RESEARCH REPORT SERIES
(Statistics #2018-12)

Statistical Methodology (2016) for Voting Rights Act, Section 203 Determinations

Eric Slud
Robert Ashmead
Patrick Joyce
Tommy Wright

Center for Statistical Research and Methodology
Research and Methodology Directorate
U.S. Census Bureau
Washington, D.C. 20233

Report Issued: December 17, 2018

Disclaimer: This report is released to inform interested parties of research and to encourage discussion. The views expressed are those of the authors and not necessarily those of the U.S. Census Bureau.

Statistical Methodology (2016) for Voting Rights Act, Section 203 Determinations

Eric Slud, Robert Ashmead, Patrick Joyce, and Tommy Wright
Center for Statistical Research and Methodology
U.S. Census Bureau

December 13, 2018

Contents

Executive Summary

As enacted in 1975 and amended in 1982 and 2006, Section 203(b) of the *Voting Rights Act of 1965* requires that a State or political subdivision in certain circumstances must provide language assistance during elections for groups of citizens who are unable to speak or understand English well enough to participate in the electoral process. These circumstances are defined in Section 203(b) in terms of specific determinations involving the sizes and proportions of designated population subgroups as measured by the most current American Community Survey (ACS) and other census data. Section 203(b) as amended prescribes that the Director of the Census Bureau shall make these determinations every 5 years, based on population estimates derived from the ACS, and allows for the use of other relevant census data. The 2016 determinations released on December 6, 2016 (https://www.census.gov/rdo/pdf/1_FRN_2016-28969.pdf) are based on 2010-2014 5-year ACS data. Although 2010 decennial Census data played a prominent role in the statistical methodology used to produce the 2011 Section 203 determinations, those data are now considered too far out of date for the 2016 determinations.

For Section 203(b), only the voting-age population (18 or over) is relevant. The law categorizes voting-age persons according to citizenship, limited English-proficiency (LEP), and illiteracy. For present purposes, the binary classifications by voting-age, citizenship, and illiteracy are each defined by the answer to a single ACS question, and LEP is defined through the answers to two ACS questions. In addition, respondents to the Census or ACS self-identify as belonging to up to 8 racial groups and 1 ethnic classification which are then used to define 'Language Minority Groups' (LMGs) for purposes of Section 203(b). These LMGs are defined by the law under broad groupings of Asian, American Indian, and Spanish languages. They are further refined by the census detailed citizenship and language proficiency categories within each of these groups. For the 2016 determinations, 68 LMGs were defined: 16 Asian, 51 American Indian or Alaska Native (AIAN), and one Hispanic LMG. Each ACS individual can contribute to the population estimates for all self-identified LMGs. (Among all sampled persons in at least one LMG, approximately 3% declare themselves in two or more LMGs.)

The nation is partitioned into roughly 8000 Jurisdictions (counties or minor civil divisions) which are the basis for Section 203(b) coverage determinations. There were 7,862 Jurisdictions in ACS 2010-2014 5-year data containing at least one voting-age respondent. Further, American Indian Areas (AIAs) are relevant to the Section 203(b) determinations. In the 2010-2015 ACS 5-year data, there are sampled persons residing in 568 unique AIAs, and those sampled persons reside in a total of 508 distinct Jurisdictions.

Section 203(b) prescribes generally that states and Jurisdictions are required to provide language assistance to voters in a language other than English for members of a LMG according to the following rules:

(i) A state must provide language assistance to voters for a LMG if the illiteracy rate among LEP citizen voting-age members of the LMG in the state exceeds the national rate of illiteracy among citizens, **and** the number of LEP voting-age citizens in the LMG is greater than 5% of the total number of voting-age citizens in the state

(ii) A Jurisdiction (county or MCD) must provide language assistance to voters for a LMG if the illiteracy rate among LEP voting-age citizens in the LMG and Jurisdiction exceeds the national rate of citizen illiteracy **and** the number of LEP voting-age citizens in the Jurisdiction and LMG is greater than either 10,000 or 5% of the total voting-age citizen population of the Jurisdiction.

(iii) All Jurisdictions containing any part of an American Indian Area (AIA) must provide language assistance to voters for a AIAN LMG if the illiteracy rate among LEP citizen voting-age AIAN persons of the LMG in the AIA exceeds the national rate of citizen illiteracy **and** the number of LEP voting-age citizens in the AIA and LMG is greater than 5% of the total voting-age citizen AIAN population of the AIA.

Special tabulations of weighted survey estimates of state, Jurisdiction, and AIA voting-age populations cross-classified by citizenship, limited English proficiency, illiteracy, and LMG are available from American Community Survey 5-year data, and could be used to create direct estimates of all of the ingredients of the criteria (i)-(iii) for determinations. However, the counts of ACS-sampled voting-age persons by Jurisdiction and LMG on which these weighted sums would be based are often quite small, so the variability (standard errors) of the direct estimates are often large compared to the estimates themselves. Moreover, the standard errors estimated by current ACS methodology are also very unreliable for population counts in small domains, the intersection of one Jurisdiction (or AIA) and LMG.

Building on the experience with statistical modeling gained in producing the 2011 determinations, development of statistical estimation methodology for the 2016 determinations has focused on model-based, *small-area estimation* techniques. Small-area estimation is a growing statistical subdiscipline devoted to enhancing the precision of estimation through the formulation of models for multiple small areas which 'borrow strength' from one another through shared statistical parameters. The main idea behind this approach is that many small domains within the same LMG may behave similarly with respect to domain proportions of citizens within the voting-age population or with respect to domain proportions of LEP persons among citizens, across Jurisdictions for fixed LMGs, and these proportions may also exhibit similar relationships with observable domain-specific variables.

The form of model chosen for the 2016 statistical estimation is a Dirichlet-multinomial model, a random-effects generalization of logistic regression models for the incidence of citizenship among voting-age persons within a domain, and for the incidence of LEP among voting-age citizens within the domain. Under this model, the characteristics of the voting-age ACS-sampled persons within each domain are viewed as the outcomes of independent multinomial trials, with each voting-age person in the domain viewed as falling into one of four mutually exclusive categories of non-citizen, LEP illiterate citizen, LEP literate citizen, or non-LEP citizen, randomly and independently of all other persons once covariates and domain-level random effects are taken into account. The underlying predictable parts of rates of citizenship and of LEP among citizens at domain level, are each modeled within LMG across Jurisdiction as logistic regressions in terms of 'synthetic' covariates consisting of the corresponding ACS directly estimated rates at the level of the state containing the domain, along with other domain-level covariates such as the proportions of college-educated, of foreign-born, and of foreign-language speakers. The outcome data used to fit this model are the directly estimated ACS domain-level citizenship, LEP, and illiterate proportions. The parameters of these multinomial trials are shared across Jurisdictions within each fixed LMG. This 'empirical Bayes' statistical framework results in estimators of citizenship, LEP, and illiterate proportions that are weighted combinations of the direct ACS survey-weighted ratio estimators and model-based parameter estimators for the corresponding quantities; and these weights heavily favor the direct estimators in large Jurisdictions.

The choices of models and predictive covariates have been assessed primarily using earlier (2008-2012) ACS data, by comparison and goodness-of-fit diagnostics against the direct ACS survey-weighted domain population estimators and also against alternative models. Among models considered, the Dirichlet-multinomial models combine the virtues of simplicity and of stability by comparison with the direct estimators. The models are fitted separately on data for each LMG, and separately for Jurisdictions and for AIAs. The models were generally chosen to be similar across LMGs, but more or fewer covariates were used according to how numerous were the Jurisdictions (or AIAs) with at least a designated minimum sample size and to how detailed a model could be fitted with numerical stability.

Beyond developing the method of point-estimation used in the 2016 determinations, this report also summarizes a methodology for estimation of the model-based mean-squared prediction errors (MSPEs), which should be compared to the unbiased direct variances. The variance estimation method, newly developed at the Census Bureau for this purpose, is a combination of Monte Carlo parametric-bootstrap re-sampling of data from the model with the replicate-weight (Successive Difference Replication) methodology regularly used in ACS data releases. The latter aspect of variance estimation is relevant in our models because the numbers of voting-age citizens within each domain are estimated directly using survey weights, with the model reflecting only the proportions of citizens, LEP and illiterate persons among the voting-age persons in the domain.

The MSPEs of the model-based estimates are compared in detail with the variances of direct survey estimates that might have been used, in order to clarify the extent to which the small-area estimation techniques have improved the precision of population estimates underlying the 2016 *Voting Rights Act Section* 203(b) coverage determinations. Overall, model-based MSPEs for LEP proportions are typically at least 25% smaller than the corresponding direct-method variances. For a small proportion of domains, MSPEs were somewhat larger than the direct-method variances. These tended to be domains with characteristics combining large sample sizes, small point estimates and small direct-method variances. For the domains of greatest relevance to the coverage determinations, those with LEP within-domain proportions near 0.05, the model-based MSPEs are found to be much smaller than the direct-method variances.

Acknowledgments

The research and implementation summarized in this Report were undertaken under the general guidance of a Technical Steering Committee including the authors, Thomas Louis as Technical Lead, and Mark Asiala and John Abowd. A broader group including that Technical Steering Committee, James Whitehorne, Colleen Joyce, Melissa Creech, David Raglin, and Patrick Cantwell, oversaw the legal and logistical aspects of the project.

1 Introduction

According to Section 203(b) of the *Voting Rights Act of 1965* beginning in 1975, as later amended in 1982 and 2006, States or political subdivisions must in certain circumstances provide language assistance during elections for groups of citizens who are unable to speak or understand English well enough to participate in the electoral process. Section 203(b) prescribes these circumstances [Appendix A] in terms of specific determinations, made by the Director of the Census Bureau, involving the sizes and proportions of designated population subgroups as measured by the decennial census and the most current available American Community Survey (ACS).

In 2016, the Director of the Census Bureau made coverage determinations for 68 specified racial/ethnic Language Minority Groups (LMGs) within roughly 8000 Jurisdictions [`https://www.census.gov/rdo/pdf/1_FRN_2016-28969.pdf`]. The Jurisdictions constitute an electorally relevant partition of the nation into counties and minor civil divisions. There were 7,862 Jurisdictions in ACS 2010-2014 5-year data containing at least one voting-age respondent. A coverage determination refers to a specific Jurisdiction-LMG pair, and multiple LMGs may be covered within a single Jurisdiction. For all Jurisdiction-LMG pairs there are at least two possible ways to be covered by Section 203. First, a LMG may meet the state-level coverage criteria, in which case all Jurisdictions in that state are covered for that LMG. Second, a LMG may meet the Jurisdiction-level coverage criteria, resulting in coverage of that specific Jurisdiction-LMG pair. Lastly, American Indian and Alaska Native (AIAN) LMGs can meet American Indian Area-level (AIA-level) coverage criteria. If an AIAN LMG meets the AIA-level coverage criteria for a certain AIA, then all Jurisdictions that contain all or part of the AIA are covered for that LMG. Specifically, the coverage criteria are:

Criteria for state-level coverage for a particular LMG:

S1 The proportion of limited English-proficient voting-age citizens in the LMG among all voting-age citizens in the state is greater than 5 percent; and

S2 The illiteracy rate among limited English-proficient voting-age citizens in the LMG in the state is greater than the national illiteracy rate.

Criteria for Jurisdiction-level coverage for a particular LMG:

J1 (a) The proportion of limited English-proficient voting-age citizens in the LMG among all voting-age citizens in the Jurisdiction (`LEPpropJur`) is greater than 5 percent; or

(b) The number of limited English-proficient citizens that are members of the LMG (`LEPtotJur`) is greater than 10,000; and

J2 The illiteracy rate among limited English-proficient voting-age citizens of that LMG in the Jurisdiction (`ILLrateJur`) is greater than the national illiteracy rate.

Criteria for AIA-level coverage for a particular AIAN LMG:

A1 The proportion of limited English-proficient voting-age citizens in the LMG among all AIAN voting-age citizens in the AIA (`LEPpropAIA`) is greater than 5 percent; and

A2 The illiteracy rate among limited English-proficient voting-age citizens of that LMG in the AIA (`ILLrateAIA`) is greater than the national illiteracy rate.

Several types of quantities are needed to form the totals and proportions used to evaluate coverage criteria at the State, Jurisdiction, and AIA levels. Specifically, the total numbers of limited English-proficient voting-age citizens and illiterate limited English-proficient voting-age citizens are required for each LMG within each State and Jurisdiction and for each AIAN LMG within each AIA. In addition, we use the total numbers of voting-age citizens within each State and Jurisdiction and of voting-age AIAN citizens within each AIA. These quantities can be estimated directly from ACS data; however, the precision of some of the estimates, especially for the limited English-proficiency proportions and illiteracy rates, can be quite poor because many of the domains are extremely small. The national illiteracy rate is computed as the number of illiterate voting-age citizens divided by the total voting-age citizens. The rate used for the 2016 coverage determinations based on ACS 2014 5-year data was 1.31%.

In an effort to improve the precision and stability of the estimates used to make the coverage determinations, the Census Bureau decided to utilize a model-based estimation method for the 2016 coverage determinations. A model-based method was also used in the 2011 Section 203(b) determinations [Joyce et al., 2012, 2014]. The basic rationale behind model-based small-domain estimation methods is that many small areas may be similar according to measured characteristics, and viewed as differing through independent random 'small domain effects'. Modeling with shared statistical parameters may allow those parameters to be estimated with an increased precision not possible for one or a few small domains. This phenomenon of gaining precision of estimation through shared parameters is often called 'borrowing strength' and is the essence of a growing statistical subdiscipline called *small-area estimation* [Rao and Molina, 2015]. The greatest gains in precision of estimation through small-area methods arise when useful predictive covariate measurements are available at the small-domain level for inclusion in regression-type models. Those aspects of small-domain differences not predictable through the 'fixed effect' covariates are modeled through independent 'random effects' from a distribution of an assumed form. The remaining sections of this report describe the rationale, model, details of implementation and assessments for the model-based method used to derive the estimates for the 2016 Section 203(b) determinations.

2 Terminology and Data

The sources of data allowed by law to be used in the coverage determinations are the ACS and comparable Census data. The ACS is an ongoing annual household survey that records information about the nation's people which is used in many different ways. The ACS releases 1-year as well as

5-year data products. The 5-year products aggregate the ACS data collected over a 5-year period, allowing increased precision of population estimates at the cost of temporal specificity. The 5-year data are particularly useful for estimating features of small geographic areas or small subgroup domains where the precision of the 1-year estimates is too poor to allow their release under Census Bureau statistical quality guidelines. For the purposes of the coverage determinations, 5-year ACS data are used specifically because of the need to estimate population subgroups in small geographic areas. At the time of estimating the models used to make the 2016 coverage determinations, the 2014 5-year ACS was the most recent dataset available and therefore was used as the data source. However, model exploration and development were done using only the 2012 ACS 5-year data, before the 2014 data were available. The use of 2010 Decennial Census data was explored in the model development process; however, 2010 data were ultimately judged to be too far out of date to provide reliable voting-age person counts for each LMG, marking a major difference from the 2011 methodology which did use 2010 Decennial Census data in producing the 2011 coverage determinations.

The Section 203(b) relevant political subdivisions, which we refer to as Jurisdictions (Jur), are Counties in most states and Minor Civil Divisions (MCDs) in eight states (CT, ME, MA, MI, NH, RI, VT, WI). In the 5-year ACS 2010-2014 data, there were 7862 Jurisdictions containing at least one sampled voting-age person, and 568 AIAs containing at least one sampled voting-age person. AIAs may intersect with multiple Jurisdictions, and single Jurisdictions may contain all or part of multiple AIAs.

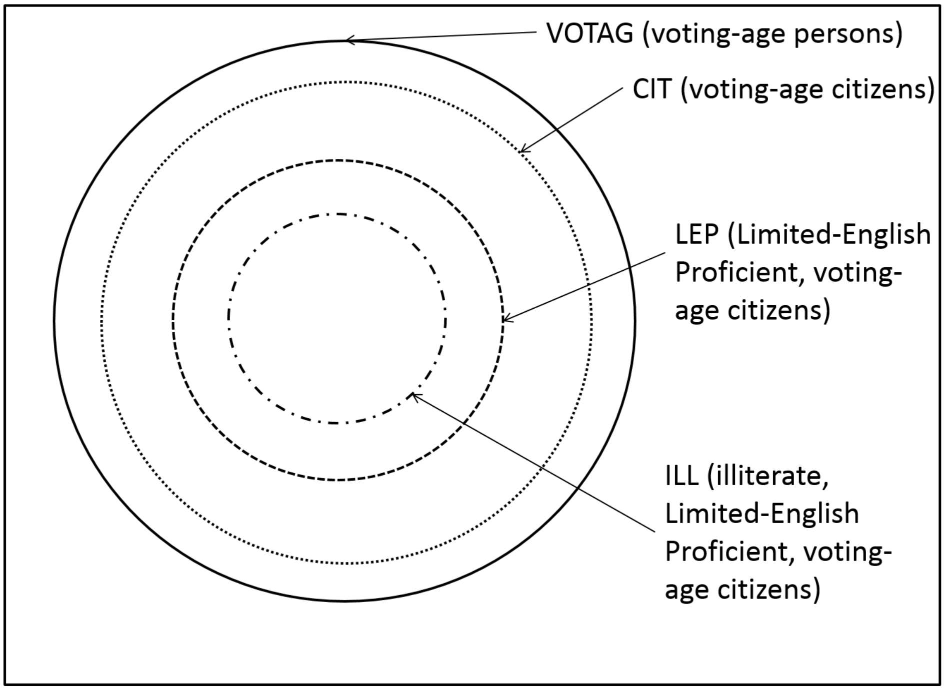
Limited English-proficiency and illiteracy indicators are derived from ACS questions. For purposes of coverage determinations, limited English-proficiency is defined as speaking a language other than English at home and speaking English "Less than Very Well". Illiteracy is defined as having less than a 5th grade education. The subgroups needed to estimate the coverage determination quantities are defined through intersections of the properties of voting-age, citizenship, limited English-proficiency, and illiteracy. Throughout this report, we refer to the relevant population subgroups by the following abbreviations:

VOTAG: Voting-age persons;
CIT: Voting-age citizens;
LEP: Limited English-proficient, voting-age citizens; and
ILL: Illiterate, limited English-proficient, voting-age citizens.

It is important to note the nested relationship among VOTAG, CIT, LEP, and ILL in this report, displayed graphically in Figure 1. Specifically, all ILL persons are LEP, all LEP persons are CIT, and all CIT persons are VOTAG. Think of these as population subgroups for a specific LMG and at a specific geographic level (state, Jurisdiction, or AIA).

In the Decennial Census and the ACS, persons self-identify into one or more racial and ethnic groups in response to race and ethnicity questions. Certain of these groups are called Language Minority Groups (LMGs) for purposes of Section 203(b); all are listed in Appendix B. There are 16 LMGs within the Asian racial group, another 51 LMGs within the AIAN racial group, and a

Figure 1: The relationship between VOTAG, CIT, LEP and ILL subgroups.



single Hispanic LMG that cuts across racial groups. People who self-identify into more than one racial/ethnic group can therefore belong to more than one LMG, although only a small proportion do, about 3% out of all LMG persons in the 2010-2014 ACS data. Table 1 displays survey-weighted estimated frequencies of multiple LMGs in these data. Coverage determinations are made separately for each LMG, so that people belonging to multiple LMGs count towards the coverage criteria for all of them.

Table 1: Frequency distribution of multiple LMG group self-identification, out of 20,338,189 voting-age persons in ACS 5-year (2010-2014) data.

