Hybrid BRR and Parametric-Bootstrap Variance Estimates for Small Domains in Large Surveys

Eric Slud [*,†] Robert Ashmead [*,†]

Abstract
Following its justification in widely cited papers (McCarthy 1969, Krewski and Rao 1981, Fay 1984, 1989), Balanced Repeated Replication (BRR) has become a standard method for variance estimation in large complex surveys, especially in the US. However, it is also known that BRR variance estimates for very small domains are unreliable. Survey point estimates for small domains are often based on empirical-Bayes small area estimation models (Rao and Molina 2015), with variances estimated through parametric-bootstrap methods. This paper presents theory and practical details for a novel hybrid method, in which variances are estimated via parametric-bootstrap replications nested within BRR weight-replications. The method is presented first in general settings where categories are modeled within larger (but sometimes still small) domains. Then the results are specialized to the Dirichlet-multinomial hierarchical model describing small outcome proportions developed in the recent estimation from 2010-2014 American Community Survey data of language-minority and English proficiency characteristics in support of alternative-language ballot assistance determinations under Section 203(b) of the *Voting Rights Act of 1965*.

Key Words: American Community Survey, Balanced Repeated Replication, Dirichlet-Multinomial, Parametric Bootstrap, Small Area Estimation, Successive Difference Replication

1. Introduction

This paper concerns the problem of variance estimation arising in small area estimation based on a large survey. Suppose that it is desired to estimate population fractions or totals in disjoint subdomains C nested within somewhat larger domains D, where D may be large enough for design-based estimation but C is generally not, so that estimation of C-within-D proportions are to be achieved via small-area prediction. We assume that there are sufficiently many comparable domains C for which area-level predictors for C or D are available so that mixed-effect small area models make sense to fit. The topic of this paper is how to estimate variances of C-within-D proportions when Balanced Repeated Replication (BRR) variances for D totals are available and the models for observed data in terms of underlying proportions are assumed to be properly specified.

To fix ideas, let the domains D be indexed by $j = 1, \ldots, m$, while disjoint sub-domains C are indexed by (c, j), for $c = 1, \ldots, K$. Let the true unknown populations of these subdomains C_{cj} be denoted by N_{cj} and that of D_j by N_j. Sampled individuals are indexed by $i \in \mathcal{S}$ and have survey weights $\underline{w} = (w_i,\ i \in \mathcal{S})$, and we assume that survey-weighted unbiased estimators

$$\hat{N}_j = \sum_{i \in \mathcal{S} \cap D_j} w_i \ , \quad \hat{N}_{cj} = \sum_{i \in \mathcal{S} \cap C_{cj}} w_i$$

based on respective sample sizes n_j, n_{cj} are available and design-consistent but that n_{cj} are too small for the estimators $\hat{N}_{cj}$ to be reliable. For simplicity, the disjoint domains D_j will be treated as poststrata, with counts n_j (but not n_{cj}) regarded as nonrandom and fixed.

[*] US Census Bureau, Center for Statist. Res. & Methodology, 4600 Silver Hill Road, Washington DC 20233

[†] This report is released to inform interested parties of ongoing research and to encourage discussion of work in progress. Any views expressed on statistical, methodological, technical, or operational issues are those of the authors and not necessarily those of the U.S. Census Bureau.

Suppose also that vectors $\mathbf{X}_j$ of predictive variables at domain level are observed or known, and that the sample-weighted estimators treated as data follow a two-level small-area model,

$$\underline{\pi}_j \sim f(\underline{p}, \theta, \mathbf{X}_j) \;, \qquad \underline{\pi}_j = (\pi_{cj}, \;\; c = 1, \ldots, K) \;, \qquad \pi_{cj} \equiv N_{cj}/N_j \qquad (1)$$

where f is a probability density of known form with probability-vector dummy argument $\underline{p}$, and its unknown parameter θ will involve regression coefficients for $\mathbf{X}_j$ that may be different for predictions applicable to each subdomain C_{cj} , leading to

$$\underline{Y}_j \;\equiv\; (Y_{cj}\,, \;\; c = 1, ..., K) \;\sim\; \text{Multinomial}(n_j, \underline{\pi}_j) \quad , \qquad Y_{cj} \equiv n_j \hat{N}_{cj}/\hat{N}_j \qquad (2)$$

The rescaled sample counts Y_{cj} for subdomain C_{cj} may not be integers, and may either be rounded or analyzed using a likelihood extended to allow non-negative non-integer data. Model (1)–(2) is interpreted conditionally given $\{n_j, \;\mathbf{X}_j\}_{j=1}^m$ and assumes that $(\underline{\pi}_j, \underline{Y}_j)$ are independent across domain-indices j. Although one might imagine a stochastic mechanism by which $\hat{N}_j$ and $(\underline{Y}_j, \underline{\pi}_j)$ could be dependent (*cf.* Remark 1 at the end of Section 1.2), the model formulation is completed here by assuming independence:

$$\hat{N}_j \quad \text{is independent of} \quad (\underline{Y}_j, \underline{\pi}_j) \qquad (3)$$

As a corollary of this assumption and the independence of $(\hat{N}_j, \underline{Y}_j, \underline{\pi}_j)$ across j, it follows immediately that $\hat{N}_j$ is independent of the maximum likelihood (ML) estimator $\hat{\theta}$ of θ.

In this setting, the targets of estimation or prediction are the subdomain population totals N_{cj} and functions of them. These predictions will be made following the Empirical Best Linear Unbiased Prediction (EBLUP) strategy (Rao and Molina 2015): first the fixed-effect parameters θ are estimated by maximum likelihood (ML) from the combined observable dataset $(\underline{Y}_j, \; j = 1, \ldots, m)$, and then the targets $N_{cj} = N_j \, \pi_{cj}$ are estimated by substituting parameter estimates $\hat{N}_j$ and $\hat{\theta}$ into $E_\theta(\pi_{cj} \,|\, \underline{Y}_j)$, in the particular form

$$\tilde{N}_{cj} \;=\; \hat{N}_j \, E_\theta(\pi_{cj} \,|\, \underline{Y}_j) \Big|_{\theta = \hat{\theta}} \quad \text{or} \quad Y_{cj} \;+\; (\hat{N}_j - n_j) \, E_\theta(\pi_{cj} \,|\, \underline{Y}_j) \Big|_{\theta = \hat{\theta}} \qquad (4)$$

(Another reasonable choice for the estimator $\tilde{N}_{cj}$ is $n_{cj} + (\hat{N}_j - n_j) \, E_\theta(\pi_{cj} \,|\, \underline{Y}_j) \Big|_{\theta = \hat{\theta}}$, and this is the estimator that was used in producing the data results of Section 3. However, the theory underlying the variance-estimation method of this paper applies only to the count estimates (4), since n_{cj} is not separately modeled in (1)-(2) while Y_{cj} is.) This setting is actually common, since large surveys are often planned to enable design-based inference at the level of aggregation of domains which correspond to our D_j. But then the desire for finer-grained survey inference requires parametric small-area models, and the variances incorporate variability of both survey-weighted and parametrically modeled components. Taking account of both types of variability simultaneously is the subject of this paper.

