RESEARCH REPORT SERIES
(Statistics #2022-06)

Statistical Methodology (2021) for Voting Rights Act, Section 203 Determinations

Eric Slud[1,2],
Carolina Franco[3,4],
Adam Hall[1],
Joseph Kang[1]

[1]Center for Statistical Research and Methodology, U.S. Census Bureau
[2]Department of Mathematics, University of Maryland College Park
[3]NORC at the University of Chicago
[4]Formerly at the Center for Statistical Research and Methodology, U.S. Census Bureau

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

Report Issued: December 20, 2022

This research was conducted under DMS Project Number 7510051. The statistics were released under DRB clearance number CBDRB-FY23-084.

Statistical Methodology (2021) for Voting Rights Act, Section 203 Determinations

Eric Slud, Carolina Franco, Adam Hall, and Joseph Kang
Center for Statistical Research and Methodology
U.S. Census Bureau

December 20, 2022

Contents

Executive Summary iii

Acknowledgments vi

1 Introduction 1

2 Terminology and Data 3

2.1 Direct Tabulated Outcomes 5

2.2 Covariates 7

2.3 Modeled Outcomes 10

2.4 Extreme Cases Where Models Degenerate 13

3 Model Classes & Modeling Choices 13

3.1 Multinomial Logit-Normal Model 16

3.2 Models Related to Dirichlet-Multinomial (DM) 17

3.3 Predictors & Model Diagnostics 18

3.3.1 Minimum Sample-Size Thresholds for Use of Geographic Units in Estimation 19

3.4 Comparisons of Fitted MLN and DM Models 21

3.5 Frequentist & Bayesian Computational Considerations 23

3.6 Model and Covariate Selection 25

3.6.1 Assessing Predictions for MLN-D Models with Covariates 25

3.6.2 Jurisdictions — Models & Covariates 31

3.6.3 American Indian Areas — Models & Covariates 32

3.6.4 Alaska Native Regional Corporations — Models without Covariates 33

3.7 Limitations of Model Assumptions 34

3.7.1 Testing an Independence Assumption via Correlations 34

4 Variance and Mean-Square Predictor Error (MSPE) Estimation 37

4.1 Comparing Direct with Model-Based Variances 37

4.2 Bayesian Posterior Variances versus SDR-based MSPEs 43

5 Summary **44**

5.1 Differences from 2016 Methodology . . . 46

5.2 Future Research Directions . . . 47

References **47**

A Section 203 of the Voting Rights Act of 1965 **52**

B Determination Flow Chart **56**

C Notations and Model Definitions **57**

C.1 Multinomial Logit-Normal Models . . . 60

C.1.1 Limiting Cases of Infinite Regression Coefficients . . . 60

C.2 Dirichlet-Multinomial Models . . . 62

C.2.1 Special Cases of DM Submodels with no Covariates . . . 63

D Frequentist vs. Bayes Prediction and Variance Estimation **64**

D.1 Relations between Different Prediction and Variance Concepts . . . 66

D.2 MSPEs of Counts from MSPEs of Random Probabilities . . . 69

D.3 Prediction Formulas . . . 70

D.4 Composing Stagewise Predictions . . . 71

D.5 Variance and MSPE Formulas . . . 72

D.6 Variance Calculations for Ratios . . . 74

E Computational Methods **76**

E.1 Adaptive Gaussian Quadrature in MLN Models . . . 76

E.1.1 Gradients and Prediction formulas for MLN Models . . . 79

E.2 Bayesian Computation in MLN Nodels . . . 80

E.3 Computational Comparison of Bayesian and Frequentist Estimates . . . 81

F Prediction Diagnostics **84**

Executive Summary

According to Section 203(b) of the *Voting Rights Act of 1965* beginning in 1975, and as amended in 1982 and 2006, states and political subdivisions must in certain circumstances make voting materials available in languages other than English. 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 decennial census and the most current American Community Survey (ACS). Section 203(b) as amended prescribes that the Director of the Census Bureau shall make these determinations every 5 years, based on the most current population estimates derived from the ACS along with relevant census data. The 2021 determinations released in December 2021 were based solely on 2015-2019 5-year ACS data. As a response to concerns about the timing of the decennial counts, which was affected by the COVID-19 pandemic, the decennial 2020 Census data were not used.

For the determinations, estimates are needed at various levels of geographic aggregation. These levels of geography include states, jurisdictions, American Indian Areas (AIAs), and Alaskan Native Regional Corporations (ANRCs). The nation is partitioned into roughly 8,000 Jurisdictions (7,859 in ACS 2015-2019 5-year data containing at least one voting-age respondent), which are Counties in most states and Minor Civil Divisions (MCDs) in the other states. Other geographic domains relevant to provisions of Section 203(b) are the American Indian Areas (AIAs), of which there are 568 with ACS respondents in 2015-2019, as well as 12 Alaska Native Regional Corporations (ANRCs). All 12 ANRCs had at least one person in the ACS sample.

For purposes of Section 203(b), only the population of voting age (18 or over) persons is relevant. Section 203(b) categorizes voting age persons according to Citizenship, Limited English Proficiency (LEP) and Illiteracy. The 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. People self-identify (in the Census or ACS) as belonging to one or more of 6 distinct racial groups, (each containing several detailed races) and 1 ethnic classification that are then used to define 73 'Language Minority Groups' (LMGs) for purposes of Section 203(b). According to the Voting Rights Act, only Asian and American Indian and Alaska Native languages and Spanish are eligible for coverage. Of the LMGs eligible in 2021, 21 are Asian, 51 are American Indian or Alaska Native (AIAN), and one is Hispanic.

Section 203(b) prescribes generally that states and political subdivisions must provide voting materials in a language other than English for members of a LMG according to the following rules:

(i) A state must do so if the illiteracy rate among citizen voting-age LEP (VACLEP) members of the LMG in the state exceeds the national rate of illiteracy among voting-age citizens (VACIT), **and** the number of VACLEP persons in the LMG is greater than 5% of the total number of VACIT in the state.

(ii) A jurisdiction must do so if the illiteracy rate among VACLEP persons in the LMG and jurisdiction exceeds the national rate of citizen illiteracy **and** the number of VACLEP LMG persons in the jurisdiction is greater than either 10,000 or 5% of the total VACIT population of the jurisdiction.

(iii) All jurisdictions (Counties or MCDs) containing any part of an American Indian Area (AIA) or Alaskan Native Regional Corporation (ANRC) must do so if an AIAN LMG has illiteracy rate among VACLEP AIAN persons of the LMG in the AIA/ANRC that exceeds the national rate of citizen illiteracy **and** the number of VACLEP AIAN persons in the AIA/ANRC and LMG is greater than 5% of the total voting-age citizen AIAN population of the AIA/ANRC.

Special tabulations of weighted survey estimates of state, jurisdiction, AIA, and ANRC voting-age populations cross-classified by citizenship, limited English proficiency, illiteracy, and LMG are available from ACS 5-year data. These tabulations could be used to create direct survey-weighted estimates of all of the ingredients of the 'triggering' 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, and the variability (standard errors) of the direct estimates are often quite large compared to the estimates themselves. Moreover, the standard errors estimated by current ACS methodology are also very unreliable for population counts in such small domains.

For reasons of estimation accuracy, starting in 2011 and again in 2016 and 2021, statistical research on the estimation methodology driving the Section 203(b) determinations has been primarily directed toward model-based 'Small Area Estimation'. Small Area Estimation is devoted to enhancing the precision of estimation through the formulation of statistical models for multiple small areas which 'borrow strength' from one another through shared statistical parameters and through use of auxiliary information.

The main idea of this approach is that geographies within the same LMG behave similarly with respect to the characteristics of interest across different geographies, and with respect to covariates.

The domains used for the small area estimation models are Jurisdictions for each of the LMGs, and AIAs or ANRCs for each of the AIAN LMGs. Statistical models are fitted separately for the different LMGs and types of geography. In addition, the complexity of the model used for a particular LMG and type of geography depends on how many distinct geographic units have ACS respondents for that LMG. This is necessary as the ACS sample for some LMGs and geographic units contains thousands of people, while for other LMGs and units of geography, the ACS sample may contain only a single person.

The general form of model chosen for the 2021 statistical estimation is a Multinomial Logit Normal (MLN) Model, formulated for the nested decreasing subpopulations of voting age persons (VOT), voting age citizens (CIT), voting age citizens who are limited English proficient (LEP), and illiterate limited English proficient voting age citizens (ILL). The MLN model is a random-effects generalization of logistic regression, in which the proportions of CIT persons within VOT, and similarly LEP within CIT and ILL within LEP, are modeled using a logit transformation and random intercepts, as well as predictive covariates.

The covariates used in modeling were computed from the same ACS dataset as the response variable but at higher levels of aggregation. One set of covariates was defined as the higher-geography-level LMG proportion of CIT within VOT, LEP within CIT, and ILL within LEP, for the portion of the State complementary to a Jurisdiction it contains, or the portion of the whole AIAN LMG complementary to an AIA. All other covariates were defined at the level of the geographic unit, without regard to LMG. One such covariate, in all geography types, was the proportion of people speaking a language other than English in the home. Covariates used in Asian and Hispanic LMGs include the proportion of Foreign-born, the average years in US for the foreign-born, and the proportions in coarse age-groups. Covariates used in various AIAN LMGs include the proportions of high-school graduates, of white nonhispanic people, and of people in poverty.

In the most detailed form of the model, the random intercepts for the CIT, LEP and ILL sub-models were jointly normally distributed and dependent. In less data-rich LMGs, the random intercepts were assumed independent. In LMGs with still less data, models of this form with reduced sets of covariates — or with none at all — were fitted. In smaller (AIAN) LMGs in which the submodel CIT rates were uniformly close to 1, or in which the LEP or ILL rates were uniformly close to 0, an even simpler form of model was fitted. This was a common-intercept beta-binomial model with no covariates or random effects, which amounts to fitting a single rate on the pooled LMG data.

The models chosen have been explored extensively in practice data analyses using ACS 2014-2018 5-year data in the same way that the model was ultimately employed on ACS 2015-2019 5-year data. The model has been assessed against the direct domain population estimators obtained from the ACS and to those obtained by a Dirichlet-Multinomial model closely related to the model used in producing 2016 determinations. Model diagnostics were used in selecting covariates for the models and in assessing the suitability of the final models chosen. These analyses are elaborated in this technical documentation. Uncertainty estimation was based on either Markov Chain Monte Carlo computation of posterior variances or a Successive Difference Replication method applied to the modeled estimates, depending on the complexity of the model.

Although all counts and proportions for Jurisdictions, AIAs, and ANRCs cross-classified by LMGs were modeled, direct ACS estimates were used for quantities at higher levels of aggregation, such as state-level estimates, or estimates by jurisdiction that are not cross-classified with LMG. In addition, direct ACS estimates of voting age persons by LMG and geography were used to translate proportion estimates from the models into corresponding population counts. The uncertainty of these direct ACS estimates of voting age person counts was taken into account when computing the variances of the corresponding ILL, LEP, and CIT counts.

Most of the determinations are the same using the model as those that would have been obtained via the direct estimators, but there are some cases in which the model would result in determination where the direct estimates would not, and vice-versa. The direct estimators can be quite volatile and unreliable for domains with small sample sizes, and there are many such domains for LMGs in ANRCs, AIAs, and even Jurisdictions. The model predictions are more stable and result in a substantial decrease in estimates with large Coefficients of Variation (CVs), e.g., CVs > 0.6, and in large overall reductions in Margins of Error (MOEs). More detailed comparisons of the CVs and MOEs are included in the technical report.

There are several ways in which the modeling approach adopted in 2021 differed from that used in 2016. First was the overall class of models chosen, Multinomial Logit Normal in place of Dirichlet-Multinomial. The MLN model has more parameters (because of the general dependence among random intercepts), which were reduced in less data-rich LMGs by assuming the three CIT, LEP and ILL random intercepts to be independent. Second was the choice to model all predictions in Geography by LMG domains (below the level of States), no matter how data-sparse. A third distinction in modeling arose because in 2016, Geography by LMG domains with sample smaller than 5 (or in some cases 3) were not used in fitting LMG-level model parameters, while all Geography by LMG domains with respondents were used in 2021. Finally, the variances of estimated totals and proportions were estimated in 2021 by a combination of Bayesian posterior variances (in the models for Jurisdictions in the largest LMGs) and replicate-weights based on repeated calculation of estimates with alternate weights, while in 2016 the variance estimates were calculated by a hybrid method combining parametric bootstrap and replicate weights.

Acknowledgments

The research and implementation summarized in this report were performed under the general supervision of a Census Bureau Technical Steering Committee including John Abowd, the former Associate Director for Research and Methodology; Eric Slud, Carolina Franco, and Tommy Wright of the Center for Statistical Research and Methodology; Joseph Kang of the Center for Optimization and Data Science; and Mark Asiala of the Decennial Statistical Studies Division. A broader group including that Technical Steering Committee, James Whitehorne of the Redistricting and Voting Rights Data Office, and others, oversaw the legal and logistical aspects of the project. Ahmed Abdulkarim Al-Attar, Joseph Kang and Xiaoyun Lu provided essential technical support with programming and software development, assisted by Patrick Campanello. This research was conducted under DMS Project Number 7510051. The statistics were released under DRB clearance number CBDRB-FY23-084.

1 Introduction

According to Section 203(b) of the *Voting Rights Act of 1965*, 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) as amended prescribes that the Director of the Census Bureau shall make these determinations every 5 years under specific rules [Appendix A] involving the sizes and proportions of designated population subgroups, based on the decennial census and the most current available American Community Survey (ACS) data. The 2021 determinations released in December 2021 were based solely on 2015-2019 5-year ACS data. In response to concerns about the timing of the decennial counts, which was affected by the COVID-19 pandemic, the decennial 2020 Census data were not used.

In 2021, the Director of the Census Bureau made coverage determinations for 73 racial/ethnic Language Minority Groups (LMGs) within roughly 8000 Jurisdictions [Census RVRDO, 2021]. The Jurisdictions constitute an electorally relevant partition of the nation into counties and minor civil divisions (MCDs). There were 7,859 Jurisdictions in ACS 2015-2019 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 three possible ways to be covered by Section 203(b). First, a LMG may meet the state-level coverage criteria, in which case all Jurisdictions in that state are covered for that LMG. [1] 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 or Alaska Native Area-level (AIAN-level) coverage criteria. If an AIAN LMG meets the AIAN-level coverage criteria for a certain AIAN area, then all Jurisdictions that contain all or part of that area 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 (`LEPprop`[2]) is greater than 5 percent; or

[1] However, if a juridiction with LMG VACLEP population less than 5% of its VACIT is in a state with statewide language coverage determination, then that jurisdiction is covered for that LMG only for statewide election materials.

[2] Throughout this report, acronyms and repeatedly used abbreviations are given in upper-case conventional font,

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

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

Criteria for AIA-level (AIA area or ANRC) 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 (`LEPprop`) is greater than 5 percent; and

A2 The illiteracy rate among limited English-proficient voting-age citizens of that LMG in the AIA (`ILLrat`) 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, AIA and ANRC 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 and ANRC. 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 and ANRC. 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 have extremely small populations. 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 2021 coverage determinations based on ACS 2019 5-year data was 1.31%.

In an effort to improve the accuracy of the estimates used to make the coverage determinations, the Census Bureau applied a model-based estimation method in producing coverage determinations in 2011 [Joyce et al., 2012, 2014] and also in 2016 [Slud et al., 2018]. 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. Underlying the modeling approach used in 2011 and 2016 Voting Rights Act (VRA) analyses,

while computer-package and computer-function names and variable names used in modeling or in formal labeling in public data files are rendered in typewriter font.

the beta-binomial and Dirichlet Multinomial models are well known in Bayes and empirical-Bayes analysis [Carlin and Louis, 2009]. An example of the use of beta-binomial in small area estimation is Aitkin et al. [2009]. For the small-area application of the multinomial logit normal models considered and ultimately adopted for the 2021 VRA statistical analyses, important precursors are the papers of Malec et al. [1997], Ghosh et al. [1998], Slud [2004] and, with a hierarchical logit normal to account for multinomial outcomes, Malec [2005]. 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 2021 Section 203(b) determinations.

2 Terminology and Data

The sources of data allowed by Section 203(b) to be used in coverage determinations are ACS and comparable census data. The ACS is an ongoing annual household survey of information about the US population that is used in many different ways. The ACS releases 1-year and 5-year data products. The 5-year products aggregate and re-weight 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 domains in which 1-year estimates are too imprecise for release under Census Bureau statistical quality guidelines. For purposes of coverage determination, 5-year ACS data are used because of the need to estimate population subgroups in small geographic areas. At the time of estimating models used to make 2021 coverage determinations, the 2015-2019 5-year ACS dataset was the most recent available and therefore served as the data source. Model exploration and development were done using 2014-2018 ACS 5-year data, before the 2019 data were released. Decennial 2020 census data were not used, a decision taken because of concerns about timeliness of the release of local-area decennial census counts, which was affected by the COVID-19 pandemic. This differed markedly from the 2011 methodology which did use 2010 decennial census data in producing 2011 coverage determinations, but decennial data were deemed too far out of date to be used in the 2016 determinations apart from having already been incorporated into national Population Estimates.

The Section 203(b) relevant political subdivisions, which we refer to as Jurisdictions (Juris), are Counties in most states and Minor Civil Divisions (MCDs) in eight states (CT, ME, MA, MI, NH, RI, VT, WI). In the ACS 2015-2019 data, there were 7,859 Jurisdictions containing at least one sampled voting-age person, 568 AIAs containing at least one sampled voting-age person, and 12 ANRCs that all had at least one sampled voting-age respondent to the ACS. 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:

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

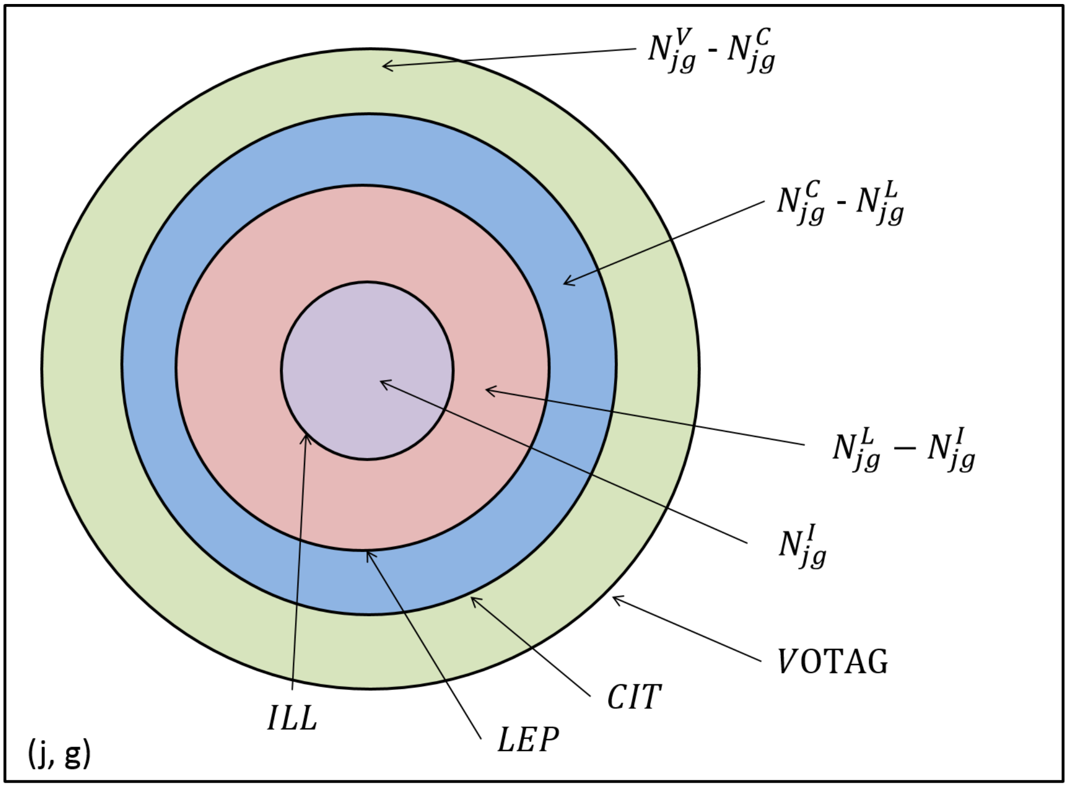
The designations VOT, CIT, LEP and ILL are referred to throughout this report as outcome *Categories*, and these same labels are also applied to the sample counts and survey-weighted estimates within population domains defined by Geography and LMG.

The nesting of subpopulations VOT (= VOTAG) , CIT, LEP, and ILL in this report, is displayed visually in Figure 1. Specifically, all ILL persons are LEP, all LEP persons are CIT, and all CIT persons are VOT. Think of these as population subgroups for a specific LMG and at a specific geographic level (State, Jurisdiction, AIA or ANRC).

In the Decennial Census and the ACS, people 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 Table 20 following Appendix A. There are 21 LMGs within the Asian racial group, 51 LMGs within the AIAN racial group, and a single Hispanic LMG that cuts across racial groups. (We number them as LMG 1–21 or AS 1–21 for Asian; LMG 22–72 or AI 1–51 for AIAN, and LMG 73 for HSP. This numbering differs from the labeling of Language Minority Groups on the Public Use Data file: on that file, they are numbered by an index `LANCOUNT` according to the following rule: `HSP = LANCOUNT 1`; `AI 1–AI 51 = LANCOUNT` 4–54; and `AS 1–AS 21 = LANCOUNT` 56–76.) 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. Of the approximately 3.7 million adults who self-identify into at least one LMG in the 2015-2019 ACS data, about 3.1% self-identify into 2 LMGs and 0.2% into 3 or more. Coverage determinations are made separately for each LMG, so that people belonging to multiple LMGs count towards the coverage criteria for all of them.

In the Tables, Figures and numerical comparisons throughout this report, we summarize estimates and Margins of Error (MOEs at the 90% confidence level) calculated either by direct survey-weighted methods or by model-based methods. All of these numbers are based on the exact numbers calculated by the methods described. However, in the publicly released data that is used in determining Section 203(b) coverage under the Voting Rights Act, all count estimates are first rounded to the nearest integer, and all estimates of proportions are first converted to percent [i.e., are multiplied by 100] and then rounded to the nearest 0.01.

Figure 1: Labels VOTAG, CIT, LEP, ILL refer to successively smaller circular regions. Arrows terminating within colored annuli and innermost circle show population-count notations N_{jg}^{A} defined in Appendix C for the multinomial categories: VOTAG non-CIT, CIT non-LEP, LEP non-ILL, ILL.



2.1 Direct Tabulated Outcomes

The outcome totals that were tabulated directly from the ACS 2015-2019 files for use in the production of coverage determinations are:

(a) unweighted counts of ACS-sampled VOT, CIT, LEP, and ILL persons in each (State, LMG), (Juris, LMG), (AIA, AIAN LMG) and (ANRC, AIAN LMG) domain;

(b) survey-weighted estimates of the total number of VOT, CIT, LEP, and ILL persons in the same domains as in (a);

(c) unweighted counts of ACS-sampled VOT, CIT, LEP, and ILL persons in each State, Juris, AIA and ANRC;

Table 1: Quantiles of coefficients of variation for ACS 2019 5-year direct estimates of total LEPs and ILL/LEP ratios in domains with estimated LEP persons ≥ 25 and respectively restricted to domains with LEP and ILL sample-sizes ≥ 2.

Estimate	Qu.25	Qu.50	Qu.75	Qu.80	Qu.90	Qu.95	# Domains
LEP in (Juris, LMG)	0.285	0.472	0.658	0.695	0.813	0.920	8727
ILL/LEP in (Juris,LMG)	0.293	0.458	0.642	0.686	0.805	0.930	2829

Note: "Qu.xx" denotes the xx percentile.