LMGs claimed	0	1	2	3	4	5	6	7	8
# of persons	16,778,586	3,448,766	104,233	5,841	541	141	65	14	2
Percent of total	82.50	16.96	0.51	0.029	.003	.001	0	0	0

3 Model Rationales and Descriptions

The quantities that we tabulated from the ACS files for use in the small area models and estimates necessary for the coverage determinations were:

(a) unweighted numbers of ACS-sampled VOTAG, CIT, LEP, and ILL persons in each (Jur, LMG);

(b) survey-weighted ACS estimates of total numbers of VOTAG, CIT, LEP, and ILL persons in each (Jur, LMG);

(c) unweighted numbers of ACS-sampled VOTAG, CIT, LEP, and ILL persons in each (AIA, AIAN LMG);

(d) survey-weighted ACS estimates of total numbers of VOTAG, CIT, LEP, and ILL persons in each (AIA, AIAN LMG);

(e) survey-weighted ACS estimates of total numbers of VOTAG and CIT persons in each (Jur);

(f) survey-weighted ACS estimates of total numbers of AIAN VOTAG and CIT persons in each (AIA); and

(g) covariates as detailed further in Section 3.1

The direct ACS survey-weighted estimators, (b) and (d), of the domain total numbers of LEP and ILL persons can be used in conjunction with (e) and (f) to estimate the coverage determination quantities. If these estimators were stable, they would be the design-based estimators of choice; however, many of these survey-weighted total estimators are based on extremely small sample sizes, and would yield estimates with large standard errors.

The coefficient of variation (CV) of a point estimate is defined as its standard error divided by the estimate, and measures its relative precision. The Census Bureau requires that the CV for a majority of the ACS estimates in each published table must be less than or equal to 0.30 for a survey to meet the Census Bureau's statistical quality standard for sampling error; and estimates with CVs greater than 0.61 are said to be unreliable. Table 2 shows that a majority of CVs for the ACS estimated total number of CITs in single (Jur, LMG) domains are quite large, whereas the CVs for the total estimated citizens in Jurisdictions are mostly small. As a result, many of the direct survey-weighted estimates using the total CITs in each (Jur, LMG) or similarly the total LEPs in the (Jur, LMG), such as those used in the calculations for estimates (b) above, will be unreliable. By contrast, the direct survey-weighted estimates of the total CITs in the Jurisdictions are typically precise, giving us confidence to use them directly in the calculations.

Figure 2 is based on 2014 ACS 5-year data and shows the CVs, point estimates, and 90% confidence intervals for direct survey-weighted (Jur, LMG) domain estimates of `LEPpropJur`, for

Table 2: Summary of coefficients of variation for ACS 2014 5-year direct survey-weighted estimates of total CITs in domains with estimated VOTAG persons ≥ 50 and estimated CITs > 0, where "Qu.xx" denotes the xx percentile.

Estimate	Min.	Qu.05	Qu.25	Median	Qu.75	Qu.95	Max.
CITs in (Jur, LMG)	0.002	0.062	0.220	0.380	0.549	0.834	4.995
CITs in Jurisdiction	0.000	0.002	0.005	0.024	0.069	0.121	0.618

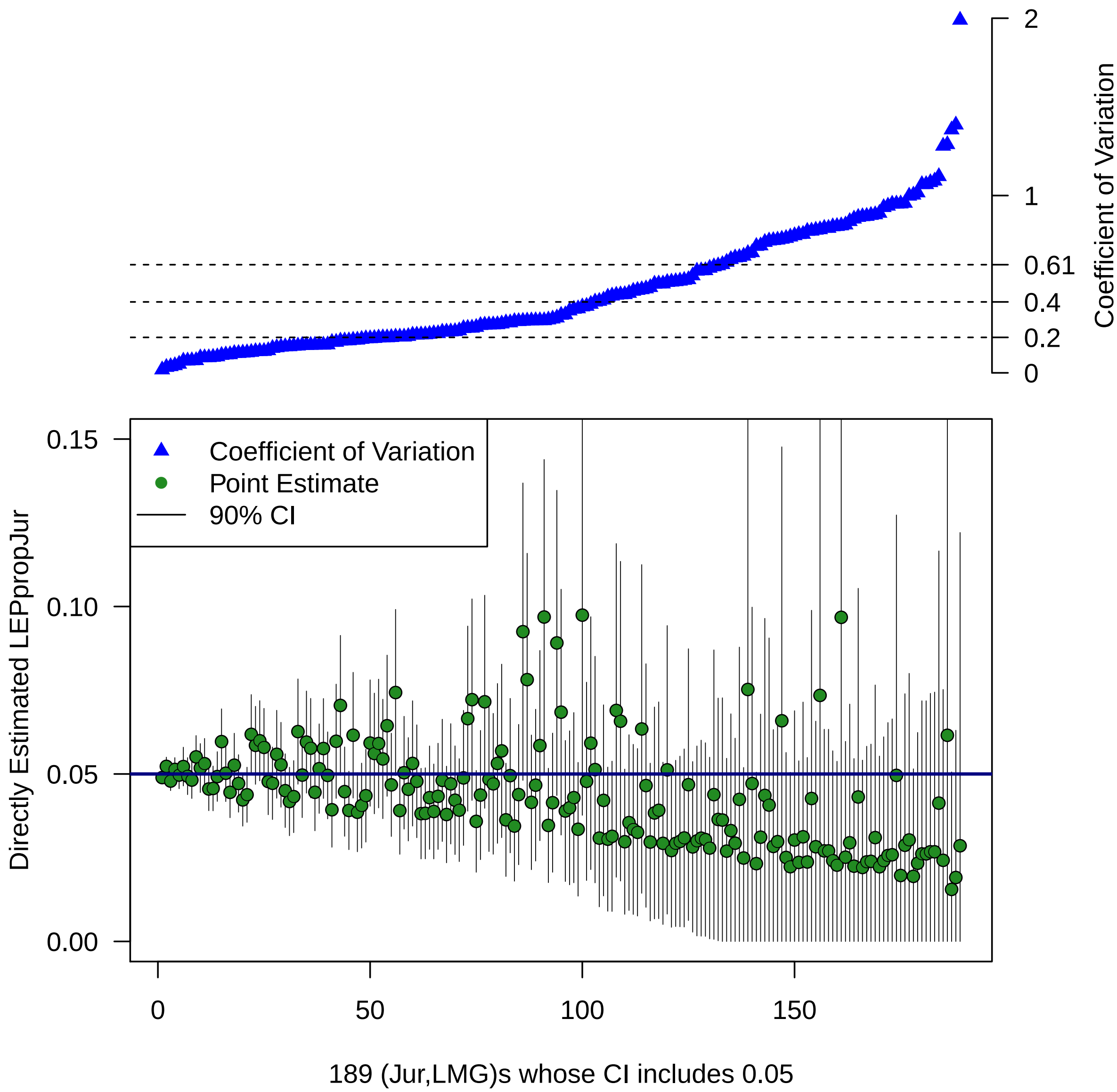
the subset of 189 domains whose 90% confidence interval includes 0.05 (the quantity and threshold for meeting determination criteria J1). Because the 90% confidence intervals for these estimates include values that meet the determination criterion (> 0.05) and other values that do not meet it, the decision for these (Jur, LMG) domains is unclear yet particularly important. Additionally, many of the 90% confidence intervals for these estimates are extremely wide, which is also reflected in the large coefficients of variation. Thus, the precision is undesirably low for many of the 189 direct survey-weighted (Jur, LMG) estimates near 0.05.

To mitigate this kind of small-domain imprecision, we have developed model-based estimators for the target domain totals in the spirit of small-area estimation [Rao and Molina, 2015]. The main idea of this approach is that many small (Jur, LMG) or (AIA, LMG) domains within the same LMG may behave similarly with respect to domain proportions of citizenship among the VOTAG population, of limited English proficiency among voting-age citizens, and of illiteracy among limited English-proficient voting-age citizens. Additionally, this similarity may be exhibited in the form of shared relationships between the outcome proportions and observable domain-specific covariates.

The unifying idea in all the models we considered was that the outcomes follow a Generalized Linear Mixed Model [Breslow and Clayton, 1993] with form and parameters shared across domains within LMG. The models use independent random effects to account for domain differences within LMG, and they are fitted separately for each LMG. The final estimates were constructed within an empirical-Bayes framework that ensured that the estimates are weighted combinations of the direct ACS survey-weighted ratio estimates and model-based estimates with substituted parameter estimates. The empirical-Bayes estimators have the feature that in areas where the direct estimate is relatively precise, the final estimate will tend to agree with the direct estimate. Alternatively, the empirical-Bayes estimator will more heavily favor the model-based estimate for areas where the direct estimate has a large standard error.

The outcomes of interest in the models, LEP and ILL totals, are not estimated directly. Instead, in each multinomial model for a particular LMG we estimate the proportion of VOTAG persons in each of four disjoint categories: not CIT, CIT but not LEP, LEP but not ILL, and ILL (depicted as successive annuli and central disk in Figure 1). These estimated proportions can then be combined with the direct estimate of the total number of VOTAG persons in the (LMG, Jur) to get an estimate of the LEP and ILL totals. Separate models were estimated for each LMG across Jurisdictions as

Figure 2: Coefficients of variation, direct estimates of LEP proportions `LEPpropJur`, and 90% confidence intervals for 2014 ACS 5-year estimates whose CI includes 0.05 .



well as across AIAs for each AIAN LMG. Guided by the principle of analyzing each LMG separately, each model was fitted from data on all persons within the LMG regardless of membership in other LMGs. As a result, one person's data may contribute to multiple LMG models. (See Table 1 for the extent of self-identifications into multiple LMGs.) This differs from the 'local majority' model strategy used in 2011 in which each person was assigned the unique largest local LMG among that person's self-identified LMGs. The only way in which models fitted to different LMGs influence one another is that the subset of fixed-effect covariates used for an LMG was chosen from a list of possible predictors according to a grouping of LMGs with similar numbers of Jurisdictions or AIAs containing ACS samples of similar sizes.

3.1 Covariates

Predictive covariates considered for use in our models consisted of population rates directly estimated from the ACS data at different levels of aggregation (using survey weights) related to citizenship, English proficiency, race/ethnicity, educational level, age, age of AIAN persons, and foreign birth, as well as average time in the United States. We explored three levels of aggregation for ACS covariates to be included in the models. At the highest level are covariates calculated by aggregating all domains within an LMG in each state. At the second level are covariates for the specific (Jur or AIA) geography across all persons regardless of LMG. The third covariate type is computed for specific domains (Jur or AIA coupled with the single LMG). We considered the following covariates, displayed by type:

State-Level Covariates

C1 Logit-transformed proportion of citizenship among voting-age persons by LMG

C2 Logit-transformed proportion of limited English-proficiency among citizens by LMG

Geography Specific Covariates

C3 Proportion of non-Hispanic White or non-Hispanic Black African-American persons among voting-age persons in Jur or AIA

C4 Proportion of no-college education among voting-age persons in Jur or AIA

C5 Average person count per housing unit among voting-age persons in Jur or AIA

C6 Average age among voting-age persons in any AIAN LMG in Jur or AIA

C7 Logit-transformed fraction of voting-age persons in Jur or AIA with no-college education

C8 Proportion of other language spoken at home among voting-age persons in Jur

C9 Average age among voting-age persons in Jur

C10 Proportion of foreign-born persons among voting-age persons in Jur

C11 Average years in US (as of 2014) among voting-age foreign-born persons in Jur

C12 Logit-transformed proportion of foreign-born persons among voting-age persons in Jur

Geography-LMG Specific Covariates

C13 Logit-transformed proportion of foreign-born persons among voting-age persons in (Jur, LMG)

C14 Average years in US by 2014 among voting-age foreign-born persons in (Jur, LMG)

C15 Logit-transformed proportion of no-college education among voting-age persons in (Jur, LMG)

The *synthetic* covariates C1 and C2, the proportions at the State level, were motivated by our objective to estimate proportions of CIT within VOTAG, LEP within CIT, and ILL within LEP. *Synthetic* survey rate-variables, in frequent Census Bureau and survey-methodology parlance [Rao and Molina, 2015, Sec. 3.2], are those defined from a level of geography higher than the one of primary interest. Such covariates were previously introduced and advocated in a small-area context, for confidence intervals of very small ACS rates by Slud [2012]. Generally, these state-level survey-weighted direct ratio estimators of CIT and LEP rates by LMG are stable but do not reflect local Jurisdiction-level variation of these rates.

The second category, *Geography Specific Covariates*, is the main source of predictive variables for the models we developed. Although we considered other ACS covariates for citizenship and LEP proportions listed in the *Geography-LMG Specific Covariates* category above, only covariate C13 was found to be usefully predictive in models for Asian or Hispanic (not AIAN) LMGs. See Sections F.1 and F.2 for exact information on the covariates used in each LMG model.

Note that all of these covariates are taken from the ACS itself, meaning that each is a survey estimate and thus subject to sampling error. Further, because both the covariates and outcomes are from the same survey data, their sampling errors may be correlated, which could complicate variance estimation. For the most part this is not an issue for the State- or geography-specific covariates because these are generally based on much larger samples than the geography-by-LMG specific outcomes. We do not here consider the effect of errors in variables along the lines of Ybarra and Lohr [2008], but may do so in future research.

3.2 Dirichlet-Multinomial Model

After extensive model exploration, *Dirichlet-multinomial* was chosen as the basic model form for all of the LMG-specific Jurisdiction models and the separate AIAN LMG-specific AIA models. In each model, the outcomes CIT, LEP, and ILL are viewed as combinations of four mutually exclusive categorical outcomes from independent multinomial trials after conditioning on given random-effect parameters. The underlying expected proportions of CIT among VOTAG, LEP among CIT, and

ILL among LEP are each parameterized as a logistic regression in terms of specified covariates, or in the case where no covariates are used, as a constant. The data used to fit the models were proportions obtained from direct survey-weighted ratio estimators along with sampled voting-age person counts. The exact form of the resulting Dirichlet-multinomial model with logistic-regression fixed effects can be seen in Appendix D along with the prediction formulas.

3.3 Model Fitting and Model Classes

The 68 LMGs vary widely on a number of characteristics relevant to the statistical model. The estimated proportions of LMG national populations that are respectively citizen, LEP, and illiterate vary widely (Table 3). In some AIAN LMGs, all or almost all sampled voting-age persons are citizens and literate. The models suitable for LMGs that fell into these extreme cases were necessarily very simple and special.

Table 3: Quantiles across LMGs of national estimates of nested CIT, LEP, and ILL proportions.

Direct Ratio Estimates	Min	Q1	Median	Q3	Max
CIT proportion among VOTAG	0.42	0.80	0.99	1.0	1.0
LEP proportion among CIT	0.00	0.01	0.02	0.19	0.52
ILL proportion among LEP	0.00	0.02	0.07	0.11	0.37

Next, LMGs vary in their size, both in the total ACS sample size nationally (approximately proportional to the population size) and in the number of areas (Jurisdictions or AIAs) where at least one person is sampled. The number of areas with positive sample for a LMG is especially important because the small area models we use are *area-level* (as opposed to unit-level), meaning the number of observations is equal to the number of areas included in the model. The number of observations has a crucial bearing upon the number of covariates that should be used in the model. The number of parameters (covariates) to use in specifying statistical models should generally increase at most proportionately and usually at a slower rate as a function of the number of observations. As a result, in some of the smallest LMGs in the data, the small number of observations made it suitable to use no covariates if no very strong predictor was available. See Table 4 for a summary across the 68 LMGs of the number of Jurisdictions and AIAs with LMG voting-age sample thresholds. The table shows that among the 68 LMGs, the median number of Jurisdictions with at least 5 sampled VOTAG persons was 83. The minimum amongst the LMGs was 8 Jurisdictions, and the maximum was 4640 Jurisdictions.

A third type of variation across LMGs was in the predictiveness of the covariates for LMG rates of citizenship and LEP. Therefore, instead of a single model used for all LMGs and geographies (Jurisdictions and AIAs), we developed classes of models depending on the characteristics of the LMG. Then, inside those classes we created levels of models depending on the number of observations and the ability of the iterative algorithms used in fitting maximum likelihood (ML) estimators to the model to meet certain convergence criteria. For full details see Appendix F.

Table 4: Summary across 68 LMGs of the number of Jurisdictions and AIAs with LMG voting-age sample thresholds, based on 5-year ACS 2010-2014 data.

	Summary across LMGs				
	Min	Q1	Median	Q3	Max
Jurisdictions with					
>0 sampled LMG VOTAG	66	265.2	554	1561	6837
≥ 3 sampled LMG VOTAG	15	64	156	605	5460
≥ 5 sampled LMG VOTAG	8	35.75	83	388.5	4640
	Summary across 51 AIAN LMGs				
	Min	Q1	Median	Q3	Max
AIAs with					
>0 sampled LMG VOTAG	4	33	55	116	349
≥ 3 sampled LMG VOTAG	1	11.5	20	54	272
≥ 5 sampled LMG VOTAG	0	7	16	36.5	241

3.4 Model Exploration

While each of the outcomes CIT, LEP, and ILL could have been fitted to separate models within each LMG, and this was done in early exploratory model-fitting, it was judged to make more sense to account for the nesting of these categories by fitting a single model using all three outcomes simultaneously.

All model exploration was done using 2012 ACS 5-year data instead of the 2014 5-year data on which the final models, estimates, and determinations were made. The three main areas of model exploration were 1) the general form of the model, 2) the specific covariates, and 3) the use of LMG numbers of Jurisdiction-level observations to result in different model classes defined by degree of model complexity. Early in the process, several general model forms were explored and compared to one another, in terms of AIC or BIC and descriptive plots of outcome proportions versus fixed-effect model-based predictions, as well as indications of proper convergence of ML estimators. Besides the Dirichlet-multinomial model, a hierarchical Beta-binomial model similar to that used in the 2011 determinations [Joyce et al., 2012, 2014] was considered. The Dirichlet-multinomial model was ultimately chosen because it was found to be easier to interpret and fit the data better in early comparisons. Within the general Dirichlet-multinomial model framework, several variations were explored. These included disaggregated models, in which models or model parameters would be estimated separately for groups of areas defined by size (small/large) or type (county/MCD). Any gains in fit from such split models turned out to be more than offset by the decrease in the ability to "borrow strength" from medium or large areas in order to estimate the attributes of smaller areas. Another variant explored was to model the ILL proportion among LEP persons separately from (CIT, LEP) within VOTAG. It was decided that this approach would have added unnecessary complexity to the process.

Two major changes in the modeling came as a result of model exploration. The first was the decision to create models for AIAN LMGs of AIAs separate from the models for those same LMGs within Jurisdictions. In the 2011 determinations [Joyce et al., 2012, 2014], model parameters concerning the Jurisdictions encompassing the AIAs were used to estimate the relevant quantities within the AIAs. However, descriptive statistics showed that in some AIAN LMGs the CIT, LEP, and ILL proportions differed markedly when calculated amongst all LMG persons from those calculated amongst only LMG persons in AIAs. Therefore, we changed our strategy to model AIAN LMGs in AIAs separately, using data only from persons living on AIAs. The second major change arose in specifying the way in which the random Jurisdiction-to-Jurisdiction variability of CIT, LEP and ILL proportions depends on sample sizes within those Jurisdictions. This can be regarded as the variability of the random-effect part of the occurrence rates within domains. In typical formulations of Dirichlet-multinomial models, the distribution of such random effects would not vary with the size of the domains, but we found improved model fit by allowing the variance of these random effects to vary inversely with the square root of sample size in each domain. (See Appendix D for additional information on this parameterization.) This change resulted from examining model diagnostics. When considering the disaggregated models mentioned above, we noticed that the concentration (inverse-variance) parameter for random effects was estimated to be much larger in larger populated areas. This pattern held across different LMGs and eventually led us to the square-root of sample size parameterization of concentration.