The variability of predictions $\tilde{N}_{cj}$ (and ratios of them) comes about in two ways: first, through $\hat{N}_j$ which is not model-based, with variability asssessed through BRR, and second, through Y_{cj} with variability assessed through parametric bootstrap under model (1)–(3). The randomness in $\hat{N}_j$ is due to the sampling mechanism, acting through the unequal weights associated with the n_j randomly selected individuals in domain j. However, the parametric statistical model (1)–(3) is assumed to govern the proportions π_{cj} of domain j population within respective subdomains C_{cj}, as well as the scaled sample counts Y_{cj}. The 'hybrid' variance estimators studied in this paper combine the two levels of error. In general, the vector of Y_{cj} counts may be stochastically dependent on $\hat{N}_j$. (See Remark 1 below for discussion of this point.)

1.1 BRR methods

BRR methods of variance estimation attempt to embody design-based quadratic-form estimators of variance for a survey-weighted point estimator in the form of a weighted sample variance of the point estimator recalculated with a series of alternative weight-columns $\underline{w}^{(r)} = (w_i^{(r)},\ i \in \mathcal{S})$ indexed by $r = 1, \ldots, R$. (In later notations, we occasionally use $w_i^{(0)} \equiv w_i$ to denote the original column of survey weights.) The perspective of Fay (1984, 1989) is that with sufficiently many replicates, corresponding to pairs either of Primary Sampling Units (PSUs) or of split samples from single PSUs, quadratic-form variance estimators can be made exactly equal to such a scaled sample variance. The underlying theoretical assumption is that weighted totals of survey attributes within paired PSUs or split-samples can be treated as independent identically distributed (*iid*) random variables. Depending on the choice of paired, split, or paired and split PSU samples, the alternative weight-columns are products of $\underline{w}$ by linear combinations of the vector $\mathbf{1}$ of 1's and of one or two columns of ±1's taken from a so-called Hadamard matrix $\{a_{j,m}\}$ of ±1's with orthonormal columns orthogonal to $\mathbf{1}$. In the case of paired PSUs or a single split PSU within selected strata, the replicate weights for $r \geq 1$ are given, following Fay (1984), as

$$w_i^{(r)} \;\equiv\; w_i \cdot (1 + \frac{1}{2}\, a_{j,r}\, (-1)^h) \;\;, \quad j = \text{Strat}(i), \;\; h = h(i)$$

where each sampled individual i belongs to a unique split or paired PSU indexed as $h(i) = 1, 2$ within a unique (pseudo-) stratum $j = \text{Strat}(i)$ that contains i. With full degrees of freedom, the number of replicates can be taken between M-1 and M-4, where M is the total number of PSUs and split half-PSUs. Design-based quadratic-form estimators of variance of survey-weighted totals – usually chosen to be unbiased, but in systematic-sampling settings this is impossible and a biased quadratic form is allowed (Fay 1989) – are expressed exactly in terms of the replicate-weight estimators by equation (5). Then the large-sample properties of the BRR variance estimators are exactly those of the underlying quadratic form estimator. Large-sample properties of nonlinear functions of the survey point-estimator are obtained as in Krewski and Rao (1981) by linearization and the Delta Method.

Software implementations of BRR in Census Bureau surveys generally allow fewer than the full number (usually many hundreds) of degrees of freedom, grouping indices j more coarsely to allow $R = 80$, but after such reduction the orthonormality relations in the resulting Hadamard-matrix columns $\{a_{j,r}\}_{j=1}^m$ are only approximate (often satisfying the equations below with an error of order 0.01–0.02): for all distinct j, j',

$$R^{-1} \sum_{r=1}^{R} a_{j,r} \;\approx\; 0 \;\;, \quad R^{-1} \sum_{r=1}^{R} a_{j,r}^2 \;\approx\; 1 \;\;, \quad R^{-1} \sum_{r=1}^{R} a_{j,r}\, a_{j',r} \;\approx\; 0$$

The large-sample properties of BRR estimators with reduced degrees of freedom are largely undocumented, as far as we are aware. Nevertheless, this is the form in which the method is generally used in large national surveys. With replicate weight-columns $\underline{w}^{(r)}$ and vector attribute Z_i, the *BRR variance-estimator* $\hat{V}^{BRR}$ for a function $g(\hat{\bar{Z}})$ of weighted (Horvitz-Thompson) survey estimators of the mean of Z_i's in a frame population of size N is

$$\hat{V}^{BRR} = \frac{4}{R} \sum_{r=1}^{R} (g(\hat{\bar{Z}}^{(r)}) - g(\hat{\bar{Z}}^{(0)}))^2 \;, \quad \hat{\bar{Z}}^{(0)} = \frac{1}{N} \sum_{i \in \mathcal{S}} w_i Z_i \;, \quad \hat{\bar{Z}}^{(r)} = \frac{1}{N} \sum_{i \in \mathcal{S}} w_i^{(r)} Z_i \tag{5}$$

A further adaptation of BRR, called Successive Difference Replication (Fay and Train 1995), is based on sorting a survey sample into segments using survey variables and treating consecutive paired segments as approximate replicates. That is the approach to variance estimation used in the American Community Survey (ACS 2012). See Wolter (2007) for a general summary of BRR methods.

1.2 Parametric Bootstrap

In survey settings involving Small Area Estimation, where model-based methods enjoy wide acceptance, variances of survey-based statistical estimators are often found using Parametric Bootstrap methods. This is too large a topic for a self-contained introduction here, and we refer to the book of Shao and Tu (1995) for background. The idea is the same as a pure Monte Carlo study of variability with a large number B of *iid* replicates within a specified model, with unknown parameters replaced by the ML estimates from actual data. (Here and below, ML estimates might be replaced by other consistent estimates with variances of the same order $1/m$, but some of the excellent properties of parametric bootstrap do seem to depend on the parameter being chosen either as the ML estimator or at least approximately from a posterior distribution, which under the regularity conditions of the Bernstein- von Mises theorem will differ at most $O(1/m)$ from the ML estimator.) While a Monte Carlo study generates replication-based estimates of expectations of complicated functions – including variances of estimators and predictors – from data of a given structure at a single fixed parameter, the parametric bootstrap is justified in large samples by the proximity of the ML estimator to the unknown parameter values governing the data.

Based on a fixed (or ML estimated) vector θ_e, and a fixed set of sample sizes n_j and domain-level covariate vectors $\mathbf{X}_j$ for $j = 1, \ldots, m,$ imagine generating sets $(\underline{\pi}_j^{(b)},\ \underline{Y}_j^{(b)},$ $j = 1, \ldots, m)$, where $\pi_{cj}^{(b)} \equiv N_{cj}^{(b)}/N_j$ and $Y_{cj}^{(b)} \equiv n_{cj}^{(b)} \hat{N}_{cj}^{(b)}/\hat{N}_j$, independently across j and *iid* from model (1)–(3) across $b = 1, \ldots B$ for a large number B:

$$\underline{\pi}_j^{(b)} \overset{indep}{\sim} f(\underline{p},\, \theta_e,\, \mathbf{X}_j) \ , \qquad \underline{Y}_j^{(b)} \overset{indep}{\sim} \text{Multinomial}(n_j,\, \underline{\pi}_j^{(b)})$$

The ML estimators of θ_e from data $(\underline{Y}_j^{(b)},\ j = 1, \ldots, m)$ conditionally given $\mathbf{X}_j,\ n_j$ for $j = 1, \ldots, m$ will be denoted $\hat{\theta}^{(b)}$, and domain-specific predictors of $\underline{\pi}_j^{(b)}$ by

$$\hat{\underline{\pi}}_j^{(b)} \ = \ E_\theta\Big(\,\underline{\pi}_j^{(b)} \,|\, \underline{Y}_j^{(b)}\Big)\ \Big|_{\theta=\hat{\theta}^{(b)}}$$