(d) survey-weighted estimates of the total number of VOT, CIT, LEP, and ILL persons in each Geography as in (c);

(e) survey-weighted estimates of totals and proportions at Geography level (State, Juris, AIAN or ANRC) of 10-15 covariates described in detail in Section 2.2.

The direct ACS survey-weighted estimators (b) of the domain total numbers of LEP and ILL persons can be used in conjunction with (d) to estimate the quantities needed for coverage VRA determinations. 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 yield estimates with large standard errors.

The coefficient of variation (CV) of a point estimate, its standard error divided by the estimate, measures relative precision. The Census Bureau requires that CVs for a majority of the key ACS survey estimates in each published table must be ≤ 0.30 to meet the Census Bureau's statistical quality standard for sampling error; and single estimates with CVs > 0.61 are said to be unreliable. Table 1 shows that a majority of CVs for the ACS estimated total number of LEPs and ILL/LEP ratios in single (Juris, LMG) domains are quite large. The CVs for the total estimated CIT persons in whole jurisdictions (not subddivided into LMGs) are mostly small: the three quartiles are $0.006, 0.025, 0.070$. In view of the large CVs displayed in Table 1, many of the direct survey-weighted estimates using the total LEPs or ILL ratios (= ILL/LEP) in each (Juris, LMG), which are used in the Criteria for Jurisdiction-level Coverage given in Section 1 above, will be unreliable. By contrast, the direct survey-weighted estimates of the total CITs in the Jurisdictions are typically precise, giving us relative confidence in using them directly in the calculations.

The domains summarized in Table 1 are a relatively small subset of all (Juris, LMG) domains. Out of 7,859 jurisdictions, averaged over LMGs, roughly 120 per LMG have estimated LEP counts ≥ 25 and LEP samples ≥ 2, and roughly 40 per LMG have estimated LEP counts ≥ 25 and ILL samples ≥ 2. The results in Table 1 are sensitive to the minimum thresholds of LEP estimates and LEP or ILL sampled persons used to restrict the domains. For example, when the defining thresholds are restricted further to estimated LEP being ≥ 50 and numbers of sampled LEP or

ILL $\geq$ 5, the numbers of domains are 5,195 for LEP and 1,327 for ILL. and the respective 0.9 quantile CVs for LEP and ILL/LEP estimates are 0.540 and 0.506. When the defining thresholds are loosened to estimated LEP being ≥ 15 and numbers of sampled LEP or ILL ≥ 1, the numbers of domains are 12,115 for LEP and 4,852 for ILL, and the respective 0.9 quantile CVs for LEP and ILL/LEP estimates are 0.968 and 1.082.

Figure 2 is based on 2019 ACS 5-year data and shows the point estimates and 90% confidence intervals for direct survey-weighted (Juris, LMG) domain estimates of the proportion `LEPprop` of LEP persons in (Juris, LMG) out of all CITs in (Juris), for the set of 226 domains whose 90% confidence interval includes 0.05 (the quantity and threshold for the determination criterion J1). The confidence intervals are plotted vertically, for domains sorted from left to right in order of increasing estimated CIT count. The interval widths have a general, but not strict, tendency to decrease with increasing domain size. 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 (Juris, LMG) domains is unclear yet particularly important. Thus, the precision is undesirably low for many of the 226 direct survey-weighted (Juris, LMG) estimates near 0.05, most of them in juridictions with CIT counts < 650. Many of the 90% confidence intervals for these estimates are extremely wide.

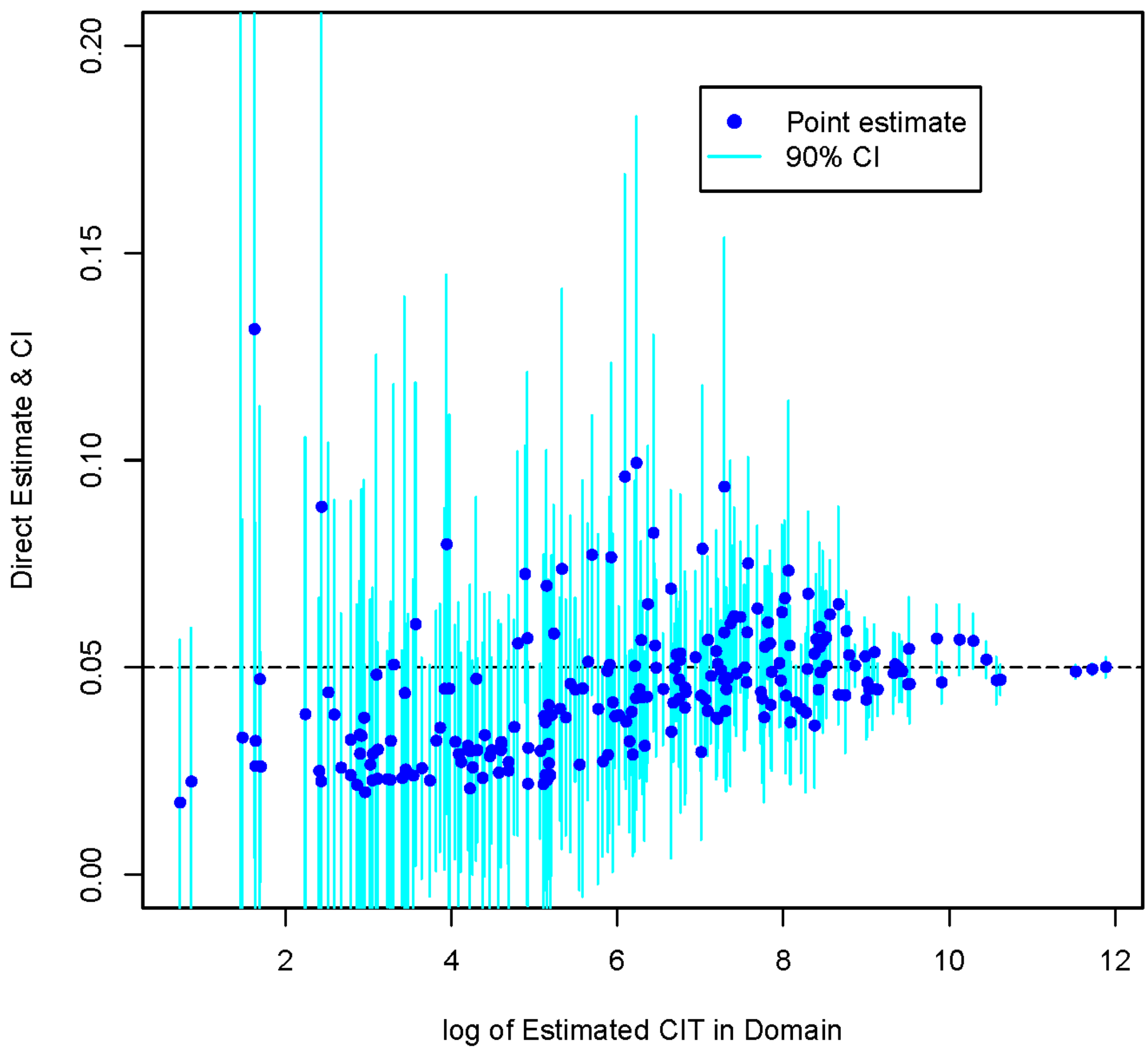
Whether using direct survey-weighted estimates as in preparing Fig. 2 or modeled estimates as in later sections of this report, the statistical analysis is particularly critical in the determinations of Voting Rights Act coverage only in the relatively small set of domains for which the ILL/LEP fraction exceeds 0.0131 and either the LEP total is close to $10,000$ or the LEP fraction (`LEPprop`) is close to 0.05. Of the 7,859 jurisdictions with data in the 73 LMGs, only 140 (Juris, LMG) domains have LEP count $< 10,000$ and ILL/LEP rate $\geq$ the national rate of 0.0131 and directly estimated 90% confidence interval for `LEPprop` containing 0.05. Of these 140 domains, 59 were determined to require language ballot assistance, by meeting the criterion that `LEPprop` > 0.05.

To mitigate the small-domain imprecision seen in Table 1 and Figure 2, we 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 (Juris, LMG) or (AIA, LMG) or (ANRC, 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. This similarity can be exhibited in the form of shared relationships between the outcome proportions and observable domain-specific covariates.

2.2 Covariates

Predictive covariates considered for use in our models consisted of population ratios directly estimated from the ACS data at different levels of aggregation (using survey weights) related to

Figure 2: Direct estimates and 90% confidence intervals of LEP proportions `LEPprop`, for the 226 (Juris, LMG) Domains from 2019 ACS 5-year estimates whose CI includes 0.05. Points are ordered and plotted by log of CIT count in Domain. Domain CIT counts range from 2 to 145,030.



citizenship, English proficiency, race/ethnicity, educational level, poverty, age, age of AIAN persons, and foreign birth, as well as average time in the United States. The levels of aggregation for ACS covariates to be included in the models were state within LMG, AIA within LMG, and geography-type (Juris or AIAN or ANRC) pooled across LMG. A fourth level of covariate aggregation, LMG-by-geography domain [(Juris, LMG) or (AIAN, LMG) or (ANRC, LMG)] was used to create possible covariates in 2016, but the covariates tallied in that way were very noisy and not found to be predictively useful for outcomes at the LMG-by-geography level. Moreover, when small-domain estimation models are formulated in terms of covariates that are estimated with standard errors of magnitude comparable to those of the outcomes, it is well known [Ybarra and Lohr, 2008] that the model predictions are biased unless the magnitude of covariate-estimation error is taken into account. Thus the LMG-domain-level covariates were *not* used in 2021. The following are the covariates used in 2021 modeling, displayed by type:

State-Level Covariates for Jurisdictions

C1 Logit-transformed fraction of citizens among VOT persons (`STC.T`)

C2 Logit-transformed fraction of limited English-proficient among CIT (`STL.T`)

Geography- (Juris- or AIA- or ANRC-) Specific Covariates

C3 Proportion of non-Hispanic White among VOT persons in geography (`WHNHSP`)

C4 Proportion of VOT persons with no college education in geography (`EDU2`)

C5 Average person count per housing unit among VOT persons in geography (`NUMPER`)

C6 Average age among VOT persons in any AIAN LMG in geography (`AGE`)

C7 Proportion of VOT persons in poverty in geography (`POV`)

C8 Proportion speaking other language at home among VOT persons in Juris (`OTHLANG`)

C9 Proportion of foreign-born persons among VOT persons in Juris (`FRNBORN`)

C10 Average years in US (as of 2019) among VOT foreign-born persons in Juris (`AvgYrs`)

AIA-Level Covariates

C11 Fraction of AIAN citizens among AIAN VOT persons (`CITrat`)

C12 Fraction of AIAN LEP persons among AIAN CIT (`LEPrat`)

C13 Fraction of AIAN ILL persons among AIAN LEP (`ILLrat`)

The synthetic covariates `STC.T` and `STL.tt`, logit-transformed State level proportions, were designed to help predict LMG-level proportions of CIT within VOT, LEP within CIT, and ILL within LEP. These covariates were calculated, for each (Juris, LMG) domain, from the complement of that domain within the State. Similarly, the untransformed proportions `CITrat, LEPrat, ILLrat` were calculated, for each (AIA, LMG) domain, from the complement of that domain within the AIA.

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 aggregation 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 higher-level survey-weighted direct ratio estimators of CIT and LEP rates are stably estimated but do not directly reflect domain-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 that are listed under *Geography-LMG Specific Covariates* above, only those listed as *Geography-Specific* were found to be usefully predictive in models for Asian or Hispanic (not AIAN) LMGs. See Section 3.6 for exact information on the covariates used in each LMG model.

All the covariates listed above 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 both for parameter estimates and predictors based on them. For the most part, this is not an important 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 consider models here for errors in variables and associated parameter estimates and predictors along the lines of Ybarra and Lohr [2008], but future Voting Rights Act methodology may do so.

2.3 Modeled Outcomes

The outcomes of interest in the models, LEP and ILL totals and ratios involving them, are not estimated directly. Instead, in a multinomial model for each LMG, we estimate the proportion of VOT persons in each of four disjoint categories: VOT non-CIT, CIT but not LEP, LEP but not ILL, and ILL (depicted as successive annuli and central disk in Figure 1). These estimated proportions are combined with the direct estimate of the total number of VOT persons in the (Juris,LMG)) domain to get an estimate of the LEP and ILL totals. Separate models were estimated for each LMG across Jurisdictions as 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. (We have already discussed in Section 2 the extent of self-identifications into multiple LMGs.) This convention, adopted also in 2016, differs from the 'local majority' modeling 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 our models fitted to different LMGs influence one another is that the subset of fixed-effect covariates used for an LMG in each geography-type was chosen from a list of possible predictors according to a grouping of LMGs with similar numbers of (Geography, LMG) domains containing ACS samples of similar sizes.

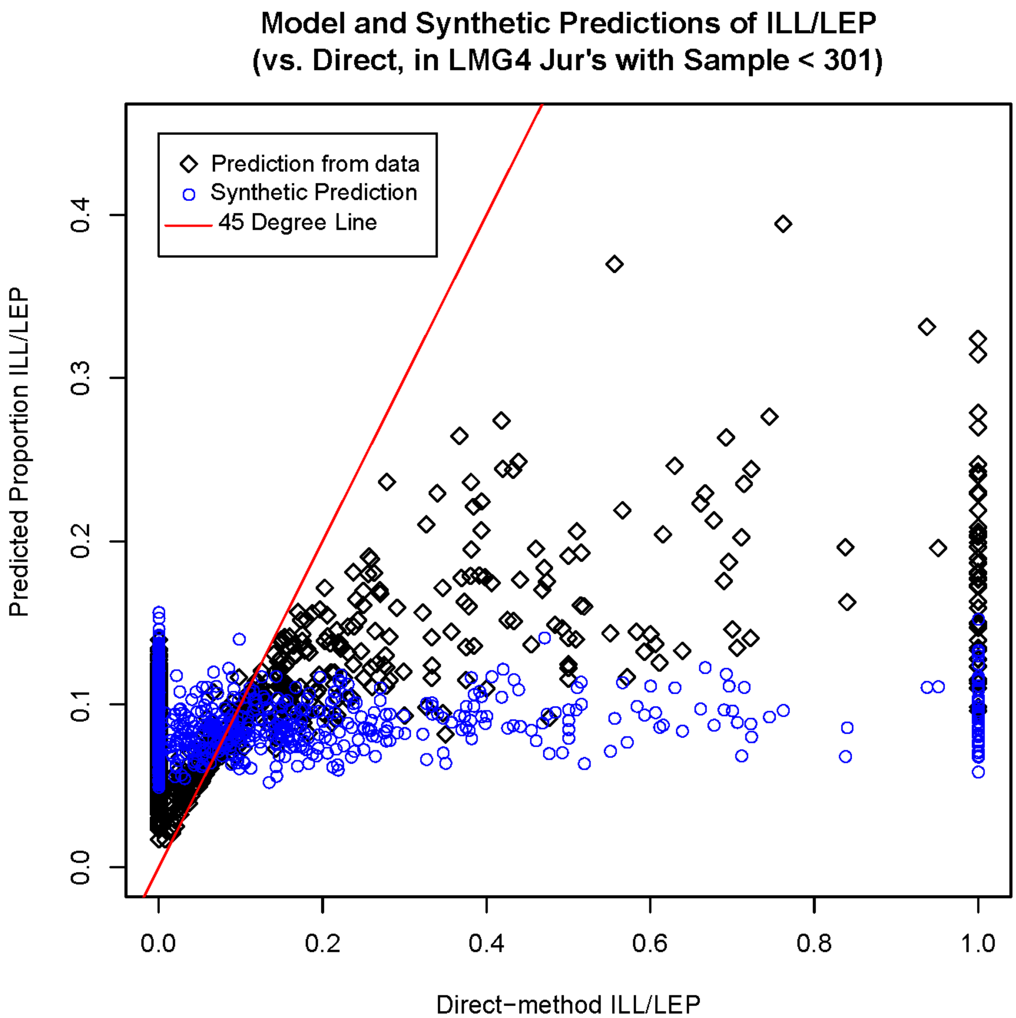
In all the models we considered, scaled, disjoint, outcome category counts follow a Generalized Linear Mixed Model [Breslow and Clayton, 1993] with form and parameters shared across domains within LMG (formula (9) in the Appendix). The models use independent random effects to account for domain differences within LMG, and they are fitted separately for each LMG.[3] Estimates from these models are either empirical-Bayes frequentist — estimating shared parameters via Maximum Likelihood on the entire sample — or Bayesian, estimating shared parameters through Monte Carlo simulated draws from their posterior distributions given the full set of observed data within each LMG. Partly for computational reasons that will be explained below, we used full-model Bayesian prediction for category counts and random effects only in the 21 most data-rich LMGs for the Jurisdiction (County/MCD) Geography-type. In addition, simple intercept-only Bayesian beta-binomial estimates were used (for all geography-types) for single-category counts (mostly CIT and ILL) in those AIAN LMGs where CIT/VOT ratios were very close to 1 or ILL/LEP ratios were very close to 0. (Section 2.4 and Table 2 below give further information about models used in LMG and outcome-category combinations with extreme CIT/VOT, LEP/CIT, or ILL/LEP ratios.)

The frequentist, empirical-Bayes estimates of outcome-category totals are weighted combinations of the direct ACS survey-weighted ratio estimates and synthetic model-based estimates in which estimated parameters have been substituted. These combination estimates have the feature of agreeing approximately with the direct estimate in domains where the direct estimate is relatively precise, but of more heavily weighting the synthetic model-based estimates in areas where the direct estimate has a large standard error. The Bayesian estimates of outcome-category totals do not have the same weighted analytical form but also agree more closely with the direct versus the synthetic estimates when the direct-estimate standard error is smaller.

Figure 3 illustrates the differences between direct-method and model-based predictions. For LMG 4, the Chinese Language Minority Group, in jurisdictions with LEP sample in the range 1–300, the Figure contrasts the model-based predictions of ILL/LEP ratios with predictions based on direct survey-weighted estimates. The model-based predictions are computed from model MLN-D described in Section 3.1 below. There are two kinds of predictions from the fitted model, the 'synthetic' ones (blue circles) that predict only from the covariates and not the outcome-data of the jurisdiction, and the fully model-based predictions (black diamonds) that use the model conditioned on the survey-weighted direct CIT, LEP and ILL estimates. The latter prediction, making fullest use of available data, is the one that is used in VRA determinations.

[3] However, because the Hispanic LMG is so much larger than all others, with estimated VOT population more than 7 times the next-largest, that LMG was modeled separately in the four geographic Census regions.

Figure 3: Model-predicted `ILL/LEP` values in domains for LMG 4 with positive LEP sample, plotted against direct estimates of `ILL/LEP`. Black diamonds are model predictions given outcome data, while blue circles are pure model predictions without regard to outcomes within domain. Red 45° line allows visual comparison of survey-weighted ratio versus modeled `ILL/LEP` estimates.



It makes sense that the model predictions using the data are closer to the 45-degree line, i.e. to the direct-estimated predictions, than the model predictions that ignore the domain-specific outcome. The model generally reflects a low rate of Illiteracy, and although the pure (synthetic) model-based predictions incorporate the effects of domain covariates, the model predictions conditioned

Table 2: Numbers of AIAN LMGs, by geography, for which `CITr = CIT/VOT, LEPr = LEP/CIT`, and `ILLr = ILL/LEP` are well-defined (denominator > 0) and extreme, each by two criteria.

Geo type	CITr $= 1$	LEPr $= 0$	ILLr $= 0$	CITr $> .995$	LEPr $< .005$	ILLr $< .005$
Juris	4	0	11	30	8	12
AIAN	14	2	22	41	11	22
ANRC	39	37	7	46	38	7

on the domain direct estimates of illiteracy and Limited English proficiency agree considerably more closely, although not perfectly, with the direct ILL/LEP estimate. The phenomenon that the model-based prediction 'shrinks' the direct estimate toward the synthetic modeled estimate is characteristic of small-area estimates that borrow strength across domains.

2.4 Extreme Cases Where Models Degenerate

One of the obstacles to developing a uniform model-based method for estimating ratios `CIT/VOT, LEP/CIT` and `ILL/LEP` at domain-level is that there are many (Geo, LMG) domains for which the direct survey-weighted estimates of these ratios are equal or very close to 0 or 1. This phenomenon occurs only in the AIAN LMGs, for all three geography types, Juris, AIA and ANRC, because samples in low-population geographic units are often very small and AIAN citizenship is nearly universal while some sampled geographic units have very low LEP proportions. Table 2 shows for each geography-type the number of AIAN LMGs (out of 51) in which the respective ratios are defined (denominator > 0) and extreme by two criteria, which means for CIT that CIT/VOT is $> 1 - 10^{-5}$ (> 0.995), and for LEP and ILL means that LEP/CIT or ILL/LEP are respectively each $< 10^{-5}$ (< 0.005).

3 Model Classes & Modeling Choices

As a a task in Small Area Estimation (SAE), the VRA Section 203 population estimates have several distinctive features, strongly influencing the choice of model. First, Section 203 imposes not a single SAE estimation task, but many: 73 for the separate LMGs in Jurisdiction geography, and 51 for the AIAN LMGs in AIA or ANRC geography. The data sizes for these different sub-tasks vary greatly, very small for some AIAN LMGs and very large for Hispanic and the largest Asian LMGs. Yet our understanding of the VRA legal requirements suggests the strong desirability of a uniform methodology for all of the (Geo-type, LMG) sub-problems, each to be analyzed using separate data. The VRA context restricts us to data with multiple distinct categorical outcomes (VOT non-CIT, CIT non-LEP, LEP non-ILL, and ILL) observed at person level and aggregated to domains consisting of geographic units within each voting-age LMG population, with covariates available only at the level of domains or higher-level aggregates (such as LMG-specific population proportions at higher-level geography, or at the level of local geography without regard to LMG). Therefore, the kinds of statistical models to be considered are regression models for discrete response-variables: individual people are treated as responding independently, conditionally within domains, so their domain-level counts are multinomial, and as is generally assumed in small-area and empirical-Bayes statistical modeling [Rao and Molina, 2015, Carlin and Louis, 2009], outcomes in the different domains have distinct independent random effects.

The literature on SAE contains many possible choices of area-level models for proportions [Jiang and Lahiri, 2006, Rao and Molina, 2015, Esteban et al., 2020] and associated estimation methods. Two of the most important model classes are beta-binomial (or for multiple categories, Dirichlet-Multinomial) [Carlin and Louis, 2009, Aitkin et al., 2009] and bivariate or multinomial logit-normal [Malec et al., 1997, Ghosh et al., 1998, Slud, 2000a, 2004, Malec, 2005, Molina et al., 2007, Franco and Bell, 2022, Koster and McElreath, 2017, McElreath, 2020], although other models mentioned by Esteban et al. [2020] have been tried. SAE literature on *benchmarking* develops the estimation of domain population totals subject to the constraint that they sum to known totals at higher levels of aggregated domains [Steorts and Ghosh, 2013, Datta et al., 2011, Pfeffermann and Tiller, 2006]. However, we viewed it as simpler in our nested-outcome setting to model the domain counts CIT within VOT, LEP within CIT, and ILL within LEP by a succession of separately parameterized random-effect binary-outcome models (binomial logit-normal or beta-binomial) rather than through multinomials subject to constraints. In each of these successive binomial models the previous (scaled) category total serves as a number of Bernoulli trials (see (iv) below for details). Another approach to such models is to replace actual scaled sample-sizes and totals by so-called *effective sample sizes* in the spirit of Kish [1987] and McAllister and Iannelli [1997], an idea implemented with Generalized Variance Functions in the Census Bureau's Small Area Income and Poverty Estimation (SAIPE) program as explained in Franco and Bell [2013]. Although we considered that approach in this VRA estimation cycle, we did not ultimately implement it for lack of time. Another active SAE research direction is the analysis of time-sequence models, that is, generalized linear models with random effects incorporating time-sequence domain-level data [Pfeffermann and Tiller, 2006, López-Vizcaíno et al., 2015]. We initially considered models of this sort in the VRA Section 203 analyses using data from successive ACS 5-year datasets, but we did not pursue this research direction due to computational difficulties and the large number of very small LMGs.