To better illustrate the comparison of models, in Table 5 we show, for selected Asian and AIAN LMGs, the Bayesian Information Criterion (BIC) and Akaike Information Criterion (AIC) for three different Jurisdiction model specifications. BIC and AIC are common metrics used in model selection and are based upon the maximized log-likelihood and model complexity. For both of these metrics smaller values are preferred. In our examination of the BIC and AIC of various models we searched for models that had sizable differences in BIC or AIC, not just those that were statistically significantly different. In Table 5, model type 1 uses a constant random effect parameterization and no covariates, model type 2 uses a constant random effect size specification and four total covariates, and model type 3 uses the square root of the sample size random effect specification and four covariates. This table shows that overall, the covariates are useful (model type 1 versus model type 2) and that the square root of sample size random effect parametrization is preferred to that of the constant parametrization (model type 3 versus model type 2). Similar comparisons, along with model diagnostics such as those in Section 5 applied to the 2012 data were used to choose the models ultimately used for the predictions.

Table 5: BIC and AIC comparison for Jurisdiction models under three different model specifications, based on 2012 ACS 5-year data, for selected LMGs.

LMG	Model Type	BIC	AIC
ASIAN_1	1	13454.9	13439.4
	2	13348.4	13312.3
	3	13079.5	13043.4
ASIAN_2	1	1351.4	1342.5
	2	1315.4	1294.5
	3	1304.6	1283.7
ASIAN_3	1	14697.2	14681.6
	2	14403.2	14366.8
	3	14226.6	14190.2
ASIAN_4	1	8897.4	8883.2
	2	8819.5	8786.3
	3	8606.8	8573.6
AIAN_1	1	2383.7	2367.2
	2	2369.2	2330.7
	3	2319.4	2280.9
AIAN_2	1	2277.5	2266.7
	2	2133.6	2108.4
	3	2085.6	2060.4
AIAN_3	1	1922.3	1908.0
	2	1880.4	1846.9
	3	1830.0	1796.5
AIAN_4	1	8768.2	8751.6
	2	8624.4	8585.8
	3	8414.7	8376.1

4 Variance and MSPE Estimation

Because the justification of model-based prediction of domain proportions in the VRA context is based largely on the reduced variability of those predictions by comparison with direct survey-weighted estimation, it was also necessary to develop methodology for estimation of expected squares of model-based mean-squared prediction errors for comparison with the variances of unbiased direct-method estimators. It is important to establish and fix terminology first. We generally refer to the statistical approximation of unknown constants (parameters and deterministic functions of them) as *estimation*. This standard terminology applies also to direct-method (approximately) unbiased estimation of domain totals and proportions in a finite-population context.

In general, mean-squared estimation errors are decomposed algebraically into estimation variance added to the square of estimation bias. When the estimates are unbiased or approximately so, as is true for the direct-method survey estimators, the bias-squared term is negligible and the mean-square discrepancies between the estimates and their targets are variances, and we refer to them in that way. By contrast, within statistical models that posit the unknown domain totals and proportions as random variables incorporating both constant parameters and random effects, statistical approximations of those target quantities as functions of data are called *predictions* and (since they may be biased) their expected squared discrepancies from their targets are called Mean Squared Prediction Errors (MPSEs). Thus in much of what follows, we compare Variances of Direct survey-weighted estimators with MSPEs of model-based predictors. When a single term is needed to refer to both measures, we sometimes use 'Variance Estimation' as in some subsection headings in place of the formally more inclusive 'MSPE Estimation'. In later theoretical sections on this topic such as Appendix G, we will maintain the MSPE terminology.

The small area models specify the proportion of CIT, LEP, and ILL persons within voting-age persons for a given LMG within each geographic area. For estimation purposes, we treat the direct survey-weighted estimates of voting-age LMG persons in each Jurisdiction as the dependent variables in data; however, these quantities are actually estimates themselves and subject to sampling variability. In order to account for both the sampling variance and the model-parameter uncertainty in making our predictions, we developed a novel method for making the estimates of variance from our model estimators.

4.1 ACS Direct Estimate Variance Estimation

The ACS utilizes a method called Successive Difference Replication (SDR) [Wolter, 1984, Fay and Train, 1995] to estimate the variance of direct estimates from the data. SDR is a type of balanced repeated replication (BRR), a more general class of replicate weight methods. In order to estimate the variance of any survey quantity in the ACS, the final survey weights along with eighty replicate weight sets are used to calculate

$$\widehat{V}(\hat{\theta}_0) = \frac{4}{80} \sum_{i=1}^{80} (\hat{\theta}_r - \hat{\theta}_0)^2, \tag{1}$$

where $\hat{\theta}_0$ is the direct estimate of the survey quantity using the final survey weights and $\hat{\theta}_r$ is the estimate using replicate weight set r. This standard version of the ACS SDR formula assumes that the sampling fraction is negligible, something that is usually but not always true for the (LMG, Jur) or (LMG, AIA) domains. See ACS [2014] for additional details about ACS variance estimation.

4.2 Parametric Bootstrap Variance Estimation

Parametric bootstrap methods are common in model-based estimation methods such as small-area estimation. The general idea is first to replace unknown model parameters with estimates from the actual data (such as maximum likelihood estimates); then to sample a large number of independent identically distributed data-replicates from the specified model. Next, for each sampled replicate of the data, the model is re-estimated and predictions made based on those model estimates. Finally, a Monte Carlo variance is calculated based on the replicates of the model estimates. Some authors, such as Carlin and Louis [2009, p. 247], suggest that for purposes of variance estimation, it might be better to account for (posterior) variability in the estimated parameters by using posterior-sampled parameter values in the parametric-bootstrap resampling of the data. In other words, a Bayesian point of view suggests incorporating variability of parameter estimates at the stage of generating replicate data samples. See Shao and Tu [2012] for additional background on the parametric bootstrap.

4.3 Hybrid BRR and Bootstrap MSPE Estimation

In order to estimate the variance of our model-based estimates, we combine BRR and parametric bootstrap ideas into a novel hybrid approach to MSPE estimation. The general idea, related to the point raised in the previous paragraph about accounting for parameter variability across bootstrap samples, is that by utilizing different sets of the replicate weights within parametric bootstrap samples, we can account for both sources of variance. Additional technical details on the methodology, including detailed computational formulas, can be found in Appendix G and in Slud and Ashmead [2017]. This is the method by which MSPE estimates for the model-based point estimates were actually calculated in the public release of *Voting Rights Act* estimates.

The mean square prediction error (MSPE) is defined as

$$\text{MSPE}(\tilde{\theta}) = E[(\tilde{\theta} - \theta)^2],$$

where $\tilde{\theta}$ is the model-based prediction of the parameter θ. In Section 6, we compare the MSPE estimates with estimated variances of the survey-weighted direct estimates.

5 Model Diagnostics

An important aspect of our class of small area models is that the prediction for a given area is a weighted average of the direct survey estimate and the regression estimate for that area. See equations (8), (9), and (10) in Section D for full mathematical details. For clarity, we use the term 'Full-Model-Prediction' (FMP) to describe the Dirichlet-multinomial prediction incorporating predicted random effects. The regression estimate uses only the fixed-effect coefficient predictors and coincides with what others call 'synthetic' estimates obtained by expressing the marginal mean as a function of parameters and then substituting the ML estimators of those parameters, so we call it the 'Marginal Mean' (MM) estimate. The target of the MM estimate is the mean outcome averaged over the unobserved random effect under the model. We use the term 'Direct Survey' (DS) in what follows for the standard survey-weighted estimator (of a total or proportion) based on the survey design and data rather than on a model. The relative weight in the FMP weighted-average given to the DS versus the MM estimate for a given geography-LMG domain is determined by the sample size in that domain together with the precision parameter in the model. The weights on these two components, derived from theoretical considerations (*cf.* Appendix D), give greater weight to the DS estimates in domains with larger sample size. Domains with very large sample size, and therefore very precise DS estimates, have FMPs that are weighted almost entirely towards the DS, relying very little on the MM estimate. By contrast, domains with small sample size rely heavily on the MM because their DS estimates are unreliable. For model diagnostics, comparisons between the DS and MM estimates and the FMPs offer valuable insights. We use only a small number among the many models as examples of the model diagnostics, illustrating the common patterns and properties found across the models. While the fitted Dirichlet-multinomial models make predictions for CIT, LEP, and ILL, only the LEP and ILL predictions from the models are used in the coverage determinations. Moreover, since the illiteracy rate criterion for coverage was generally met regardless of the estimated model, we focus our attention on LEP, the most sensitive modeled quantity.

5.1 Residual Analysis

We begin with examples of the relation between the DS and the MM estimates. Figure 3 shows both for estimates of the LEP proportion among VOTAGs each plotted against their respective sample sizes for a particular Asian LMG Jurisdiction model. This figure illustrates a few features of the data and models. First, notice that the variation in the direct survey predictions themselves decreases as the sample size increases. For small sample sizes, there is increased sampling variability and there are many more extreme data points, both low and high. Second, notice that the MM estimates are much less variable than their DS counterparts, and the extreme DS estimates (near 1.0 or 0.0) are the ones from which the corresponding MM estimates are most different.

Figure 4 plots the MM estimate against the DS estimate for both the CIT among VOTAG and LEP among CIT proportions for the 2014 Jurisdiction model for the same medium-sized Asian LMG as in Fig. 3. This figure shows that the covariates are useful in predicting the direct estimates. Specifically with respect to LEP proportions, the figure shows that when DS estimates are extremely large (near 1.0), their MM estimates are smaller, while extremely small DS estimates (near 0.0) are smaller than their corresponding MM estimates. Thus for LEP proportions, differences between MM and DS estimates tend to be larger when the DS estimates are below 0.1 or above 0.4.

Figure 5 shows the residuals between the DS and MM estimates for proportions both of CIT within VOTAG and LEP within CIT. The MM estimates are more accurate in predicting the DS value for areas with larger sample size.

Figure 3: Plots of Direct Survey (DS) and Marginal Mean (MM) estimates of LEP proportion within VOTAG each versus sample size by Jurisdiction, for the 2014 Jurisdiction Model on a medium-sized Asian LMG, restricted to Jurisdictions with sample ≥ 5.

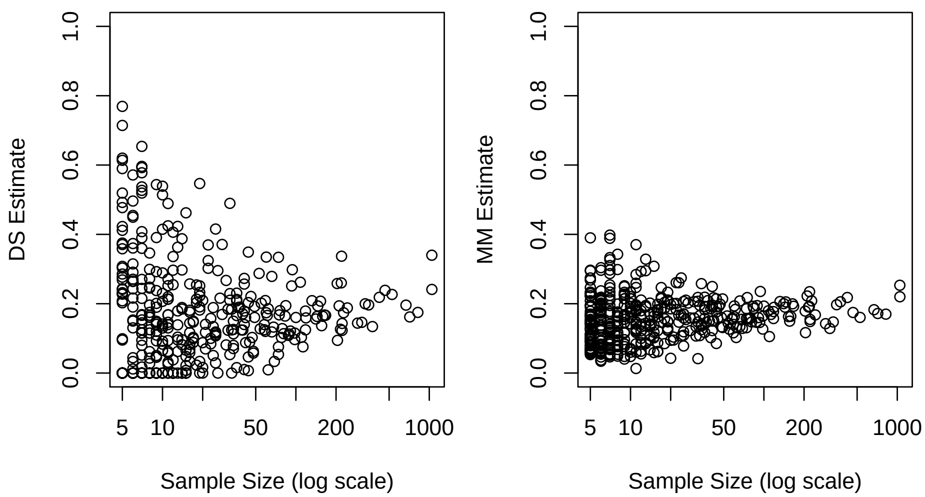


Figure 4: Estimates of proportions of CIT within VOTAG and LEP within CIT, Direct Survey (DS) versus Marginal Mean (MM) estimates for the 2014 Jurisdiction Model on the same Asian LMG and Jurisdictions. (with sample ≥ 5) as in Figure 3.

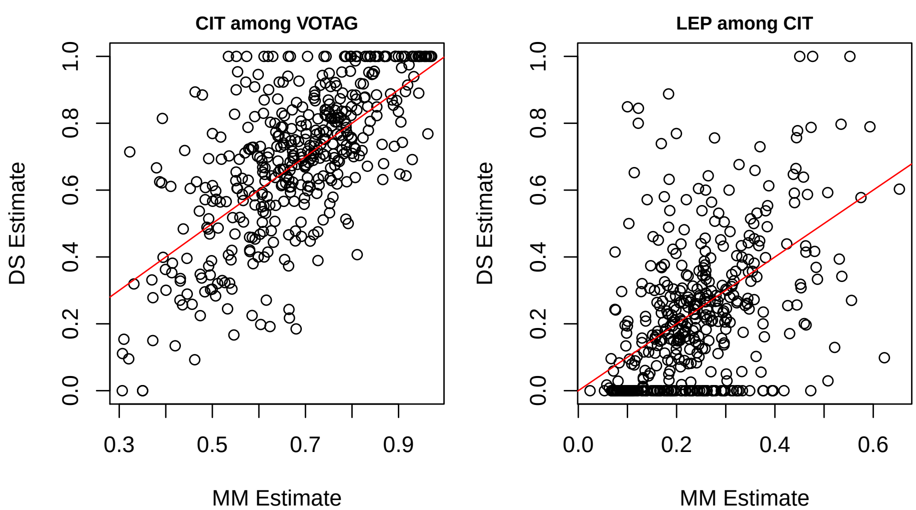
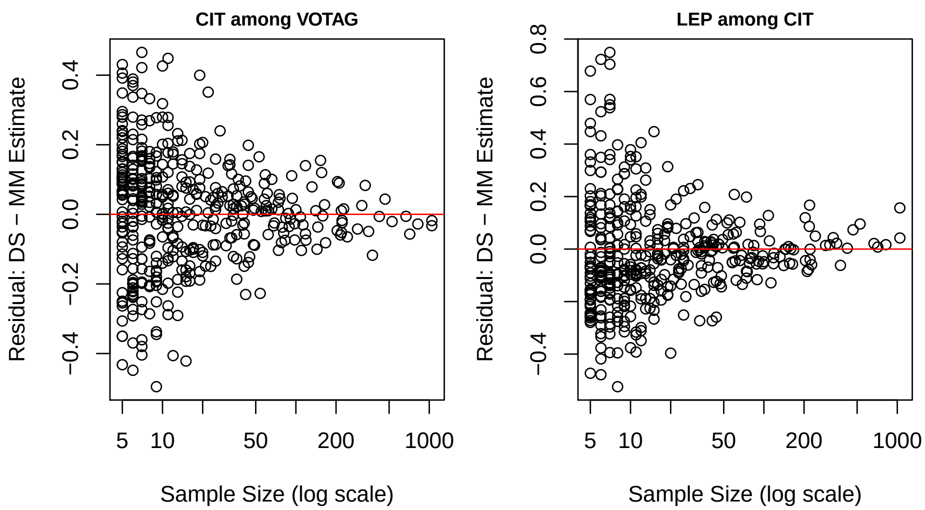


Figure 5: Direct Survey minus Marginal Mean estimate differences for CIT proportions within VOTAG and for LEP within CIT versus sample size, for 2014 Jurisdiction model on the same Asian LMG and Jurisdictions (with sample ≥ 5) as in Figure 3.



5.2 Lack-of-Fit Diagnostics

We attempted to develop models that accurately describe the observed data. Lack-of-fit methodology based on a parametric bootstrap [Ashmead and Slud, 2017] provided one set of tools useful in assessing the quality of that description. The rationale behind these methods was that if we assume our fitted models are correct and simulate values from them, then they should look similar to the actual observed results. In practice we simulate from the models many times and provide the reference distribution for a statistic codifying an important feature. Then we calculate the same statistic for the observed data. Referring the observed statistic (called a 'diagnostic') to the distribution of the statistic from the parametric bootstraps provides a formal comparison. The central region of the distribution can be regarded as a probable range for the statistic under the model, and an observed statistic falling there is deemed compatible with the model. If the observed statistic is extreme or completely outside the range of observed values for the distribution, that indicates that the assumed model is not consistent with the observed data.

Figures 6 and 7 summarize the sum of squared errors lack of fit diagnostic (given by formula (45) in Appendix H), specifically for the LEP proportion among voting-age persons in the models. The statistic of interest resembles the sum of the squared differences between the DS and MM estimates of LEP proportion for each area, but assesses mostly the random-effect aspects of the model. This diagnostic does not speak to the validity of the model prediction for any single area or areas, but rather to the variability of predictions of the model in general. See Appendix H for additional technical details.

In Figures 6 and 7 we plot the lack of fit quantile (47) for all LMGs for which a model was used respectively in Jurisdictions and in AIAs, separately within the AIAN and Asian/Hispanic LMGs. The LMGs are first sorted by the quantiles of their lack of fit statistics each within its own bootstrapped reference distribution. The points are plotted with y-coordinates in sequence (1 to 17 for Asian/Hispanic, 18-68 for AIAN) and with their quantiles as x-coordinates. Points to the left of the 0.025 dashed line indicate observed statistics smaller than those typically simulated by the assumed model. Points to the right of the 0.975 dashed line indicate observed statistic values larger those typically simulated by the assumed model. Overall, we find that the great majority of models do not show evidence of lack of fit according to this measure. Within the set of AIAN LMGs, both figures show the quantiles increasing approximately linearly with sequence numbers, as one would expect for properly specified models with ordered $\text{Uniform}(0, 1)$ p-values, and the number of extreme AIAN-LMG quantiles (5 in Fig. 6 and 3 in Fig. 7) are hardly larger than the expected number $51 \cdot 0.05$. However, an excess of models among the Asian and Hispanic LMGs in Figure 6 showed evidence of lack of fit by their extreme values, 7 intead of the $17 \cdot 0.05 \approx 1$ that would be expected if all models were correct. These extreme LMGs tended to be large, and the diagnostic is more sensitive for LMGs containing areas with especially large sample sizes. These diagnostics show that there is room for improving the models for Asian/Hispanic LMGs in the future.

Figure 6: Lack of fit summary for predicted LEP counts in Jurisdiction models. Diagnostic (45) bootstrap quantile is plotted on horizontal axis, and numerical sequence (1 to 17 in Asian/Hispanic and 18 to 68 in AIAN) plotted on vertical axis in increasing order of the quantiles.