By the law of large numbers, empirical averages such as

$$B^{-1} \sum_{b=1}^{B} \ h(\hat{\theta}^{(b)},\, \hat{\underline{\pi}}_j^{(b)}) \ \ , \quad B^{-1} \sum_{b=1}^{B} \ \Big(\hat{\underline{\pi}}_j^{(b)} \ - \ \underline{\pi}_j^{(b)}\Big)^{\otimes 2}$$

(where $v^{\otimes 2}$ for a vector v denotes $v\,v'$, and v' denotes transpose) estimate well-defined population quantities that are functions of (the fixed data $n_j,\ \mathbf{X}_j$ and) the model parameter θ_e. Parametric-bootstrap estimators would generally estimate the unconditional expectations of quantities $h(\hat{\theta}^{(b)},\, \hat{\underline{\pi}}_j^{(b)})$, but in our setting it is important to emphasize that they can estimate only expectations conditionally given the design-based total-estimates $\hat{N}_j$. The need for a hybrid estimator becomes clear when we try to estimate mean-squared prediction error (MSPE) from empirical averages of expressions involving estimates $\hat{N}_j$. In that case the two kinds of variability (design-based for $\hat{N}_j$ and model-based for $\hat{\underline{\pi}}_j$) interact.

For each $r = 1, \ldots, R$, alternative replicate estimators

$$\hat{N}_j^{(r)} \ = \sum_{i \in \mathcal{S} \cap D_j} w_i^{(r)} \ \ , \quad \hat{N}_{cj}^{(r)} \ = \sum_{i \in \mathcal{S} \cap C_{cj}} w_i^{(r)}$$

are generated, from which replicate ML estimators $\hat{\theta}^{(r)}$ are fitted to model (1)–(2) on data $n_j\, \hat{N}_{cj}^{(r)}/\hat{N}_j^{(r)}$. Then parametric bootstrap samples are drawn based on $\theta_e \equiv \hat{\theta}^{(r)}$, also for $r = 0$ which corresponds to the case of the original data Y_{cj} based on estimates $\hat{N}_{cj}$ with weights w_i, and all bootstrap loops are nested within loops of replicate-weights indexed by r. Since assumption (3) implies that $\underline{\pi}_j$ and $\underline{Y}_j$ are stochastically independent of $\hat{N}_j$, it follows that $\underline{\pi}_j^{(b)}$, $\underline{Y}_j^{(b)}$ and $\hat{\theta}^{(b)}$ are stochastically independent of $\hat{N}_j$ and $\hat{N}_j^{(r)}$ for fixed θ_e. However, the bootstrapped quantities $\underline{Y}_j^{(b)}$, $\underline{\pi}_j^{(b)}$ drawn in this way will generally depend on the index r through $\theta_e = \hat{\theta}^{(r)}$. For this reason, we generated the bootstrapped data $\underline{\pi}_j^{(b)}$, $\underline{Y}_j^{(b)}$ nested within loops of replicate-weights indexed by r. These conventions on nested parametric bootstrapping are indicated notationally by placing an asterisk and 'r' wherever bootstrap indices (b) appear. Thus the bootstrapped quantities are denoted $\underline{\pi}_j^{*(r,b)}$, $Y_{cj}^{*(r,b)}$, $\hat{\theta}^{*(r,b)}$, etc., where $r = 0, 1, \ldots, R$. The estimated domain counts $\hat{N}_j^{(r)}$ are not bootstrapped, and the weighted replicate subdomain estimates $\hat{N}_{cj}^{(r)}$ are used only in the rescaled variables $n_j\, \hat{N}_{cj}^{(r)}/\hat{N}_j^{(r)}$ and in the resulting ML estimates $\hat{\theta}^{(r)}$ which play the role of true parameters for parametric-bootstrap under the r'th weight-replicates.

Remark 1 *Especially when the number of sampled individuals within D_j is very small, it is easy to see that the design-based ratios $\hat{N}_{cj}/\hat{N}_j$ may be dependent on $\hat{N}_j$. The randomness of $\hat{N}_{cj}$ depends both on the number of sampled D_j persons in C_{cj} but also on the weights associated with those individuals. The form of model (2) does not depend on the estimated numbers $\hat{N}_{cj}$ of sampled C_{cj} persons, and we have imposed an assumption (3) of independence and done our bootstrap sampling nested within fixed $\hat{N}_j^{(r)}$. Then $(\underline{\pi}_j^{*(r,b)}, \underline{Y}_j^{*(r,b)})$ is related to $\hat{N}_j^{(r)}$ only through dependence of both on $\hat{\theta}^{(r)}$. The validity of assumption (3) may be investigated by exploratory data-analytic comparison of the degree of dependence of $(Y_{cj},\, \hat{N}_j)$ versus that of $(Y_{cj}^{*(r,b)},\, \hat{N}^{(r)})$, but we have not done that in the real application to ACS data in this paper.* □

2. Hybrid BRR and Bootstrap

The idea behind the nested replicate-bootstrap loops described in the previous section is to estimate the variability of predictors $\tilde{N}_j^c$ (defined in the second way within (4)) in terms of levels of error from $\hat{N}_j$ in estimating N_j and from $\hat{\underline{\pi}}_j$ in predicting $\underline{\pi}_j$. Since the estimates $\tilde{N}_{cj}$ aim to predict $N_{cj} = N_j \cdot \pi_{cj}$, which according to our formulation (1) contains π_{cj} as a random domain effect, the overall measure of prediction error to be estimated is the Mean Squared Prediction Error, $\text{MSPE}_{cj} = E(\tilde{N}_{cj} - N_j\, \pi_{cj})^2$. Here and below, the expectation is defined both over the design and the model, conditionally given n_j and the covariates. This MSPE is decomposed as in Analysis of Variance, based on the idea of estimating the separate errors at N_j and π_{cj} level. Using $\tilde{N}_{cj} = Y_{cj} + (\hat{N}_j - n_j)\hat{\pi}_{cj}$, we find

$$\tilde{N}_{cj} - \hat{N}_j\, \pi_{cj} = \{Y_{cj} - n_j\hat{\pi}_{cj} + N_j(\hat{\pi}_{cj} - \pi_{cj})\} + (\hat{N}_j - N_j)(\hat{\pi}_{cj} - \pi_{cj})$$

so that by (3) and unbiasedness of $\hat{N}_{cj}$ for N_{cj},

$$E\Big(\tilde{N}_{cj} - \hat{N}_j\, \pi_{cj}\Big)^2 = E\Big(Y_{cj} - n_j\hat{\pi}_{cj} + N_j(\hat{\pi}_{cj} - \pi_{cj})\Big)^2 + \text{Var}(\hat{N}_j)\, E(\pi_{cj} - \hat{\pi}_{cj})^2$$

and similarly

$$E\Big(\tilde{N}_{cj} - N_j\, \pi_{cj}\Big)^2 = E\Big(Y_{cj} - n_j\hat{\pi}_{cj} + N_j(\hat{\pi}_{cj} - \pi_{cj})\Big)^2 + \text{Var}(\hat{N}_j)\, E((\hat{\pi}_{cj})^2)$$

so that (by subtracting the first of these last two equations from the second), $\text{MSPE}_{cj} =$

$$E\Big(\tilde{N}_{cj} - N_j\,\pi_{cj}\Big)^2 = E\Big(\tilde{N}_{cj} - \hat{N}_j\,\pi_{cj})\Big)^2 + \text{Var}(\hat{N}_j)\,E\Big((\hat{\pi}_{cj})^2 - (\pi_{cj} - \hat{\pi}_{cj})^2\Big) \tag{6}$$

The purpose of the four preceding displayed equations was to express MSPE_{cj} in terms of bootstrap residuals

$$e^{*(r,b)}_{cj} = \tilde{N}^{*(r,b)}_{cj} - \hat{N}^{(r)}_j\,\pi^{*(r,b)}_{cj} \qquad \text{and} \qquad \epsilon^{*(r,b)}_{cj} = \hat{\pi}^{*(r,b)}_{cj} - \pi^{*(r,b)}_{cj} \tag{7}$$

instead of expressions involving N_j, which is not known and not modeled. This step is important, since the residual expression $\tilde{N}_{cj} - N_j\,\pi_{cj}$ cannot be bootstrapped and although $\hat{N}^{(r)}_j$ is computed from weight-replicates for purposes of design-based variance estimation, these quantities are not independent replicates of the unknown N_j across r .