The model classes considered in the 2021 VRA determinations shared several common features. Within each LMG (labeled g) and Geography type (with geographic units indexed by j for Jurisdiction or a for AIA or ANRC[4]), the basic data consist of the voting-age sample-size n_{jg}^V and the direct survey-weighted estimates $\hat{N}_{jg}^A$ or $\hat{N}_{ag}^A$ of the numbers of persons in the 4 nested decreasing

[4]For the rest of this section, we use only the jurisdiction geography-level notation j, although the analogous considerations and models apply to AIAN and ANRC models indexed by a for AIAN VOT-person counts.

categories $A = V, C, L, I$ (respectively, V =VOT, C =CIT, L =LEP, and I =ILL), as described in Section 2, together with various geographic-level covariates $\underline{X}_{jg}$ for LMG g described in Section 2.2. For these data, the outcome variables to be modeled are true population counts N^A_{jg} in the $A = V, C, L, I$ categories. It was decided on the basis of experience in previous VRA cycles:

(i) to treat n^V_{jg} as fixed and known;

(ii) to obtain N^V_{jg} solely in terms of the direct estimate $\hat{N}^V_{jg}$ and the jurisdiction voting-age citizen total N^C_{j+} from its direct estimate $\hat{N}^C_{j+}$;

(iii) to model the proportions $N^A_{jg}/N^V_{jg} \equiv \pi^A_{jg}$ for $A = C, L, I$ as random variables between $0, 1$ jointly independent of $\hat{N}^V_{jg}$;

(iv) to model $(n^V_{jg}/\hat{N}^V_{jg}) \cdot (\hat{N}^V_{jg} - \hat{N}^C_{jg}, \hat{N}^C_{jg} - \hat{N}^L_{jg}, \hat{N}^L_{jg} - \hat{N}^I_{jg}, \hat{N}^I_{jg})$ as a Multinomial random vector with n^V_{jg} trials and probabilities $\underline{\pi}_{jg} \equiv (1 - \pi^C_{jg}, \pi^C_{jg} - \pi^L_{jg}, \pi^L_{jg} - \pi^I_{jg}, \pi^I_{jg})$; where

(v) the random probability-vector $\underline{\pi}_{jg}$ is in turn modeled in terms of the covariates $\underline{X}_{jg}$.

The random outcome-variables $\big(Y_{jk,g},\ \ k = 0, 1, 2, 3\big)$ respectively equal to $\big(n^V_{jg} \cdot \hat{N}^A_{jg}/\hat{N}^V_{jg}, A = V, C, L, I\big)$ have successive differences modeled in (iv) as 'Multinomial' although their $k = 1, 2, 3$ components will usually not be integers. This is done because the amount of statistical information in the data, as reflected in variances of parameter estimates, is more reliably reflected by the sample size n^V_{jg} (partitioned into categories $Y_{jk,g}$) than by the population size N^V_{jg} estimated by $\hat{N}^V_{jg}$. When sample sizes are large, the likelihood terms are essentially the same as if the entries $Y_{jk,g}$ were rounded to integers; otherwise the 'multinomial likelihood' terms are not exact likelihood terms of any model, and the parameter estimation technique is not exactly Maximum Likelihood, but rather analogous to it. To our knowledge, there is no precise mathematical theory to say that the resulting parameter estimates behave approximately (with limiting normal distribution and variances estimated by the negative Hessian of log-likelihood with respect to unknown parameters) as we expect by analogy with estimates from multinomial random-effect models. The same rough analogy to multinomial likelihood theory has been followed previously in the VRA modeling done in 2011 and 2016. See Section C for further elaboration of notations and models.

The models of type (i)-(v) actively studied in the 2021 cycle of statistical analysis for Voting Rights Act determinations fell into two main types, of which we chose one for production, as explained below. The two types of model differed only in the distributional form of the random outcome-probabilities $\underline{\pi}_{jg}$ and parametric regression form expressing them in terms of covariates $\underline{X}_{jg}$. These two model classes were respectively the Multivariate Logit-Normal models described in Sec. 3.1 and App. C.1 and the Dirichlet-Multinomial models described in Sec. 3.2 and App. C.2.

Beyond the choice of parametric statistical models within which to estimate parameters, there were many data-handling and model development choices to make in creating production estimates based on business rules specifying which data subsets and covariates would be used. For example, model fitting in 2016 was based only on LMG data in geographic units for which VOT sample size was at least as large as a threshold of `minsamp` $= 1, 3$, or 5 according to LMG data-set size. The idea was to make use only of sufficiently reliable data, rather than data in geographies with tiny sample size, when there was enough aggregate data in geographies with samples at least of size `minsamp`. In the 2021 VRA data analyses, we assessed (in Sec. 3.6 below) whether such a `minsamp` threshold made a difference to the quality of prediction of population counts N^A_{jg} and if so what threshold to choose. Similar assessments were made, in Sections 3.6.2–3.6.4, to develop business rules — acoording to the numbers of geographies with adequate sample within each LMG and geography-type — for how many covariates and which ones would be used in fitting the MLN-class models that were eventually chosen (with `minsamp` $= 1$) for production.

Choices of computational techniques for estimation and prediction were also necessary. These involved selecting specific forms of models for computational tractability, based on accuracy and on amount and timing of computational effort. Some of these choices, made according to the data richness for each LMG and geography-type, were: whether to use a fully parameterized MLN or a slightly simpler model; use of a frequentist (Maximum-Likelihood) or Bayesian (Markov-Chain Monte-Carlo) method of parameter estimation and prediction of populations and margins of error; the method of numerical integration to use in frequentist estimation, and the method of initializeation in the Bayesian computations; and finally, what kind model to use as a fallback in the relatively small number of (Geo, LMG) combinations where the parameter estimates in the otherwise chosen models failed to converge. These computational choices are discussed in Section 3.5 and Appendix E.

3.1 Multinomial Logit-Normal Model

One version of the model (i)-(v) in Section 3 above is expressed within LMG g (notation for which is suppressed, since models are developed separately for each LMG) as:

$$Y_{j,k} \sim \text{Binom}\big(Y_{j,k-1}\,, h(\eta_{j,k} + u_{j,k})\big) \quad \text{given} \qquad (Y_{j,t-1},\, \eta_{j,t},\, u_{j,t},\;\; t = 1, \ldots, k) \tag{1}$$

for geographic units j within LMG g, where $h(x) \equiv e^x/(1+e^x)$, and for $1 \le k \le 3$,

$$\eta_{j,k} \equiv \underline{X}_j' \, \beta^{(k)} \quad , \qquad \underline{u}_j = (u_{j,1},\, u_{j,2},\, u_{j,3})^{tr} \sim \mathcal{N}(\underline{0}\,,\, \Sigma)$$

where $\beta^{(k)}$ for $k = 1, 2, 3$ are unknown coefficient vectors (some entries of which may be structural zeroes), and Σ is a general unknown 3×3 covariance matrix. This model is referred to as the *full Multinomial Logit-Normal* (MLN-F), with the random probabilities π^A_j in (iv) defined by:

$$\pi^C_j = h(\eta_{j,1} + u_{j,1}) \;\; , \quad \pi^L_j = \pi^C_j \cdot h(\eta_{j,2} + u_{j,2}) \;\; , \quad \pi^I_j = \pi^L_j \cdot h(\eta_{j,3} + u_{j,3}) \tag{1'}$$

These random probabilities are defined through random-intercept logistic expressions in the stages $k = 1, 2, 3$, where $\eta_{j,k}$ are the *regression* terms and $u_{j,k}$ are *random effects*. The dimensions of the vectors $\beta^{(k)}$, including structural zeroes, is the same as the number of components d of the covariates vectors $\underline{X}_j$. Models of this sort with a different regression parameterization have been proposed beginning with Aitchison and Shen [1980] and are now standard in categorical data analysis [Agresti, 2013]. Such models have also been used before in small area estimation problems [Molina et al., 2007]. Precursors for univariate logit-normal small area estimation include Efron and Morris [1975], Ghosh et al. [1998], and Slud [2000a]. Applications of the Multivariate logit model to many different social science problems, with multinomial-outcome regression parameterization as in Agresti, can be found in Koster and McElreath [2017], McElreath [2020]. The cascaded logit parameterization of the Dirichlet-Multinomial probabilities used in this report (i.e., in ($1'$) above and in formula (19) in the Appendix) was previously used in the 2016 VRA statistical analysis [Slud et al., 2018].

As discussed fully in Appendix C.1, this model has $|\mathcal{I}_1| + |\mathcal{I}_2| + |\mathcal{I}_3| + 6$ unknown parameters, where $\mathcal{I}_k$ for $k = 1, 2, 3$ is the set of coefficient index positions in $\beta^{(k)}$ that are *not* structural 0's, and $|\mathcal{I}_k|$ denotes its cardinality. These models always include intercepts, so that $1 \in \mathcal{I}_k$ and $|\mathcal{I}_k| \geq 1$.

A simplified *diagonal*-covariance version MLN-D of the Multivariate Logit-Normal model is defined in exactly the same way as in (1), with Σ a diagonal matrix with diagonal elements σ_k^2. In that model, which has 3 fewer parameters than MLN-F, the 3 stagewise random-intercept logistic regression models (1) are decoupled in the sense that $u_{j,k}$ are independent across $k = 1, 2, 3$, and therefore π_j^C, π_j^L/π_j^C, π_j^I/π_j^L are independent. This independence turns out to be a great computational simplification, since the k'th stage unknown parameters $\beta^{(k)}, \sigma_k^2$ are different across the three model stages and can be estimated separately via Maximum Likelihood from data ($Y_{j,k}$, $\underline{X}_j$) given $Y_{j,k-1}$ in the three conditional-model stages $k = 1, 2, 3$.

The MLN-F and MLN-D models make sense and can be fitted from data even when one or more of the model-stages (1) involve $\beta^{(k)}$ in which only the intercept is not structurally 0. We will use these intercept-only random-effect stagewise models in many LMGs containing relatively few geographic units j with nonempty data-samples, especially in fitting CIT (stage $k = 1$) models in LMGs where nearly all VOT persons are citizens and in fitting ILL (stage $k = 3$) models in LMGs where only a tiny fraction of LEP persons is illiterate.

3.2 Models Related to Dirichlet-Multinomial (DM)

Appendix C.2 explains that a different model of the form (i)-(v), the Dirichlet-Multinomial [Carlin and Louis, 2009, p. 284], was used in 2016 in estimation and production for VRA determinations[5].

[5] Actually, that model was modified in 2016 production with the slight variation that the parameter τ was replaced by the term $\tau_0 \cdot \sqrt{n_j^V}$ depending on sample size, with τ_0 a parameter to be estimated.

In that formulation, the random probabilities $N_j^A/N_j^V \equiv \pi_j^A$ within LMG g were modeled as

$$\underline{\pi}_j \sim \text{Dirichlet}\big(\tau,\ (1-h(\eta_{j,1}),\ h(\eta_{j,1})\,(1-h(\eta_{j,2}),\ h(\eta_{j,1})\,h(\eta_{j,2})\,(1-h(\eta_{j,3}),\ h(\eta_{j,1})\,h(\eta_{j,2})\,h(\eta_{j,3}))\big) \tag{2}$$

based on the Dirichlet probability distribution [Carlin and Louis, 2009, p. 425], a generalization of the beta distribution to a number of categories greater than 2. It is shown in Appendix C.2 that the Dirichlet-Multinomial (2) for $\underline{\pi}_j$ is equivalent (for $\tau_1 = \tau_2 = \tau_2 = \tau$) to the stagewise model

$$1 - \pi_{j,1} \sim \text{Beta}\big(\tau_1 v_1, \tau_1(1-v_1))\big) \quad , \qquad 1 - \frac{\pi_{j,2}}{1-\pi_{j,1}} \sim \text{Beta}\big(\tau_2 v_1 v_2, \tau_2 v_1(1-v_2)\big) \ ,$$

$$1 - \frac{\pi_{j,3}}{1-\pi_{j,1}-\pi_{j,2}} \sim \text{Beta}\big(\tau_3\, v_1 v_2 v_3, \tau_3\, v_1 v_2(1-v_3)\big) \quad \text{are independent, for} \quad k = 1,2,3 \tag{3}$$

for unknown (*dispersion*) parameters (τ_1, τ_2, τ_3), where $v_k \equiv h(\eta_{j,k}) = \exp(\eta_{j,k})/(1+\exp(\eta_{j,k}))$. Here we have allowed the dispersion parameters τ_k for $k = 1,2,3$ to be general and distinct. This model is more general than (2), with number of parameters $|\mathcal{I}_1|+|\mathcal{I}_2|+|\mathcal{I}_3|+3$, the same as the MLN-D model. Like the MLN-D model, this version (3) of DM makes the unknown parameters $(\tau_k,\ \beta^{(k)})$ in each model stage distinct from those of the other stages, and the corresponding likelihoods factor into the product of likelihoods for separate stages, which leads to the helpful computational simplification that the stagewise parameters can be estimated separately.

In the DM model, like the MLN, it makes sense to estimate stagewise parameters in data-sparse LMGs in models without covariates. Either when n_j^V is small or $Y_{j,k}/Y_{j,k-1}$ is too close to 0 or 1, we fit models in which the intercept is the only nonzero entry ($\{1\} = \mathcal{I}_k$) of $\beta^{(k)}$.

There is one further special case of a simplified DM model that is ultimately used as the fallback in stagewise models for data-sparse $Y_{j,k}$ where random-effects MLN or DM models cannot be made to converge. In such LMGs and stages, we use the Jeffreys-prior beta-binomial

$$Y_{j,k} \sim \text{Binom}(Y_{j,k-1},\ \gamma_{j,k}) \quad , \qquad \gamma_{j,k} \sim \text{Beta}(0.5,\ 0.5)$$

3.3 Predictors & Model Diagnostics

We briefly sketch here the model diagnostics, elaborated in Appendix F, that we use in subsequent Sections of the report in assessing the quality of fit of different models and of different specifications of particular chosen models. In all cases, we assess the models through their predictions. It is important to remark that within a given LMG and geography-type, we produce estimates only for domains (Geo, LMG) within which there is at least one sampled voting-age person in ACS 2015-2019 data, and we construct diagnostic tests from this set of jurisdictions.

It is well-tested and accepted that direct survey-weighted ACS estimates of population in large domains are reliable and roughly unbiased: these are the work-horse estimates of the ACS. Therefore, within each LMG the quality of fit of models is assessed by comparing predicted total populations (within CIT, LEP, ILL) against the direct ACS estimates over large aggregates of jurisdictions for which some sample-size variable or geography-level covariate falls in specified ranges. Different models or model specifications are compared via discrepancy between their estimates and direct ACS estimates for the resulting aggregated domains. We use two primary measures of discrepancy `Delta` between model predictions $\tilde{N}$ and the ACS direct estimates $\hat{N}$ for each such domain: (a) the percent relative difference $100 \cdot (\tilde{N} - \hat{N})/\hat{N}$ (denoted `PctRel`Δ in table captions below), and (b) the standardized relative difference $(\tilde{N} - \hat{N})/SE(\hat{N})$ (denoted `Stdizd`Δ in tables), where $SE(\hat{N})$ is the square-root of the 'Successive Difference Replication' or SDR-estimator of variance of the direct estimate $\hat{N}$ described in Appendix D.5, formula (48). This standard-error estimate is the one generally used by ACS, and for large domains the standardized difference $(\hat{N} - N_{true})/SE(\hat{N})$ is distributed approximately as a standard normal, so the `Stdizd`Δ discrepancy referred to a $\mathcal{N}(0, 1)$ distribution roughly measures departure from the null hypothesis that the $\tilde{N}$ prediction and $\hat{N}$ estimate are the same. (However, since all of our model-based predictions make use of and are conditioned on the direct estimate, $\tilde{N}$ and $\hat{N}$ will always be positively correlated, and we should expect `Stdizd`Δ to be smaller in absolute value than a $\mathcal{N}(0, 1)$ deviate.)

3.3.1 Minimum Sample-Size Thresholds for Use of Geographic Units in Estimation

As a preliminary application of our model diagnostics, we discuss the choice of whether to set a minimum VOT sample-size threshold `minsamp` for geographic units to have their data included in the fitting of model parameters. Such a threshold *was* imposed for larger LMGs in the 2016 VRA cycle of statistical analyses [Slud et al., 2018]: in those analyses, within larger LMGs and each geography-type (Jurisdiction or AIA), only those geographic units with VOT sample-size at least 3 or 5 were included in parameter-estimation, although the fitted parameters were then used in creating predictions for all geographic units. The idea was that survey-weighted observations were much more stable and less noisy in those geographic units with sample-sizes greater than 1, although it was recognized that the models would necessarily be less relevant to tiny geographic units if units with the smallest sample sizes were excluded from parameter estmation. In the 2021 VRA cycle, we decided the question of whether `minsamp` should ever be > 1 by calculating diagnostics for quality of predictions with MLN-D models in selected large LMGs with Jurisdiction geography-type when `minsamp` is taken to be 5 versus the default of 1.

We conducted analyses concerning `minsamp` on LMGs 2, 11, 13, 17, 18, 19, 20, 22, 26, 51, 57, 59, 61 (from the whole list of 73 LMGs) in Jurisdiction geography-type for which MLN-D models were ultimately fitted (in some cases, with no covariates) at all 3 model stages (CIT, LEP, ILL). We found that restricting the jurisdictions used in parameter estimation by `minsamp`> 1 (and then predicting from the model in all jurisdictions with positive VOT sample) appeared to improve the quality of some predictions while making others worse. For example, Table 3 shows the diagnostics of discrepancy for the LEP predictions in LMG2 both with `minsamp` values of 1 and 5. In this Table, `njuris` is the number of jurisdictions in LMG2 that fall in the designated `minsamp` range, with `Pop-size` the number of sampled persons in those jurisdictions, rounded to 4 significant digits.

Table 3: LEP Model Diagnostics by VOT Sample Size with Min. Sample Size 5 or 1, in LMG 2

Measure	minsamp	Interval of VOT sample-size in Jurisdiction					
		(0,4]	(4,12]	(12,25]	(25,50]	(50,200]	201+
njuris		323	104	40	21	24	5
Pop-size		1,295	2,230	2,293	2,463	10,140	17,910
Delta	5	866.40	196.60	-48.20	44.20	-219.30	-22.76
PctRelΔ		66.90	8.82	-2.10	1.79	-2.16	-0.13
StdizdΔ		4.53	0.74	-0.18	0.16	-0.38	-0.03
Delta	1	611.90	9.69	-144.00	-47.50	-319.90	-64.40
PctRelΔ		47.20	0.43	-6.23	-1.93	-3.15	-0.36
StdizdΔ		3.20	0.04	-0.53	-0.17	-0.55	-0.08

Notes: njuris is the number and Pop-size the direct-estimated population of jurisdictions with sample-size in indicated intervals; Delta is model-estimated minus direct-estimated pop-count; PctRelΔ is 100 times Delta over direct pop-count; StdizdΔ is Delta divided by direct-estimated standard deviation of pop-count.

Table 4: LEP Relative Delta in LMG 13 jurisdictions, for 3 different minsamp values

Measure	minsamp	Interval of VOT sample-size in Jurisdiction					
		(0,4]	(4,12]	(12,25]	(25,50]	(50,200]	201+
njuris		362	82	35	13	7	1
Pop-size		421	779	836	437	1,366	841
PctRelΔ	1	54.3	-18.39	-6.69	-1.90	-3.63	-1.36
PctRelΔ	3	72.9	-13.22	-1.89	3.69	-2.73	-1.23
PctRelΔ	5	91.3	-9.31	0.72	6.90	-2.15	-1.08

Notes: Measures defined under Notes for Table 3.

With minsamp= 5 versus 1, the absolute value of standardized discrepancy StdizdΔ is meaningfully larger in the smallest sample-size bin, somewhat larger in the second-smallest, and slightly smaller in other bins. PctRelΔ shows a similar pattern. Omitting jurisdictions with low sample sizes from model fitting adversely affects the LEP model predictions in low sample-size jurisdictions for most LMGs. The pattern is fairly clear that more stringent sample size restrictions (larger minsamp) lead to worse discrepancies in the lowest-sample bin. Table 4 shows this with LEP relative discrepancies (PctRelΔ) in LMG 13 with minsamp values of $1, 3, 5$. There is a similar pattern for the ILL predictions as a function of minsamp (not shown) as for LEP predictions.

There are isolated exceptions to these patterns in the LMGs. LEP predictions in LMG 67 had reduced absolute `PctRel`Δ values with `minsamp` of 5 versus 1 in smaller-sample bins, but larger values in higher-sample bins. With ILL predictions, the same kind of reversal happens in LMGs 2 and 19. In these anomalous cases, the effects are not large and do not always persist as `minsamp` moves from 1 to 3 to 5, and the numbers of sampled jurisdictions are rather small. `PctRel`Δ values likely have high variances in these and other cases.

A reasonable interpretation of these results is that for most LMGs, the low-sample jurisdiction estimates contain information not accounted for by the models fitted to jurisdictions with more sample. Models that exclude this information may predict outcomes in higher sample jurisdictions slightly better. We decided based on these remarks to fit the models using data from all jurisdictions with sample (i.e., with `minsamp`= 1). Nevertheless, the suggestion here that models do not qualitatively capture jurisdiction-size differences is borne out in analyses presented in later sections, and represent a direction for improving the models in future research.

3.4 Comparisons of Fitted MLN and DM Models

The two different model types considered in this research were Multinomial Logit Normal (MLN) and a kind of staged Dirichlet-Multinomial (DM) which is better regarded as a stacked or staged Beta-Binomial (see Section 3.2 and Appendix C.2.) Both are multinomial regression models with a similar logistic parameterization of stagewise conditional expectations, and when the MLN model is restricted to have independent stagewise random effects (MLN-D) the two models have exactly the same parameter dimension, equal to the total number of nonzero stagewise regression coefficients ($\beta^{(k)},\ k = 1, 2, 3$) plus three (respectively for σ_k in MLN and τ_k in DM) controlling random-effect variances. The model DM bears some similarity to the model used in the 2016 VRA cycle [Slud et al., 2018], but in the present research we did not allow the DM k'th stage parameter τ_k to vary with n^V_{jg} sample size as was done in 2016.

The principle of comparison between the MLN and DM models is sketched in Appendix F and in Section 3.3 above. We assess the relative quality of models by comparing the patterns and magnitudes of discrepancies between the respective model predictions and the direct survey-weighted estimators of population subgroups defined by CIT, LEP or ILL over aggregated jursdictions within the LMG. The jurisdictions are aggregated into table cells by combining all jurisdictions with a covariate (`VOTsmp, EDU2`, etc.) falling in a designated range of values. When the LMG population of the aggregated jurisdictions is large enough (generally hundreds to thousands is enough), the direct estimators form a reliable unbiased target to which the model predictions should be close.

As argued below in Section 3.5 using Table 6, the MLN-D model usually passes tests of model adequacy versus the more richly parameterized MLN-F that allows correlated stagewise random effects. Therefore, it seemed reasonable to compare model MLN and DM strictly based on MLN-D predictions. Because the LEP and ILL (Juris, LMG) domain-population predictions are most

Table 5: Measures of relative (PctRelΔ) and standardized (StdizdΔ) discrepancy between each of MLN-D and DM model predictions versus direct estimates of (Juris, LMG) LEP and ILL populations aggregated to VOT-sample-size (VOTsmp) classes, for LMGs 4 and 10. For each population-group and LMG, the direct-estimated population size is also shown, rounded to 4 significant digits.