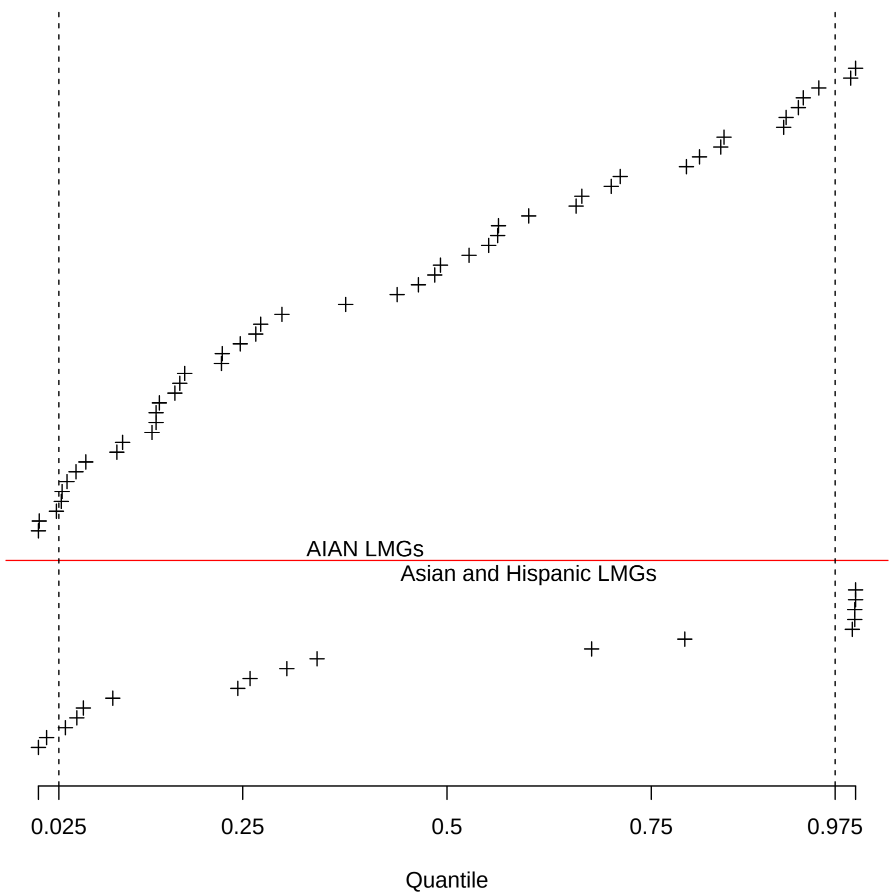
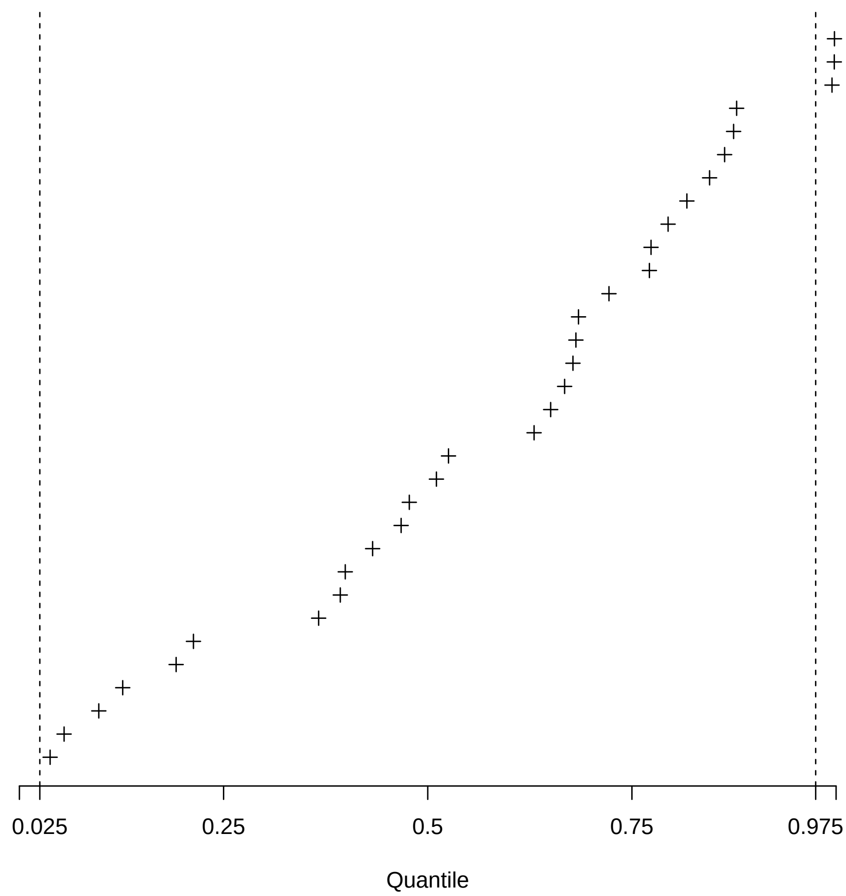


Figure 7: Lack of fit summary for predicted LEP counts in AIA models. Diagnostic (45) bootstrap quantile is plotted on horizontal axis, and numerical sequence (1 to 51) of AIAN plotted on vertical axis in increasing order of the quantiles.



6 Results Summary

In this section, we focus on results comparing variances for the proportions of LEP to VOTAG persons in the Jurisdiction (`LEPpropJur`) or AIA (`LEPpropAIA`), specifically comparing estimated MSPEs for the modeled estimates versus variances for the DS estimates. We do not discuss the estimated variance of the illiteracy rate (`ILLrateJur` or `ILLrateAIA`), because as mentioned previously, the locally domain-estimated illiteracy rate is almost always greater than the national rate. The overall conclusion is that the small area models reduce the estimated variance, especially in those Jurisdictions or AIAs with large DS variance estimates or point estimate near the 0.05 threshold. We detail our findings in the following tables and figures.

Table 6 shows the estimated percent variance reduction of the LEP proportion (`LEPpropJur`) for Jurisdictions among all LMGs for which a model was produced. The Jurisdictions are first divided into deciles by their ACS directly estimated variance, then within those deciles we calculate the median and (95th, 75th, 25th, 5th) percentiles of the estimated variance reduction. From the median reduction, this table shows that in general the variance of model-predicted LEP proportions was less than those of their DS counterparts, and the reduction tended to be systematically larger for Jurisdictions with larger ACS (DS estimated) variance. Those with the largest DS variance (the Top Decile) had the greatest reductions in variance. However, in some cases (at least 5%, across the board for all deciles of DS Variance, and at least 25% for the bottom four DS Variance deciles), the model-based predictions had a larger estimated MSPE than the DS variance estimates.

Table 6: Estimated percent variance reduction of predicted LEP proportion for Jurisdictions (`LEPpropJur`) by decile of direct ACS SDR variance estimates.

	# of (AIA, LMG)	Percent Variance Reduction				
	Estimates	95^{th} Percentile	75^{th} Percentile	Median	25^{th} Percentile	5^{th} Percentile
Top Decile*	1768	98.26	90.30	73.35	40.91	-32.27
2nd Decile	1767	98.40	87.53	68.29	31.95	-69.66
3rd Decile	1768	97.70	83.19	58.38	17.67	-108.30
4th Decile	1767	97.74	82.25	54.25	12.48	-152.09
5th Decile	1768	97.87	82.91	53.86	7.36	-171.17
6th Decile	1767	97.50	81.24	48.65	7.07	-188.06
7th Decile	1768	97.28	79.83	47.89	-0.56	-250.49
8th Decile	1767	96.00	76.50	42.61	-18.54	-388.35
9th Decile	1768	96.89	75.97	34.33	-49.88	-606.54
Bottom Decile	1768	97.67	77.15	25.12	-115.67	-1249.01

*Decile of largest ACS SDR variance estimates. Percent variance reduction calculated only among Jurisdictions with non-zero point and variance estimates.

Table 7 gives the estimated percent variance reduction of the LEP proportion (`LEPpropAIA`) for AIAs among all AIAN LMGs for which a model was produced. It shows a pattern similar to that of Table 6. The median variance reduction is largest for the those with the largest DS variances and decreases with the DS variance. While there are some estimates for which the MSPE of modeled predictions was larger, the majority of such estimates saw a reduction in variance.

Table 7: Percent variance reduction of estimated LEP proportion for AIAs (`LEPpropAIA`) by decile of direct (SDR) ACS variance estimates.

	# of (AIA, LMG)	Percent Variance Reduction				
	Estimates	95^{th} Percentile	75^{th} Percentile	Median	25^{th} Percentile	5^{th} Percentile
Top Decile*	61	97.62	92.93	80.40	44.24	-67.48
2nd Decile	60	95.62	87.05	54.17	10.84	-180.99
3rd Decile	60	96.92	84.23	60.05	-15.89	-120.41
4th Decile	61	95.59	84.53	58.94	-4.87	-162.55
5th Decile	60	98.77	93.29	58.10	-27.32	-350.97
6th Decile	60	98.50	89.15	71.32	6.52	-376.28
7th Decile	61	95.91	66.20	30.80	-94.14	-555.53
8th Decile	60	98.95	84.86	28.94	-116.90	-1519.83
9th Decile	60	98.56	81.45	39.32	-146.99	-1269.56
Bottom Decile	61	96.65	62.46	-14.00	-142.30	-1048.36

*Decile of largest ACS SDR variance estimates. Percent variance reduction calculated only among AIAs with non-zero point and variance estimates.

In order to take into account not only an estimate's variance but also its value, we calculated the CV for the LEP proportion (`LEPpropJur`, `LEPpropAIA`), both for the model predictors and the direct estimators, for each of the modeled domains, (Jur, LMG) or (AIA, LMG), in Table 8. The Table shows that the models dramatically reduced the number of domains with extremely large (> 1) CVs and modestly reduced the number of domains with large (> 0.6) CVs.

Table 8: Number of modeled domains with non-zero SDR CVs which have large LEP proportion (`LEPpropJur`, `LEPpropAIA`) CVs.

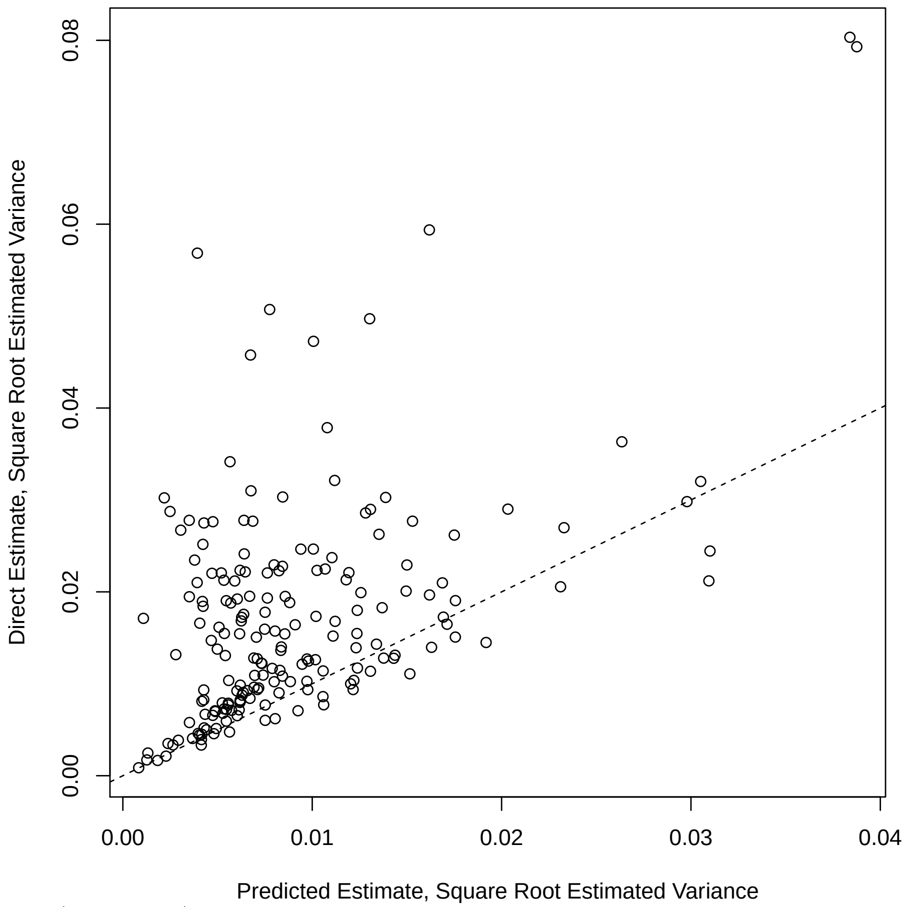
	Total Estimates	SDR-CV >0.6	MSPE-CV >0.6	SDR-CV >1	MSPE-CV >1
Asian, Hispanic Jurs	14461	8014	5543	4826	524
AIAN Jurs	3215	2618	1278	1914	301
AIAN AIAs	604	402	244	214	99

NB: For comparability, CVs for both the SDR and MSPE are calculated using modeled point estimates.

In Figure 2 we showed the CVs for 189 (Jur, LMG) domain estimates of LEP rate whose 90% confidence interval includes 0.05. In Figure 8 we plot the square root of the estimated variance of these same (Jur, LMG) domains for both the direct estimate and the model prediction to illustrate the variance reduction. Of the 189 (Jur, LMG)s estimates, 182 were from LMGs for which a model was fitted. For 161 of the 189 (85%) of the (Jur, LMG) domain LEP proportion estimates, the estimated variance of the model predicted proportion was smaller than that of the directly estimated proportion. Additionally, many of the variance reductions were large compared with the percentages by which MSPEs (rarely) exceeded the SDR variances.

We provide several further summaries of the comparisons in variances of coverage-relevant

Figure 8: Comparison of square root of the estimated variance for 189* (Jur,LMG) domain estimates of LEP proportion (`LEPpropJur`) whose 90% directly estimated CI includes 0.05



*182 of 189 (Jur, LMG) domain estimates were from LMGs for which we are able to fit a model

domain-level estimates between the direct SDR variances and those MSPEs that resulted from the Dirichlet-multinomial small-area estimation techniques actually used along with the variance estimation technique described in Section 4.3. These comparisons apply only to those (LMG,Jur) or (LMG,AIA) domains for which the Dirichlet-multinomial models could be fitted, according to the steps described in Appendix F.1. As we have seen in Tables 6 and 7, the small-area estimation methodology resulted in reduced variances for the most part, but in some domains the variances actually increased. We characterize in greater detail those domains and estimates for which variances increased and decreased, in order to understand better the effect that the small-area estimation methodology had on the precision of the statistical decisions leading to determinations of coverage under Section 203(b) of the *Voting Rights Act.*

Table 9 shows the medians of MSPE over SDR variance ratios for LEP-proportions (`LEPpropJur`) across all LMGs for which Jurisdiction models were fit, separately grouped by Asian and Hispanic LMGs and by AIAN LMGs. Within each cell defined by model-based point estimates and CVs (defined from the model-based point estimates and SDR Variances), we calculate the median ratios to quantify variance improvement for subsets of Jurisdictions. Ratios less than one show a decrease in typical estimated variance for the model-based estimates compared with the direct estimates. The table shows that the model-based MSPEs improve noticeably over the SDR variances of LEP-proportion estimates, except in Jurisdictions where the point estimates are particularly small. The improvement tends to increase with the point estimate and for each range of estimates with the CV. In the final row of each of the upper and lower tables, which includes the Jurisdictions with estimated LEP-proportions near the 0.05 threshold, the MSPEs are considerably lower than SDR variances. Similar patterns are observed in tables analogous to Table 9 (not shown) in which median ratios are replaced by upper quartiles, and where medians or upper quartiles of Jurisdiction-specific ratios (MSPE over SDR variance) are replaced by ratios of the corresponding quantiles of MSPEs divided by quantiles of SDR variances.

Another cross-tabulation of variance comparisons makes the patterns clearer. In Table 10 we show the ratios of upper quartiles of MSPEs over upper quartiles of SDR-Variances for LEP-proportions (`LEPpropJur`), within cells of domains defined through ranges of model-based point estimates cross-classified with VOTAG sample-sizes. Note that in this table the ratios of quantiles tend to increase from left to right, since sample sizes are generally inversely related to CV. Also, the ratio of MSPE and variance quantiles tend to be smaller and less variable across cells than the quantiles of MSPE over variance ratios. In Table 10, larger sample sizes are associated with large Jurisdictions, where SDR variances were generally small. The reduction in MSPE versus SDR variance in these large Jurisdictions, while systematic, was smaller than in Jurisdictions with larger SDR variances.

The Jurisdictions likely to meet the *Voting Rights Act* Section 203(b) coverage criteria tended to be those, as seen in Figure 2, where the LEP-proportion confidence interval covered the threshold value of 0.05. The last rows of each upper and lower Table in Tables 9 and 10 all show dramatic reduction in MSPE compared with SDR-variance, substantiating that among Jurisdictions with larger LEP-proportion point estimates there was greater accuracy in assessing the coverage criteria using model-based estimates compared with the direct survey estimates.

Table 9: Medians of ratios of MSPEs over SDR variances for estimates of LEP proportion (`LEPpropJur`) for LMGs with fitted models, cross-classified by model-based point estimate and CV (using SDR variance). **Upper table:** Asian & Hispanic LMGs; **Lower table:** AIAN LMGs. Cell entries are median ratios followed (in parentheses) by numbers of Jurisdictions in the cell.

Asian & Hispanic LMGs	CV range				
Point Estimate Range	(0,0.2]	(0.2,0.4]	(0.4,0.6]	(0.6,1]	(1,2]
(0, 0.002]	0.996 (318)	1.329 (1659)	1.050 (1860)	0.595 (2525)	0.160 (4044)
(0.002, 0.005]	0.824 (294)	0.866 (466)	0.678 (386)	0.439 (444)	0.156 (582)
(0.005, 0.01]	0.715 (216)	0.710 (237)	0.493 (136)	0.382 (135)	0.138 (142)
(0.01, 1]	0.751 (512)	0.605 (275)	0.494 (88)	0.305 (84)	0.119 (58)

AIAN LMGs	CV range				
Point Estimate Range	(0,0.3]	(0.3,0.6]	(0.6,1]	(1,2]	
(0,.001]	1.808 (82)	1.603 (420)	0.710 (654)	0.058 (1756)	
(.001,.002]	1.809 (5)	1.156 (24)	0.551 (21)	0.039 (106)	
(.002,1]	0.588 (32)	0.664 (34)	0.414 (29)	0.045 (52)	

Table 10: Ratios of Upper Quartiles (Q3) of MSPEs over Q3 of SDR-Variances for LEP-proportions (`LEPpropJur`), for LMGs with fitted models, cross-classified by (model-based) point estimate and VOTAG sample-size. **Upper table:** Asian & Hispanic LMGs; **Lower Table:** AIAN LMGs. Cell entries are ratios of Q3's followed (in parentheses) by numbers of Jurisdictions in the cell.

Asian & Hisp LMGs	Sample-size range				
Point Estimate Range	$(0,6]$	$(6,10]$	$(10,20]$	$(20,50]$	$(50,\infty]$
(0,.002]	0.190	0.426	0.576	0.874	0.840
	(3542)	(1445)	(1891)	(1802)	(1726)
(.002,.005]	0.165	0.350	0.427	0.552	0.747
	(476)	(233)	(266)	(388)	(809)
(.005,.01]	0.211	0.255	0.468	0.472	0.582
	(113)	(59)	(75)	(141)	(478)
(.01,1]	0.188	0.187	0.417	0.593	0.632
	(34)	(25)	(55)	(94)	(809)

AIAN LMGs	Sample-size range				
Point Estimate Range	$(0,6]$	$(6,10]$	$(10,20]$	$(20,50]$	$(50,\infty]$
(0,.001]	0.030	0.085	0.162	0.330	0.792
	(924)	(360)	(435)	(512)	(681)
(.001,.002]	0.034	0.076	0.174	0.174	0.802
	(75)	(18)	(8)	(19)	(36)
(.002,1]	0.030	0.108	0.056	0.1514	0.508
	(31)	(10)	(9)	(10)	(87)

References

American Community Survey Design and Methodology. Technical report, U.S. Census Bureau, 2014. Available online at https://www2.census.gov/programs-surveys/acs/methodology/design_and_methodology/acs_design_methodology_report_2014.pdf.

Robert Ashmead and Eric Slud. Small Area Model Diagnostics and Validation with Applications to the Voting Rights Act, Section 203. In *Proceedings of the Survey Research Methods Section, American Statistical Association, Alexandria, VA*, pages 1515–1523, 2017.

N.E. Breslow and D.G. Clayton. Approximate Inference in Generalized Linear Mixed Models. *Journal of the American Statistical Association*, pages 9–25, 1993.

Bradley P. Carlin and Thomas A. Louis. *Bayesian Methods for Data Analysis*. Taylor & Francis, Chapman & Hall, 2009.

Robert E. Fay and George F. Train. Aspects of Survey and Model-Based Postcensal Estimation of Income and Poverty Characteristics for States and Counties. In *American Statistical Association Proceedings of the Government Statistics Section*, 1995.