To develop the decomposition of MSPE_{cj}, using the notation $e_{jc} \equiv \tilde{N}_{cj} - \hat{N}_j\,\pi_{cj}$ and the independence (3), we expand

$$E(e^2_{cj}) = E(e_{cj} - E(e_{cj}\,|\,\hat{N}_j))^2 + E\Big(E(e_{cj}\,|\,\hat{N}_j) - E(e_{cj})\Big)^2 + (E(e_{cj}))^2 \tag{8}$$

To estimate the three terms in this decomposition from the nested replicate-weight bootstrap, we define respective *Within*, *Between* and *Bias*^{2} terms, using the notations

$$\bar{e}^{*(r+)}_{cj} \equiv \frac{1}{B}\sum_{b=1}^{B} e^{*(r,b)}_{cj}\;, \qquad \text{Within}(\{e^{*(r,b)}_{cj}\}) = \frac{1}{R(B-1)}\sum_{r=1}^{R}\sum_{b=1}^{B}(e^{*(r,b)}_{cj} - \bar{e}^{*(r+)}_{cj})^2$$

$$\text{Betw}(\{e^{*(r,b)}_{cj}\}) = \frac{4}{R}\sum_{r=1}^{R}(\bar{e}^{*(r+)}_{cj} - \bar{e}^{*(0+)}_{cj})^2\;, \qquad \text{Bias}(\{e^{*(r,b)}_{cj}\}) = \frac{1}{R}\sum_{r=1}^{R}\bar{e}^{*(r+)}_{cj}$$

The *Within* estimator represents an unconditional design-based estimator of the bootstrapped conditional variance of the residual $e^{*(r,b)}_{cj}$ given $\hat{N}^{(r)}$. The idea in the *Between* term is to estimate the r'th replicate of $E(e_{cj}\,|\,\hat{N}_j)$, viewed as a nonlinear function of $\hat{N}_j$, via the bootstrap-averaged within-r residuals $\bar{e}^{*(r+)}$, and to use the replicates to estimate the unconditional variance of $E(e_{cj}\,|\,\hat{N}_j)$ by the standard BRR recipe (5). The summed terms on the right-hand side of (8) are estimated by

$$\hat{E}(e^2_{cj}) = \text{Within}(\{e^{*(r,b)}_{cj}\}) + \text{Betw}(\{e^{*(r,b)}_{cj}\}) + \Big(\text{Bias}(\{e^{*(r,b)}_{cj}\})\Big)^2 \tag{9}$$

where the laws of large numbers underlying the Parametric Bootstrap imply the consistency

$$\frac{1}{B-1}\sum_{b=1}^{B}(e^{*(r,b)}_{cj} - \bar{e}^{*(r+)}_{cj})^2 \xrightarrow{P} \text{Var}(e_{jc}\,|\,\hat{N}_j = \hat{N}^{(r)}_j)\;, \qquad \bar{e}^{*(r+)}_{cj} \xrightarrow{P} E(e_{cj}\,|\,\hat{N}_j = \hat{N}^{(r)}_j)$$

for fixed r and $B \to \infty$. The probability limits in this last display are then regarded as nonlinear functions of the replicate-weight domain j population estimates which when averaged over the full set of $r = 1, \ldots, R$ replicates are approximately equal to the design expectations over $\hat{N}_j$.

Remark 2 *There is no theoretical justification for treating weight-replicated estimates $\hat{N}^{(r)}_j$ as independent identically distributed variables across r. A law of large numbers over indices r might be justified for large domains by viewing $\hat{N}^{(r)}_j$ as a weighted sum* *of many independent terms corresponding to primary sampling units (PSUs). For small domains and reduced sets of replicates, this approximation is likely not to be very good.*

There is another justification for estimates based on averaging over r. In settings where the number m of domains is large and the model (1)–(3) holds, under some assumptions the estimates $\hat{\theta}^{(r)}$ can be shown to be close in the sense of convergence in probability to the true parameter values θ. The dependence of residuals e^c_j on $\hat{N}_j$ when $\hat{\theta}^{(r)}$ estimates are replaced by θ is linear, so that the MSPE terms are linear or quadratic, and in that case the justifications of large-sample BRR consistency in Krewski and Rao (1981) hold, when R and D_j are large, just as they do for

$$\hat{V}^{BRR}(\hat{N}_j) \; = \; \frac{4}{R}\sum_{r=1}^{R} (\hat{N}_j^{(r)} - \hat{N}_j)^2 \qquad \Box$$

The reasoning of the previous paragraphs justifies the accuracy for large B, large domains D_j and large sets of distinct replicates, of the estimation of mean-squared error $E(e_{cj})^2$ by (9). Similar reasoning support the accuracy (for large B, R, and D_j) of

$$\hat{E}(\epsilon_{cj})^2 \; = \; \text{Within}(\{\epsilon_{cj}^{*(r,b)}\}) \, + \, \text{Betw}(\{\epsilon_{cj}^{*(r,b)}\}) \, + \, \Big(\text{Bias}(\{\epsilon_{cj}^{*(r,b)}\})\Big)^2 \qquad (10)$$

as an estimator of $E(\epsilon_{cj})^2$. Averaging over both replicate and bootstrap iterations provides a similarly justified estimator

$$\hat{E}((\hat{\pi}_{cj})^2) \; = \; \frac{1}{BR}\sum_{r=1}^{R}\sum_{b=1}^{B} (\hat{\pi}_{cj}^{*(r,b)})^2 \qquad \text{for} \qquad E(\hat{\pi}_{cj})^2$$

These results, together with the formula (6) for $\text{MSPE}_{cj} = E(\tilde{N}_{cj} - N_j\pi_{cj})^2$, establish

Proposition 1 *For large B, R, and D_j, subject to the caveats of Remark 2, the estimator*

$$\widehat{MSPE}_{cj} \; = \; \textit{Within}(\{e_{cj}^{*(r,b)}\}) \, + \, \textit{Betw}(\{e_{cj}^{*(r,b)}\}) \, + \, \Big(\textit{Bias}(\{e_{cj}^{*(r,b)}\})\Big)^2 \, + \, \hat{V}^{BRR}(\hat{N}_j)\cdot$$

$$\Big\{\frac{1}{BR}\sum_{r=1}^{R}\sum_{b=1}^{B}(\hat{\pi}_{cj}^{*(r,b)})^2 - \textit{Within}(\{\epsilon_{cj}^{*(r,b)}\}) - \textit{Betw}(\{\epsilon_{cj}^{*(r,b)}\}) - \Big(\textit{Bias}(\{\epsilon_{cj}^{*(r,b)}\})\Big)^2\Big\}$$

is an accurate estimator of MSPE$_{cj}$.