Measure	Pop.gp	LMG	Model	Interval of VOT sample size in Jurisdiction					
				(0,4]	(4,12]	(12,25]	(25,50]	50,200]	201+
Pop-size	LEP	4	MLN-D	5,704	10,220	12,460	15,660	44,500	869,800
PctRelΔ				9.02	-6.72	-6.09	0.14	-0.58	-0.01
StdizdΔ				1.30	-1.36	-1.21	0.03	-0.24	-0.03
PctRelΔ			DM	49.80	16.73	8.29	10.86	4.35	0.22
StdizdΔ				7.17	3.39	1.65	2.60	1.82	0.43
Pop-size		10	MLN-D	4,225	5,051	5,061	5,535	18,950	14,790
PctRelΔ				15.12	-1.51	-3.67	-3.71	-1.56	-0.42
StdizdΔ				1.87	-0.20	-0.47	-0.53	-0.38	-0.09
PctRelΔ			DM	35.00	12.62	7.30	3.77	1.83	0.66
StdizdΔ				4.32	1.69	0.93	0.54	0.45	0.14
Pop-size	ILL	4	MLN-D	503	909	991	1,429	3,662	94,440
PctRelΔ				27.89	1.57	5.40	2.25	0.36	-0.26
StdizdΔ				1.50	0.11	0.35	0.17	0.05	-0.17
PctRelΔ			DM	67.38	22.65	15.14	9.89	4.53	0.14
StdizdΔ				3.63	1.54	0.98	0.74	0.60	0.09
Pop-size		10	MLN-D	1,052	1,232	1,160	1,613	5,104	4,940
PctRelΔ				18.45	3.66	2.18	-9.83	-3.07	-2.59
StdizdΔ				1.14	0.26	0.14	-0.73	-0.42	-0.34
PctRelΔ			DM	38.88	15.74	10.30	0.95	1.26	-0.03
StdizdΔ				2.41	1.13	0.67	0.07	0.17	0.00

Notes: Pop.gp is population subgroup LEP or ILL within aggregated jurisdictions with sample-sizes in indicated intervals. MLN-D is multinomial logit-normal model with independent random efects; DM is Dirichlet-multinomial model. See Notes under Table 3 for Measure definitions.

relevant to the VRA determinations, we exhibit model comparisons only for LEP and ILL population-groups within (Juris, LMG) domains. After examining many such comparisons, we found that in general the MLN-D predictions — which line up nearly perfectly with the DM predictions, jurisdiction by jurisdiction within each LMG — are slightly superior, by our criteria described in Section 3.3. Table 5 shows this for LMGs 4 and 10, but the pattern of results is very similar for other LMGs in Jurisdiction geography. The conclusion that MLN was the superior and more flexible model was secure enough that we continued the research on the VRA project using it alone.

3.5 Frequentist & Bayesian Computational Considerations

Up to this point, we settled on using the Multinomial Logit-Normal class of models in statistical modeling and prediction for this cycle of Voting Rights Act determinations. There were still some decisions to make whether the model would be parameterized in its most general (MLN-F) form, with general random-effects covariance matrix Σ, or whether that covariance matrix would be restricted to be diagonal (the MLN-D model). There was also a decision to make regarding the computational and conceptual strategy for analyzing the model, whether by a frequentist numerical likelihood maximization (described in Appendix E.1) or with a Bayesian estimation of the posterior distribution for unknown parameters via Markov Chain Monte Carlo (Appendix E.2). These two decisions turned out to be closely related: both methods turned out to be valid approaches to estimating unknown parameters predicting domain-level population counts and Margins of Error (square roots of MSPEs) for the outcome categories CIT, LEP and ILL, but their relative computational efficiency depended crucially on whether the MLN-D model was adequate within an LMG or whether the added complexity of random-effect covariances made an important difference to the predicted (Geo, LMG) population counts.

It is shown in Table 21 of Appendix Section E.3 that the computation times for frequentist estimation of MLN-D model parameters, together with predictions of (Geo, LMG) population category (CIT, LEP, ILL) counts and the 80 analogous calculations with replicate weights used to estimate Mean-Square Prediction Errors, are extremely rapid, less than 11 minutes for each of the large Jurisdiction-geography LMGs (3–10 and the NE region of the Hispanic LMG) displayed there. The same table also showed that the corresponding times to run a complete suite of Bayesian MCMC computations for the same outputs took from 40 to 100 times as long. (Multiply the times shown in the first row of Table 21 by 4, since the times shown there were for the average of 4 parallel simulated Markov chains needed to confirm convergence.) Finally, the second row of Table 21 (plus 1/81 times the third row, which is nearly negligible) provides the run-times for frequentist estimation of the full MLN-F model estimation and predictions, but those times — ranging from 6 to 40 minutes — while far less than the Bayes MCMC run-times, do *not* include the many runs based on replicate weights that would be needed to estimate MSPEs. While the parameter estimates computed from the Bayes and frequentist analyses of MLN-F for the displayed LMGs were close but not identical, Table 22 in Appendix E.3 demonstrated that the predictions from the two methods of analysis were indistinguishably close. The primary computational conclusions from Appendix E.3 are twofold. The first is that for larger LMGs with many sampled jurisdictions, the Bayesian MCMC analysis was the fastest and most reliable way of deriving all needed parameter estimates, population category predictions and MSPEs from the MLN-F model. However, the second conclusion was that if the MLN-D model were deemed adequate, then the results could be computed at least 40 times faster by a frequentist Adaptive Gaussian Quadrature likelihood-maximization method.

The reason to focus on computation times for the model fits is not the time required to compute final model fits and predictions, but rather to allow time for many alternative model fits along the way to selecting covariates and to decide by the fitting of alternative models whether there should be a minimum threshold (`minsamp`) for geographic-unit sample size to allow the sample in a geographic unit to contribute to parameter estimation. We described the process of deciding on `minsamp` in Section 3.3.1 and on the covariates to use in different LMGs and geographies in Section 3.6 below. Since the difference between computational resources required for MLN-D versus MLN-F model fits is large, it is important to justify why the higher-dimensional model MLN-F was needed only in the largest LMGs with Jurisdiction geography-type, while MLN-D was adequate everywhere else. Table 6 shows the pattern. Among the displayed LMGs, with Jurisdiction geography and `njuris` denoting the number of jurisdictions with positive sample for each LMG, the larger and more significant chi-square likelihood ratio test (LRT) statistics (equal to twice the difference between the MLN-F and MLN-D maximized log-likelihoods, referred to the 3 degree of freedom percentage point $\chi^2_{3,\,0.05} = 7.81$) tend to occur in large LMGs. There are exceptions, and only the LRT statistics for MLN-D versus MLN-F in LMGs 9 and 73 are extremely significant, but the larger LMGs do tend to have larger LRT chi-square values. The pattern here suggests that few LMGs will really have clearly better accuracy from predictions with MLN-F instead of MLN-D. This pattern is borne out in the smaller Jurisdiction-geography LMGs and in the AIA-geography LMGs: the more parsimonious MLN-D is statistically adequate except for LMGs with largest `njuris`.

Table 6: AGQ Maximized LogLik_F for MLN-F model (rounded to nearest 100), LRT χ^2_3 (MLN-F vs. MLN-D), number of Jurisdictions with sample, and β-coefficient dimension, in 9 selected LMGs.

	LMG:	3	4	5	6	7	8	9	10	73
$100{\cdot}\text{logLik}_F$		-178	-3,369	-1,833	-160	-61	-556	-1,035	-138	-3,373
LR χ^2_3		7.6	5.6	9.6	9.2	8.8	4.4	22.6	7.2	27.4
$\dim(\beta)$		13	12	12	9	9	12	12	12	13
`njuris`		989	3,309	4,004	742	845	3,029	3,126	1,052	1,405

Notes: LR is Likelihood Ratio statistic; `njuris` is number of jurisdictions with sample in LMG; MLN-F differ from MLN-D multinomial logit-normal models in allowing dependent random effects. LMG 73 denotes NE-region Hispanic.

3.6 Model and Covariate Selection

A large number of MLN models were fitted for each geography type (Juris, AIA and ANRC). There was no formal variable selection methodology based on hypothesis testing or AIC. However, variables were screened based on standardized coefficients found significant using model likelihoods, by fitting preliminary MLN-D models in each LMG using a common maximal set of useful variables and then re-fitting the model using only the ones that were significant for the LMG in a second model-fitting pass. These potentially useful covariates included C1–C10 described in Section 2.2, with particular attention to the covariates C8–C10 (`FRNBORN, OTHLANG`, and `AvgYears`) for Jurisdictions and C1, C2, C4 (`STL.C, STL.L` or `LEPrat, EDU2`) for all geographies. At that point convergence was assessed, and when legitimate MLEs were found, the set of variables with significant coefficients was recorded. The variables that appeared predominantly in MLN-D models were: `EDU2, STL.T, OTHLANG, FRNBORN, AvgYrs` in Jurisdiction geography models, and AIA-wide LEP and ILL rates (`LEPrat, ILLrat`) in AIA geography-type. These were the variables appearing most often in the final models, and also the variables most often strongly significant (with absolute standardized coefficients greater than 4 in Jurisdiction geography or greater than 3 in AIAs.)

These sets of useful covariates were compared across all LMGs within major LMG groupings (Asian/Hisp and AIAN-Large and AIAN-Small) for each geography-type. Within these categories a few subsets of LMGs were found with mostly the same significant covariates (based on the 2014-2018 data on which these steps were performed) for the final production round of model-fitting on 2015-2019 data. Groups of LMGs were formed — the business rules described in detail below — either to fit MLN-F with a full set of covariates, MLN-D with a full set of covariates, MLN-D with a much reduced set of covariates, MLN-D with no covariates, or the no-covariate Beta-Binomial (described in Appendix Section C.1.1 and C.2.1) when MLN-D convergence failed. Such failures of convergence happened especially in LMG cases where the single-stage ratios CIT/VOT were close to 1 or LEP/CIT or ILL/LEP were close to 0. In the 2016 VRA cycle [Slud et al., 2018], model selection and business rules were based on a similar but less elaborate and less fully documented screening procedure, while the methodology of variable selection in 2011 VRA analysis, used for geographic-unit grouping in lieu of geography-level regression, was still less formal [Joyce et al., 2012, 2014]. In general, we attempted to use the same covariates for as many LMGs as possible.

3.6.1 Assessing Predictions for MLN-D Models with Covariates

Up to this point, we found general support through the relative and standardized discrepancy diagnostics for the choice of model MLN over DM, for the adequacy of MLN-D versus MLN-F in all but the largest LMGs, and for analysis using `minsamp` = 1, i.e., allowing jurisdictions to contribute data to parameter estimation regardless of sample size. In addition, comparison of execution times and for computation of MSPEs along with predictions supported the decision that Bayesian model-fitting techniques should be used for fitting MLN-F in the largest LMGs (for

Jurisdiction geography-type only). In this section, we use prediction diagnostics to explore model adequacy of the chosen models and the effectiveness of the covariate-sets chosen for them.

As described in Section 2.2, the covariates used in the MLN models are based on characteristics observed for geographic units without regard to LMG membership: either specific outcome rates (C1–C2, `CITrat, LEPrat, ILLrat`) or geographic data (C3–C10) in jurisdictions or AIAs without regard to outcome. The covariates are weak in the sense that they are not specific to LMG members within geographic units, and characteristics of LMG members are generally not well predicted by those of their non-LMG neighbors. Indeed, we show in this section that the measures of relative and standardized discrepancy between predicted and direct-estimated counts only weakly indicate the value of covariates. However, the correlations between predicted and direct-estimated (Geo, LMG) outcome rates indicate a clear benefit of using the covariates we have over none at all.

Table 7 displays results in two LMGs (both with covariates `STL.T, EDU2, FRNBORN, AvgYears`) for their full models versus intercept-only models. In LMG 2, the LEP full model does not improve much upon the intercept-only models, and for all but two VOT sample size bins, the full LEP model seems worse. A similar pattern holds for LEP predictions in other small to medium sized LMGs. The full model's performance looks better in some larger LMGs. In LMG 4, for example, we see in the Table that the full model outperforms the intercept-only model in the lowest VOT sample size bin, while PctRelΔ increases in most other bins.

In many cases, inclusion or exclusion of a single covariate makes a noticeable difference in the performance of the model with covariates. For example, in LMG 4 the LEP model without the predictor `FRNBORN` performs about as well as the intercept-only model. Omitting the other predictor variables has a smaller impact on PctRelΔ. In most AIAN and some Asian LMGs, most VOT sample size bins have absolute PctRelΔ larger for the intercept-only model than for the full model, but the differences are generally small.

A visual summary across many LMGs is given in Figure 4, which display log absolute PctRelΔ for LEP over all LMGs where the MLN-D model can be fitted. This includes all Asian LMGs and all AIAN LMGs except for LMG 25, 39, 48, 60, 69, and 71. We separate 'Large' from 'Small' Asian LMGs by whether they have more than 700 sampled jurisdictions. In these plots, the log absolute PctRelΔ is plotted at the left of each size-interval facet, with log absolute PctRelΔ for the intercept-only model plotted at the right. Points from the same LMG are colored and connected by lines of the same color. An upward-sloping line indicates an LMG with full model predictions differs less (from direct estimates) than the intercept-only model, and therefore with better predictions by the full model in that LMG and size-class. A downward sloping line indicates better performance by the intercept-only model. In general, the full models for LEP tend to perform better relative to intercept-only models in large Asian LMGs, such as LMG 4, than they do in smaller Asian LMGs, such as LMG 2. There is considerable variability in the performance of full models for AIAN LMGs, with some outperforming the intercept-only models, some underperforming them, and some with little visible difference. When the full model does in larger LMGs seem to work well with respect

Table 7: Relative Discrepancy PctRelΔ for LEP outcome on jurisdictions aggregated into VOT sample-size classes for LMGs 2, 4. PctRelΔ is shown for MLN-D models with full set of covariates, with single covariates removed, and with Intercept-only.

LMG		VOT sample-size intervals					
		(0,4]	(4,12]	(12,25]	(25,50]	(50,200]	200+
2	njuris	323	104	40	21	24	5
	Estd. Pop.	1,295	2,230	2,293	2,463	10,140	17,910
	Full Model	47.25	0.43	-6.23	-1.93	-3.15	-0.36
	Intercept-Only	36.26	1.59	-2.80	-0.57	-2.56	-0.54
	STL.T omitted	47.56	0.31	-6.26	-1.74	-2.99	-0.36
	EDU2 omitted	35.89	1.07	-2.93	-0.42	-2.68	-0.48
4	njuris	1,824	612	277	207	214	175
	Estd. Pop.	5,704	10,220	12,460	15,660	44,500	869,800
	Full Model	9.02	-6.72	-6.09	0.14	-0.58	-0.01
	Intercept-Only	12.37	-4.17	-4.78	-0.21	-0.55	-0.07
	STL.T omitted	10.52	-6.57	-5.76	-0.14	-0.59	-0.02
	EDU2 omitted	4.65	-7.28	-5.75	0.49	-0.13	0.00
	FRNBORN omitted	11.81	-4.43	-5.42	0.04	-0.68	-0.06

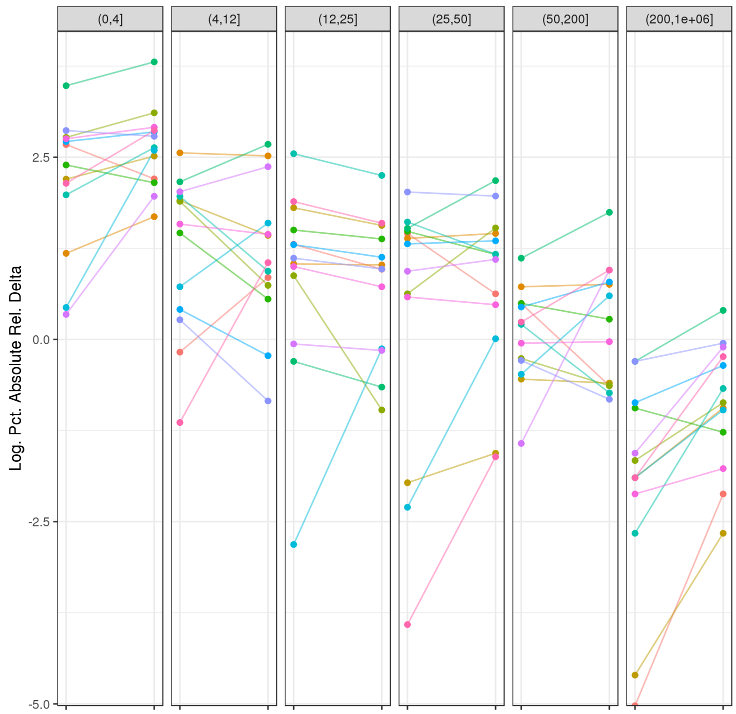
Note: Estd. Pop. is the (rounded) direct estimate of LEP count in jurisdictions of indicated size range.

to the partition of jurisdictions by sample size, that effect mostly occurs in the lowest VOT sample size bin. Plots like Figure 4 for small Asian LMGs, not shown, have mostly rather flat or slightly downward-sloping lines, with a few LMGs showing steeper upward- or downward-sloping lines.

There are fewer ILL observations in our sample, and therefore fewer LMGs with sufficient sample to fit the MLN-D model with covariates. When the ILL model can be fit with covariates in an LMG, its number of Jurisdictions with LEP sample is generally much smaller and the ILL models have fewer covariates than LEP for the same LMG. Covariates are less significant (as indicated by their model-standardized coefficients), less predictive and the full-covariate models versus intercept-only often do worse compared to the LEP full models.

Figure 4: Lineplots Large Asian and all-AIAN

Large Asian Jurisdictions



AIAN Jurisdictions

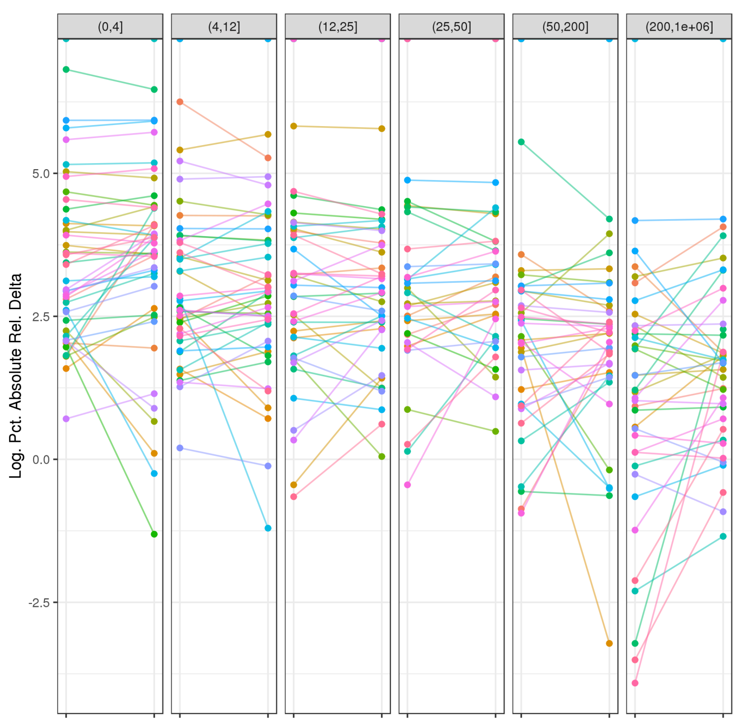


Table 8: Relative Discrepancy PctRelΔ for LEP in LMG 2 on jurisdictions aggregated respectively into EDU2 and FRNBORN intervals. PctRelΔ is shown for MLN-D models with full set of covariates, with single covariates removed, and with Intercept-only.

	EDU2 intervals				
	$(0, 0.356]$	$(0.356, 0.43]$	$(0.43, 0.491]$	$(0.491, 0.557]$	$(0.557, 1]$
njuris	220	142	82	46	27
Estd. Pop. (rounded)	7,383	6,852	16,730	4,337	1,028
Full Model	-0.94	-2.28	0.96	0.46	8.96
Intercept-Only	5.37	-3.32	-0.06	-2.62	2.50
EDU2 omitted	4.94	-3.65	0.14	-2.16	0.86
STL.T omitted	-1.01	-1.88	0.91	0.41	9.87

	FRNBORN intervals				
	$(0, 0.06]$	$(0.06, 0.1]$	$(0.1, 0.15]$	$(0.15, 0.25]$	$(0.25, 0.66]$
njuris	106	105	103	101	102
Estd. Pop. (rounded)	367	1,841	1,930	5,865	26,330
Full Model	46.90	3.06	4.20	-5.66	0.26
Intercept-Only	30.13	2.64	4.90	-3.49	0.08
FRNBORN omitted	43.79	2.55	3.76	-5.55	0.35
STL.T omitted	45.04	2.14	4.45	-5.19	0.31
EDU2 omitted	30.07	2.65	3.98	-4.22	0.25

Notes: njuris and PctRelΔ defined in Notes under Table 3.

To counter the impression that the covariates have no value, we exhibit their positive effect on model predictions in two ways. First, partitioning the jurisdictions by intervals of values of significant variables like EDU or FRNBORN in a model allows the value of these covariates to come into clearer focus. In Table 8 and other tables not shown, the PctRelΔ discrepancy measures are much better for the full model including the covariate than for intercept-only models. Again this effect is strongest for the large Asian LMGs but persists for some of the larger AIAN LMGs.

The diagnostic reported in the first half of Table 8 shows that the PctRelΔ relative-discrepancy of LEP predictions from the full model is much less (in columns 1,2 and 4 although not 3 and 5) as compared with predictions from the model omitting EDU2. The model omitting STL.T performs as well (and better in column 3) than the full model. These results accord with internal model-based Wald tests for significance of coefficients. In LMG2, the CIT stage of the MLN-D model uses covariates FRNBORN, AvgYears; the LEP stage of the model uses STL.T, EDU2; and the ILL model uses EDU2. Of these covariates, only FRNBORN, AvgYears for CIT and EDU2 for LEP have model-standardized estimated coefficients that are significant (all strongly so, with Wald-test p-values < 0.00015). We see that aggregating jurisdictions according to similar values of the strong

LEP-model covariate results in worse predictions when that strong covariate (`EDU2`) is dropped from the model, but improved predictions when a very insignificant covariate (`STL.T`) is dropped.

In the second half of Table 8, jurisdictions are aggregated by similar values of the `FRNBORN` covariate, which along with the closely related covariate `AvgYears` is highly significant in the CIT model; but does not appear in the LEP stage model. In this setting, as when the jurisdictions were aggregated by the VOT sample-size, the Full model LEP predictions are actually worse than those of the intercept-only model. The overall conclusion from these jurisdiction-grouping relative-discrepancy diagnostics seems to be that stronger covariates can be felt in the improved quality of predictions when the grouping variable is itself the strong covariate, but perhaps not otherwise.

Table 9: Correlations (`corr`) between model-predicted and direct-estimated LEP/VOT and ILL/VOT proportions, over the `njuris` jurisdictions with direct LEP ≤ 1000 or ILL ≤ 100, respectively for LMGs with ≥ 50 jurisdictions with LEP or ILL sample.