Patrick M. Joyce, Donald Malec, Roderick J.A. Little, and Aaron Gilary. Statistical Modeling Methodology for the Voting Rights Act Section 203 Language Assistance Determinations. *U.S. Census Bureau. Center for Statistical Research & Methodology Research Report Series*, Statistics #2012-02, 2012. Available online at https://www.census.gov/srd/papers/pdf/rrs2012-02.pdf.

Patrick M. Joyce, Donald Malec, Roderick J.A. Little, Aaron Gilary, Alfredo Navarro, and Mark E. Asiala. Statistical Modeling Methodology for the Voting Rights Act Section 203 Language Assistance Determinations. *Journal of the American Statistical Association*, 109(505):36–47, 2014.

H.L. Oh and Fritz J. Scheuren. Weighting adjustment for unit non-response. in incomplete data in sample surveys.(eds wg madow, i. olkin and db rubin.) pp. 143–184, 1983.

J.N.K Rao and Isabel Molina. *Small-Area Estimation*. Wiley, 2nd edition, 2015.

Jun Shao and Dongsheng Tu. *The Jackknife and Bootstrap*. Springer Science & Business Media, 2012.

Eric Slud and Robert Ashmead. Hybrid BRR and Parametric-Bootstrap Variance Estimates for Small Domains in Large Surveys. In *Proceedings of the Survey Research Methods Section, American Statistical Association, Alexandria, VA*, pages 1716–1730, 2017.

Eric V. Slud. Assessment of Zeroes in Survey-estimated Tables via Small-area Confidence Bounds. *Journal of the Indian Society for Agricultural Statistics*, 66:157–169, 2012.

Kirk M. Wolter. An Investigation of Some Estimators of Variance for Systematic Sampling. *Journal of the American Statistical Association*, pages 781–790, 1984.

Lynn M.R. Ybarra and Sharon L. Lohr. Small Area Estimation when Auxiliary Information is Measured with Error. *Biometrika*, 95(4):919–931, 2008.

A Section 203 of the Voting Rights Act of 1965

(a) Congressional findings and declaration of policy

The Congress finds that, through the use of various practices and procedures, citizens of language minorities have been effectively excluded from participation in the electoral process. Among other factors, the denial of the right to vote of such minority group citizens is ordinarily directly related to the unequal educational opportunities afforded them resulting in high illiteracy and low voting participation. The Congress declares that, in order to enforce the guarantees of the fourteenth and fifteenth amendments to the United States Constitution, it is necessary to eliminate such discrimination by prohibiting these practices, and by prescribing other remedial devices.

(b) Bilingual voting materials requirement

(1) Generally

Before August 6, 2032, no covered State or political subdivision shall provide voting materials only in the English language.

(2) Covered States and political subdivisions

(A) Generally

A State or political subdivision is a covered State or political subdivision for the purposes of this subsection if the Director of the Census determines, based on the 2010 American Community Survey census data and subsequent American Community Survey data in 5-year increments, or comparable census data, that –

(i)(I) more than 5 percent of the citizens of voting age of such State or political subdivision are members of a single language minority and are limited English-proficient;
(II) more than 10,000 of the citizens of voting age of such political subdivision are members of a single language minority and are limited English-proficient; or
(III) in the case of a political subdivision that contains all or any part of an Indian reservation, more than 5 percent of the American Indian or Alaska Native citizens of voting age within the Indian reservation are members of a single language minority and are limited English-proficient; and

(ii) the illiteracy rate of the citizens in the language minority as a group is higher than the national illiteracy rate.

(B) Exception

The prohibitions of this subsection do not apply in any political subdivision that has less than 5 percent voting-age limited English-proficient citizens of each language minority which comprises over 5 percent of the statewide limited English-proficient population of voting-age citizens, unless the political subdivision is a covered political subdivision independently from its State.

(3) Definitions

As used in this section–

(A) *the term "voting materials" means registration or voting notices, forms, instructions, assistance, or other materials or information relating to the electoral process, including ballots;*

(B) *the term "limited English-proficient" means unable to speak or understand English adequately enough to participate in the electoral process;*

(C) *the term "Indian reservation" means any area that is an American Indian or Alaska Native area, as defined by the Census Bureau for the purposes of the 1990 decennial census;*

(D) *the term "citizens" means citizens of the United States; and*

(E) *the term "illiteracy" means the failure to complete the 5th primary grade.*

(4) Special Rule

The determinations of the Director of the Census under this subsection shall be effective upon publication in the Federal Register and shall not be subject to review in any court.

(c)-(d) [Not given here. See U.S. Code, Title 52, Subtitle I, Chapter 105, §10503.]

(e) Definitions

For purposes of this section, the term "language minorities" or "language minority group" means persons who are American Indian, Asian American, Alaska Native, or of Spanish heritage.

B Language Minority Groups

Table 11: 2016 Language Minority Groups.

Hispanic	Houma
Asian Indian	Iroquois
Bangladeshi	Kiowa
Cambodian	Lumbee
Chinese	Menominee
Filipino	Mexican American Indian
Hmong	Navajo
Indonesian	Osage
Japanese	Ottawa
Korean	Paiute
Laotian	Pima
Malaysian	Potawatomi
Pakistani	Pueblo
Sri Lankan	Puget Sound Salish
Thai	Seminole
Vietnamese	Shoshone
Other Asian	Sioux
Apache	South American Indian
Arapaho	Spanish American Indian
Blackfeet	Tohono O'Odham
Canadian and French Indian	Ute
Central American Indian	Yakama
Cherokee	Yaqui
Cheyenne	Yuman
Chickasaw	All other AI tribes
Chippewa	AI tribes, not specified
Choctaw	Alaska Athabascan
Colville	Aleut
Comanche	Inupiat
Cree	Tlingit-Haida
Creek	Tsimshian
Crow	Yup'ik
Delaware	Alaskan Native Tribes, not specified
Hopi	AI or AN tribes, not specified

C Determination Flow Chart

How the Law Prescribes the Determination of Covered Areas under the Language Minority Provisions of Section 203 of the Voting Rights Act

For State, County, County Subdivision, and American Indian and Alaska Native Areas:

Total Population

Total Population for Minority Groups:

- Spanish/Hispanic/Latino
- AIAN (American Indian/Alaskan Native)
- AIAN tribal group (examples include Cherokee, Iroquois, Sioux)
- Asian
- Asian group (examples include Chinese, Filipino, Japanese)

↓

For each Minority Group:

Total Voting Age Population

Total Voting Age Citizens

↓

For each Minority Group:

Total Voting Age Citizens who are Limited-English Proficient

↓

For each Minority Group:

Total Voting Age Citizens who have Less Than a 5th Grade Education

Total Voting Age Citizens who are Limited-English Proficient and have Less Than a 5th Grade Education

↓

Determinations are then computed for each language minority group based on the following:

If more than 5% of voting age citizens are limited-English proficient,

OR

If more than 10,000 voting age citizens are limited-English proficient,

AND

The rate of total voting age citizens who are limited-English proficient and have less than a 5th grade education is higher than the national rate (1.16 for the 2011 release and 1.31 for the 2016 release).

THEN:

The state, county, or county subdivision under consideration is covered for that specific minority group of Section 203 of the Voting Rights Act.

There is a special computation for American Indian or Alaska Native areas or other tribal lands:

If more than 5% of the American Indian or Alaska Native voting age citizens belonging to an American Indian/Alaska Native tribe are limited-English proficient

AND

The rate of those voting age citizens who are limited-English proficient and have less than a 5th grade education is higher than the national rate (1.16 for the 2011 release and 1.31 for the 2016 release).

THEN:

Any political subdivision(s) in which that AIA/ANA is located is covered under Section 203 of the Voting Rights Act.

D Notations and Model Definitions

The primary quantities tabulated from the ACS files are, for Jurisdictions j, AIAs a, and LMGs g:

$n^V_{jg},\ n^C_{jg},\ n^L_{jg},\ n^I_{jg}$	=	the numbers of ACS sampled persons (respectively VOTAG, CIT, LEP, ILL) in Jurisdiction j and LMG g
$\hat{N}^V_{jg},\ \hat{N}^C_{jg},\ \hat{N}^L_{jg},\ \hat{N}^I_{jg}$	=	the ACS survey-weighted estimates of total persons (respectively VOTAG, CIT, LEP, ILL) in Jurisdiction j and LMG g
$n^V_{ag},\ n^C_{ag},\ n^L_{ag},\ n^I_{ag}$	=	the numbers of ACS persons (respectively VOTAG, CIT, LEP, ILL) sampled in AIA a and LMG g
$\hat{N}^V_{ag},\ \hat{N}^C_{ag},\ \hat{N}^L_{ag},\ \hat{N}^I_{ag}$	=	the ACS survey-weighted estimates of total persons (respectively VOTAG, CIT, LEP, ILL) in AIA a and LMG g.

Also, let $\hat{N}^C_{j+}$ denote the ACS survey-weighted estimate of the total number of voting-age citizens in Jurisdiction j, including both LMG and non-LMG persons. Similarly let $\hat{N}^C_{a,+}$ denote the ACS survey-weighted estimate of the total number of AIA voting-age citizens in AIA a. Note that the set $\mathcal{A}$ of AIA labels a is assumed to be distinct from the set $\mathcal{J}$ of Jurisdiction labels, and these labels distinguish the AIA and Jurisdiction notations.

Let $(N^V_{jg}, N^C_{jg}, N^L_{jg}, N^I_{jg})$, $(N^V_{ag}, N^C_{ag}, N^L_{ag}, N^I_{ag})$, N^C_{j+}, and N^C_{a+} respectively denote the true population totals for each of the ACS survey-weighted estimates above. The quantities needed for the determination criteria can be expressed in terms of these true population totals. For the Jurisdiction level determination criteria, the primary targets of estimation are the LMG by Jurisdiction LEP totals, LEP proportions, and ILL rates, denoted as

$$N^L_{jg} \quad , \qquad \frac{N^L_{jg}}{N^C_{j+}} \quad , \qquad \text{and} \qquad \frac{N^I_{jg}}{N^L_{jg}}. \tag{2}$$

These can be estimated using direct ACS survey-weighted totals, respectively by

$$\hat{N}^L_{jg} \quad , \qquad \frac{\hat{N}^L_{jg}}{\hat{N}^C_{j+}} \quad , \qquad \text{and} \qquad \frac{\hat{N}^I_{jg}}{\hat{N}^L_{jg}} \tag{3}$$

Similarly, the AIA-level determination criteria utilize the quantities

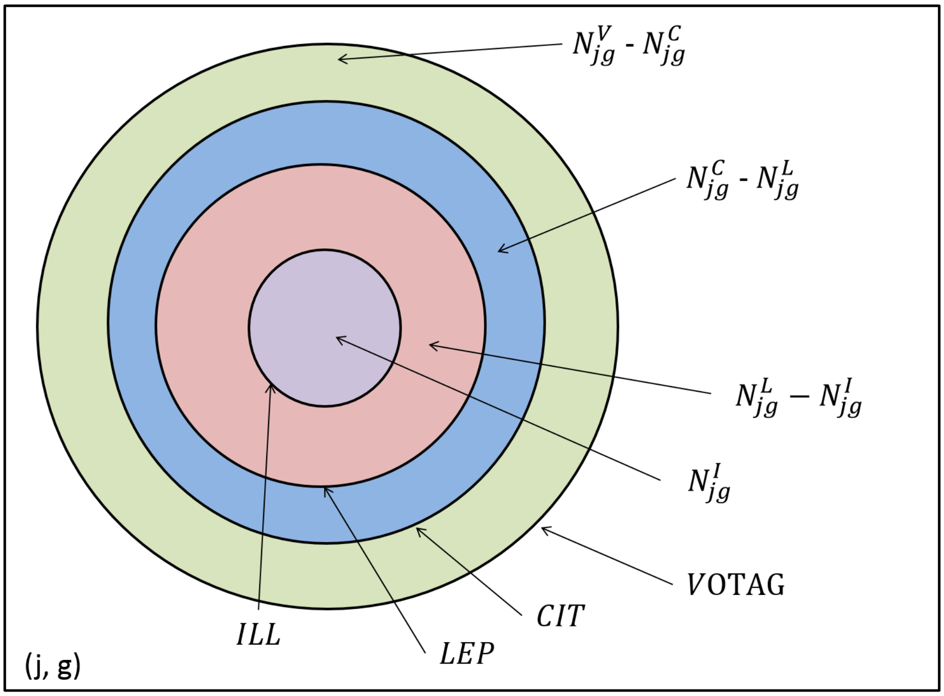
$$\frac{N^L_{ag}}{N^C_{a+}} \qquad \text{and} \qquad \frac{N^I_{ag}}{N^L_{ag}}. \tag{4}$$

Lastly, the state level targets of estimation determining *Voting Rights Act* Section 203(b) coverage are

$$\frac{\sum_{j\in\mathcal{S}} N^L_{jg}}{\sum_{j\in\mathcal{S}} N^C_{j+}} \qquad \text{and} \qquad \frac{\sum_{j\in\mathcal{S}} N^I_{jg}}{\sum_{j\in\mathcal{S}} N^L_{jg}}, \tag{5}$$

where $j \in \mathcal{S}$ ranges over all Jurisdictions in a given state $\mathcal{S}$. The AIA and state level quantities (4) and (5) can also be estimated by replacing each of their constituent totals by its direct survey-weighted estimate.

Figure 9: Illustration of multinomial categories and population counts.



Labels VOTAG, CIT, LEP, ILL refer as in Fig. 1 to successively smaller circular regions. Arrows terminating within colored annuli and innermost circle show population-count notations for the multinomial categories: VOTAG non-CIT, CIT non-LEP, LEP non-ILL, ILL.

If these estimators were stable, they would be the design-based estimators of choice. However, many of these weighted-total estimators are based on sampled numbers of persons in ACS that are very small. So we are driven to develop model-based estimators for the quantities (2), (4), and (5) in the spirit of small-area estimation [Rao and Molina, 2015]. The main idea of this approach is that for a given LMG g many small areas (Jurisdictions or AIAs) may behave similarly with respect to the proportions of CIT, LEP, and ILL persons among VOTAG persons (e.g. N_{jg}^C/N_{jg}^V, N_{jg}^L/N_{jg}^V, N_{jg}^I/N_{jg}^V) within the LMG. In addition, observable domain-specific covariates may be used to predict these same proportions. Therefore, we model the differences across Jurisdiction or AIA for a fixed LMG as being partly random and partly predictable in terms of covariates $\mathbf{X}$. The empirical-Bayes framework results in within-LMG estimators of citizenship, LEP, and illiteracy proportions that are weighted combinations of the direct estimators and model-based parameter estimators, where the weights heavily favor the direct estimators in domains with a large sample size and the model-based estimator in domains with a small sample size.

Now we turn to the general form of the small area models used to make the 2016 determination estimates. The same form of model was used for Jurisdiction estimates and separately for AIA estimates. In the formulas below, we use the notation for Jurisdictions (e.g. n_{jg}^V, N_{jg}^L, etc.); however, the same general model form applies to AIAs and could be written by replacing all j subscripts with a subscripts.

The nested characteristics (CIT, LEP, ILL) of the n_{jg}^V sampled persons in (j, g) are viewed, conditionally given random-effect parameters, as the outcomes of n_{jg}^V independent multinomial trials with four possible mutually exclusive outcomes: non-citizen, non-LEP citizen, literate LEP citizen, and LEP illiterate citizen. From the direct VOTAG population estimate $\hat{N}_{jg}^V$ along with estimates of the proportions of these mutually exclusive categories, we estimate the determination-relevant quantities $(N_{jg}^C, N_{jg}^L, N_{jg}^I)$. The direct survey-weighted (DS) estimates of multinomial outcome proportions, along with number of sampled voting-age persons in (j, g), are treated as the observed data. Then the *Dirichlet-multinomial model* with logistic-regression fixed effects can be written as:

$$\frac{n_{jg}^V}{\hat{N}_{jg}^V} \left(\hat{N}_{jg}^V - \hat{N}_{jg}^C,\ \hat{N}_{jg}^C - \hat{N}_{jg}^L,\ \hat{N}_{jg}^L - \hat{N}_{jg}^I,\ \hat{N}_{jg}^I \right) \sim \text{Multinom}(n_{jg}^V\,;\ \underline{\omega}_{jg}), \tag{6}$$

where $\underline{\omega}_{jg} = (\omega_{1,jg},\ \omega_{2,jg},\ \omega_{3,jg}, \omega_{4,jg})$ and

$$\underline{\omega}_{jg} \sim \text{Dirichlet}\Big(\tau_{jg}, \big((1-\mu_{jg}),\ \mu_{jg}(1-\nu_{jg}),\ \mu_{jg}\nu_{jg}(1-\rho_g),\ \mu_{jg}\nu_{jg}\rho_g\big)\Big), \tag{7}$$

$$\tau_{jg} = \tau_{o_g} \sqrt{n_{jg}^V}\ , \qquad (1-\mu_{jg}) + \mu_{jg}(1-\nu_{jg}) + \mu_{jg}\nu_{jg}(1-\rho_g) + \mu_{jg}\nu_{jg}\rho_g = 1,$$

with

$$\mu_{jg} \ = \ \frac{\exp(\beta_g'\,\mathbf{X}^{C(g)})}{1+\exp(\beta_g'\,\mathbf{X}^{C(g)})} \quad , \qquad \nu_{jg} \ = \ \frac{\exp(\gamma_g'\,\mathbf{X}^{L(g)})}{1+\exp(\gamma_g'\,\mathbf{X}^{L(g)})}.$$

Here $\mathbf{X}_j^{C(g)}$ and $\mathbf{X}_j^{L(g)}$ respectively denote covariate vectors for LMG g used to predict proportions of CIT within VOTAG and of LEP within CIT.