The terms in this formula multiplying $\hat{V}^{BRR}$ are grouped as 'Between' terms together with *Betw*$(\{e_{cj}^{*(r,b)}\})$. Then the three types of terms are interpreted as follows: *Within* terms for e residuals reflect the average (over replicate-weights) contribution to mean-squared e residuals of bootstrap iterations b within r; *Bias-Squared* terms for e reflect the contribution to variability of averaged e residuals and measure errors in proper centering across r; and all other terms viewed as *Between* terms reflect variability due to variance contributions from sampling differences arising from different replicate weights.

It is evident that the MSPE estimates obtained by the hybrid method described here are available only for domains j with positive sampled and true counts n_j, N_j. Thus, in the next section, jurisdictions j without sample n_j are out of scope for estimates and variance predictions. Only a purely synthetic or model-based prediction of $\underline{\pi}_j$ could be hoped for in such jurisdictions.

Remark 3 *There is a further modification of the variance estimation formula which can be implemented, for computational savings, by basing the BRR loop on a subset of L weight-replicates sampled from the full set $\{1, \ldots, R\}$ of available replicates. (In our application below to Hispanic data from ACS 2010-2014, $R = 80$ and $L = 40$.) In this case, the replicate-weight indices r range over L values selected as a simple random sample $\mathcal{L}$ of L elements from 1 to R. Then the BRR variance formula changes to*

$$\hat{V}_L^{BRR}(\hat{N}_j) \;=\; 4\,\frac{L-1}{L}\,Var(\{\hat{N}_j^{(r)}\}_{r\in\mathcal{L}}) \,+\, 4\Big(\frac{1}{L}\sum_{r\in\mathcal{L}} \hat{N}_j^{(r)} \;-\; \hat{N}_j\Big)^2$$

The Betw *terms in Prop. 1 change similarly, and the other averages $R^{-1}\sum_{r=1}^{R}$ in the final variance formula, including those in* Within *and* Bias*, change to $L^{-1}\sum_{r\in\mathcal{L}}$.* □

3. Implementation on ACS 2010-2014 Data

Section 203(b) of the *Voting Rights Act of 1965* as amended in 1975 requires that US states and political subdivisions must in certain circumstances make voting materials available in languages other than English. These circumstances are specified in terms of the sizes and proportions of designated population subgroups measured by the decennial census and most current available American Community Survey (ACS), as determined from these data sources by the Director of the Census Bureau. See `https://www.census.gov/rdo/data/voting_rights_determination_file.html` for additional details.

The population subgroups whose estimated sizes figure into the law are the numbers of voting-age persons within states and 'jurisdictions' and American Indian or Alaska Native (AIAN) voting-age persons within American Indian Areas (AIAs), who also fall into one of 68 specified racial/ethnic Language Minority Groups (LMGs) and are either Citizens, Limited English Proficient (LEP) Citizens, or Illiterate LEP Citizens. The 'jurisdictions' are the political units which run local elections, counties except in 8 'MCD states' where Minor Civil Divisions (MCDs) are the relevant units. The sizes of the Citizen, Citizen LEP, and illiterate Citizen LEP subpopulations of the intersections of the political subdivisions and LMGs range from quite large (as, for example, Hispanics within large cities or large border-state counties) to very small (for the smaller Asian LMGs, or AIAN tribal groups outside the areas near their AIAs or reservation areas). For this reason, the estimates of population sizes and fractions within these small population domains, which is how we refer to these intersections, may be based on very small sample-sizes in the 5-year 2010-2014 direct ACS survey-weighted estimates.

The methods of this paper were developed in connection with the Voting Rights Act Section 203(b) determinations recently released (Dec. 5, 2016) by the Census Bureau based on ACS 2010-2014 data. Related methods for Section 203(b) determinations in 2011, developed by Joyce et al. (2014), were based on ACS 2005-2009 data but also made direct use of decennial 2010 Census totals N_j . The estimation of population proportions in the categories of Citizen, LEP Citizen, and illiterate LEP Citizen within the universe of voting-age Language Minority Group persons was done in 2016 using a Dirichlet-Multinomial model which is a special case of the model (1)-(3) given in the Introduction. We introduce the model in a slightly simplified setting, considering only a single LMG (Hispanic, the largest one) and as domains D_j only the 6837 jurisdictions (2897 counties and 3940 MCDs) in which ACS sampled at least one Hispanic voting-age person in 2010-2014, and only three c categories: non-citizen (c=1), non-LEP citizen (c=2) and LEP citizen (c=3). The Voting Rights Act Section 203(b) requires not only the estimation of the LEP citizen population in each LMG, both as a count and as a proportion of the overall citizen population, within jurisdictions, but also other counts and proportions in other political subdivisions. In this section, and in the data results, we restrict attention to estimation of LEP citizen counts for the Hispanic LMG.

The Dirichlet-Multinomial model used in this application (for each LMG) had the form (1)–(3) in which (1) is specified as

$$(\pi_{1j},\, \pi_{2j},\, \pi_{3j}) \;\sim\; \text{Dirichlet}(\tau\, \sqrt{n}_j,\, (1-\mu_j,\, \mu_j(1-\nu_j),\, \mu_j\nu_j)) \tag{11}$$

where $\text{Dirichlet}(m, (p_1, p_2, p_3))$ is the density on 3-entry probability vectors which is proportional to $x_1^{mp_1}\, x_2^{mp_2}\, (1-x_1-x_2)^{mp_3}$, the unknown parameter is $\theta = (\tau, \beta, \gamma)$, and where domain-level parameters $\mu_j,\, \nu_j$ are expressed in terms of coefficient vectors $\beta,\, \gamma$ of the same dimension d as the predictive domain-level covariates $\mathbf{X}_j$ in the form

$$\mu_j \;=\; \exp(\beta'\,\mathbf{X}_j)\,/\,(1+\exp(\beta'\,\mathbf{X}_j)) \quad , \qquad \nu_j \;=\; \exp(\gamma'\,\mathbf{X}_j)\,/\,(1+\exp(\gamma'\,\mathbf{X}_j))$$

For this model, the conditional expectations used in predictors take the explicit BLUP form

$$E_\theta(\pi_{3j}\,|\,\underline{Y}_j,\, \mathbf{X}_j) \;=\; (\tau\,\sqrt{n}_j\,\mu_j\,\nu_j \;+\; Y_{3j})\Big/(\tau\,\sqrt{n}_j \;+\; n_j) \tag{12}$$

after which the predictors $\hat{\pi}_{3j}$ are given by the same formula with ML estimates $\hat{\mu}_j,\, \hat{\nu}_j$ obtained by substituting respective MLEs $\hat{\beta},\, \hat{\gamma}$ for $\beta,\, \gamma$ in the defining formulas for μ_j, ν_j. The predictor variables $\mathbf{X}_j$ used in fitting this Hispanic LMG model are: State-level rates of citizenship within Hispanic voting-age population and of LEP within Hispanic citizens; Jurisdiction-level proportions of Hispanic persons who are foreign-born, or have educational level less than high-school; Jurisdiction average number of years Hispanic persons have been in the US; Jurisdiction proportions of all voting-age persons with less than high school education or who are white-or-black non-Hispanic. All rates and proportions in this covariate list were logit-transformed, but other averages and counts were not transformed.