LEP	LMG	1	2	3	4	5	6	7	8	9	10
	`corr`	0.110	0.207	0.129	0.091	0.131	0.057	0.139	0.145	0.095	0.064
	`njuris`	864	173	386	1,119	1,090	290	156	550	935	376
	LMG	11	12	13	14	15	16	17	18	20	22
	`corr`	0.163	0.169	0.155	0.074	0.101	0.160	0.151	0.216	0.211	0.022
	`njuris`	84	356	95	534	960	589	59	219	171	111
	LMG	26	27	44	45	56	57	63	64	72	
	`corr`	0.059	0.006	0.197	0.023	0.060	0.258	0.001	0.080	0.069	
	`njuris`	244	120	212	65	227	78	264	212	583	
ILL	LMG	1	2	3	4	5	6	8	9	10	
	`corr`	0.090	0.143	0.075	0.114	0.019	0.132	0.056	0.127	0.075	
	`njuris`	317	50	207	454	212	185	86	248	221	
	LMG	12	14	15	16	18	44	72			
	`corr`	0.058	0.050	0.088	0.104	0.057	-0.068	-0.128			
	`njuris`	104	177	494	217	111	81	152			

This subsection argues only that the covariates have some value, not that they are strong predictors. To assess this fairly, we turn to another diagnostic. In models with covariates, we compare *synthetic* LEP rate predictions $\tilde{\pi}^L_{jg}$ versus the direct-estimated ratios $\hat{N}^L_{jg}/\hat{N}^V_{jg}$. *Synthetic* means that the domain (j, g) predictor is generated from parameters of the fitted model and domain covariate values using as though that domain had not been sampled. These predictions reflect whether the model itself has any value in distinguishing LEP outcomes of different jurisdictions for the same LMG. The Intercept-only model ignores the covariates and predicts the same LEP/VOT ratio for the LMG in all jurisdictions. So, to the extent that the synthetic $\tilde{\pi}^{A}_{jg}$ predictions for $A = \texttt{LEP}, \texttt{ILL}$ show positive correlation with the direct-estimated observed rates, they show predictive value in the same way weak but significant linear-regression predictors do. In Table 9, the correlations are generally low but large enough to reflect useful models, and scatterplots bear this out. The ILL correlations are smaller than LEP correlations, and in some AIAN LMGs are so small that the intercept-only model is clearly warranted for ILL in those LMGs. However, LMG 72 is anomalous and hard to interpret, with ILL predictions and direct ILL/VOT ratios negatively correlated.

3.6.2 Jurisdictions — Models & Covariates

In this subsection and later ones on AIAs and ANRCs, it is shown how the general approach described above to the inclusion of covariates in the MLN model translated into a grouping of LMGs for each geography-type and a specific set of covariate coefficients to fit in MLN-D or MLN-F models for each group. The guiding principle was that larger LMGs (with larger sample and larger population within geographic units) can make greater use of more predictive covariates, and that the covariates selected should apply to all LMGs within a few main groupings.

In the Jurisdiction geography, it was necessary to screen covariates for models in 21 Asian and 51 AIAN LMGs and — because of our decision to split the Hispanic LMG into four Regional sub-LMGs for purposes of model-fitting and prediction — 4 more models for the Hispanic LMG. Each model consisted either of a 3-stage combined MLN-F model with correlated random effects or an MLN-D model in which the three CIT, LEP and ILL models could be fitted independently. It was decided, for reasons elaborated in Section 3.5, that a Bayesian MCMC model-fitting approach would be used only on large LMGs, and only within the Jurisdiction geography-type, and the criterion for 'large' was that the number of within-LMG sampled jurisdictions be larger than 700. There were 21 such LMGs (including 13 Asian LMGs, 7 AIAN, and all 4 regional Hispanic LMGs) all with total direct-estimated population of at least $88,000$.

A first stage of automated screening for covariates was done with fixed-effect logistic regression models, first with single covariates; then with combinations chosen through greedy stepwise selection (removing one variable at a time), primarily in Asian LMGs; and finally, in parallel and independently across LMGs, for random-intercept logistic models. The sets of frequently chosen covariates differed slightly for CIT, LEP and ILL. In decreasing order of importance for each outcome-group, covariates `FRNBORN, STC.T, EDU2, OTHLANG, AvgYrs, NumPer, WHNHSP` were dominant in CIT models; `EDU2, OTHLANG, STL.T, FRNBORN, NumPer, WHNHSP, AGEGP2, AGEGP3`[6] in LEP models; and `EDU2, OTHLANG, STI.T, WHNHSP, AGEGP3` in ILL models. In some LMGs, especially the largest ones, stepwise selection included more variables; but in some smaller AIAN LMGs, the

[6] There were 3 Age-group variables corresponding to age-intervals 18–44, 45–64, 65+.

random-effect MLN-D models with more than 1 or 2 covariates did not converge, We separated out a group of AIAN LMGs (nearly all small) where either CIT/VOT direct-estimated ratios were > 0.9991 or ILL/LEP direct-estimated ratios were $= 0$), and in that group declared that both the CIT and ILL MLN-D stage models would be fitted intercept-only.

The rules for inclusion of covariates in models for different LMGs can be understood as a function of `njuris`, the count of within-LMG jurisdictions with VOT sample, as follows:

- *Large Asian LMGs*: in the 6 Asian LMGs with `njuris` > 2000, the CIT model was fitted with covariates `STC.T, EDU2, FRNBORN, AvgYrs`; the LEP model with `STL.T, EDU2, FRNBORN`; and the ILL model with `EDU2, NumPer`

- *Medium-sized Asian LMGs*: in the 8 Asian LMGs with `njuris` between 600 and 2000, and with direct-estimated VOT population $> 75,000$, the CIT model was fitted with covariates `EDU2, FRNBORN, AvgYrs`; the LEP model with `EDU2, FRNBORN`; and the ILL model with `EDU2`

- *Small Asian LMGs*: in the 7 Asian LMGs with `njuris` between 100 and 600, the CIT model was fitted with covariates `FRNBORN, AvgYrs`; the LEP model with `STL.T, EDU2`; and the ILL model with `EDU2`

- *Regional sub-parts of Hisp LMG* : the CIT model was fitted with covariates `STC.T, EDU2, FRNBORN, AvgYrs`; the LEP model with `STL.T, EDU2, OTHLANG, AGEGP2, AGEGP3`; and the ILL model with `EDU2` for Regions 1,2 (Northeast, Midwest) and `STI.T, EDU2, AGEGP3` for Regions 3,4 (South, West)

- *High CIT or Low ILL rate AIAN LMGs*: in the 37 AIAN LMGs with direct-estimated CIT/VOT > 0.9991 or ILL $= 0$, the CIT and ILL models were intercept-only, and the LEP model was fitted with covariates `STL.T,OTHLANG`, to which were added `EDU2, POV` for LMGs with `njuris` > 500

- *Other AIAN LMGs*: CIT model fitted with covariates `STC.T,Reg4`, plus `POV_AG2` when `njuris` > 500; LEP model fitted with covariates `STL.T, OTHLANG`, plus `POV, Reg4` when `njuris` > 500; ILL model fitted with `OTHLANG`, plus `EDU2` when `njuris > 500`

3.6.3 American Indian Areas — Models & Covariates

In the AIA geography data, it was necessary to screen covariates for 51 models. None of the LMGs has more than 350 AIAs with VOT sample. Moreover for AIAN LMGs in AIAs, very few AIAs account for most of the sample: in 28 of the 51 AIAN LMGs, more than half of the AIA sample falls within 3 AIAs. Therefore only the MLN-D model (with independent stagewise random-effects) was considered, which means that the models took the form of 51 separate triplets of models for CIT, LEP and ILL outcome rates, each in terms of the next higher level (respectively VOT, CIT, LEP) in the nested outcome hierarchy of Fig. 1. Not only did many of the LMGs have extremely small numbers (less than 50) of AIAs with sampled persons, many of the LMGs also had CIT/VOT ratios extremely close to 1 and/or ILL/LEP ratios extremely close to 0. For the CIT and ILL models in all such LMGs, a random-intercept logistic (binomial logit-normal) intercept-only form was chosen.

Preliminary investigation of models in LMGs for which CIT and ILL rates were extreme showed that many MLN-D models did not converge, and in those LMGs we fell back on beta-binomial models. In LMGs without such extreme rates, we found that only a few predictive covariates were useful, judging by model-based tests of significance of coefficients and related stepwise model selection, those covariates were `LEPrat` and occasionally `WHNHSP, EDU2, POV`. The result of these investigations was the following set of rules assigning covariates to LMGs:

- *Beta-binomial models*: if CIT/LEP for an LMG in was ≥ 0.995 or was ≥ 0.95 when also ILL/LEP ≤ 0.001, the CIT model was fitted as beta-binomial (with $\text{Beta}(1/2, 1/2)$ prior); if LEP/CIT was ≤ 0.0025 for all AIAN persons in AIAs in the LMG, the LEP model was fitted as beta-binomial; and if the total of LEP sampled AIAN persons was < 3 or ILL/LEP was ≤ 0.001 or both ILL/LEP ≤ 0.003 and at most 5 LEP AIAN persons were sampled, then the ILL model was beta-binomial

- *Other CIT models*: in other LMGs, the CIT model was MLN-D with the single covariate `CITrat` if that model's estimates converged, and otherwise the model was intercept-only

- *Other LEP models*: in AIAN LMGs 9, 32, 43, 49, 51, use covariates `LEPrat, WHNHSP`; in AIAN LMG 50 use `LEPrat, POV`; and in other LMGs with ILL/LEP > 0.0025 use `LEPrat` and change to intercept-only if < 40 CIT persons were sampled or if the model fails to converge or the `LEPrat` coefficient is not significant

- *Other ILL models*: in AIAN LMGs 24, 44 use covariates `ILLrat, EDU2`; in AIAN LMG use `ILLrat, WHNHSP`; and otherwise use the single covariate `ILLrat`, reverting to intercept-only if the model is nonconvergent or the coefficient non-significant or fewer than 20 LEP AIAN persons were sampled

- *Intercept-only models*: in all cases using an intercept-only model, switch to beta-binomial either if the model is nonconvergent or the estimated random-effect σ parameter is $< .001$

3.6.4 Alaska Native Regional Corporations — Models without Covariates

For ANRC geography, it was necessary to screen covariates for 51 models. Since there were only 12 distinct geographic units, we considered only the MLN-D model for predictions, and we rapidly found that no covariates were so different between single ANRCs that they helped with predictions. So the only further modeling choice was to decide which of the model stages for the 51 AIAN

LMGs could be fitted with an intercept-only logit-normal model and which must be handled with the backup beta-binomial model. The rule adopted to distinguish between these two cases is:

- *Beta-binomial models for LMGs*: for LMGs in which fewer than 8 ANRCs have VOT sample,the CIT, LEP and ILL models are all beta-binomial; in addition, for LMGs with CIT/VOT ≥ 0.996, the CIT model is beta-binomial; for LMGs with LEP/CIT ≤ 0.002 the LEP model is beta-binomial; and for LMGs with ILL/LEP ≤ 0.002 the ILL model is beta-binomial

- *Intercept-only models*: for all other LMG and CIT, LEP, ILL model combinations, the model is MLN-D intercept-only; except if that model fails to converge or random-effect σ parameter is estimated $< .001$, the model reverts to beta-binomial

3.7 Limitations of Model Assumptions

The model assumptions used in the VRA statistical analysis have several limitations with respect to the available data. Two in particular can be assessed and improved upon in future VRA cycles, with the objective of providing firmer theoretical underpinnings to the model-based estimates used in the production of statistical estimates on which VRA determinations are based. The first limitation has already been mentioned: we used multinomial regression models — models assuming integer-valued count data — for the non-integer derived quantities $Y_{jk,g} = (n^V_{jg}/\hat{N}_{jg}) \cdot \hat{N}^A_{jg}$ (k=1 for A=C, 2 for A=L, and 3 for A=I) introduced in assumption (iv) of Section 3 to partition the (j, g) sample-size n^V_j into categories A. A second important assumption that can be challenged is the independence (of $\hat{N}^V_j$) from ratios $\underline{\pi}_{jg} = (N^A_{jg}/N^V_{jg}, \quad A = C, L, I)$ assumed in (iii) of Sec. 3 and in (7) of Appendix C. This assumption is used both in model formulation and variance estimation, and was used similarly in the 2016 VRA cycle. A preliminary assessment of this assumption is contained in the small correlational study summarized in the following paragraphs and tables.

3.7.1 Testing an Independence Assumption via Correlations

There are two different ways to obtain estimated correlations between values $\hat{N}^V_j$ and (estimated) fractions $\hat{\pi}^A_j$ within a fixed LMG g. Since the independence assumption (7) of Appendix C applies only to the random variables for a fixed geographic unit j, the most relevant way to find correlations is to re-compute $\hat{N}_{j,(r)}$, $\hat{\pi}^A_{j,(r)}$ for each of the $r = 1, \ldots, 80$ SDR weight-replicates used in standard ACS calculations of variances and covariances and take the correlation across these two vectors of 80 numbers. A second way to look at dependence (without considering weight-replications) is to take the correlations across all geographic units j (with positive VOT sample-size) within LMG g. This second type of correlation has a somewhat different interpretation.

First, Table 10 summarizes the jurisdiction-specific correlations across SDR replicate estimates. For each of 12 LMGs and each of the populations A = C (CIT), L (LEP), I (ILL) in all jurisdictions

j that had positive VOT sample-size in ACS 2015-2019, we computed the correlations

$$\frac{\sum_{r=1}^{80} (\hat{N}_{j,(r)} - \bar{N}_j)\, (\hat{\pi}^A_{j,(r)} - \bar{\pi}^A_j)}{\left[\, \sum_{r=1}^{80} (\hat{N}_{j,(r)} - \bar{N}_j)^2 \, \sum_{r=1}^{80} (\hat{\pi}^A_{j,(r)} - \bar{\pi}^A_j)^2 \,\right]^{1/2}} \quad , \quad \bar{N}_j \;=\; \frac{1}{80} \sum_{r=1}^{80} \hat{N}_{j,(r)} \quad , \quad \bar{\pi}^A_j \;=\; \frac{1}{80} \sum_{r=1}^{80} \hat{\pi}^A_{j,(r)} \tag{4}$$

Quartiles of these computed correlations for each of 12 LMGs, across all jurisdictions with VOT sample for the LMG, are displayed in the first 3 columns of the table, with the number of such jurisdictions for the LMG in the 4th column. In each LMG, there are some jurisdictions, tallied in the 7th column of the table, for which these estimated correlations are large, in the sense that their absolute values are ≥ 0.25. However, the VOT sample sizes in those large-absolute-correlation jurisdictions are for the most part extremely small. We take Table 10 to be a rough confirmation that for each jurisdiction in each LMG other than LMG 72, the VOT estimated total $\hat{N}_{jg}$ is approximately uncorrelated with the model-predictions $\hat{\pi}^A_{jg}$ for the ratios $\pi^A_{jg} \equiv N^A_{jg}/N^V_{jg}$. Apart from LMG 72, the lower and upper quartiles bracket 0, typically with absolute values of size 0.12 or smaller. For some reason we cannot fully explain, LMG 72 is different, showing systematically positive correlations that are not small. Indeed, in LMG 72, correlations greater than 0.25 occur respectively in 2747 and 2430 jurisdictions out of 4399 for LEP and ILL, with correlations < -0.25, respectively occurring only 27 and 28 times.

One reason why LMG 72 *might* be different is that it is a catch-all category, comprised of American Indian or Alaska Native persons from many different unspecified tribes (other than those accounted for in other LMGs). In that sense, the correlations (4) within LMG 72 might be viewed as correlations across many distinct, individually small, tribal groups. That is a kind of correlation that we can approximate by looking at correlations of $\hat{N}^V_j$ and $\hat{\pi}^A_j$ across sampled jurisdictions within LMG. Table 11 shows correlations of that type, for $A = C, L, I$. However, these correlations across geographic units do not directly address the within-jurisdiction independence-assumption that was the main topic of this subsection.

The strong positive correlations seen across weight-replicates in LMG 72 are a concerning anomaly. We investigated this further by breaking down the cross-replicate correlations for all LMGs separately for jurisdictions in each of 6 VOT sample-size classes, respectively those with VOTsmp in the ranges 1, 2–3, 4–6, 7–10, 11–40, 41–100, and 101+. Remarkably, although this does not occur in any other of the 11 LMG's in Table 10, the cross-replicate correlations for LMG 72 in each outcome group CIT, LEP, ILL become progressively stronger for the larger jurisdiction sample-sizes. This may be an oddity of ACS weighting that applies especially strongly to members of this catch-all AIAN group LMG 72, but it may also be true of AIAN LMGs that, like LMG 72, have a particularly large proportion of their population falling in jurisdictions with small sample size.

Table 10: Summary of Correlations between $\hat{N}_{j,(r)}$, $\hat{\pi}^A_{j,(r)}$ across weight-replicates $r = 1, \ldots, 80$, for CIT, LEP and ILL totals in Jurisdictions j within indicated LMGs. Cols. 1-3 are quantiles across j of cross-replicate correlations. Col. 4 is the number of jurisdictions with ACS sample within LMG. Cols. 5-7 refer to jurisdictions with absolute cross-replicate correlations ≥ 0.25: the median and Q3 of the sample sizes for correlations ≥ 0.25 and the number of such jurisdictions.

Total	LMG	Cross-Replicate Correlations				Large-Corr. Jurisdictions		
		Q1.cor	med.cor	Q3.cor	Num.Jur	med.samp	Q3.samp	Num.Jur
CIT	2	-0.109	0.009	0.121	517	5	14.25	136
	4	-0.137	0.006	0.147	3309	5	14.5	943
	7	-0.078	0.019	0.151	845	3	7	206
	8	-0.070	0.040	0.162	3029	3	8	675
	11	-0.073	0.022	0.130	461	3	8	118
	13	-0.083	0.005	0.108	500	4	8	117
	15	-0.103	0.016	0.150	2374	5	14	611
	22	-0.066	0.026	0.125	1274	1	2	152
	24	-0.059	0.027	0.131	1905	1	1	259
	27	-0.070	0.014	0.107	3830	1	2	350
	61	-0.067	0.007	0.116	301	1	2	34
	72	0.253	0.339	0.497	4399	4	11	3353
LEP	2	-0.089	0.011	0.140	517	6	15	121
	4	-0.105	0.015	0.161	3309	5	13	854
	7	-0.105	0.003	0.118	845	2	6	168
	8	-0.115	-0.013	0.097	3029	2	7	506
	11	-0.101	-0.01	0.089	461	2	7.5	83
	13	-0.093	0.003	0.118	500	4	7	106
	15	-0.102	0.019	0.171	2374	5	11	621
	22	-0.084	0.000	0.092	1274	1	3	168
	24	-0.068	0.024	0.120	1905	1	1.5	251
	27	-0.073	0.008	0.096	3830	1	3	383
	61	-0.075	-0.003	0.106	301	1	5.5	42
	72	0.214	0.289	0.396	4399	4	11	2774
ILL	2	-0.083	0.008	0.131	517	5	11	92
	4	-0.104	0.005	0.133	3309	4	10	725
	7	-0.091	-0.002	0.099	845	1	4	104
	8	-0.095	0.002	0.105	3029	2	6	405
	11	-0.091	-0.009	0.086	461	1	6.75	70
	13	-0.080	0.003	0.093	500	3	6.5	91
	15	-0.095	0.013	0.145	2374	4	9	520
	22	-0.079	0.005	0.101	1274	1	3	161
	24	-0.056	0.016	0.100	1905	1	2	240
	27	-0.076	0.005	0.092	3830	1	4	364
	61	-0.086	0.010	0.087	301	1	2	37
	72	0.183	0.267	0.371	4399	5	13	2458

Table 11: Correlations between $\hat{N}_j$, $\hat{\pi}_j^A$ for $A = C, L, I$ across j within LMG.

LMG	2	4	7	8	11	13	15	22	24	27	61	72
CIT	0.02	0.03	-0.01	0.02	-0.07	0.06	0.05	0.01	0.00	NA	-0.07	-0.19
LEP	0.13	0.14	0.25	0.02	0.03	0.17	0.15	0.13	0.07	0.03	0.14	0.17
ILL	0.04	0.10	-0.01	-0.02	-0.02	0.07	0.09	0.05	0.07	0.03	0.16	0.21

4 Variance and Mean-Square Predictor Error (MSPE) Estimation

At least 3 different kinds of variance or mean-square prediction-error (MSPE) estimates are produced to quantify variability in our statistical predictions for (Geo, LMG) population subgroups. All are described in full details, both conceptually and computationally, in the Appendix D in sections D.1, D.2, D.5 and D.6. Briefly, these are the standard successive difference replication (SDR) variance estimates used by ACS for direct survey-wieghted estimates, and two methods that combine the SDR variances for the direct estimate of the VOT population $\hat{N}_{jg}^V$ in (Geo, LMG) with model-based Bayesian or frequentist MSPEs. The Bayesian method relies on Markov-Chain Monte-Carlo (MCMC) simulations estimating posterior variances for model predictions, while the frequentist method applies the SDR idea to model-based predictions recalculated for 80 different weight-replicates. Some comparisons between these two different types of model-based MSPEs calculated on a selection of LMGs are given in Section 4.2 below. The model-based MSPE estimation methods adopted in this VRA cycle differ from a method based on the model-based Parametric Bootstrap [Slud and Ashmead, 2017] used in the 2016 VRA cycle [Slud et al., 2018].

4.1 Comparing Direct with Model-Based Variances

Within each of the 7,859 jurisdictions, we aim to estimate values for 73 LMGs, leading to a total of 573,707 estimands for each of 6 variables: the VOT, CIT, LEP and ILL counts), the proportion of the total citizen[7] population within a geography that is LEP (`LEPprop`), and the (Geo, LMG) domain ratio ILL/LEP. We estimate the same variables within the 568 AIAs and 12 ANRCs in 51 Native American LMGs, for a total of 28,968 and 612 estimands respectively.

The tables and figures in this section will focus on quantities related to LEP and ILL, as these are the most relevant for the VRA Section 203(b) coverage determinations. Unfortunately, as we can see in the following tables, the 2015-2019 5 Year ACS has small samples of voting age citizens for many LMG / geography combinations we are interested in.

[7]citizen AIAN, for AIA and ANRC geographies

Table 12 shows the numbers of (Geo, LMG) domains with indicated ranges of ACS 2015-19 sample-size for jurisdictions, AIAs, and ANRCs. Domains with no sampled VOT persons have no direct or model-based estimates. Domains with very small samples have extremely imprecise direct estimates, and we fit models in an attempt to reduce this estimation error. Estimated standard deviations (SDs or MSPEs) and Coefficients of Variation (CVs) measure the quality of these estimates. For our purposes, the most important variables we estimate are LEP and ILL/LEP, since these are used to make coverage determinations. This section compares CVs and SDs associated with the direct and model-based estimates of LEP and ILL/LEP.