In (7), the parameter τ_{jg} measures how concentrated the Dirichlet distribution is and can be viewed as a prior sample size associated with the model, inversely related to the variance of the random $\omega_{k,jg}$ terms. The term μ_{jg} models the citizenship proportion within voting-age persons in (j, g) like a logistic regression, and similarly ν_{jg} is a logistic-regression fixed-effect term for the LEP-proportion among (j, g) voting-age citizens. The parameter ρ_g denotes the illiteracy proportion among LEP voting-age citizens. Unlike the proportions for citizenship and LEP, we do not attempt to model this proportion with a logistic regression. The parameterization of the concentration term $\tau_{jg} = \tau_{o_g} \sqrt{n_{jg}^V}$ was selected after assessing model diagnostics. See Section 3.4 for additional information about this modeling choice. The parameters $\beta_g, \gamma_g, \rho_g,$ and τ_{o_g} are jointly estimated by maximum likelihood from the data. These estimated parameters are substituted into best-predictors for $\underline{\omega}_{jg}$, in order to make so-called empirical or 'plug-in' predictions from the model. The predictions consist of convex combinations of the 'Direct Survey' (DS) and the 'Marginal Mean' (MM) estimates for that area. These estimates, previously mentioned in Section 5, are defined as follows:

$$\underline{\text{DS}}_{jg} = \left(\frac{\hat{N}^V_{jg} - \hat{N}^C_{jg}}{\hat{N}^V_{jg}} \, , \, \frac{\hat{N}^C_{jg} - \hat{N}^L_{jg}}{\hat{N}^V_{jg}} \, , \, \frac{\hat{N}^L_{jg} - \hat{N}^I_{jg}}{\hat{N}^V_{jg}} \, , \, \frac{\hat{N}^I_{jg}}{\hat{N}^V_{jg}} \right); \tag{8}$$

$$\underline{\text{MM}}_{jg} = \left(1 - \widehat{\mu}_{jg}, \, \widehat{\mu}_{jg}(1 - \widehat{\nu}_{jg}), \, \widehat{\mu}_{jg}\widehat{\nu}_{jg}(1 - \widehat{\rho}_g), \, \widehat{\mu}_{jg}\widehat{\nu}_{jg}\widehat{\rho}_g \right), \tag{9}$$

where

$$\hat{\mu}_{jg} \; = \; \frac{\exp(\hat{\beta}'_g \, \mathbf{X}^{C(g)})}{1 + \exp(\hat{\beta}'_g \, \mathbf{X}^{C(g)})} \quad , \qquad \hat{\nu}_{jg} \; = \; \frac{\exp(\hat{\gamma}'_g \, \mathbf{X}^{L(g)})}{1 + \exp(\hat{\gamma}'_g \, \mathbf{X}^{L(g)})}.$$

The final prediction vector $\underline{\hat{\omega}}_{jg}$, earlier called 'Full-Model-Predictions' (FMPs) are convex combinations of (8) and (9) weighted by a function of the sample size and concentration parameter:

$$\underline{\hat{\omega}}_{jg} \; = \; \frac{n^V_{jg}}{n^V_{jg} + \hat{\tau}_{o_g}\sqrt{n^V_{jg}}} \left(\underline{\text{DS}}_{jg} \right) \quad + \quad \frac{\hat{\tau}_{o_g}\sqrt{n^V_{jg}}}{n^V_{jg} + \hat{\tau}_{o_g}\sqrt{n^V_{jg}}} \left(\underline{\text{MM}}_{jg} \right). \tag{10}$$

These predictors $\underline{\hat{\omega}}_{jg}$ can be viewed as posterior expectations of $\underline{\omega}_{jg}$ under the model given the data $(\hat{N}^V_{jg}, \, \hat{N}^C_{jg}, \, \hat{N}^L_{jg}, \, \hat{N}^I_{jg})$. We derive predictors for totals N^C_{jg}, N^L_{jg}, and N^I_{jg}, in the form

$$\begin{aligned} \tilde{N}^C_{jg} \; &= \; n^C_{jg} \; + \; (\hat{N}^V_{jg} - n^V_{jg}) \, (1 - \hat{\omega}_{1,jg}) \\ \tilde{N}^L_{jg} \; &= \; n^L_{jg} \; + \; (\hat{N}^V_{jg} - n^V_{jg}) \, (\hat{\omega}_{3,jg} + \hat{\omega}_{4,jg}) \\ \tilde{N}^I_{jg} \; &= \; n^I_{jg} \; + \; (\hat{N}^V_{jg} - n^V_{jg}) \, \hat{\omega}_{4,jg} \end{aligned} \tag{11}$$

These predictors together with the DS estimates $\hat{N}^C_{j+}$ (or $\hat{N}^C_{a+}$, for AIAs) are then used to predict quantities related to the determination criteria. The predictors of the state-level quantities (5) are simply obtained by summing, separately in numerator and denominator, the predictors of the particular state's constituent Jurisdiction-level predictors.

E Reference Summary of Notations and Models

Table 12: Index sets and ranges used in this report.

Notation	Interpretation
$j \in \mathcal{J} = \{1, \ldots, 7862\}$	Disjoint Jurisdictions, including 2919 counties and 4943 Minor Civil Divisions with ACS sample in 2010-2014
$g \in \{1, \ldots, 68\}$	Language Minority Groups: 16 Asian, 51 AIAN, and Hispanic
$a \in \mathcal{A} = \{1, \ldots, 568\}$	American Indian Areas (AIAs)
$A \in \{V, C, L, I\}$	Person characteristics: V=VOTAG (18+), C= Citizen 18+ (CIT), L = Limited English-proficient CIT (LEP), I = Illiterate LEP (ILL)

Table 13: Primary quantities tabulated from ACS files, for Jurisdictions j, AIAs a, and LMGs g.

Notation	**Interpretation**
$n_{jg}^V,\ n_{jg}^C,\ n_{jg}^L,\ n_{jg}^I$	The numbers of ACS sampled persons (respectively VOTAG, CIT, LEP, ILL) in Jurisdiction j and LMG g
$\hat{N}_{jg}^V,\ \hat{N}_{jg}^C,\ \hat{N}_{jg}^L,\ \hat{N}_{jg}^I$	The ACS survey-weighted estimates of total persons (respectively VOTAG, CIT, LEP, ILL) in Jurisdiction j and LMG g
$n_{ag}^V,\ n_{ag}^C,\ n_{ag}^L,\ n_{ag}^I$	The numbers of ACS persons (respectively VOTAG, CIT, LEP, ILL) sampled in AIA a, and LMG g
$\hat{N}_{ag}^V,\ \hat{N}_{ag}^C,\ \hat{N}_{ag}^L,\ \hat{N}_{ag}^I$	The ACS survey-weighted estimates of total persons (respectively VOTAG,CIT, LEP, ILL) in AIA a and LMG g
$\hat{N}_{j+}^C$	The ACS survey-weighted estimate of total number of CIT persons (LMG and non-LMG) in Jurisdiction j
$\hat{N}_{a+}^C$	The ACS survey-weighted estimate of total number of CIT AIAN persons (LMG and non-LMG) in AIA a
$N_{jg}^A,\ N_{ag}^A$	The true totals of persons with characteristic A=V,C,L,I in Jurisdiction j or in AIA a within LMG g
$N_{j+}^C,\ N_{a+}^C$	The ACS survey-weighted estimate of total number of CIT persons or AIAN persons, both LMG and non-LMG, in Jurisdiction j or AIA a

From this point on, we no longer elaborate notations for AIA areas, simply noting that all models for Jurisdictions j within LMG g have counterparts for AIAs a within AIAN LMGs g.

Table 14: Observable quantities figuring in the Dirchlet-multinomial models.

Notation	**Interpretation**
$\mathbf{X}_j^{C(g)},\ \mathbf{X}_j^{L(g)}$	Vectors of covariates from those described in Sec. 3.1 respectively used to model CIT rate within VOTAG and LEP rate within CIT for LMG g.
$Y_{jg}^A\ =\ n_{jg}^V\,\hat{N}_{jg}^A/\hat{N}_{jg}^V$	Ratio-estimator-scaled sample sizes, the observations modeled in the Dirichlet-multinomial model, for $A = C, L, I$

Table 15: Statistical parameters and intermediate modeled quantities within the Dirichlet-multinomial model.

Notation	Interpretation
μ_{jg}, ν_{jg},	Rates of CIT within VOTAG, LEP within CIT, for Jurisdiction j in LMG g
ρ_g	Rate of illiteracy within LEP, assumed not to vary by j within LMG g
β_g, γ_g	Regression coefficients applying respectively to $\mathbf{X}_j^{C(g)}$ and $\mathbf{X}_j^{L(g)}$ in models within LMG g for μ_{jg} and ν_{jg}
τ_{o_g}	Dispersion-factor parameter for Dirichlet multinomial model for scaled sample proportions Y_{jg}^A within LMG g
$\tau_{jg} \;=\; \tau_{o_g}\,(n_{jg}^V)^{1/2}$	The dispersion in Dirichlet multinomial model for scaled sample proportions Y_{jg}^A within LMG g
π_{jg}^A, $A = C$, L, I	The proportions N_{jg}^A/N_{jg}^V of true voting-age population in nested categories $A = C, L, I$, modeled as unobserved random effects within Jurisdiction j, LMG g
$\underline{\omega}_{jg} \;=\; \{\omega_{k,jg}\}_{k=1}^4$	The modeled random-effect vector $(1-\pi_{jg}^C,\, \pi_{jg}^C - \pi_{jg}^L,\, \pi_{jg}^L - \pi_{jg}^I,\, \pi_{jg}^I)$ of proportions in disjoint categories in Jurisdiction j, LMG g

The Dirichlet-multinomial model can now be summarized using these notations as:

$$\Big(n_{jg}^V - Y_{jg}^C,\, Y_{jg}^L - Y_{jg}^L,\, Y_{jg}^L - Y_{jg}^I,\, Y_{jg}^I\Big) \;\sim\; \text{Multinomial}(n_{jg}^V,\, \underline{\omega}_{jg})$$

$$\underline{\omega}_{jg} \;\sim\; \text{Dirichlet}\Big(\tau_{jg},\; (1-\mu_{jg}, \mu_{jg}(1-\nu_{jg}), \mu_{jg}\nu_{jg}(1-\rho_g), \mu_{jg}\nu_{jg}\rho_g)\Big)$$

$$\tau_{jg} = \tau_{o_g}\sqrt{n_{jg}^V}\;\;, \qquad (1-\mu_{jg}) + \mu_{jg}(1-\nu_{jg}) + \mu_{jg}\nu_{jg}(1-\rho_g) + \mu_{jg}\nu_{jg}\rho_g = 1$$

$$\mu_{jg} \;=\; \frac{\exp(\beta_g'\,\mathbf{X}^{C(g)})}{1+\exp(\beta_g'\,\mathbf{X}^{C(g)})} \quad , \qquad \nu_{jg} \;=\; \frac{\exp(\gamma_g'\,\mathbf{X}^{L(g)})}{1+\exp(\gamma_g'\,\mathbf{X}^{L(g)})}.$$

Table 16: Dirichlet-multinomial estimated parameters and model-based predictors.

Notation	**Interpretation**
$\hat{\beta}_g,\ \hat{\gamma}_g,\ \hat{\tau}_{o_g},\ \hat{\rho}_g$	MLEs of model parameters from data $\{Y^A_{jg},\ A = C, L, I,\ j \in \mathcal{J}\}$
$\hat{\mu}_{jg},\ \hat{\nu}_{jg}$	Domain-specific rate predictors for π^A_{jg}, obtained from the formulas for μ_{jg}, ν_{jg} by substituting MLEs $\hat{\beta}_g$, $\hat{\gamma}_g$ for β_g, γ_g
$\underline{\hat{\omega}}_{jg} = \{\hat{\omega}_{k,jg}\}_{k=1}^4$	Predictors of random-effect vector $\underline{\omega}_{jg}$ for Jurisdiction j, LMG g
$\hat{\pi}^A_{jg}$, $A = C,\ L,\ I$	Model-based predictors of nested random effects π^A_{jg} for Jurisdiction j, LMG g $\hat{\pi}^C_{jg} = \hat{\mu}_{jg},\ \hat{\pi}^L_{jg} = \hat{\mu}_{jg}\hat{\nu}_{jg},\ \hat{\pi}^I_{jg} = \hat{\mu}_{jg}\hat{\nu}_{jg}\hat{\rho}_g$
$(\tilde{N}^A_{jg},\ A = C, L, I)$	Model-based predictors of true population counts N^A_{jg} for Jurisdiction j, LMG g $\tilde{N}^A_{jg} \ = \ n^A_{jg} + \{\hat{N}^V_{jg} - n^V_{jg}\}\, \hat{\pi}^A_{jg},\ A = C, L, I$
$(\tilde{N}^A_{jg,alt},\ A = C, L, I)$	Alternative predictors of true population counts N^A_{jg} for Jurisdiction j, LMG g $\tilde{N}^A_{jg,alt} \ = \ Y^A_{jg} + \{\hat{N}^V_{jg} - n^V_{jg}\}\, \hat{\pi}^A_{jg},\ \ A = C, L, I$

Table 17: Notations related to the replication (BRR or SDR) and parametric bootstrap steps in hybrid variance estimation. The notations are defined below with respect to LMG indices g, but in Appendix F the g sub-indices are suppressed for simplicity.

$\hat{N}_{jg}^{V(r)}$	Version of estimates $\hat{N}_{jg}^{V}$ calculated with r'th set of replicate SDR weights
$\pi_{jg}^{A(r)}$	Version of π_{jg}^{A} recalculated from Dirichlet-multinomial model with $(\hat{\beta}_g, \hat{\gamma}_g, \hat{\tau}_{o_g}, \hat{\mu}_{jg}, \hat{\nu}_{jg},\ \hat{\rho}_g)$ replacing $(\beta_j, \gamma_g, \tau_{o_g}, \rho_g, \mu_{jg}, \nu_{jg})$
$\hat{\beta}_g^{(r)}, \hat{\gamma}_g^{(r)}, \hat{\tau}_{o_g}^{(r)}, \hat{\rho}_g^{(r)}, \hat{\mu}_{jg}^{(r)}, \hat{\nu}_{jg}^{(r)}$	ML parameter estimates calculated from data $n_j^V \cdot \pi_{jg}^{A(r)}$ replacing Y_{jg}^A in Dirichlet-multinomial model
$\{\omega_{jg,k}^{*(b,r)}\}_{k=1}^{4},\ \ \pi_{jg}^{*A(b,r)}$	Version of $\omega_{j,k}$ and π_j^A simulated in the b'th bootstrap iteration for the r'th replicate weight according to Dirichlet-multinomial model with parameters $\hat{\beta}_g^{(r)}, \hat{\gamma}_g^{(r)} \hat{\tau}_{jg}^{(r)}, \hat{\rho}_g^{(r)}, \hat{\mu}_{jg}^{(r)}, \hat{\nu}_{jg}^{(r)}$
$Y_{jg}^{*A(b,r)},\ A = C, L, I$	Bootstrap-simulated version of scaled sample sizes for use with r'th rep-weights, with $(1 - Y_{jg}^{*C(b,r)}, Y_{jg}^{*C(b,r)} - Y_{jg}^{*L(b,r)}, Y_{jg}^{*L(b,r)} - Y_{jg}^{*I(b,r)}, Y_{jg}^{*I(b,r)})$ generated as Multinomial$(n_{jg}^V,\ \underline{\omega}_{jg}^{*(b,r)})$
$\hat{\beta}_g^{*(b,r)}, \hat{\gamma}_g^{*(b,r)}, \hat{\tau}_{o_g}^{*(b,r)}, \hat{\rho}_g^{*(b,r)}$	ML parameter estimates derived from the Dirichlet-multinomial model with data $\{Y_{jg}^{*A(b,r)},\ A = C, L, I\}$
$\tilde{N}_{jg,alt}^{*A(b,r)}$	Bootstrap-based predictors of true population counts N_{jg}^A for Jurisdiction j, LMG g $(\tilde{N}_{jg,alt}^{*A(b,r)} = Y_{jg}^{*A(b,r)} + \{\hat{N}_{jg}^{V(r)} - n_{jg}^V\}\, \hat{\pi}_{jg}^{*A(b,r)},\ \ A = C, L, I)$
$\tilde{e}_{jg}^{*A(b,r)}$	$\tilde{N}_{jg,alt}^{*A(b,r)} - \hat{N}_{jg}^{V(r)} \pi_{jg}^{*A(b,r)}$ bootstrap prediction errors
$\texttt{Bias}_{jg}^A$	Replicate-weighted and bootstrapped bias term estimated from $\{e_{jg}^{*A(b,r)}\}_{b,r}$
$\texttt{Within}_{jg}^A$	Replicate-weighted average of within-bootstrap variances
$e_{jg}^{*IR(b,r)}$	Bootstrap-within-replicate prediction error for ratio estimator of ILLIT rate within LEP in Jurisdiction j, LMG g, $e_{jg}^{*IR(b,r)} = \tilde{N}_{jg,alt}^{*I(b,r)} / \tilde{N}_{jg,alt}^{*L(b,r)} - \pi_{jg}^{*I(b,r)} / \pi_{jg}^{*L(b,r)}$

F Model Estimation Details

As was discussed earlier, LMGs vary greatly in their number of observations (i.e., of areas in the LMG with at least a minimum sample size), and also in the quality of predictions from covariates within LMG of the domain-level CIT and LEP proportions. As a result, no single model works well for every LMG. We attempted to fit generally similar models to all LMGs while maintaining some differences among LMGs due to the number of observations and to the covariates that turned out to be predictive of domain CIT and LEP proportions within the LMG. In LMGs with no sampled non-citizens and no sampled ILL persons nationally, we fit reduced models with the assumption that the corresponding model parameters μ and ρ were equal to 1 and 0 respectively.

The numerical thresholds determined for the "Model Convergence Criteria" in the final form of the models implemented on 2014 data, were the result of extensive testing on 2012 data. We found that models exhibiting minimum eigenvalues of the log-likelihood Hessian matrix less than 10^{-4} or maximum fitted coefficient as large as 12 or maximum estimated standard error of 6 or more were exhibiting questionable convergence, and we excluded those model fits from consideration. In the algorithm described below, that resulted either in re-fitting the model with fewer predictor variables or (in case the model with no predictors was already being considered) in excluding modeling for that domain (LMG, geography) , where "geography" refers either to Jurisdiction or to AIA. The failure of convergence criteria in this sense tended to occur only for LMGs with very few observations, i.e. few geographic units with at least a specified minimum number of sampled voting-age LMG persons. However, some of the excluded cases did in fact have convergent but anomalous models, associated with small Fisher information and very poorly estimated parameters, and we thought it quite proper to exclude them on that account. The minimum-observation criteria and model-convergence criteria summarized below were in a sense ad hoc business rules, but represented our best experience from 2012 data in distinguishing useful from ill-fitted models, and worked equally well for models based on Jurisdictions as for those based on AIAs.

F.1 Jurisdiction Model Algorithm

For Jurisdiction models, we first divided the LMGs into three broader racial groups: Asian LMGs, AIAN LMGs, and the Hispanic LMG. Next, within these three general groups of LMGs, we developed three levels of the model with varying number of covariates. For each class, we created a *full, medium,* and *no-covariate* model level. For each LMG, we attempted to fit the model starting with the *full* option. Based on specified fit metrics and a minimum observation criterion, we accepted or rejected the fit of the *full* model. If we rejected the fit, we continued the model fitting process and attempted the *medium* model. Again, we accepted or rejected the fit of the model based on fit metric and minimum observation criteria. If we rejected the model fit of the *medium* model, we repeated the process using the no-covariate model. For a given LMG, if the fit of this final level of model was rejected, then we determined that no small area model was possible and used the direct survey-weighted estimates in the estimation of the quantities used in the coverage determinations.

Minimum Observation Criteria:

- *Full* model: At least 50 observations (areas) with ≥ 5 voting-age LMG persons
- *Medium* model: At least 50 observations (areas) with ≥ 3 voting-age LMG persons
- *No-covariate* model: At least 25 observations (areas) with ≥ 1 voting-age LMG persons

Model Convergence Criteria:

- The optimization function used to compute the MLE returned a "converged" code
- The minimum eigenvalue of the Hessian matrix $> 1.e-4$
- The absolute value of all model parameters is ≤ 11.5
- The estimated standard error of all model parameters is ≤ 6

Jurisdiction Level Models:

Asian LMGs

- *Full Model*:
  - Citizenship Predictors: (C13, C14)
  - LEP Predictors: (C13, C15, C2)
  - Observations used to fit model: Jurisdictions with ≥ 5 voting-age LMG persons
- *Medium Model*:
  - Citizenship Predictors: (C13)
  - LEP Predictors: (C15)
  - Observations used to fit model: Jurisdictions with ≥ 3 voting-age LMG persons
- *No-covariate Model*:
  - Citizenship Predictors: none
  - LEP Predictors: none
  - Observations used to fit model: Jurisdictions with ≥ 1 voting-age LMG persons

AIAN LMGs

- *Full Model*:
  - Citizenship Predictors: (C13, C14)
  - LEP Predictors: (C6, C15)

- Observations used to fit model: Jurisdictions with ≥ 5 voting-age LMG persons

- *Medium Model*:
  - Citizenship Predictors: (C13)
  - LEP Predictors: (C15)
  - Observations used to fit model: Jurisdictions with ≥ 3 voting-age LMG persons

- *No-covariate Model*:
  - Citizenship Predictors: none
  - LEP Predictors: none
  - Observations used to fit model: Jurisdictions with ≥ 1 voting-age LMG persons

Hispanic LMG

- *Full Model*:
  - Citizenship Predictors: (C13, C14)
  - LEP Predictors: (C3, C5, C13, C15, C1, C2)
  - Observations used to fit model: Jurisdictions with ≥ 5 voting-age LMG persons

The following are the final Jurisdiction models used for each of the corresponding LMGs after applying the Jurisdiction model algorithm.

