ASM}=(\tilde{V}_i^{GVF.RB}+\tilde{V}_i^{GVF.HBY}+\tilde{V}_i^{DEFF})/3$ as a data pooling to obtain the final smoothed variance estimate. The proposed smoothing methods simplify the smoothing procedure for practical users as they don’t need other complicated GVF models or auxiliary variables for the sampling variance modeling. Also the proposed smoothing procedure can be easily implemented in practice.

References

Beaumont, J.-F., and C. Bocci (2016), “Small area estimation in the Labour Force Survey“, paper presented at the Advisory Committee on Statistical Methods, May 2016, Statistics Canada.

Dick, P. (1995), “Modelling net undercoverage in the (1991) Canadian Census”, *Survey Methodology*, 21, 1, pp. 45-54.

Fay, R.E., and R. A. Herriot (1979), “Estimates of income for small places: an application of James-Stein procedures to census data”, *Journal of the American Statistical Association*, 74, pp. 269-277.

Gambino, J.G. (2009), “Design effect caveats”, *The American Statistician*, 63, pp. 141-146.

Hidiroglou, M.A., Beaumont, J.-F., and W. Yung (2019), “Development of a small area estimation system at Statistics Canada”, *Survey Methodology*, 45, No. 1, pp. 101-126.

Lesage, E., Beaumont, J.F., and C. Bocci (2021), “Two local diagnostics to evaluate the efficiency of the empirical best predictor under the Fay-Herriot model”, Survey Methodology, 47, No. 2, pp. 279-297.

Liu, B., Lahiri, P., and G. Kalton (2014), “Hierarchical Bayes modeling of survey weighted small area proportions”, *Survey Methodology*, 40, No. 1, pp. 1-13.

Rivest, L.P., and E. Belmonte (2000), “A conditional mean squared error of small area estimators”, *Survey Methodology*, 26, pp. 67-78.

Rivest, L.P., and N. Vandal (2002), “Mean squared error estimation for small areas when the small area variances are estimated”, *Proceedings of the International Conference on Recent Advances in Survey Sampling*, July 13, 2002, Ottawa, Canada.

Rao, J.N.K., and I. Molina (2015), *Small Area Estimation*, 2nd Edition, New York: John Wiley & Sons.

Souza, D.F., Moura, F.A.S., and H. S. Migon (2009), “Small area population prediction via hierarchical models”, *Survey Methodology*, 35, pp. 203-214.

Sugasawa, S., Tamae, H., and T. Kubokawa (2017), “Bayesian estimators for small area models shrinking both means and variances”, *Scandinavian Journal of Statistics*, 44, pp. 150-167.

Wang, J., and W. A. Fuller (2003), “The mean square error of small area predictors constructed with estimated area variances”, *Journal of the American Statistical Association*, 98, pp. 716-723.

You, Y. (2008), “An integrated modeling approach to unemployment rate estimation for sub-provincial areas of Canada”, *Survey Methodology*, 34, 1, pp. 19-27.

You, Y. (2021), “Small area estimation using Fay-Herriot model with sampling variance smoothing and modeling”, *Survey Methodology*, 47, pp. 361-370.

You, Y., and B. Chapman (2006), “Small area estimation using area level models and estimated sampling variances”, *Survey Methodology*, 32, pp. 97-103.

You, Y., Rao, J.N.K., and M. Hidiroglou (2013), “On the performance of self benchmarked small area estimators under the Fay-Herriot area level model”, *Survey Methodology*, 39, pp. 217-229.