3.1 Data Results

Model fitting and validation of (11) and (2) in the data application are discussed in Ashmead and Slud (2017). In this section, we exhibit varous aspects of the variances of the model-based estimator $\tilde{N}_{3j}$ of LEP citizen Hispanic population counts obtained by the 'hybrid' variance estimation methodology of Section 2. Recall that the BRR method of survey-weighted variance estimation used in the ACS is the SDR method of Fay and Train (1995). We compare the variances estimated by the new hybrid method with the direct SDR variance estimators $\hat{V}^{BRR}(\hat{N}_{3j})$ that would have been used if direct ACS-weighted estimators $\hat{N}_{3j}$ had been used to estimate N_{3j}, showing that the latter would have had excessively large Coefficients of Variation (CVs). Although CVs were generally much better using the model-based estimates, the improvement was not universal, and drastic differences between model-based variance improvements are seen in large versus small jurisdictions. For model-based variances, we will display the proportions due respectively to the Within and Between terms, noting that the Bias-squared terms are small.

3.2 CV's and Variances for Direct and Model-based Estimators

The main reason for estimating domain-level LEP citizen populations within each LMG through a parametric model with cross-domain shared parameters is that the variances of direct survey-weighted estimators are too large. To quantify this effect, Table 1 displays the proportion of jurisdiction-within-Hispanic estimated CVs within various ranges, for estimators of total numbers of LEP Hispanic citizens. The ranges include a small-CV

Table 1: Fractions of 3671 jurisdictions with positive number of sampled LEP citizens in which direct-method and model-based estimators of CV fall within indicated ranges.

CV Range	(0, .2]	(.2, .3]	(.3, .4]	(.4, .5]	(.5, .61]	(.61, 1]	(1,∞]
Direct CV	0.168	0.104	0.089	0.089	0.093	0.317	0.140
Model CV	0.234	0.151	0.141	0.111	0.108	0.221	0.034

bracket (those less than 20%) and two moderate-range CV's (20-30% and 30-40%). CV's of 50% are usually considered large, and those 61% or more excessively large. (According to ACS quality guidelines, tables whose entries have median CV > 0.61 are held back from public release in American Factfinder.) The first row of Table 1 shows the distribution of the direct-method CV, that is, the square-root of the SDR variance of the direct survey-weighted estimator, divided by the value of that estimatior. This row of CV's makes sense only for those jurisdictions in which the direct survey estimator was non-zero, i.e., only for those with at least one sampled Hispanic LEP citizen. The second row contains ranges of CVs for the model-based estimator also restricted to those sample jurisdictions, for comparability. Although generally much better than the direct-method CVs, the model-based CVs are also frequently large. For example, while 64% of the direct CVs are greater than 0.4, the corresponding proportion for model-based CVs is 47%.

Because of the sharing of model parameters across jurisdictions, the model-based estimators and variances are positive (and generally, small) even in those jurisdictions in which no Hispanic LEP citizens were sampled. Thus, in the 3166 jurisdictions in which Hispanic voting-age persons were sampled but none were LEP citizens, the rounded 90$^{\text{th}}$ percentile of the estimated jurisdiction totals of Hispanic LEP citizens was 5, while the 90$^{\text{th}}$ percentile of the proportion of Hispanic LEP citizens out of all citizens was 0.15%. The latter proportion should be compared with the national LEP proportion of 4.6% among Hispanic voting-age citizens. The jurisdictions with no Hispanic LEP citizen sample at all, represent the extreme of the tendency for small jurisdictions with only a few sampled Hispanic LEP citizens to yield model-based variance estimators for Hispanic LEP citizen estimates that are larger than the corresponding direct-method SDR Variance estimates.

To obtain further perspective, consider Figure 1 in which the log ratios of MSPEs over SDR direct-method variances are plotted against the logarithms of direct SDR variances. The Figure shows first that the MSPE/(SDR Variance) ratio is usually less than 1, as can be visualized through the lowess line falling below the level 1. However, there are numerous jurisdictions where this ratio is > 2. However, the great majority of jurisdictions where this occurs have SDR Variances of 250 or less (so standard deviations of 16 or less). In the larger jurisdictions, where logarithm of SDR Variance is 10 or more, the great bulk have MSPE/(SDR Variance) < 1 and nearly all have ratio < 1.5. Moreover, the coloring of points progressing from green (the smallest LEP-total jurisdictions) through yellow toward red (the jurisdictions with largest LEP totals) shows several clear tendencies. First, the jurisdictions with very small sample sizes or LEP count estimates have especially small ratios of MSPE over SDR variance. Second, the jurisdictions with large variances tend also to have large LEP-count estimates, showing very little MSPE improvement versus SDR variance. Finally, the great majority of jurisdictions with intermediate variance (say from 100 to $30,000$) have MSPE much smaller than SDR variance; but within this group, it tends to be the jurisdictions with smallest LEP-count that have MSPEs notably worse than SDR variance.

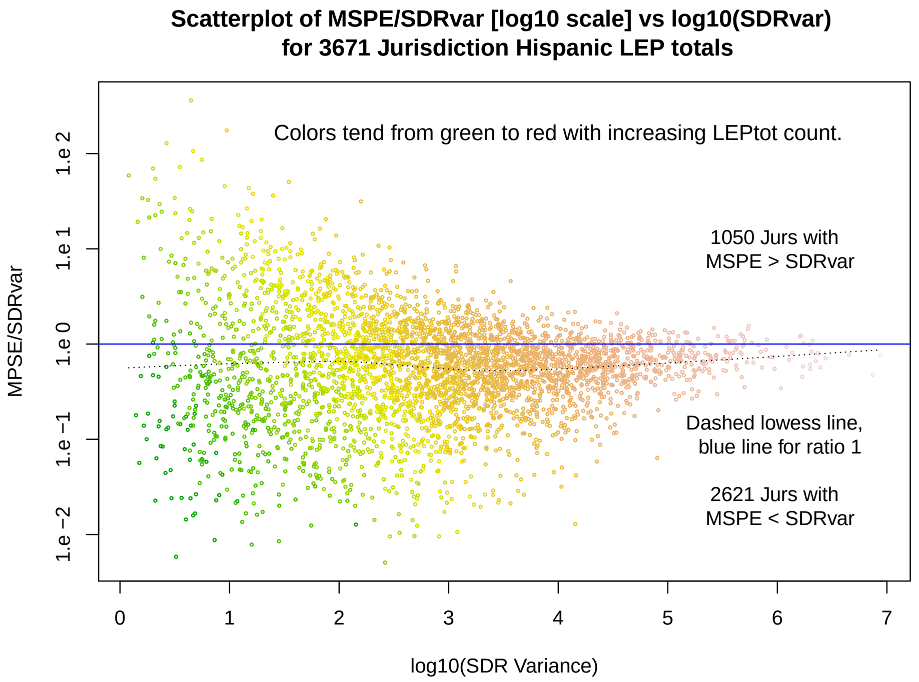

Figure 1: Plot of MSPE/SDR Variance (on $\log_{10}$ scale) vs. $\log_{10}$(SDR Var), for the 3671 jurisdictions with 2010-2014 ACS-sampled LEP Hispanic citizens. Color tends from green to red with increasing model-estimated LEP count. Dashed line is plotted lowess curve.