Full Model LMGs:
Asian Indian, Bangladeshi, Cambodian, Chinese, Filipino, Hmong, Indonesian, Japanese, Korean, Laotian, Malaysian, Pakistani, Sri Lankan, Thai, Vietnamese, Other Asian, Apache, Blackfeet, Cherokee, Chippewa, Choctaw, Creek, Iroquois, Lumbee, Mexican American Indian, Navajo, Pueblo, Sioux, South American Indian, All other AI tribes, AI tribes not specified, Tlingit-Haida, AI or AN tribes not specified, Hispanic.

Medium Model LMGs:
Canadian and French Indian, Central American Indian, Chickasaw, Comanche, Cree, Crow, Delaware, Hopi, Menominee, Osage, Ottawa, Paiute, Potawatomi, Seminole, Shoshone, Spanish American Indian, Yaqui, Alaska Athabascan, Inupiat, Alaskan Native Tribes not specified

No-covariate Model LMGs:
Arapaho, Colville, Houma, Pima, Puget Sound Salish, Tohono O'Odham, Yakama, Yuman, Aleut, Tsimshian, Yup'ik

No Model (Direct Estimate) LMGs:
Cheyenne, Kiowa, Ute

F.2 AIA Model Algorithm

For AIAN AIA models, we used a similar but simplified approach to the Jurisdiction model algorithm. For each AIAN LMG, we started with a *medium* model that included two covariates, and attempted to fit that model. If the model met certain fit and observation criteria, it was accepted. If not, we attempted to fit a *no-covariate* model using similar criteria. If neither model level met the criteria for a given LMG, we determined that no small area model was possible and used the direct survey-weighted estimates in the estimation of the determination criteria.

Minimum Observation Criteria:

- *Medium Model*: At least 80 observations (areas) with ≥ 1 voting-age LMG persons
- *No-covariate Model*: At least 25 observations (areas) with ≥ 1 voting-age LMG persons

Model Convergence Criteria:

- The optimization function used to compute the MLE returned a "converged" code
- The minimum eigenvalue of the Hessian matrix $> 1.e - 4$
- The absolute value of all parameters is ≤ 11.5
- The estimated standard error of all model parameters is ≤ 6

Jurisdiction Model Levels:

- *Medium Model*:
  - Citizenship Predictors: (C6)
  - LEP Predictors: (C4)
  - Observations used to fit model: Jurisdictions with ≥ 1 voting-age LMG persons
- *No-covariate Model*:
  - Citizenship Predictors: none
  - LEP Predictors: none
  - Observations used to fit model: Jurisdictions with ≥ 1 voting-age LMG persons

The following are the final AIA models used for each of the corresponding LMGs after applying the AIA model algorithm.

Medium Model LMGs:
All other AI tribes, AI tribes not specified, Aleut, Inupiat, AI or AN tribes not specified

No-covariate Model LMGs:
Apache, Canadian and French Indian Cherokee, Chippewa, Choctaw, Cree, Delaware, Hopi, Iroquois, Kiowa, Mexican American Indian, Navajo, Paiute, Pima, Potawatomi, Pueblo, Puget Sound Salish, Seminole, Shoshone, Sioux, South American Indian, Ute, Yuman, Alaska Athabascan, Tlingit-Haida, Yup'ik, Alaskan Native Tribes not specified

No Model (Direct Estimate) LMGs:
Arapaho, Blackfeet, Central American Indian, Cheyenne, Chickasaw, Colville, Comanche, Creek, Crow, Houma, Lumbee, Menominee, Osage, Ottawa, Spanish American Indian, Tohono O'Odham, Yakama, Yaqui, Tsimshian

G Hybrid BRR-Bootstrap MSPE Estimation Methodology, Technical Details

First, we define a few additional notations for convenience. Letting the superscript $A = C, L$, or I refer generically to one of the CIT, LEP, or ILL levels, recall the notation

$$N_{jg}^A \quad , \qquad \hat{N}_{jg}^A \quad , \qquad \tilde{N}_{jg}^A \quad , \qquad n_{jg}^A \qquad \text{for} \quad A \in \{C, L, I\}.$$

Define the scaled sampled counts as

$$Y_{jg}^A \; = \; n_{jg}^V \, \frac{\hat{N}_{jg}^A}{\hat{N}_{jg}^V} \qquad \text{for} \quad A \in \{C, L, I\}.$$

Then define random effect parameters and estimators that are aligned with CIT, LEP, ILL:

$$\pi_{jg}^C = 1 - \omega_{jg,1} \quad , \qquad \pi_{jg}^L = \omega_{jg,3} + \omega_{jg,4} \quad , \qquad \pi_{jg}^I = \omega_{jg,4}$$

$$\hat{\pi}_{jg}^C = 1 - \hat{\omega}_{jg,1} \quad , \qquad \hat{\pi}_{jg}^L = \hat{\omega}_{jg,3} + \hat{\omega}_{jg,4} \quad , \qquad \hat{\pi}_{jg}^I = \hat{\omega}_{jg,4},$$

and refer to these as π_{jg}^A or $\hat{\pi}_{jg}^A$ for $A \in \{C, L, I\}$.

To simplify notation we drop the LMG subscript g from all mathematical notation and assume that the LMG is fixed throughout. In our framework N_j^V refers to the (unobserved) voting-age population in LMG g in Jurisdiction j, and is treated as a nonrandom quantity. Populations N_j^A for $A = C, L, I$ are the true but unobserved voting-age population in (j, g) respectively CIT, LEP, and ILL, and these populations are treated as unobserved random variables. Equivalently, the random-effect proportions $N_j^A / N_j^V \equiv \pi_j^A$ are converted to probabilities of disjoint categories, creating the probability-vector of mixed-effect parameters $\underline{\omega}_j \; = \; (\omega_{1j}, \omega_{2j}, \omega_{3j}, \omega_{4j})$ modeled in terms of predictors in (7) and appearing as underlying category probabilities in (6). The assignment of the N_j^V persons into nested categories C, L, I is viewed as being the result of multinomial randomization at the person level (as in the 'pseudo-randomization' model of Oh and Scheuren [1983]) superimposed onto the fixed finite population N_j^V. The variables N_j^V are viewed nonrandom, while N_j^A / N_j^V are random. Similarly, the sample sizes n_j^V are treated as nonrandom[1], while the partition of the sample into C, L, I individuals, along with all of the survey-weighted population estimators $\hat{N}_j^V$, $\hat{N}_j^A$, are treated as random.

The form of the predictions of the CIT, LEP, and ILL totals is given in equation (11) by $\tilde{N}_j^A \; = \; n_j^A + (\hat{N}_j^V - n_j^V)\, \hat{\pi}_j^A$. The idea behind the hybrid BRR and Bootstrap variance method is to estimate the variability of predictors $\tilde{N}_j^A$ in terms of levels of error from $\hat{N}_j^V$ in estimating N_j^V and from $\hat{\pi}_j^A$ in predicting π_j^A for $A = \{C, L, I\}$. The estimates $\tilde{N}_j^A$ aim to predict $N_j^A = N_j^V \cdot \pi_j^A$,

[1] This is somewhat artificial, since the ACS is certainly not stratified on Jurisdiction within LMG. However, the BRR idea applied in obtaining SDR variances effectively treats the sample sizes n_j^V as fixed and describes the sampling variability only through varying the weights of the fixed set of sampled individuals within (j, g).

which according to our model contains π_j^A as a random domain effect. The overall measure of prediction error to be estimated is the Mean Squared Prediction Error

$$\text{MSPE}(\tilde{N}_j^A) = E(\tilde{N}_j^A \,-\, N_j^V\, \pi_j^A)^2. \tag{12}$$

The expectation is defined both over the ACS sampling design and the model, conditionally given n_j^V and the covariates. This MSPE is decomposed as in Analysis of Variance, based on the idea of estimating the separate errors at the N_j^V and π_j^A levels.

The MSPE (12) is actually estimated in a modified form replacing $\tilde{N}_j^A \;=\; n_j^A + (\hat{N}_j^V - n_j^V)\,\hat{\pi}_j^A$ with $\tilde{N}_{j,alt}^A \;=\; Y_j^A + (\hat{N}_j^V - n_j^V)\,\hat{\pi}_j^A$, that is, replacing the sampled counts n_j^A in the prediction formula with the survey-weighted, scaled counts Y_j^A. This is done because we do not have a design-based variance calculation for n_j^A, nor do we model the joint distribution of n_j^A and Y_j^A. For larger Jurisdictions, $\text{MSPE}(\tilde{N}_{j,alt}^A)$ will provide a close approximation of $\text{MSPE}(\tilde{N}_j^A)$ because n_j^A and Y_j^A will be very similar, while for smaller Jurisdictions the replacement of $\tilde{N}_j^A$ by $\tilde{N}_{j,alt}^A$ represents an alternative choice of predictor but not the only reasonable choice.

A key assumption underlying the development of our hybrid variance estimation theory is that sample survey estimates of the LMG voting-age totals for a Jurisdiction or AIA are independent of the scaled sample sizes Y_j^A and the random effects π_j^A.

$$\textbf{Assumption:} \quad \hat{N}_j^V \;\text{ and }\; \big((Y_j^C, Y_j^L, Y_j^I), (\pi_j^C, \pi_j^L, \pi_j^I)\big) \;\text{ are independent for all }\; j. \tag{13}$$

This assumption requires some discussion. First, the estimator $\hat{N}_j^V$ of the unknown constant population size N_j^V is defined purely through the ACS random sample of size n_j^V, using survey weights that we also treat as fixed, despite the fact that they are developed in multiple stages of nonresponse adjustment. The random proportions π_j^A and associated domain sample sizes n_j^A and ratios N_j^A/N_j^V are regarded as the result of a population-level random assignment of (voting-age) persons and their associated weights to nested categories C, L, I, resulting in estimators $\hat{N}_j^A/\hat{N}_j^V$. In that sense the independence of sampling from category-assignment makes plausible the independence of $\hat{N}_j^V$ from $\{n_j^A, N_j^A/N_j^V\}$. The approximate independence of $\hat{N}_j^V$ from $\hat{N}_j^A/\hat{N}_j^V$ is less clear, since the latter does involve the randomness of sampling with respect to the rather complicated system of individual weights; however even if the set of weights of the n_j^V sampled individuals were fixed (and $\hat{N}_j^V$ is defined as their sum), the ratio-estimator $\hat{N}_j^A/\hat{N}_j^V$ involves the proportion of that weight-sum assigned to individuals in category A, and in that sense seems much more associated with the random category-assignment than with the overall sum of n_j^V sampled weights. So far, the only provable assertion related to this assumption is that if the N_j^V population members (including those n_j^V that are sampled) each have random probability vector $\{\omega_j\}_{j=1}^4$ of falling into the categories $VOTAG \cap CIT^c, CIT \cap LEP^c, LEP \cap ILLIT^c$, and $ILLIT$, then $cov(Y_j^A, \hat{N}_j^V) = 0$.

Further developments are based on the decomposition

$$\tilde{N}_{j,alt}^A \,-\, \hat{N}_j^V\,\pi_j^A \;=\; \big\{Y_j^A - n_j^V\hat{\pi}_j^A \,+\, N_j^V(\hat{\pi}_j^A - \pi_j^A)\big\} \,+\, (\hat{N}_j^V - N_j^V)(\hat{\pi}_j^A - \pi_j^A) \tag{14}$$

Under Assumption (13) and the unbiasedness of $\hat{N}_j^V$ for N_j^V,

$$E\Big(\tilde{N}_{j,alt}^A \,-\, \hat{N}_j^V\,\pi_j^A\Big)^2 \,=\, E\Big(Y_j^A - n_j^V\,\hat{\pi}_j^A \,+\, N_j^V\,(\hat{\pi}_j^A - \pi_j^A)\Big)^2 + \mathrm{Var}(\hat{N}_j^V)\,E((\pi_j^A - \hat{\pi}_j^A)^2), \tag{15}$$

and similarly

$$E\Big(\tilde{N}_{j,alt}^A \,-\, N_j^V\,\pi_j^A\Big)^2 \,=\, E\Big(Y_j^A - n_j^V\,\hat{\pi}_j^A + N_j^V\,(\hat{\pi}_j^A - \pi_j^A)\Big)^2 + \mathrm{Var}(\hat{N}_j^V)\,E((\hat{\pi}_j^A)^2). \tag{16}$$

Then taking the difference between equations (16) and (15), we get

$$\begin{aligned}
\mathrm{MSPE}(\tilde{N}_{j,alt}^A) &= E\Big(\tilde{N}_{j,alt}^A \,-\, N_j^V\,\pi_j^A\Big)^2 \\
&= E\Big(Y_j^A - n_j^V\,\hat{\pi}_j^A + N_j^V\,(\hat{\pi}_j^A - \pi_j^A)\Big)^2 + \mathrm{Var}(\hat{N}_j^V)\,E((\hat{\pi}_j^A)^2) \qquad (17)\\
&= E\Big(\tilde{N}_{j,alt}^A \,-\, \hat{N}_j^V\,\pi_j^A\Big)^2 - \mathrm{Var}(\hat{N}_j^V)\,E((\pi_j^A - \hat{\pi}_j^A)^2) + \mathrm{Var}(\hat{N}_j^V)\,E((\hat{\pi}_j^A)^2) \qquad (18)\\
&= E\Big(\tilde{N}_{j,alt}^A \,-\, \hat{N}_j^V\,\pi_j^A\Big)^2 + \mathrm{Var}(\hat{N}_j^V)\,E\Big((\hat{\pi}_j^A)^2 - (\pi_j^A - \hat{\pi}_j^A)^2\Big). \qquad (19)
\end{aligned}$$

The method of drawing parametric bootstrap samples begins with first estimating the maximum likelihood estimated parameters $\hat{\theta}^{(r)}$ for each of $r = 0, 1, \ldots, R$ sets of unit-level replicate weights, where $R \leq 80$ and $r = 0$ denotes the set of final weights (those used for point estimation). Each set of weights r corresponds to a set of estimates $(Y_j^{C(r)}, Y_j^{L(r)}, Y_j^{I(r)}), \hat{N}_j^{V(r)}$ generated by using the r-replicate weights for the same set of n_j^V sampled voting-age individuals in the j'th Jurisdiction.[2] Then, for each r and the maximum likelihood estimated parameters $\hat{\theta}^{(r)}$, B independent sets of variables $(\underline{\omega}_j^{*(b,r)}, (Y_j^{A*(b,r)},\ A = C, L, I))$ are generated, according to the models

$$\underline{\omega}_j^{*(b,r)} \sim \mathrm{Dirichlet}\Big(\hat{\tau}_j^{(r)}, (1-\hat{\mu}_j^{(r)}),\, \hat{\mu}_j^{(r)}(1-\hat{\nu}_j^{(r)}),\, \hat{\mu}_j^{(r)}\hat{\nu}_j^{(r)}(1-\hat{\rho}^{(r)}),\, \hat{\mu}_j^{(r)}\hat{\nu}_j^{(r)}\hat{\rho}^{(r)}\Big)$$

and

$$(n_j^V - Y_j^{*C(b,r)},\, Y_j^{*C(b,r)} - Y_j^{*L(b,r)},\, Y_j^{*L(b,r)} - Y_j^{*I(b,r)},\, Y_j^{*I(b,r)}) \ \sim \mathrm{Multinom}(n_j^V,\, \underline{\omega}_j^{*(b,r)}).$$

From these bootstrapped quantities, new maximum-likelihood bootstrap parameter estimates are made from the data $(\mathbf{X}_j^C, \mathbf{X}_j^L, Y_j^{*A(b,r)},\ A = C, L, I,)$, and are denoted $\hat{\theta}_j^{*(b,r)}$. These lead to bootstrapped estimates $\hat{\omega}_{j,k}^{*(b,r)},\ k = 1, \ldots, 4,$ from a formula like (10), and to predictions

[2] A few of the unit replicate weights generated by SDR in ACS turn out to be negative, and some of the generated 'estimates' $\hat{N}_j^{V(r)}$ in small Jurisdictions for some LMGs are either 0 or less than the fixed sample sizes n_j^V. When $\hat{N}_j^{V(r)} \leq 0$, the values n_j^V, $Y_j^{*A(b,r)}$ are set to 0; when $0 < \hat{N}_j^{V(r)} < n_j^V$, the value n_j^V is left alone but $\hat{N}_j^{V(r)}$ is replaced by n_j^V. Thus, the values n_j^V that we have treated notationally as constant over all weight and bootstrap replicates, do depend on r through being forced to 0 when $\hat{N}_j^{V(r)} \leq 0$.

$$
\begin{pmatrix} \tilde{N}^{*C(b,r)}_{j,alt} \\ \tilde{N}^{*L(b,r)}_{j,alt} \\ \tilde{N}^{*I(b,r)}_{j,alt} \end{pmatrix} = \begin{pmatrix} Y^{*C(b,r)}_{j} \\ Y^{*L(b,r)}_{j} \\ Y^{*I(b,r)}_{j} \end{pmatrix} + (\hat{N}^{V(r)}_j - n^V_j) \begin{pmatrix} 1 - \hat{\omega}^{*(b,r)}_{j,1} \\ \hat{\omega}^{*(b,r)}_{j,3} + \hat{\omega}^{*(b,r)}_{j,4} \\ \hat{\omega}^{*(b,r)}_{j,4} \end{pmatrix}. \tag{20}
$$

The computing formulas for (19) are defined as follows. Denote $\left(\pi^{*C(b,r)}_j, \pi^{*L(b,r)}_j, \pi^{*I(b,r)}_j\right) = \left(1 - \omega^{*(b,r)}_{j,1}, \omega^{*(b,r)}_{j,3} + \omega^{*(b,r)}_{j,4}, \omega^{*(b,r)}_{j,4}\right)$, and analogously for the estimated versions $\hat{\pi}^{*A(b,r)}_j$. Define the bootstrapped prediction errors

$$
e^{*A(b,r)}_j \;=\; \tilde{N}^{*A(b,r)}_{j,alt} \;-\; \hat{N}^{V(r)}_j \pi^{*A(b,r)}_j \quad , \qquad \text{for} \quad A = C, L, I. \tag{21}
$$

First, to estimate the mean of $\tilde{N}^A_{j,alt} \;-\; \hat{N}^V_j\, \pi^A_j$, we define a replicate-weighted and bootstrapped bias term by

$$
\texttt{Bias}^A_j \;=\; \frac{1}{(R+1)} \sum_{r=0}^{R} \frac{1}{B} \sum_{b=1}^{B} e^{*A(b,r)}_j \quad , \qquad \text{for} \quad A = C, L, I. \tag{22}
$$

Next we develop an estimator of 'within' variance (expected conditional variance given $\hat{N}^V_j$) for $\tilde{N}^A_{j,alt} \;-\; \hat{N}^V_j \pi^A_j$, equal to

$$
E\Big(\tilde{N}^A_{j,alt} \;-\; \hat{N}^V_j \pi^A_j \;-\; E(\tilde{N}^A_{j,alt} \;-\; \hat{N}^V_j \pi^A_j | \, \hat{N}^V_j)\Big)^2.
$$

This estimator, the 'within' terms of the MSPE estimate, is given by the replicate-weighted average of the bootstrap variances,

$$
\texttt{Within}^A_j \;=\; \frac{1}{R+1} \sum_{r=0}^{R} \mathrm{var}(\{e^{*A(b,r)}_j\}_{b=1}^{B}) \quad , \qquad \text{for} \quad A = C, L, I \tag{23}
$$

where $\mathrm{var}(\mathbf{w})$ denotes the sample variance of the entries of the finite-dimensional vector $\mathbf{w}$. The third term contributing to the MSPE estimate has as its target

$$
E\left(E(\tilde{N}^A_{j,alt} \;-\; \hat{N}^V_j \pi^A_j \,\Big|\, \hat{N}^V_j) \;-\; E(\tilde{N}^A_j - \hat{N}^A_j \pi^A_j)\right)^2.
$$