Table 12: Rounded numbers of (Geo, LMG) domains with indicated ranges of ACS 2015-19 VOT sample size, for Jurisdiction, AIA, and ANRC geographies

Sample	0	1	2-5	6-10	11-20	21-50	51-100	101-250	251-500	501+
Juris	495,100	20,670	27,210	7,911	5,253	4,333	2,005	1,688	688	895
AIA	24,790	1,456	1,358	373	282	275	160	192	51	31
ANRC	353	66	84	27	27	19	10	10	5	11

Table 13: Percent of (Juris, LMG) and (AIA, AIAN LMG) domains with $CV < 0.6$ for variables LEP and ILLrat, by sample size, based on M= Model-based or D=Direct estimates

Sample Size	1	2-5	6-10	11-20	21-50	51-100	101-250	251-500	501-1000	1001+
LEP (M)	0.2	5.8	24.3	46.1	70.0	82.5	88.5	94.3	97.4	99.8
LEP (D)	0.0	0.3	2.4	9.4	30.4	61.0	80.9	91.0	95.0	99.6
ILL/LEP (M)	41.7	40.1	34.4	35.2	38.4	50.9	66.9	81.0	93.0	97.3
ILL/LEP (D)	0.3	1.8	3.4	4.6	7.3	12.7	23.1	46.2	68.1	93.7
AIA										
LEP (M)	0.0	14.0	50.7	63.1	70.5	67.5	77.1	74.5	73.7	83.3
LEP (D)	0.0	0.2	0.5	1.4	6.2	18.1	38.0	51.0	63.2	91.7
ILL/LEP (M)	54.5	76.1	71.3	69.9	73.8	77.5	80.2	88.2	94.7	100.0
ILL/LEP (D)	0.0	0.2	0.8	1.8	2.2	7.5	10.9	7.8	10.5	16.7

Broadly, we consider four types of estimates: (i) direct survey-weighted estimates with SDs estimated using Successive Difference Replication (SDR); (ii) Bayesian estimates from the multinomial logit normal (MLN-F) model with general random-effect covariances, with SDs estimated as posterior standard deviations; (iii) frequentist estimates from the multinomial logit-normal(MLN-D) models with random effects assumed independent and SDs estimated using SDR applied to model-based predictions; and (iv) Beta-Binomial model estimates with no covariates. Types (ii)-(iv) are model-based estimates. Each (Geo, LMG) domain with VOT sample has one direct and one model

Table 14: Rounded count of (Juris, LMG) domains with estimable CV by variable × estimate-type

Variable	Direct	Model
LEP	18,600	78,260
ILL/LEP	5,527	78,200

based estimate, the latter of type (ii) in data-rich domains and of type (iii) [sometimes combined with (iv)] in domains with fewer sampled domains. The primary comparisons of this Section are of estimated SDs of direct versus model-based estimates for LEP counts and ILL/LEP ratios.

The range of uncertainties of estimates is displayed in terms of their coefficients of variation (CVs). The Census Bureau's quality standards for ACS require a majority of published key estimates to maintain a CV < 0.3. Estimates with CVs greater than 0.61 are deemed unreliable. In Table 13, the percentage of estimates acceptable for ACS, with CVs < 0.6) is seen to be low, due in large part to the small sample sizes available for smaller LMGs and jurisdictions or AIAs in the ACS. The percentage of acceptable model-based estimates is noticeably higher than that of acceptable direct estimates in all sample-size classes. (The same holds for ANRCs, not shown.)

Another motivation for producing model-based estimates is that CVs cannot be calculated for domains were the sample is too small to estimate both the standard deviation and mean. This often implies that CVs are not estimable using direct estimates, but are using models. The numbers of (Juris, LMG) domains with estimable CVs in this sense are shown in Table 14.

For (Juris, LMG) domains with LEP $< 10,000$ to influence the Section 203(b) determinations, the ratio of the estimate of LEP to the total voting age citizen population of that geography (AIAN citizen population for AIA or ANRC) must be greater than 0.05. There are only 5 domains for jurisdictions and 2 for AIAs that meet this criterion and have sample sizes of 5 or less. Hence, outcomes in the lowest two sample size categories will have limited practical impact.

Advantages of the model estimates are also shown in Table 15 comparing the distribution of model and direct CV estimates in (Juris, LMG) domains where both can be calculated. The median

Table 15: Quantiles of CVs for LEP and ILL/LEP variables in (Juris, LMG) domains where both direct and model CVs can be estimated and are not 0

Variable	Est. Type	Q5	Q25	Median	Q75	Q95
LEP	Direct	0.16	0.49	0.79	0.98	1.28
LEP	Model	0.13	0.38	0.61	0.90	1.38
ILL/LEP	Direct	0.14	0.36	0.60	0.89	1.33
ILL/LEP	Model	0.15	0.31	0.41	0.56	0.75

model-based CVs are lower than the median direct CVs for both LEP, ILL/LEP in the (Juris, LMG) domains where both types of CVs can be estimated. This is another sense in which model-based CVs typically outperform the direct CVs.

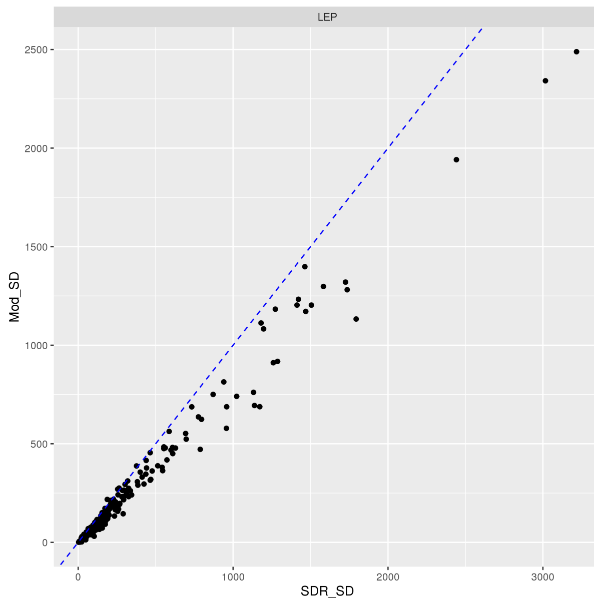
We next address frequency and improvement of the estimates in single domains, by comparing direct to model-based estimates of SDs and CV. Larger standard deviations (SDs) or CVs imply a higher degree of uncertainty, and therefore worse estimation performance. For most comparisons, we omit any estimates with a standard deviation of zero. These cases are not easily interpretable, as all of our estimates have at least some uncertainty. In many cases, we also omit estimates based on sample size less than 5, both because estimates based on extremely small samples are noisy and generally unreliable, and also because samples this small only rarely result in Section 203(b) determinations.

The scatterplots in Figure 5 show the relationship between Direct (SDR) and Model-based SDs for single (Juris, LMG) domains for the variables LEP and ILL/LEP. In each, the SDR and model-based SDs are equal along the dashed blue line. Points below [above] the line indicate (Juris, LMG) domains for which the direct estimated SD is larger [smaller] than the corresponding model-based estimate. Domains for which either the SDR or Model SD are estimated as zero are not meaningful, and have been omitted. In order to reduce visual clutter, and emphasize relevance to VRA determinations, the plots are restricted to domains with LEPprop > 0.05. In both plots, the points tend to cluster on or below the dashed blue line; that is, for the great majority of (Juris, LMG) domains, the model-based estimates have smaller estimated SDs than the direct estimates.

Tables 16 and 17 make the same point by displaying quantiles of the percent reduction of model versus direct CVs by decile of the direct estimated CV. The tables tally only (Juris, LMG) domains for which both the model and SDR SDs are positive. In both tables, the median percent reduction of CV is generally positive, meaning that the model-based CV is usually lower than the direct CV. For LEP estimates, the median reduction is similar in most of the direct CV bins, except in the largest deciles which show the largest reduction. The range of the LEP CV reductions is also seen to increase as the direct CV increases. In contrast, the median reduction of the ILL/LEP CVs seems to increase with the decile of the direct CVs: the higher the direct CV, the larger appears the benefit of the model-based estimate. Similar patterns hold for both LEP and ILLrat in (AIA, AIAN LMG) domains.

Figure 5: Nonzero SDR vs Model SDs for (Juris, LMG) domains where `LEPprop` > 0.05. (Axis labels are: SDR_SD for SDR SDs and Mod_SD for SDs of model-based estimates.)

LEP (Jurisdictions)



ILL/LEP (Jurisdictions)

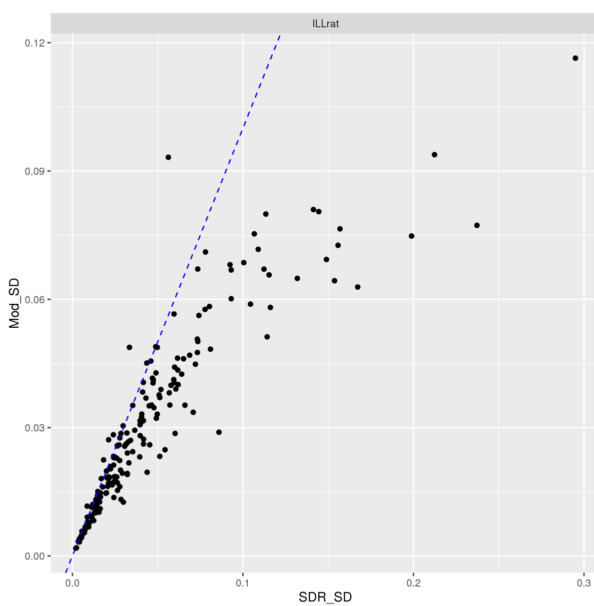


Table 16: Quantiles of percent CV reduction of model vs. direct CV of LEP in (Juris, LMG) domains with sample size ≥ 5 and nonzero SDs, displayed by decile of direct CV

	CV Decile	# (Juris,LMG)	Qu.05	Qu.25	Median	Qu.75	Qu.95
1	(0.01, 0.25]	1,849	-7.38	6.86	15.51	23.94	34.52
2	(0.25, 0.41]	1,809	-11.69	4.68	15.82	26.22	38.61
3	(0.41, 0.55]	1,826	-17.77	4.36	17.61	29.81	44.41
4	(0.55 ,0.68]	1,786	-19.43	3.37	17.06	29.41	45.74
5	(0.68, 0.79]	1,542	-24.47	3.62	18.61	32.16	49.01
6	(0.79, 0.89]	1,154	-26.25	7.06	21.50	35.62	50.26
7	(0.89, 0.95]	1,104	-30.50	10.34	24.22	36.95	52.16
8	(0.95, 1.01]	1,039	-32.51	9.79	26.72	38.85	54.36
9	(1.01, 1.13]	1,023	-23.39	14.24	30.91	42.17	58.58
10	(1.13,11.30]	924	-6.92	29.62	44.22	57.54	74.05

Table 17: Quantiles of percent CV reduction of model vs. direct CV of ILL/LEP in (Juris, LMG) domains with sample size ≥ 5 and nonzero SDs, displayed by decile of direct CV

	CV Decile	# (Juris, LMG)	Qu.05	Qu.25	Median	Qu.75	Qu.95
1	(0.01, 0.21]	436	-405.1	-13.09	9.39	18.22	29.55
2	(0.21, 0.32]	467	-175.3	-15.58	13.79	24.47	32.86
3	(0.32, 0.41]	422	-67.72	-4.69	16.77	28.18	38.97
4	(0.41, 0.50]	477	-32.96	8.51	21.62	32.08	43.32
5	(0.50, 0.60]	501	-29.66	9.70	25.77	35.55	48.71
6	(0.60, 0.71]	490	-10.18	18.75	33.98	43.87	54.68
7	(0.71, 0.82]	441	1.13	22.21	37.22	48.56	59.59
8	(0.82, 0.95]	505	9.10	30.06	41.39	53.01	65.18
9	(0.95, 1.10]	460	19.05	37.49	45.89	57.78	69.10
10	(1.10,12.60]	478	36.68	50.35	61.21	71.06	82.30

Throughout, we have focused on VOT sample sizes in ACS 2015-2019. Yet even if the VOT sample is large, in the (Geo, LMG) domains where this sample includes no citizens or no LEP citizens, it is not possible to produce direct estimates or SDR SDs. However, model estimates can still be produced by using covariates and outcomes from other domains with the same LMG. Model estimates for LEP when there is no CIT sample, or for ILL when there is no LEP sample, are called 'synthetic'. Counts of domains with synthetic estimates are displayed by VOT sample size in Table 18. The great bulk of such estimates are in domains with small VOT sample sizes. Because synthetic estimates have no direct estimates to compare against, we summarize the differences between synthetic model-based estimates and non-synthetic model-based estimates in Table 19.

Table 18: Number of (Juris, LMG) domains (rounded to 4 significant digits) with synthetic estimates for LEP or ILL/LEP and model SD > 0, displayed by VOT sample size

VOTsmp	1	2-5	6-10	11-20	21-50	51-100	101-250	251-500	501+
LEP	2,566	1,302	69	12	2	0	0	0	0
ILL/LEP	27,500	22,900	5,251	2,561	1,272	291	100	15	2

Table 19: Quantiles of Positive Model-based CVs by Synthetic Status, for (Juris, LMG) domains

Variable	Synthetic	Qu.05	Qu.25	Median	Qu.75	Qu.95
LEP	FALSE	0.31	0.70	0.98	1.33	2.48
LEP	TRUE	0.76	0.99	1.20	1.40	1.91
ILL/LEP	FALSE	0.21	0.43	0.61	0.75	1.12
ILL/LEP	TRUE	0.21	0.45	0.67	0.93	1.95

The general pattern is that purely synthetic LEP estimates exhibit a reduced range, being slightly higher than non-synthetic LEP estimates below the median, but lower than non-synthetic LEP estimates above the median. Lower quantiles of synthetic ILLrat CVs are similar to those of non-synthetic ILL/LEP CVs, but higher quantiles (particularly at or above the median) are notably larger for synthetic as opposed to non-synthetic estimates. At the median, synthetic ILL/LEP and LEP estimates have higher model CVs than non-synthetic estimates. In this way, the synthetic ILLrat CVs show an *increased* range, in contrast to the LEP CVs. The upshot is that the synthetic CVs seem to have different distributions than their non-synthetic counterparts, based on their central tendencies and the shape of their tails.

Inclusion determinations for Section 203(b) of the Voting Rights Act depend crucially on the number of voters with limited English proficiency (the LEP count) in each (Geo, LMG) domain, and the proportion of those voters who are illiterate (ILL/LEP). The diagnostics shown above all support the conclusion that the model estimates are on average more precise than the direct SDR estimates of standard deviation for both LEP and ILL/LEP. The Section 203(b) model-based determinations are more accurate in that sense than they would have been if based on direct survey-weighted estimates.

4.2 Bayesian Posterior Variances versus SDR-based MSPEs

It is important to verify that the variances or MSPE estimates for model-based predictors agree closely when calculated by different methods *for the same model*. We have verified in Section 3.5 (Table 6, with the exceptions of LMGs 9 and 73) model MLN-D adequately captures the outcome data and the additional MLN-F parameters allowing dependence of random effects are not necessary. We verified for the same LMGs 3–10 as in Table 6 and outcomes LEP and ILL that the differently calculated Bayes-posterior derived MLN-F MSPE estimates track closely with those calculated by the replication-based frequentist MLN-D MSPE estimates. However, there is an interesting pattern by which the Bayes SDs (root MSPEs) tend to overestimate small (0–20) values by comparison with the frequentist SDR-based estimates; the Bayes versus frequentist SD estimates are just as often slightly lower or higher for medium-low SDs (say ranging from 20–50); Bayes SD estimates are lower for medium-high SDs (50–300); and the Bayes estimates are a tiny bit smaller but very close for large SD values. These ranges apply to LEP-count SDs, but a similar pattern with scaled-down ranges applies to ILL-count SDs. In fact the estimated SDs track together fairly closely in all the LMGs investigated, with the largest but still tolerable differences seen in LMG4, which is illustrated in Figure 6 for LEP.

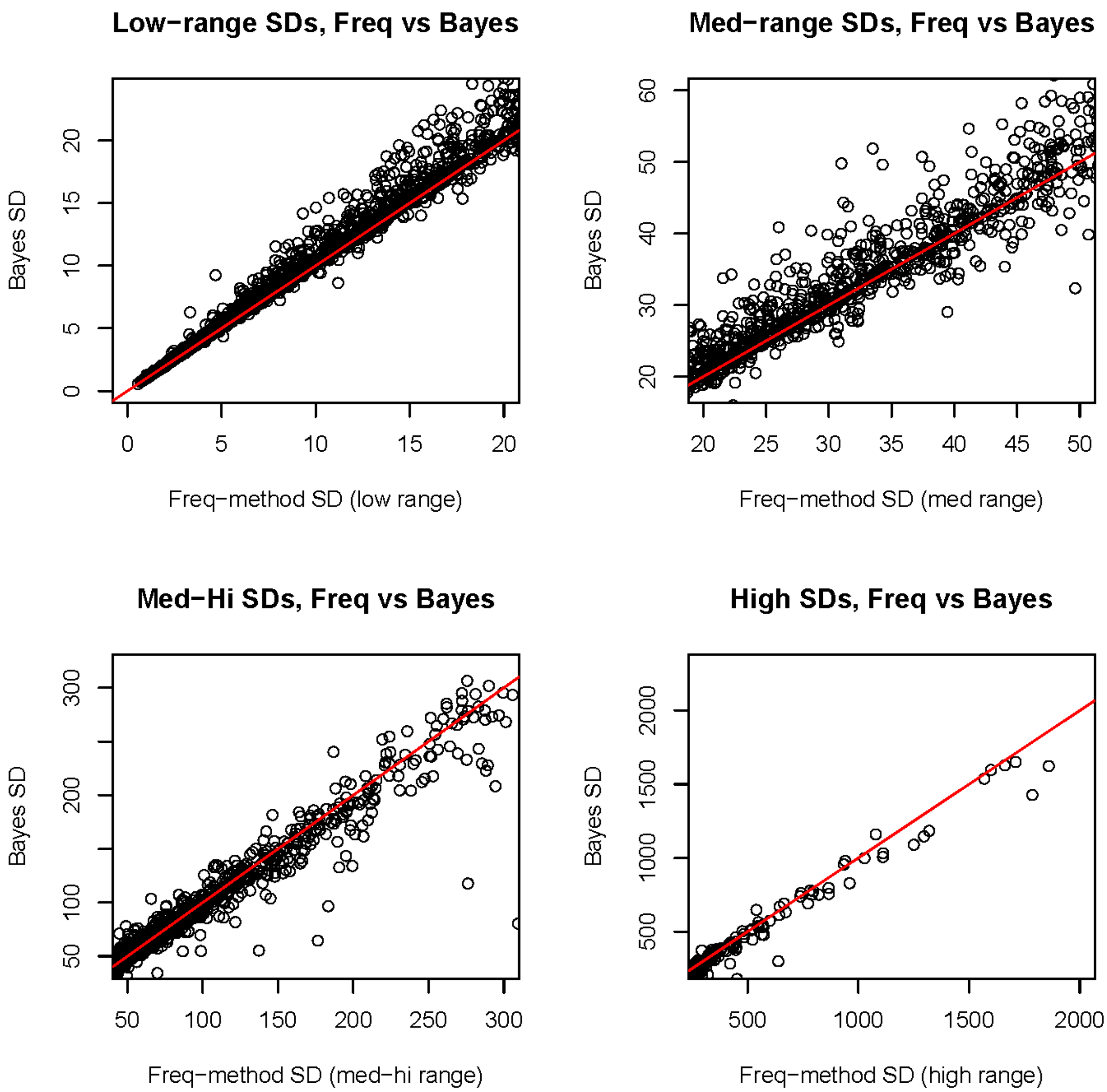
5 Summary

Since 2011, the first year in which Voting Rights Act Section 203(b) determinations were made using a model-based statistical analysis, the Census Bureau has pursued methodological development and innovation in these analyses. Somewhat different models were used in each of the 2011, 2016 and 2021 cycles, although the basic approach and guiding philosophy of Small Area Estimation [Rao and Molina, 2015] has been maintained. That philosophy is to improve the quality of estimates for geographic units in each LMG by 'borrowing strength' from the similarity of relationships between covariates in the other geographic units and the corresponding population-subgroup (CIT, LEP, ILL) fractions in those units.

The present report has extensively described the statistical models and computational methods used in producing the estimates of LMG population subgroups and the ratios of them used in the Section 203(b) determinations. Evaluative analyses in Section 4.1 demonstrate, as similar analyses in previous cycles also did, that the model-based analyses provide estimates with smaller variances (and in that sense, better) than the 'direct' estimates that could have been generated by standard ACS survey weighting. Different evaluative analyses in Section 3.6 described assessments showing that the estimates are nevertheless reliable in conforming to ACS direct estimates of LEP and ILL populations within LMGs over aggregates of jurisdictions. Nevertheless, the accuracy of the model predictions is so far limited by the noisy relationship between jurisdiction-wide covariates and the sizes of LMG-specific population subgroups.

The following two subsections detail the differences in methodology between the 2016 VRA statistical analyses and the present ones, and suggest some directions in which future development of VRA Section 203(b) statistical methods may be fruitfully pursued.

Figure 6: Plots of estimated SDs by Frequentist SDR versus Bayesian MCMC for jurisdiction LEP counts in LMG4 for four ranges of successively larger SD estimates. Red lines are 45°.



5.1 Differences from 2016 Methodology

The analytical methods used in the 2021 VRA estimates are different from those used in previous cycles in choices of model, of predictive covariates, of inclusion criteria for small-sample geographies to contribute to model parameter estimates, of method of computation of model predictions and estimates of variances (MSPEs), and of model assessment. The models used in this cycle are Multinomial Logit-Normal (MLN), chosen in preference to the Dirichlet-Multinomial (DM) models used in 2016 and the related DM models (See Sections 3, 3.1 and 3.2 for model definitions and descriptions and section 3.4 for the selection of MLN for data production in this estimation cycle.) The definitions and choice of potential predictive covariates were discussed in Section 2.2: these differed from the ones considered in 2016 mainly by excluding survey-based covariates defined from the same geography/LMG domains in which estimates were desired, on the grounds that such covariates are themselves noisy and erratic outcomes of survey data-collection. One initial objective of new research in this cycle was to develop new covariates based on historical (previous ACS) data, but there was not time to develop new methodology for the large-scale analysis of random-effect models of this sort. We had also hoped to make use of decennial-census-related administrative records as covariates, but COVID-related delays in decennial 2020 data publication schedules made that impossible. Development of methods based on such improved predictive data must await future research, some directions of which are sketched in Section 5.2 below.

Other details of methodological choices differing from those of 2016 have been described throughout this report. Section 3.3.1 explained and justified the choice to use all possible data in parameter estimation, regardless of the smallness of sample-size in the geographic unit where it was collected. This choice differed from 2016, where minimum sample-size thresholds were used to include data in estimation. Bayesian estimation methods, including MCMC-based estimates of Variance (MSPE) were introduced in this research for two reasons. First, we had hoped to analyze data using more general models flexible enough to incorporate past ACS data, and for some such models the computation of frequentist estimates would have been intractable. Second, we found (*cf.* Section 3.5) that in the 'Full' MLN models suitable for the largest LMGs (MLN-F), with dependent random effects, the computation of variances would have been too computationally burdensome by the frequentist methods found for simpler MLN models, and the Bayesian computations were more practical. We view this combination of techniques as a strength of the current research effort. The frequentist parametric-bootstrap method used in 2016 to estimate variances would not have been computationally feasible this time around because of the large number of repeated MLN-F estimates that it would have required.

The methods we developed in this research project to assess models, covariate choices (in Section 3.6.1) and model-assumptions (in Section 3.7.1) were innovative and different from assessments used in previous cycles of VRA research, and contributed to the quality of our final data product.

5.2 Future Research Directions

The research done and documented in this report has left incomplete several steps that could be developed further to produce more flexible models and efficient computer code using existing data. One such step is to expand the study begun in Section 3.7.1 and possibly generalize the independence assumption [(iii) in Section 3] used in variance estimation. Another kind of extension would be to further explore the recording and interaction of covariates incorporating VOT sample-size classes, since our model assessments suggest that the geographic units with smallest sample sizes are not well fitted by models developed for domains with large sample. If the production of estimates were again to be based on MLN-F models, there is room for numerical-analysis improvements in the optimization of multiparameter mixed-effect likelihoods that might allow frequentist Adaptive Gaussian Quadrature computations with MLN-F to speed up to the point of supplanting the Bayesian MCMC methods used in the 2021 cycle for larger LMGs. Finally, even within the current research cycle we contemplated replacing direct SDR variances in small domains with Generalized Variance Functions (GVFs) [Wolter, 2007], curve-fitted approximations to variances as a function of domain sample sizes within distinct geography types. Work in this direction was begun by Xiaoyun Lu as part of the 2021 research but did not advance far enough to play a role in the final product. It would make sense to carry these GVF investigations forward in future VRA cycles.