3.3 *Model-based Variance Improvements by Size*

Two additional displays usefully show the improvement of MSPE over SDR Variance and the magnitude of MSPE as a function of jurisdiction size. The first partitions the ACS 2010-2014 sample (of all voting-age Hispanic persons in each jurisdiction) by size-classes. While Figure 1 showed comparisons of variances of estimators of LEP Hispanic citizen totals by individual jurisdiction, Figure 2 compares selected quantiles (0.25, 0.5, 0.75, 0.8, .0.9) of these same variances among jurisdictions with sample, for three different size-classes and for all jurisdictions pooled together. This comparison strikingly shows both that the MSPE quantiles are much reduced versus SDR Variances in all size classes, but also that the reduction is greater in the small-size jurisdictions than in the large ones, and greater for the above-median quantiles than the below-median. The larger model-based reductions for small-size jurisdictions does not contradict our comments summarizing Figure 1, in view of the color (LEP-total) progression and the very large numbers of small-variance (usually small-size) jurisdictions.

A second way to exhibit variances of model-based estimated totals is to plot them or their CVs against the estimated weight $\sqrt{n_j}/(\hat{\tau} + \sqrt{n_j})$ applying to the direct estimator Y_{3j}/n_j in equation (12). This weight increases monotonically with domain sample size n_j, and approaches 1 as n_j gets large. Figure 3 plots MSPE-based CV versus weight, for jurisdictions in which 10 or more voting-age persons were sampled in ACS 2010-2014 and in which the rounded estimate of Hispanic LEP citizen count was at least 1. The picture strongly shows the pattern of CV decrease with direct-estimate weight, narrowing down to a CV close to 0 as the sample-size becomes large and the weight approaches 1. Similar figures, in which the restriction to sample-size of 10 or more is dropped, are more cluttered but show the same pattern.

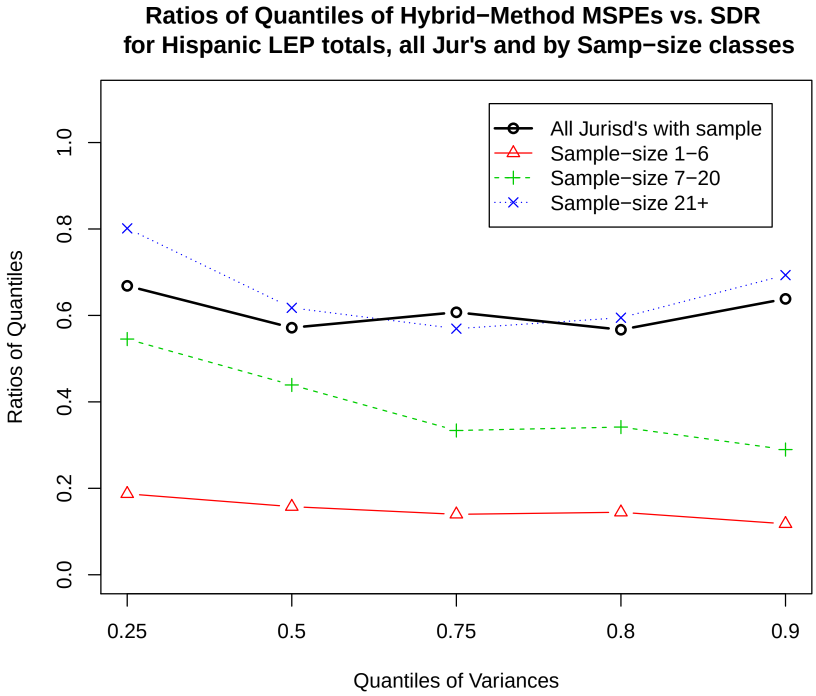

Figure 2: Ratios of selected quantiles of MSPEs over quantiles of SDR Variances, among jurisdictions with ACS-sampled LEP Hispanic citizens in 2010-2014, by sample-size class

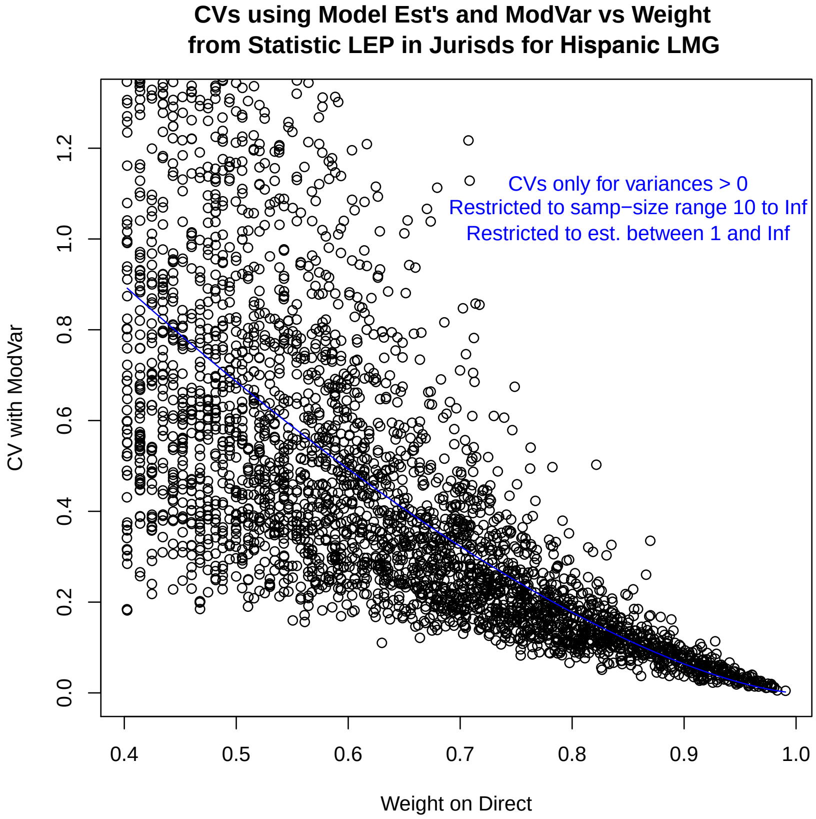

Figure 3: MSPE-based CVs for estimated Hispanic LEP totals versus weight of predictor (12) on direct estimator, plotted for jurisdictions with positive variance, rounded LEP count estimate ≥ 1, and sample-size ≥ 10