To derive the estimator, we view $E(\tilde{N}^A_{j,alt} \,|\, \hat{N}^V_j)$ as a nonlinear function of $\hat{N}^V_j$ into which we can plug replicate-weighted estimates $\hat{N}^{(r)}_j$. The expression inside the square term in the last displayed equation is estimated by bootstrap averages, while the outer expected square is a variance estimated via the SDR formula, providing the 'between' terms of the MSPE estimates,

$$
\texttt{Between}^A_j \;=\; \frac{4}{R} \sum_{r=1}^{R} (e^{*A(+,r)}_j - e^{*A(+,0)}_j)^2 \quad , \qquad \text{for} \quad A = \{C, L, I\} \tag{24}
$$

where

$$e_j^{*A(+,r)} \;\equiv\; \frac{1}{B}\sum_{b=1}^{B} e_j^{*A(b,r)}.$$

Finally, the quantity $\mathrm{Var}(\hat{N}_j^V)\, E\Big((\hat{\pi}_j^A)^2 - (\pi_j^A - \hat{\pi}_j^A)^2\Big)$ in equation (19) is estimated via replication in the term

$$\widehat{\mathrm{Var}}(\hat{N}_j^V) = \frac{4}{R}\sum_{r=1}^{80}\Big(\hat{N}_j^{V(r)} - \hat{N}_j^V\Big)^2 \tag{25}$$

and via replicate and bootstrap averaging in the term

$$\mathtt{DiffSq}_j^A = \frac{1}{BR}\sum_{r=1}^{R}\sum_{b=1}^{B}\ \Big((\hat{\pi}_j^{*A(b,r)})^2 - (\pi_j^{*A(b,r)} - \hat{\pi}_j^{*A(b,r)})^2\Big). \tag{26}$$

Putting equations (22) – (26) together gives us the estimate of (19).

$$\widehat{\mathrm{MSPE}}(\tilde{N}_{j,alt}^A) = (\mathtt{Bias}_j^A)^2 + \mathtt{Within}_j^A + \mathtt{Between}_j^A + \widehat{\mathrm{Var}}(\hat{N}_j^V)\ \cdot\ \mathtt{DiffSq}_j^A. \tag{27}$$

To reduce computation time, in the case of some LMGs, we calculated these terms using a smaller number $K < 80$ of the replicate weight sets. Denote these indices by $r_1, \ldots, r_K \in \{1, \ldots, 80\}$ and assume they are sampled randomly (equiprobably, without replacement). When this is done, formulas (22) and (23) change only by replacing the average $(R+1)^{-1}\sum_{r=0}^{R}$ by $(K+1)^{-1}\sum_{k=0}^{K}$ and all of the r indices by r_k. However, formula (24) changes further if replicate-weight columns are sampled in this way. To see how, note that

$$\frac{4}{R}\sum_{r=1}^{R}\ \Big(e_j^{*A(+,r)} - e_j^{*A(+,0)}\Big)^2 \;=\; \frac{4}{R}\sum_{r=1}^{R}\ \Big(e_j^{*A(+,r)} - \bar{e}_j^{*A}\Big)^2 \;+\; 4\,\Big(\bar{e}_j^{*A} \;-\; e_j^{*A(+,0)}\Big)^2$$

where $\bar{e}_j^{*A} \;=\; R^{-1}\sum_{r=1}^{R} e_j^{*A(+,r)}$, and the right-hand side of the last expression

$$=\; 4\,\frac{R-1}{R}\,\mathrm{var}(\{e_j^{*A(+,r)}\}_r) \;+\; 4\,(\bar{e}_j^{*A} \;-\; e_j^{*A(+,0)})^2. \tag{28}$$

By simple random sampling variance formulas, it is easy to check that (28) is equal to the expectation over samples $\{r_1, \ldots, r_K\}$ of

$$4\,\frac{R-1}{R}\,\mathrm{var}(\{e_j^{*A(+,r_k)}\}_{k=1}^K) \;+\; 4\,\Big(\frac{1}{K}\sum_{k=1}^{K} e_j^{*A(+,r_k)} \;-\; e_j^{*A(+,0)}\Big)^2 \;-\; \frac{4R(R-K)}{R^2\,K}\,\mathrm{var}(\{e_j^{*A(+,r_k)}\}_{k=1}^K)$$

$$=\; 4\,\frac{K-1}{K}\,\mathrm{var}(\{e_j^{*A(+,r_k)}\}_{k=1}^K) \;+\; 4\,\Big(\frac{1}{K}\sum_{k=1}^{K} e_j^{*A(+,r_k)} \;-\; e_j^{*A(+,0)}\Big)^2. \tag{29}$$

This last expression, for $A = C, L, I$, replaces formula (24) when $K < R$ replicate-weight columns are sampled equiprobably without replacement.

So far, we have only discussed the estimated MSPE for estimated totals $(\tilde{N}^C_{j,alt}, \tilde{N}^L_{j,alt}, \tilde{N}^I_{j,alt})$, but we also need to estimate the MSPE for the LEP and ILL proportions respectively. We estimate $\text{MSPE}(\tilde{N}^L_{j,alt}/\hat{N}^C_{j+})$ by the approximation $\text{MSPE}(\tilde{N}^L_{j,alt})\Big/(\hat{N}^C_{j+})^2$. Therefore our estimate is

$$\widehat{\text{MSPE}}\Big(\tilde{N}^L_{j,alt}\Big/\hat{N}^C_{j+}\Big) = \widehat{\text{MSPE}}(\tilde{N}^L_{j,alt})\Big/(\hat{N}^C_{j+})^2. \tag{30}$$

The denominator of the LEP proportion estimate in (30) is the count of all citizens in a Jurisdiction, and it is generally estimated much more accurately than the LEP count within the LMG. A refinement of the approximate formula could be calculated by linearizing the ratio estimator, but because the term contributed by the variance of the denominator is generally so much smaller the term from (27) with $A = L$, we have not found it worthwhile to do so.

Next, to estimate $\text{MSPE}(\tilde{N}^I_{j,alt}\Big/\tilde{N}^L_{j,alt})$ we first note that

$$\begin{aligned}\text{MSPE}(\tilde{N}^I_{j,alt}\Big/\tilde{N}^L_{j,alt}) &= E\Big(\frac{\tilde{N}^I_{j,alt}}{\tilde{N}^L_{j,alt}} - \frac{N^V_j\pi^I_j}{N^V_j\pi^L_j}\Big)^2 \\ &= E\Big(\frac{\tilde{N}^I_{j,alt}}{\tilde{N}^L_{j,alt}} - \frac{\pi^I_j}{\pi^L_j}\Big)^2,\end{aligned} \tag{31}$$

so that N^V_j cancels in the numerator and denominator of the second term. We can then create an estimate from the bootstrap replicates in a similar way to (27). Define a prediction error equation analogous to Equation (21)

$$e_j^{*\text{IR}(b,r)} = \frac{\tilde{N}^{*I(b,r)}_{j,alt}}{\tilde{N}^{*L(b,r)}_{j,alt}} - \frac{\pi_j^{*I(b,r)}}{\pi_j^{*L(b,r)}}. \tag{32}$$

Using $e_j^{*\text{IR}(b,r)}$ in place of $e_j^{*\text{A}(b,r)}$, calculate the analogous Bias, Within, and Between terms in equations (22)–(24) and denote them $\texttt{Bias}^{IR}_j$, $\texttt{Within}^{IR}_j$, and $\texttt{Between}^{IR}_j$. The MSPE estimate becomes

$$\widehat{\text{MSPE}}(\tilde{N}^I_{j,alt}\Big/\tilde{N}^L_{j,alt}) = (\texttt{Bias}^{IR}_j)^2 + \texttt{Within}^{IR}_j + \texttt{Between}^{IR}_j. \tag{33}$$

In equation (31) both numerator and denominator are highly variable, and therefore the linearizatition of the ratio would not give an accurate approximation and is not needed because the factor N^V_j cancels out from numerator and denominator of the target Illiteracy rate. However, we do use a linearized form to estimate the MSPE of the illiteracy proportion for calculating MSPEs of aggregated (State or National) totals, because the terms N^V_j do not cancel out of the target aggregated illiteracy rate estimators.

Lastly, we exhibit the formulas for estimating the MSPEs of state and national predictions, which are based on aggregations of Jurisdiction-level predictions. Let the set $\mathcal{S}$ denote the Jurisdictions in a particular aggregation, such as a state. Then one goal is to estimate MSPE $\left(\sum_{j\in\mathcal{S}} \tilde{N}^{A}_{j,alt} \right)$.

Moving next to the estimation of MSPEs of state-level ratios, define

$$\tilde{\theta}^{\mathrm{IR}}_{\mathcal{S}} = \frac{\sum_{j\in\mathcal{S}} \tilde{N}^{I}_{j,alt}}{\sum_{j\in\mathcal{S}} \tilde{N}^{L}_{j,alt}}. \tag{34}$$

Then

$$\begin{aligned}
\mathrm{MSPE}(\tilde{\theta}^{\mathrm{IR}}_{\mathcal{S}}) &= E\left(\tilde{\theta}^{\mathrm{IR}}_{\mathcal{S}} - \frac{\sum_{j\in\mathcal{S}} N^V_j \pi^I_j}{\sum_{j\in\mathcal{S}} N^V_j \pi^L_j}\right)^2 \\
&= E\left(\sum_{j\in\mathcal{S}} \left(\tilde{\theta}^{\mathrm{IR}}_{\mathcal{S}} N^V_j \pi^L_j - N^V_j \pi^I_j\right) / \sum_{j\in\mathcal{S}} N^V_j \pi^L_j\right)^2 \\
&= E\left(\left(\sum_{j\in\mathcal{S}} \left\{\tilde{N}^I_{j,alt} - \tilde{\theta}^{\mathrm{IR}}_{\mathcal{S}} \tilde{N}^L_{j,alt}\right\} - \sum_{j\in\mathcal{S}} N^V_j \left(\pi^I_j - \tilde{\theta}^{\mathrm{IR}}_{\mathcal{S}} \pi^L_j\right)\right) / \sum_{j\in\mathcal{S}} N^V_j \pi^L_j\right)^2.
\end{aligned}$$

By linearization of the ratio estimator, this can be approximated by

$$\begin{aligned}
&\approx E\left(\sum_{j\in\mathcal{S}} \left(\left\{\tilde{N}^I_{j,alt} - \tilde{\theta}^{\mathrm{IR}}_{\mathcal{S}} \tilde{N}^L_{j,alt}\right\} - N^V_j \left(\pi^I_j - \tilde{\theta}^{\mathrm{IR}}_{\mathcal{S}} \pi^L_j\right)\right)\right)^2 / \left(\sum_{j\in\mathcal{S}} N^V_j \pi^L_j\right)^2 \\
&= E\left(\sum_{j\in\mathcal{S}} \tilde{N}^{IR}_{j,alt} - \pi^{IR}_j N^V_j\right)^2 / \left(\sum_{j\in\mathcal{S}} N^V_j \pi^L_j\right)^2 \\
&= \mathrm{MSPE}\left(\sum_{j\in\mathcal{S}} \tilde{N}^{IR}_{j,alt}\right) / \left(\sum_{j\in\mathcal{S}} N^V_j \pi^L_j\right)^2,
\end{aligned} \tag{35}$$

where

$$\tilde{N}^{IR}_{j,alt} = \tilde{N}^I_{j,alt} - \tilde{\theta}^{\mathrm{IR}}_{\mathcal{S}} \tilde{N}^L_{j,alt} \quad , \qquad \pi^{IR}_j = \pi^I_j - \tilde{\theta}^{\mathrm{IR}}_{\mathcal{S}} \pi^L_j.$$

Therefore we also need to estimate MSPE $\left(\sum_{j\in\mathcal{S}} \tilde{N}^{IR}_{j,alt}\right)$. In general, for $A = C, L, I, IR$, we find as in deriving (19) that

$$
\begin{aligned}
\text{MSPE}\left(\sum_{j\in\mathcal{S}}\tilde{N}^A_{j,alt}\right) &= E\left(\sum_{j\in\mathcal{S}}\left(\tilde{N}^A_{j,alt}-N^V_j\pi^A_j\right)\right)^2 \\
&= E\left(\sum_{j\in\mathcal{S}}\left(\tilde{N}^A_{j,alt}-\hat{N}^V_j\pi^A_j\right)\right)^2 + E\left(\sum_{j\in\mathcal{S}}\left(\hat{N}^V_j-N^V_j\right)\hat{\pi}^A_j\right)^2 \\
&\quad - E\left(\sum_{j\in\mathcal{S}}\left(\hat{N}^V_j-N^V_j\right)\left(\hat{\pi}^A_j-\pi^A_j\right)\right)^2. \qquad (36)
\end{aligned}
$$

To estimate MSPE $\left(\sum_{j\in\mathcal{S}}\tilde{N}^A_{j,alt}\right)$, we define prediction-error variables $\epsilon^{*A(b,r_k)}_{\mathcal{S}}$ analogous to the Jurisdiction-level prediction errors (21) by

$$
\epsilon^{*A(b,r_k)}_{\mathcal{S}} = \sum_{j\in\mathcal{S}} e^{*A(b,r_k)}_j = \sum_{j\in\mathcal{S}} \tilde{N}^{*A(b,r_k)}_{j,alt} - \hat{N}^{V(r_k)}_j \pi^{*A(b,r_k)}_j, \quad \text{for} \quad A=\{C,L,I,IR\}, \qquad (37)
$$

where $b=1,\ldots,B$ indexes the B bootstrap iterations and we sample $K\leq 80$ of the replicate weight sets and denote the indices by $r_1,\ldots,r_K\in\{1,\ldots,80\}$.

The first term of MSPE $\left(\sum_{j\in\mathcal{S}}\tilde{N}^A_{j,alt}\right)$, $E\left(\sum_{j\in\mathcal{S}}\left(\tilde{N}^A_{j,alt}-\hat{N}^V_j\pi^A_j\right)\right)^2$, can be decomposed as in (27) into Bias, Within, and Between terms estimated by

$$
\texttt{Bias}^A_{\mathcal{S}} = \frac{1}{B(K+1)}\sum_{k=0}^{K}\sum_{b=1}^{B}\epsilon^{*A(b,r_k)}_{\mathcal{S}} \qquad (38)
$$

$$
\texttt{Within}^A_{\mathcal{S}} = \frac{1}{K+1}\sum_{k=0}^{K} var\left(\left\{\epsilon^{*A(b,r_k)}_{\mathcal{S}}\right\}_{b=1}^{B}\right) \qquad (39)
$$

$$
\texttt{Between}^A_{\mathcal{S}} = \frac{4(K-1)}{K} var\left(\epsilon^{*A(+,r_k)}_{\mathcal{S}}\right) + 4\left(\frac{1}{KB}\sum_{k=1}^{K}\sum_{b=1}^{B}\left(\epsilon^{*A(b,r_k)}_{\mathcal{S}}-\epsilon^{*A(b,0)}_{\mathcal{S}}\right)\right)^2. \qquad (40)
$$

The formula for $\texttt{Between}^A_{\mathcal{S}}$ has two terms, as in (27), because of the sampling of K replicate weight sets. The final two terms of (36) can be estimated by bootstrap equivalents.

$$
\texttt{Term1}^A_{\mathcal{S}} = \frac{4}{KB}\sum_{k=1}^{K}\sum_{b=1}^{B}\left(\sum_{j\in\mathcal{S}}\hat{\pi}^{*A(b,r)}_j(\hat{N}^{V(r)}_j-\hat{N}^V_j)\right)^2. \qquad (41)
$$

$$\texttt{Term2}_{\mathcal{S}}^{A} = \frac{4}{KB}\sum_{k=1}^{K}\sum_{b=1}^{B}\left(\sum_{j\in\mathcal{S}}(\hat{\pi}_j^{*A(b,r)} - \pi_j^{*A(b,r)})(\hat{N}_j^{V(r)} - \hat{N}_j^{V})\right)^2. \tag{42}$$

Finally, this leads to the estimator of MSPE $\left(\sum_{j\in\mathcal{S}}\tilde{N}_{j,alt}^{A}\right)$ (36), which is given by

$$\widehat{\text{MSPE}}\left(\sum_{j\in\mathcal{S}}\tilde{N}_{j,alt}^{A}\right) = \left(\texttt{Bias}_{\mathcal{S}}^{A}\right)^2 + \texttt{Within}_{\mathcal{S}}^{A} + \texttt{Between}_{\mathcal{S}}^{A} + \texttt{Term1}_{\mathcal{S}}^{A} + \texttt{Term2}_{\mathcal{S}}^{A}. \tag{43}$$

Using the approximation in equation (35), we estimate $\text{MSPE}(\tilde{\theta}_{\mathcal{S}}^{\text{IR}})$ by

$$\widehat{\text{MSPE}}(\tilde{\theta}_{\mathcal{S}}^{\text{IR}}) = \widehat{\text{MSPE}}\left(\sum_{j\in\mathcal{S}}\tilde{N}_{j,alt}^{IR}\right) / \left(\sum_{j\in\mathcal{S}}\hat{N}_j^{V}\hat{\pi}_j^{L}\right)^2. \tag{44}$$

H Lack of Fit Diagnostic, Technical Details

For a given LMG g and a given set of areas (Jurisdictions or AIAs) indexed by i let $\hat{\theta}_{i,\text{MM}}$ and $\hat{\theta}_{i,\text{DS}}$ denote the MM and DS estimates (5) of a parameter of interest θ_i. In our context, the most important parameter of interest is the LEP proportion among voting-age citizens in LMG g and area i, but the considerations of this Section apply equally well to the CIT or ILL proportions among voting-age persons. Let $\mathcal{A} = \{i : n_i^V > 0\}$ denote the set of all areas i that have at least a single sampled person in LMG g.

Using the assumed model and fitted model parameters, take B parametric bootstraps from the model for each of the areas assuming a fixed voting-age sample size of n_i. For each of the B bootstraps re-estimate the model parameters along with the MM and DS estimates for each area. Denote the MM and DS estimates for area i for bootstrap b as $\hat{\theta}_{i,\text{MM}}^{(b)}$ and $\hat{\theta}_{i,\text{DS}}^{(b)}$ respectively.

In order to compute the bootstrap lack of fit diagnostic, do the following:

[1] For the observed data and fitted model, compute:

$$\widehat{SS} = \sum_{i\in\mathcal{A}}\left(\left(\hat{\theta}_{i,\text{MM}} - \hat{\theta}_{i,\text{DS}}\right)\sqrt{n_i^V}\right)^2. \tag{45}$$

[2] Compute the empirical distribution for the parametric-bootstrap replicates:

$$\widehat{SS}^{(b)} = \sum_{i\in\mathcal{A}}\left(\left((\hat{\theta}_{i,\text{MM}}^{(b)} - \hat{\theta}_{i,\text{DS}}^{(b)}\right)\sqrt{n_i^V}\right)^2 \tag{46}$$

for $b = 1, \ldots, B$ using the fitted model.

[3] Compare the observed statistic $\widehat{SS}$ to the reference distribution $\widehat{SS}^{(b)}$ by calculating the bootstrap quantile

$$Q \;=\; \frac{1}{B}\sum_{b=1}^{B} I_{[\widehat{SS}\geq\widehat{SS}^{(b)}]}. \tag{47}$$

Extreme values of Q near 0 (or 1), indicate that the amount of variation in the observed data is respectively less than (greater than) the parametrically bootstrapped data and indicate a lack of fit. See [Ashmead and Slud, 2017] for additional discussion and generalizations of this parametric bootstrap lack of fit diagnostic.