Other models and directions of research are also promising for the next VRA research cycles to support Section 203(b) determinations in 2026 and 2031. Primary among them is the development of predictive covariates from administrative records to be used together with decennial census data to supplement ACS data. The difficult issue is that Language Minority Groups are based on race and national origin not directly included in most administrative records (with the exception of some AIAN classifications), so that administrative records would have to be linked to the ACS and/or decennial census to enable ACS covariates and CIT, LEP and ILL outcomes to be used together in local (Geo, LMG) domains. If the linkage problems could be solved, the resulting unit or local-area covariates would likely become much more powerful as predictors of CIT, LEP, and ILL status than they are now. Linkage of administrative records to ACS does not so far run afoul of newly instituted Differential Privacy procedures that inject noise into granular local-area census data, but linkage to decennial census microdata does, so utilization of census data in future VRA cycles will require new ideas.

Past ACS data could be another new source of predictive data. Recent past nonoverlapping ACS releases would provide additional predictive covariates in most if not all jurisdictions and AIAs and could enable time-series extensions to the Multinomial Logit-Normal models used in 2021 production. Estimation within such models incorporating random domain-effects would present formidable computational challenges, but on a smaller scale such Bayesian-hierarchical time series models and MCMC are becoming feasible [López-Vizcaíno et al., 2015]. Slightly less ambitiously, future research might develop models similar to the MLN models used here but incorporating past ACS data (possibly recoded together with current data on the same domains) as covariates.

References

Agresti, Alan. *Categorical Data Analysis*. John Wiley, 3rd edition, 2013.

Aitchison, J. and Shen, S. Logistic normal distributions: some properties and uses. *Biometrika*, 67:261–272, 1980.

Aitkin, M., Liu, C., and Chadwick, T. Bayesian model comparison and model averaging in small-area estimation. *Annals of Applied Statistics*, 3:199–221, 2009.

Ash, Stephen. Using successive difference replication for estimating variances. *Survey Methodology*, 40:147–160, 2014.

Bickel, Peter and Kleijn, B. The semiparametric Bernstein-von Mises theorem. *Annals of Statistics*, 40:206–237, 2012.

Breslow, N.E. and Clayton, D.G. Approximate inference in generalized linear mixed models. *Journal of the American Statistical Association*, pages 9–25, 1993.

Brown, L., Cai, T., and DasGupta, A. Interval estimation for a binomial proportion. *Statistical Science*, 16:101–117, 2001.

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

Census RVRDO. Section 203 language determinations, redistricting and voting rights data office. Technical report, U.S. Census Bureau, 2021. Available online at `https://www.census.gov/programs-surveys/decennial-census/about/voting-rights/voting-rights-determination-file.html`.

Datta, G., Ghosh, M., Steorts, R., and Maples, J. Bayesian benchmarking with applications to small-area estimatio. *TEST*, 20:574–588, 2011.

Efron, B. and Morris, C. Data analysis using Stein's estimator and its generalizations. *Journal of the American Statistical Association*, 70:311–319, 1975.

Esteban, Maria, Lombardía, Maria, López-Vizcaíno, Esther, Morales, Domingo, and Pérez, Agustín. Small-area estimation of proportions under area-level compositional mixed models. *TEST*, 29:793–818, 2020.

Fay, Robert E. and Dippo, C.S. Theory and application of replicate weighting for variance calculations. In *Proceedings of the Survey Research Methods Section of the American Statistical Association*, pages 212–217, 1989.

Fay, Robert E. and Train, George. Aspects of survey and model-based postcensal estimation of income and poverty characteristics for states and counties. In *Proceedings of the Section on Government Statistics, American Statistical Association, Alexandria, VA*, pages 154–159, 1995.

Franco, Carolina and Bell, William. Applying bivariate binomial/logit normal models to small area estimation. In *Proceedings of the Section on Survey Research Methods, American Statistical Association, Alexandria, VA*, pages 690–702, 2013.

Franco, Carolina and Bell, William. Using American Community Survey data to improve estimates from smaller US surveys through bivariate small area estimation models. *Journal of Survey Statistics and Methodology*, 10:225–247, 2022.

Franco, Carolina, Little, Roderick, Louis, Thomas, and Slud, Eric. Comparative study of confidence intervals for proportions in complex sample surveys. *Journal of Survey Statistics and Methodology*, 7:334–364, 2019.

Ghosh, M., Natarajan, K., Stroud, T., and Carlin, B. Generalized linear models for small-area estimation. *Journal of the American Statistical Association*, 93:273–282, 1998.

Jiang, Jiming and Lahiri, Partha. Mixed model prediction and small area estimation, with discussion. *TEST*, 15:1–96, 2006.

Joyce, Patrick M., Malec, Donald, Little, Roderick J.A., and Gilary, Aaron. 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*, #2012-02, 2012. Available online at https://www.census.gov/srd/papers/pdf/rrs2012-02.pdf.

Joyce, Patrick M., Malec, Donald, Little, Roderick J.A., Gilary, Aaron, Navarro, Alfredo, and Asiala, Mark E. Statistical modeling methodology for the Voting Rights Act Section 203 language assistance determinations. *Journal of the American Statistical Association*, 109:36–47, 2014.

Kish, Leslie. Weighting in Deft^2. *The Survey Statistician*, 1987.

Koster, J. and McElreath, R. The multinomial analysis of behavior. *Behavioral Ecology & Sociobiology*, 71:138, 2017.

Krewski, D. and Rao, J.N.K. Inference from stratified samples: Properties of the linearization, jackknife and balanced repeated replication methods. *Annals of Statistics*, 9:1010–1019, 1981.

López-Vizcaíno, E., Lombard´a, M., and Morales, D. Small area estimation of labour force indicators under a multinomial model with correlated time and area effects. *Journal of the Royal Statistical Society, Series A*, 178:535–565, 2015.

Malec, Don, Sedransk, Joseph, Moriarty, Christopher, and LeClere, Felicia. Small area inference for binoary variables in the National Health Interview Survey. *Journal of the ASmerican Statistical Association*, 92:815–827, 1997.

Malec, Donald. Small area estimation from the American Community Survey using a hierarchical logistic model of persons and housing units. *Journal of Official Statistics*, 21:411–432, 2005.

McAllister, M. and Iannelli, J. Bayesian stock assessment using catch-at-age data and the sampling-importance resam[pling algorithm. *Canadian Journal of Fisheries and Aquatic Science*, 54, 1997.

McElreath, R. *Statistician Rethinking: A Bayesian Course with Examples in R and Stan*. Chapman & Hall, 2020.

Molina, I., Saei, A., and Lombard´a, M. Small area estimates of labour force participation under a multinomial logit mixed model. *Journal of the Royal Statistical Society, Series A*, 170:975–1000, 2007.

Pfeffermann, D. and Tiller, R. Small area estimation with state-space models subject to benchmark constraints. *Journal of the American Statistical Association*, 101:1387–1397, 2006.

Pinheiro, José and Bates, Douglas. Approximations of the loglikelihood function in the nonlinear mixed effects model. *Journal of Computational & Graphical Statistics*, 4:12–35, 1995.

Pinheiro, José and Chao, Douglas. Efficient Laplacian and adaptive Gaussian quadrature algorithms for multilevel generalized linear mixed models. *Journal of Computational & Graphical Statistics*, 15:58–81, 2006.

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

Slud, E. Models for simulation and comparison of SAIPE analyses. Technical report, US Census Bureau, 2000a. SAIPE Working Paper, Available online at https://www.census.gov/library/working-papers/2000/demo/slud-01.html.

Slud, Eric V. Accurate calculation and maximization of loglikelihood for mixed logistic regression. Technical report, U.S. Census Bureau, 2000b. SAIPE Working Paper, Available online at https://www.census.gov/library/working-papers/2000/demo/slud-02.html.

Slud, Eric V. Small area estimation errors in SAIPE using GLM versus FH models. In *Proceedings of the Survey Research Methods Section of the American Statistical Association, Alexandria, VA*, pages 4402–4409, 2004.

Slud, Eric V. Assessment of zeroes in survey-estimated tables via small-area confidence bounds. *Journal of the Indian Society for Agricultural Statistics*, 66:157–169, 2012.

Slud, Eric V. and Ashmead, Robert. 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.

Slud, Eric V., Ashmead, Robert, Joyce, Patrick, and Wright, Tommy. Statistical modeling methodology (2016) for Voting Rights Act, Section 203 determinations. *U.S. Census Bureau. Center for Statistical Research & Methodology Research Report Series*, #2018-12, 2018. Available online at https://www.census.gov/srd/papers/pdf/rrs2012-02.pdf.

STAN Development Team. STAN user's guide version 2.24. Technical report, 2019. Available at https://mc-stan.org/docs/2_24/stan-users-guide-2_24.pdf.

Steorts, Rebecca and Ghosh, Malay. Estimation of mean squared errors of benchmarked empirical bayes estimators. *Statistica Sinica*, 23:749–767, 2013.

Stroup, Walter. *Generalized Linear Mixed Models*. CRC Press, 2012.

van der Vaart, Aad. *Asymptotic Statistics*. Cambridge University Press, 1998, Ch. 10.

Wolter, Kirk. *Variance Estimation*. Springer, 2nd edition, 2007.

Ybarra, Lynn M.R. and Lohr, Sharon L. 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

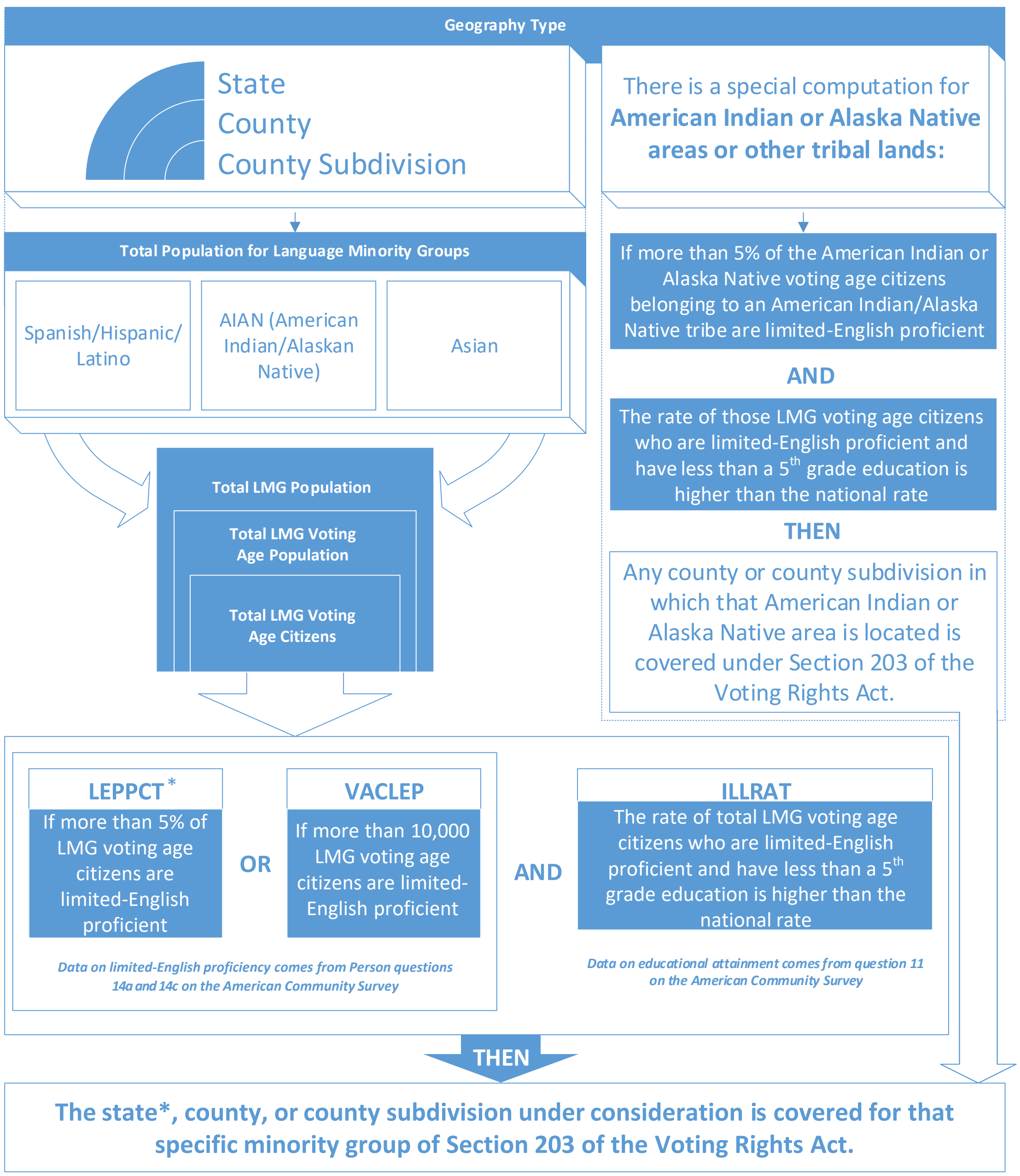
For purposes of this section, the term "language minorities" or "language minority group" means persons who are American Indian, Asian American, Alaska Native, or of Spanish heritage.

Table 20: 2021 Language Minority Groups. LMG 1–21 = AS 1–21, LMG 22-52 = AI 1–51.

HSP. Hispanic	AI 16. Delaware
AS 1. Asian Indian	AI 17. Hopi
AS 2. Bangladeshi	AI 18. Houma
AS 3. Cambodian	AI 19. Iroquois
AS 4, Chinese	AI 20. Kiowa
AS 5. Filipino	AI 21. Lumbee
AS 6. Hmong	AI 22. Menominee
AS 7. Indonesian	AI 23. Mexican American Indian
AS 8. Japanese	AI 24. Navajo
AS 9. Korean	AI 25. Osage
AS 10. Laotian	AI 26. Ottawa
AS 11. Malaysian	AI 27. Paiute
AS 12. Pakistani	AI 28. Pima
AS 13. Sri Lankan	AI 29. Potawatomi
AS 14. Thai	AI 30. Pueblo
AS 15. Vietnamese	AI 31. Puget Sound Salish
AS 16. Other Asian	AI 32. Seminole
AS 17. Bhutanese	AI 33. Shoshone
AS 18. Burmese	AI 34. Sioux
AS 19. Mongolian	AI 35. South American Indian
AS 20. Nepalese	AI 36. Spanish American Indian
AS 21. Okinawan	AI 37. Tohono O'Odham
AI 1. Apache	AI 38. Ute
AI 2. Arapaho	AI 39. Yakama
AI 3. Blackfeet	AI 40. Yaqui
AI 4. Canadian & French Indian	AI 41. Yuman
AI 5. Central American Indian	AI 42. All other AI tribes
AI 6. Cherokee	AI 43. AI tribes, not specified
AI 7. Cheyenne	AI 44. Alaska Athabascan
AI8. Chickasaw	AI 45. Aleut
AI 9. Chippewa	AI 46. Inupiat
AI 10. Choctaw	AI 47. Tlingit-Haida
AI 11. Colville	AI 48. Tsimshian
AI 12. Comanche	AI 49. Yup'ik
AI 13. Cree	AI 50. Alaskan Native tribes, not specified
AI 14. Creek	AI 51. AI or AN tribes, not specified
AI 15. Crow	

B 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



*Statewide coverage only occurs under the 5% trigger, it does not use the 10,000 trigger.

C Notations and Model Definitions

The models considered in this report all share the same data-structure within each geography by LMG domain. In this Section, the LMG index is g (fixed, and later suppressed because all models are defined within a single LMG), and the geography index is $j \in \mathcal{J}_g = \mathcal{J}$ for sampled jurisdictions. (Analogous notations apply with index a for AIAN or ANRC geography within LMG.) Let $n_{jg} = n_{jg}^V$ denote the number of respondent (voting-age person, or VOT) records in the (Geo, LMG) domain (j, g). For each respondent $i \in (j, g)$, with survey weight w_i, the data consist of indicators $y_{i,jg}^A$ for the nested decreasing categories $A = V, C, L, I$ respectively denoting Voting-age persons (VOT), Voting-age Citizens (CIT), Limited English Proficiency CIT (LEP), and Illiterate LEP (ILL). Thus, $1 \equiv y_{i,jg}^V \geq y_{i,jg}^C \geq y_{i,jg}^L \geq y_{i,jg}^I \geq 0$ for all $i \in (j, g)$. Corresponding to each (j, g), there is also a covariate-vector $\underline{X}_{jg}$ with dimension d_g and coordinates that may depend on g. All models defined below treat the covariates $\underline{X}_{jg}$ as nonrandom, or equivalently, condition on them.

The counts of VOT respondents and the direct survey-weighted estimate of total VOT population in domain (j, g) are respectively

$$n_{jg} = \sum_{i \in (j,g)} 1 \quad , \qquad \hat{N}_{jg} = \hat{N}_{jg}^V = \sum_{i \in (j,g)} w_i$$

The unweighted counts and survey-weighted totals in the domain in the VOT, CIT, LEP, ILL categories are defined as

$$n_j^A \equiv \sum_{i \in (j,g)} y_{i,jg}^A \quad , \quad \hat{N}_j^A \equiv \sum_{i \in (j,g)} w_i \, y_{i,jg}^A \quad A = V, C, L, I \tag{5}$$

where the g index is everywhere suppressed from now on because models are defined and fitted separately for different LMGs g.

In what follows, the VOT respondent-count $n_j = n_j^V$ and population estimates $N_j = N_j^V$ are regarded as fixed and not modeled, and the category population estimates N_j^A are treated as data subject to modeling assumptions. Because of the discrete nature of the respondent counts, we model the imputed subsets of the sample n_j^V falling in the CIT, LEP, and ILL categories, accounting for the survey weights by defining pro-rated respondent counts $Y_{j,k}, \ k = 0, 1, 2, 3$:

$$\underline{Y}_j = (Y_{j,0}, Y_{j,1}, Y_{j,2}, Y_{j,3}) = \frac{n_j}{\hat{N}_j^V} \cdot (\hat{N}_j^V, \hat{N}_j^C, \hat{N}_j^L, \hat{N}_j^I) \tag{6}$$

Throughout this report, our models assume about $\underline{Y}_j$ that

$$\text{the counts} \quad \underline{Y}_j \quad \text{are independent of} \quad n_j^V, \ \hat{N}_j^V \tag{7}$$

Equivalently, all models concerning scaled outcome-counts $\underline{Y}_j$ (or the disjoint counts $\underline{W}_j$ defined next) are conditioned on fixed values of n_j, $\hat{N}_j^V$ for all domains $j \in \mathcal{J}$.

A second way of representing the nested decreasing category totals within n_j is to make the categories disjoint, i.e., to define quadruples $\underline{W}_j = (W_{j,k},\ k = 1, \ldots, 4)$ of nonnegative category counts $W_{j,k} \equiv Y_{j,k-1} - Y_{j,k} \geq 0$ for the four disjoint categories of VOT non-CIT, CIT non-LEP, LEP non-ILL, and ILL persons, within the (j, g) domain. Figure 1 in the main text exhibits these four categories pictorially. Although the data vectors $\underline{Y}_j$ and $\underline{W}_j$ do not have integer entries, we formulate models as though they do, as was done also in Joyce et al. [2014] and Slud et al. [2018].

Our models all have a mixed-effect, generalized-linear multinomial form. This means that there is a random vector $\underline{\pi}_j = (\pi_{j,1}, \ldots, \pi_{j,4})$ of 4 category probabilities, depending on covariates $\mathbf{X}_j$, on three vectors of d-dimensional regression coefficients $\beta^{(k)}$, and on further random-effect variables that are independent and identically distributed across geographic indices $j \in \mathcal{J}$. The distributional form of the random effects will be different in the two classes of models that we consder, but the category probabilities depend on $\mathbf{X}_j$ and on coefficients $\beta^{(k)}$ only through the quantities

$$\eta_{j,k} \ \equiv \ \mathbf{X}_j' \, \beta^{(k)} \quad , \qquad k = 1, 2, 3 \tag{8}$$

The nonrandom part of the model expresses $\pi_{j,k}$ in similar but not identical ways through the logistic function $h(w) \equiv \texttt{plogis}(w) = e^w/(1 + e^w)$ and $\eta_{j,k}$. Conditionally given $\underline{\pi}_j$, the pro-rated respondent counts are modeled as

$$\underline{W}_j \ \equiv \ (n_j - Y_{j,1},\, Y_{j,1} - Y_{j,2},\, Y_{j,2} - Y_{j,3},\, Y_{j,3}) \ \sim \ \text{Multinom}\big(n_j,\, \underline{\pi}_j\big) \tag{9}$$

As a matter of modeling strategy, the vectors $\underline{\pi}_j$ of random category-probabilities are regarded as a feature of the overall population in domain (j, g) regardless of the size of the n^V_{jg} sample. The randomness in $\underline{\pi}_j$ is our model for the differences between domains (j, g) for different j, but the model (9) reflects our idea that the total population counts in the VOT, CIT, LEP and ILL categories respectively satisfy

$$N^C_j \ \approx \ N^V_j \cdot (1 - \pi_{j,1}) \ , \quad N^L_j \ \approx \ N^V_j \cdot (\pi_{j,3} + \pi_{j,4}) \ , \quad N^I_j \ \approx \ N^V_j \cdot \pi_{j,4} \tag{10}$$

The true-population category totals N^A_j are denoted without the hats that appear in their approximately unbiased survey-weighted estimates $\hat{N}^A_j$. The quantities to be predicted within models (9) are the totals N^A_j, $A = C, L, I$, in (10). The ratios N^L_j/N^C_j and N^I_j/N^L_j respectively define the within-domain proportion of voting-age citizens who are LEP and of LEP persons who are ILL. Within our models in which N^A_j counts are random, we augment the independence assumption (7) to say

$$\text{counts and random probabilities} \quad (\underline{Y}_j,\, \underline{\pi}_j) \quad \text{are independent of} \quad n^V_j, \ \hat{N}^V_j \tag{7$'$}$$

In order to explain different forms of the multinomial models, we make use of a standard property of multinomial random variables

$$\underline{M} = (M_1, M_2, M_3, M_4) \ \sim \ \text{Multinom}\big(n,\, (p_1, p_2, p_3, p_4)\big)$$

The property, which is actually equivalent to the multinomial distribution, is that

$$M_1 \sim \text{Binom}(n,\, p_1), \qquad M_2 \sim \text{Binom}(n - M_1,\, \frac{p_2}{1-p_1}) \quad \text{given} \;\; M_1$$

$$M_3 \sim \text{Binom}(n - M_1 - M_2,\, \frac{p_3}{1-p_1-p_2}) \quad \text{given} \;\; M_1, M_2 \tag{11}$$

The two classes of models we consider in the Voting Rights Act analyses are Multinomial Logit-Normal and Dirichlet-Multinomial, respectively described in detail in Sections C.1 and C.2 below. In both, the four category probabilities $(\pi_{j,k},\ k = 1,2,3,4)$ for each $j \in \mathcal{J}$ are randomized versions of the probability vector

$$p(v_1, v_2, v_3) \;=\; (p_1, p_2, p_3, p_4) \;\equiv\; \big(1 - v_1,\; v_1\,(1 - v_2),\; v_1\, v_2\,(1 - v_3),\; v_1\, v_2\, v_3\big) \tag{12}$$

where

$$v_k \;=\; v_{j,k} \;\equiv\; h(\eta_{j,k}) \;=\; \frac{\exp(X_j'\,\beta^{(k)})}{1 + \exp(X_j'\,\beta^{(k)})} \quad , \qquad k = 1,2,3 \tag{13}$$

These quantities $v_k \;=\; \sum_{b=k+1}^{4} p_b$ for $k = 1,2,3$ are respectively interpreted as approximate ratios CIT/VOT, LEP/CIT, and ILL/LEP for the (j,g) domain in LMG g, before introducing random effects into each type of model.