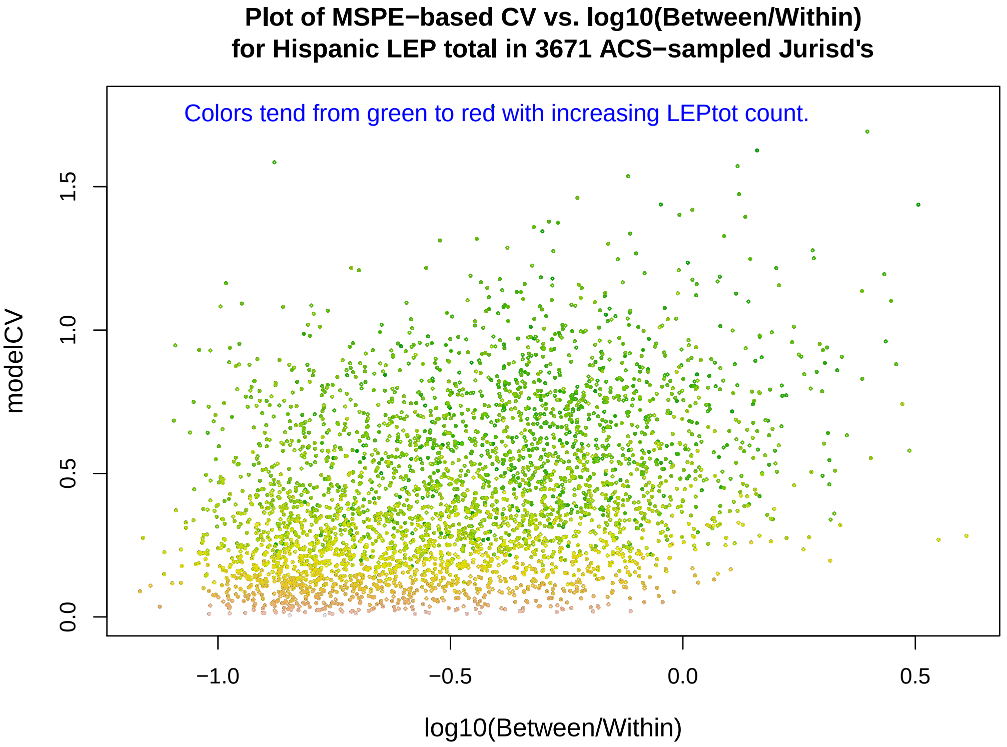

Figure 4: Plot of MSPE-based CV versus $\log_{10}$(Between/Within) for estimated Hispanic citizen LEP totals in 3671 jurisdictions with positive Hispanic sample. Colors range from green to red with increasing estimated LEP count .

Another aspect of Figure 3 deserves comment. The blue plotted curve, generated by a least squares fit of $\log(\mathrm{CV})$ to $\log(\mathrm{Wt})$, shows an unexpectedly simple relation between CV_j versus the sample-size-dependent weight $\mathrm{Wt}_j = \sqrt{n_j}/(\sqrt{n_j}+\hat{\tau})$. From the simplicity of the fitted curve, the blue line suggests a simple Generalized Variance Function.

3.4 Relative Contribution of Variance Components

Immediately following Proposition 1, the terms in the hybrid MSPE/Variance formula given there were sorted into three categories: *Within*, *Between*, and *Bias-squared*. The relative magnitudes of these three types of variance terms for the outcome estimate of Hispanic LEP citizen totals in the jurisdictions with Hispanic sample, vary considerably with LEP count. In Figure 4, we exhibit the MSPE-based Coefficient of Variation as a function of the base-10 logarithm of the ratio Between/Within. The points are plotted in colors progressively tending from green to red as (estimated) LEP count increases. The plot shows that, at least in the Hispanic LMG, the Within variance-component is generally the largest one (since more than 90% of the points lie to the left of the log-ratio value 0), and that the Between/Within ratio has a slight tendency to increase with increasing LEP count, while the CV has a marked tendency to decrease with increasing LEP count, with Between fraction generally larger when the CV is smaller. (The Between and Within components of variance separately have a tendency to increase with LEP count, but the slight increase of the Between/Within ratio shows that Between component increases slightly faster with LEP count than the Within does.) For the points in this plot, we did not show the *Bias-squared* term since it was always relatively small, with maximum value 0.113.

4. Summary and Future Research

This research was first undertaken as part of the technical support for the estimation of Language Minority Group population components and ratios required for determinations of alternative language election materials mandated for states and political subdivisions under the Voting Rights Act, Section 203(b). The methods described here were used in estimating variances for all model-based population estimates[1] released together with the 2016 Voting Rights Act determinations on December 5, 2016. These variances were released July 25, 2017 by the Census Bureau. Full technical documentation of the point and variance estimation in this application will be publicly released in the near future.

The hybrid MSPE and variance estimation method introduced in this paper has theoretical validity established in Proposition 1 and has been shown in the paper's data application (within the Hispanic Language Minority Group) to provide generally smaller and more acceptable variance and CV estimates than would have been possible for direct (model-free) estimates from ACS data. Accurate variance and MSPE estimation enhances the credibility of model-based estimates of the population components required by the Voting Rights Act Section 203(b). The regression-type models (11) and (2) used in the model-based estimates are assessed in the related research report of Ashmead and Slud (2017). In addition, the variance estimation method studied here rests in part on an assumption (3) which, while plausible, needs to be checked further in the data application, as mentioned in Remark 1.

The variance estimation method advanced here may be useful in other contexts where Small Area Estimation is used to estimate population subdomains within domains that are themselves large enough to support direct survey-weighted (Horvitz-Thompson) estimates. It applies broadly in the setting where design-based estimates are adequate for domains within which conditional probabilities of falling in subdomains can be parametrically modeled and are needed because direct estimates of subdomain totals are too inaccurate. Another example of survey inference with this combination of design-based domain estimates and model-based subdomain proportions can be found in Thibaudeau et al. (2017), with data application to the Survey of Income and Program Participation.

We suggest three promising directions of further research related to the hybrid variance estimation method of this paper. First, it seems likely that a similar nested-loop variance estimation procedure could be developed under similar assumptions with the outer loop provided by a jackknife or bootstrap procedure in place of BRR. Second, combined balanced-replicate and parametric-bootstrap variance estimation would benefit from further research extending to cases where the independence assumption (3) does not hold. And third, in the Voting Rights Act datasets extending to other Language Minority Groups, further research developing Generalized Variance Functions in the spirit of Figure 3 would be valuable.

[1] For some smaller Language Minority Groups, data were too sparse to fit models, and in those cases the released population estimates were calculated via direct survey-weighted totals from ACS 2010-2014 data, with variances estimated by the direct SDR method regularly used by ACS.

REFERENCES

American Community Survey (2012) documentation, Chapter 12 Variance Estimation.

Ashmead, R. and Slud, E. (2017), Small area model diagnostics and validation with applications to the Voting Rights Act Section 203, Census Bureau preprint.

Census Bureau (2016), VRA Statistical Methodology Summary, `https://www.census.gov/rdo/pdf/3_VRA_Statistical_Methodology_Summary_V7.pdf`.

Fay, R. (1984), Some properties of estimators of variance based on replication methods, *Proc. Amer. Statist. Assoc., Survey Res. Methods Section*, 495-500.

Fay, R. (1989), Theory and application of replicate weighting for variance calculations, *Proc. Amer. Statist. Assoc., Survey Res. Methods Section*, 212-217.

Fay, R. and Train, G. (1995), Aspects of survey and model-based post-censal estimation of income and poverty characteristics for states and counties, Census Bureau Report.

Joyce, P., Malec, D., Little, R., Gilary, A., Navarro, A., and Asiala, M. (2014), Statistical modeling methodology for the Voting Rights Act Section 203 language assistance determinations, *Jour. Amer. Statistical Assoc.*, 109 (505), 36-47.

Krewski, D. and Rao, J. (1981), Inference from stratified samples: properties of the linearization, jackknife and balanced repeated replication methods, *Annals of Statistics* **9**, 1010-1019.

McCarthy (1969) Pseudoreplication: half-samples, *International Statistical Review* **37**, 239-264.

Rao, J. and Molina, I. (2015) **Small Area Estimation**, 2nd ed., Wiley.

Shao, J. and Tu, Y. (1995), **The Bootstrap and Jackknife**, Springer.

Thibaudeau, Y., Slud, E. and Gottschalck, A. (2017) Modeling log-linear conditional probabilities for estimation in surveys, *Annals of Applied Statistics* **11**, 680-697.

Wolter, K. (2007) **Introduction to Variance Estimation**, 2nd ed., Springer.