In the Multinomial Logit-Normal class of models, the random effects for geographic index $j \in \mathcal{J}$, denoted $\underline{u}_j \;=\; (u_{j,k},\ k = 1,2,3)$ are assumed jointly normal with unknown variance parameters, and the randomized vector $\underline{\pi}_j$ of disjoint-category probabilities is given in terms of the function p defined in (12) by

$$\textbf{MLN:} \qquad \underline{\pi}_j \;\equiv\; p\big(h(\eta_{j,1} + u_{j,1}),\, h(\eta_{j,2} + u_{j,2}),\, h(\eta_{j,3} + u_{j,3})\big) \tag{14}$$

In the Dirichlet-Multinomial class of models, the randomized vector $\underline{\pi}_j$ is given by:

$$\textbf{DM:} \qquad 1 - \pi_{j,1} \sim \text{Beta}\big(\tau_1 v_1, \tau_1(1 - v_1))\big) \quad , \qquad 1 - \frac{\pi_{j,2}}{1 - \pi_{j,1}} \sim \text{Beta}\big(\tau_2 v_1 v_2, \tau_2 v_1(1 - v_2)\big) \;,$$

$$1 - \frac{\pi_{j,3}}{1 - \pi_{j,1} - \pi_{j,2}} \;\sim\; \text{Beta}\big(\tau_3\, v_1 v_2 v_3, \tau_3\, v_1 v_2(1 - v_3)\big) \quad \text{are independent, for} \quad k = 1,2,3 \tag{15}$$

for unknown (*dispersion*) parameters (τ_1, τ_2, τ_3), where as in (13), $v_k \equiv h(\eta_{j,k})$.

In both types of model, the probabilities

$$\alpha_{j,k} \;=\; \sum_{b=k+1}^{4} \pi_{j,b} \Big/ \sum_{b=k}^{4} \pi_{j,b} \quad , \qquad j \in \mathcal{J},\;\; k = 1,2,3 \tag{16}$$

play a special role as the logistic rates $v_{j,k} = h(\eta_{j,k})$ with a random effect. In the MLN models, $\alpha_{j,k} \;\equiv\; h(\eta_{j,k} + u_{j,k})$, while in the DM model the random probabilities $\alpha_{j,k}$ are precisely the quantities assumed independent and beta-distributed in (15) and have expectations $v_{j,k}$.

C.1 Multinomial Logit-Normal Models

The Multinomial Logit-Normal (MLN) Model is defined as in (9) and (14) with random effects $\underline{u}_j$ independent identically distributed across j satisfying

$$\begin{pmatrix} u_{j,1} \\ u_{j,2} \\ u_{j,3} \end{pmatrix} = \begin{pmatrix} \sigma_1 & 0 & 0 \\ a_1 & \sigma_2 & 0 \\ a_2 & a_3 & \sigma_3 \end{pmatrix} \begin{pmatrix} z_{j,1} \\ z_{j,2} \\ z_{j,3} \end{pmatrix} \equiv A\, \underline{z}_j \;, \qquad z_{j,k} \overset{iid}{\sim} \mathcal{N}(0,1) \tag{17}$$

so that

$$\sigma_k^2 = \text{Var}(u_{j,k}), \qquad k \;=\; 1,\, 2,\, 3$$

and

$$E(u_{j,2}\,|\,u_{j,1}) \;=\; \frac{a_1}{\sigma_1}\, u_{j,1}\;, \qquad E(u_{j,3}\,|\,u_{j,1},\, u_{j,2}) \;=\; (a_2 \;-\; \frac{a_1 a_3}{\sigma_2})\, u_{j,1} \;+\; \frac{a_3}{\sigma_2}\, u_{j,2} \tag{18}$$

By Choleski decomposition, every covariance matrix Σ for $\underline{u}_j$ can be written in the form $V = A\,A'$ with lower-triangular A. The parameterization (17) is useful in later sections in simplifying the form of the conditional distribution of $Y_{j,k}$ given $(Y_{j,t},\; t < k)$.

The MLN model given in (9), (14) and (17) has an alternate, equivalent expression as a cascaded random-intercept logistic regression model

$$Y_{j,k} \;\sim\; \text{Binom}\big(Y_{j,k-1}\;, h(\eta_{j,k} + u_{j,k})\big) \quad \text{given} \qquad (Y_{j,t-1},\, \eta_{j,t},\, u_{j,t},\;\; t = 1, \ldots, k) \tag{19}$$

Equivalence between (19) and the MLN model (9) is an immediate consequence of the multinomial re-expression (11).

The dimension of the unknown parameter $\vartheta = (\beta^{(1)}, \beta^{(2)}, \beta^{(3)}, a_1, a_2, a_3, \sigma_1, \sigma_2, \sigma_3)$ in the 'full' MLN model (denoted **MLN-F**) is $3d + 6$. Parameter reductions can be achieved structurally by specifying in advance which index subsets $I_1, I_2, I_3 \subset \{1, \ldots, d\}$ will contain the non-zero coefficients of $\beta^{(1)}$, $\beta^{(2)}$, $\beta^{(3)}$. The parameter-dimension for the regression coefficients is then $|I_1| + |I_2| + |I_3|$. There are 6 variance parameters, reducing to 3 if we restrict to independent random effects by assuming a_1, a_2, a_3 equal to 0. When we make this reduction of the MLN model to have diagonal random-effect covariance matrix, the model is denoted **MLN-D**. In MLN-D, the three random effects $(u_{j,k},\;\, k = 1, 2, 3)$ are independent for each $j \in \mathcal{J}$. In that case, each of the three random-intercept logistic regression models in (19) depends on a separate and independent random effect $u_{j,k}$, and the one indexed by k depends only on parameters $\beta^{(k)}, \sigma_k$. For that reason, the whole model likelihood factorizes into separate likelihoods for separate sets of parameters $(\beta^{(k)}, \sigma_k)$, and they can be maximized separately, resulting in considerable computational simplification.

C.1.1 Limiting Cases of Infinite Regression Coefficients

Under model MLN-D with $Y_{j,k}$ very close to 0 or to $Y_{j,k-1}$, the estimated value of σ_k can become very large along with some or all of the estimated $\beta^{(k)}$ components. To understand this phenomenon in modeling terms, consider the case where there are no covariates (with nonzero regression coefficients), so that each of the models (19) in a single domain has the form

$$Y \;\sim\; \text{Binom}\big(m,\, h(\eta \,+\, \sigma\, z)\big) \quad , \qquad z \sim \mathcal{N}(0,1) \tag{20}$$

The log-likelihood contribution for Y then has the form $\log g(m, Y, \eta, \sigma)$, apart from an additive constant not depending on (η, σ), where

$$g(m, r, \eta, \sigma) \;\equiv\; \int h(\eta + \sigma z)^r \, (1 - h(\eta + \sigma z))^{m-r} \, \phi(z)\, dz \tag{21}$$

and $\phi(z)$ is the standard normal density function $e^{-z^2/2}/\sqrt{2\pi}$. An easy calculation shows that

$$\frac{\partial}{\partial \eta}\, g(m, r, \eta, \sigma) \;=\; r\, g(m+1, r, \eta, \sigma) \;-\; (m - r)\, g(m+1, r+1, \eta, \sigma) \tag{22}$$

and a similar calculation followed by an integration-by-parts shows

$$\frac{\partial}{\partial \sigma}\, g(m, r, \eta, \sigma) \;=\; \int h(\eta{+}\sigma z)^r \, (1{-}h(\eta{+}\sigma z))^{m-r} \big[r\,(1{-}h(\eta{+}\sigma z)) - (m{-}r)\, h(\eta{+}\sigma z)\big]\, z\, \phi(z)\, dz \tag{23}$$

$$= \; \sigma\, \frac{\partial}{\partial \eta} \big[r\, g(m+1, r, \eta, \sigma) \;-\; (m-r)\, g(m+1, r+1, \eta, \sigma)\big] \;=\; \sigma\, \frac{\partial^2}{\partial \eta^2}\, g(m, r, \eta, \sigma)$$

$$= \sigma\big[r^2\, g(m{+}2, r, \eta, \sigma) - (2r(m{-}r){+}m)\, g(m{+}2, r{+}1, \eta, \sigma) + (m{-}r)^2\, g(m{+}2, r{+}2, \eta, \sigma)\big] \tag{24}$$

Since we are interested in this subsection in understanding the loglikelihood when $r = m$ or $r = 0$, we note that (22) is strictly positive for $r = m > 0$, and

$$g(m, r, \eta, \sigma) \;\equiv\; g(m, m-r, -\eta, \sigma) \quad , \qquad \text{for all} \quad 0 \le r < m$$

When $Y_{j,k} = Y_{j,k-1}$, we see that the conditional MLN-D loglikelihood contribution (without any covariates) from $Y_{j,k}$ given $Y_{j,k-1}$ is strictly increasing in $\eta = \beta_1^{(k)}$ and for large values of η is strictly decreasing in σ_k. This last assertion follows from the expression (24) rewritten for $r = m$ as

$$\frac{\partial}{\partial \sigma}\, g(m, m, \eta, \sigma) \;=\; \sigma m \int h(\eta + \sigma z)^m (1 - h(\eta + \sigma z))\, \big[m - (m+1) h(\eta + \sigma z)\big]\, \phi(z)\, dz$$

and as $\eta \to \infty$, by the dominated convergence theorem

$$(1 - h(\eta))^{-1}\, \frac{\partial}{\partial \sigma}\, g(m, m, \eta, \sigma) \;\longrightarrow\; -\sigma\, m \quad \text{as} \quad \eta \to \infty \tag{25}$$

Suppose that for a fixed k, for all $j \in \mathcal{J}$, $Y_{j,k} = Y_{j,k-1}$. Consider the MLN-D loglikelihood with no covariates, for that same fixed k for joint data (19) over all $j \in \mathcal{J}$ such that $n_j > 0$. Starting from any initial parameter values ($\beta_1^{(k)}$, σ_k, $k = 1, 2, 3$), by (22) the log-likelihood is strictly increased by making all three $\beta_1^{(k)}$ values large and positive; after that, by (25) and (22) increasing $\beta_1^{(k)}$ and decreasing $\sigma_k > 0$ further increases the loglikelhood. Thus, when $Y_{j,k} = Y_{j,k-1}$ for all $j \in \mathcal{J}$, the MLEs for $\beta_1^{(k)}$ and σ_k diverge to ∞ and 0 respectively in this case. Similarly, if $Y_{j,k} = 0$ for all $j \in \mathcal{J}$ for some $k \geq 1$, the corresponding MLN-D MLEs for $\beta_1^{(k)}$ and σ_k diverge to $-\infty$ and 0 respectively. These effects were seen numerically wherever $Y_{j,k} = Y_{j,k-1}$ for all $j \in \mathcal{J}$ (for a fixed $k \geq 1$) or $Y_{j,k} = 0$ for all $j \in \mathcal{J}$ for some fixed $k \geq 1$. However, the form of (25) does imply, and we confirmed numerically, that for very large absolute values of $\beta_1^{(k)}$ the log-likelihood approaches 0 and the further effect on the log-likelihood of changes in σ_k are minuscule.

The MLE divergence found in this subsection occurred numerically also in some LMGs and k-values where $\sum_j Y_{j,k-1}/\sum_j Y_{j,k}$ ratios were very close but not identical to 1 or 0. In LMGs and k for which the ratios are so extreme (*cf.* Table 2), these observations motivate the use of a simplified beta-binomial model to replace single stages k within (19). Those simplified models given in Section C.2.1 below.

C.2 Dirichlet-Multinomial Models

The *Beta-Binomial* model [Carlin and Louis, 2009, p. 55] for a random count Y has the form

$$Y \sim \text{Binom}(n, \omega) \quad \text{given} \quad \omega \ , \qquad \omega \sim \text{Beta}\big(\tau\, a,\, \tau\,(1-a)\big)$$

where $0 < a < 1$ and $\tau > 0$ are unknown scalar parameters. The Beta random variable ω, with mean a and variance $a(1-a)/(\tau+1)$, serves as a random success-probability for n Bernoulli trials. The so-called *dispersion* parameter τ controls the variability of ω.

The DM Model (9) and (15) is a random-effect multinomial regression model, which after re-expression (11) of the multinomial, becomes a series of three cascaded beta-binomial models

$$Y_{j,k} \sim \text{Binom}(Y_{j,k-1},\, \alpha_{j,k}) \qquad \text{given} \quad (Y_{j,b}, \ \ b < k) \tag{26}$$

with independent random conditional category-inclusion probabilities $\alpha_{j,1} = 1 - \pi_{j,1}$ for CIT, $\alpha_{j,2} \equiv 1 - \pi_{j,2}/(1 - \pi_{j,1})$ for LEP given CIT, and $\alpha_{j,3} \equiv 1 - \pi_{j,3}/(1 - \pi_{j,1} - \pi_{j,2})$ for ILL given LEP. (Recall that the $\alpha_{j,k}$ notations were first defined in equation (16).) The data consist of $Y_{j,0} = n_j$, $\{Y_{j,k} : \ k = 1, 2, 3\}_j$ and $\mathbf{X}_j$ for all domains (j, g) in LMG g, and the unknown parameters are $\beta^{(k)} \in \mathbb{R}^d$, $\tau_k > 0$ for $k = 1, 2, 3$. Allowing for the possibility of restricting the nonzero coefficients of $\beta^{(k)}$ to the index-set $\mathcal{I}_k$ for $k = 1, 2, 3$, the parameter dimension is $|\mathcal{I}_1| + |\mathcal{I}_2| + |\mathcal{I}_3| + 3$. This resembles the cascaded beta-binomial model used in VRA analysis by Joyce et al. [2014] in 2011, except that in that analysis, the covariates entered not through regression but through a grouping of data across domains within LMGs.

A variant of the DM model would define dispersion parameters as $\tau_1 = \tau$, $\tau_2 = \tau \cdot (1 - p_1)$, $\tau_3 = \tau \cdot (p_3 + p_4)$, reducing the parameter dimension by 2. That form of the model is Dirichlet Multinomial [Carlin and Louis, 2009, p. 284], because a Dirichlet(τ, (p_1, p_2, p_3, p_4)) distributed random probability vector $(\omega_1, \omega_2, \omega_3, \omega_4)$ has the property that ω_1, $\omega_2/(1-\omega_1)$, $\omega_3/(1-\omega_1-\omega_2)$ are jointly independent Beta$(\tau p_1, \tau(1-p_1))$, Beta$(\tau p_2, \tau(p_3+p_4))$, Beta$(\tau p_3, \tau p_4)$ random variables. Thus, if $\underline{\pi}_j$ were Dirichlet(τ, $p(v_1, v_2, v_3)$) distributed, the cascaded models (15) and (26) with $\tau_1 = \tau$, $\tau_2 = \tau \cdot (1 - p_1)$, $\tau_3 = \tau \cdot (p_3 + p_4)$ would follow. However, the model (26) with general (τ_1, τ_2, τ_3) is not a Dirichlet model for $\underline{\pi}_j$. A different form of Dirichlet-Multinomial with τ replaced in domain j by $\tau + \tau_0 \cdot \sqrt{n_j}$ was the model used by Slud et al. [2018] in the 2016 VRA cycle.

An immediate consequence of the unusual regression parameterization in model (15) is that

$$E(\alpha_{j,k}) \;=\; E\big(\sum_{b=k+1}^{4} \pi_{j,b} / \sum_{b=k}^{4} \pi_{j,b} \big) \;=\; E\big(1 - \; \pi_{j,k} / \sum_{b=k}^{4} \pi_{j,b} \big) \;=\; h(\eta_{j,k}) \qquad \text{for} \quad k = 1, 2, 3$$

and therefore in the DM model with general τ_k,

$$E\big(Y_{j,k} \,|\, Y_{j,k-1}\big) \;=\; Y_{j,k-1} \,\cdot\, h(\eta_{j,k})$$

exactly as in the MLN model. Moreover, like MLN-D, the DM model has the simplifying property that the conditional likelihood of $Y_{k,i}$ given $(Y_{j,t} : t < k)$ depends only on the parameters $\beta^{(k)}$, τ_k. These conditional likelihoods factor into expressions for separate k that depend on disjoint sets of parameters, and the maximum likelihood estimators of $\beta^{(k)}$, τ_k can be found through maximization of Beta-Binomial conditional log-likelihoods (separate for $k = 1, 2, 3$) of the form

$$\sum_{j \in \mathcal{J}} \log \Big(\frac{\Gamma(\tau_k) \cdot \Gamma(\tau_k h(\eta_{j,k}) + Y_{j,k}) \, \cdot \Gamma(\tau_k (1 - h(\eta_{j,k})) + Y_{j,k-1} - Y_{j,k})}{\Gamma(\tau_k \, h(\eta_{j,k})) \, \cdot \, \Gamma(\tau_k \, (1 - h(\eta_{j,k}))) \cdot \Gamma(\tau_{j,k} + Y_{j,k-1})} \Big) \tag{27}$$

The parameterization of the DM model with general τ_k has the same dimension as the corresponding MLN-D model with the same sets $\mathcal{I}_k$ of structurally non-zero regression coefficients in $\beta^{(k)}$. The dispersion parameters $1/(1 + \tau_k)$ play a similar role in DM models as the variances σ_k^2 do in MLN-D, controlling the variance of the unmodeled random effects on outcomes $Y_{j,k}$ given $Y_{j,k-1}$. The Dirichlet-Multinomial model with $\tau_k \equiv \tau$ has two fewer parameters (than MLN-D or DM with general τ_k), and the general-covariance model MLN-F has three more parameters.

C.2.1 Special Cases of DM Submodels with no Covariates

The DM models, like the MLN models, are fitted with more covariates (i.e., larger sets $\mathcal{I}_1$, $\mathcal{I}_2$, $\mathcal{I}_3$ of regression coefficients that are not set to be structurally 0) in LMGs g with a lot of data, i.e., with many domains (j, g) in which there were more than a few respondents. In the most data-sparse

LMGs (mostly AIAN LMGs), either few covariates or none are used. Each of the staged DM models in (26) along with (15) then takes the simplified form of a Beta-binomial model

$$Y_{j,k} \sim \text{Binom}\big(Y_{j,k-1},\, \alpha_{j,k}\big) \quad \text{given} \quad (Y_{j,t-1},\, \alpha_{j,t},\;\; 1 \le t \le k)$$

where the jointly independent random success probabilities $\alpha_{j,k}$ across (j,k) are given by

$$\alpha_{j,k} \sim \text{Beta}\big(\tau_k \cdot \mu_{k-1} \cdot h(\underline{X}_j'\, \beta_1^{(k)}),\, \tau_k \cdot \mu_{k-1} \cdot (1 - h(\underline{X}_j'\, \beta_1^{(k)})\big) \;, \qquad \mu_b \;=\; \prod_{t=1}^{b} h(\underline{X}_j'\, \beta_1^{(t)})$$

In the extreme case where either $\sum_j (Y_{j,k-1} - Y_{j,k}) \;=\; 0$ or $\sum_j Y_{j,k} \;=\; 0$, as in Section C.1.1 the maximum likelihood estimates for $\beta_1^{(k)}$ within the DM model would diverge, repectively to $+\infty$ or to $-\infty$, and τ_k would become indeterminate. To avoid that numerical instability, we change the form of the stage k model. When either $\sum_j (Y_{j,k-1} - Y_{j,k})$ or $\sum_j Y_{j,k}$ is close enough to 0, either within MLN-D or DM models, we allow no covariates and use a Beta-binomial model

$$Y_{j,k} \sim \text{Binom}\big(Y_{j,k-1},\, \gamma_{j,k}\big) \;, \qquad \gamma_{j,k} \sim \text{Beta}(0.5,\, 0.5) \tag{28}$$

This choice of $\text{Beta}(0.5, 0.5)$ prior for the binomial success probability is called the *Jeffreys prior* [Carlin and Louis, 2009, p. 39]. It has well-known properties of providing numerically stable posterior confidence intervals (*credible intervals*) with good frequentist coverage probabilities, even in extreme settings like those we consider with posterior success probabilities γ_k nearly degenerate at 0 or 1 [Brown et al., 2001, Franco et al., 2019]. The fixed choice of prior removes the unknown parameters τ_k, $\beta_1^{(k)}$ from the estimation problem, leaving only the predictor of the random effect γ_k to feed into the prediction of $Y_{j,k}$ from $Y_{j,k-1}$.

D Frequentist vs. Bayes Prediction and Variance Estimation

Within the field of Small Area Estimation, the methods of prediction and variance estimation differ in frequentist Empirical-Bayes analysis and Bayesian analysis. We maintain a terminological distinction in both kinds of analysis: whenever a function of data is used to approximate a fixed unknown statistical parameter, that function or statistic is called an *estimator*; if the quantity being approximated involves an unobserved random variable (which is also how Bayesians view unknown parameters) and may also involve fixed unknown parameters, then the statistic is called a *predictor*. 'Estimate' is a term that is used more generically and loosely, as in reference to statistics published regarding domain population counts and ratios and their margins of error. Throughout this report, all model-based population estimates involve latent random effects and are therefore 'predictors'. Margin-of-error estimates that are conditional on data are therefore also predictors, but statistics approximating unconditional margins of error can be called estimators.

Schematically, the setting has the following ingredients: a set of fixed covariates $\{\underline{X}_j\}_{j\in\mathcal{J}}$ indexed by domains j; data $\{\underline{Y}_j\}_{j\in\mathcal{J}}$ independent across domain-index j; latent (unobservable) domain-specific random effects $\underline{\pi}_j = q(\theta, \underline{X}_j, \underline{z}_j)$ (expressed in terms of *iid* random variables $\underline{z}_j$ whose distributions contain no unknown parameters). The observations $\underline{Y}_j$ conditionally given $\underline{\pi}_j, \underline{z}_j, \theta$ are independent across $j \in \mathcal{J}$ and have distributions depending only on $\underline{\pi}_j$. In the Empirical-Bayes approach [Carlin and Louis, 2009, Rao and Molina, 2015], parameters θ are estimated (usually by Maximum Likelihood or a related method) using the marginal likelihood for $\underline{Y}_j$ in terms of θ, and the objects $\underline{\pi}_j$ of prediction are estimated as $\hat{\underline{\pi}}_j \equiv E\big(\underline{\pi}_j \,\big|\, \underline{Y}_j, \theta\big)$ by substituting an estimator $\hat{\theta}$ for the unknown θ. In both the MLN and DM models, these conditional expectations $E\big(\underline{\pi}_j \,\big|\, \underline{Y}_j, \theta\big)$ have a very specific and tractable form, based on which the integral formula (E.1.1) in the MLN-F model and (42) for the DM model are based:

$$E_\vartheta\big(\underline{\pi}_{j,k} \,|\, \underline{Y}_j = \underline{m}_j\big) \;=\; \frac{E\big(q_k(\vartheta, \underline{X}_j, \underline{z}_j) \prod_{b=1}^4 q_b(\vartheta, \underline{X}_j, \underline{z}_j)^{r_{j,b}}\big)}{E\big(\prod_{b=1}^4 q_b(\vartheta, \underline{X}_j, \underline{z}_j)^{r_{j,b}}\big)} \tag{29}$$

where

$$(r_{j,1}, \ldots, r_{j,4}) \;\equiv\; (m_{j,0} - m_{j,1},\, m_{j,1} - m_{j,2},\, m_{j,2} - m_{j,3},\, m_{j,3})$$

This formula (29) follows immediately from the multinomial random-effects model (9).

In the Bayes approach to predicting $\underline{\pi}_j$, parameters θ are predicted from the posterior distribution for θ given the data $\mathcal{D} \equiv \{(\underline{Y}_j, \hat{N}_j^V\}_{j\in\mathcal{J}}$, and that posterior is usually obtained from an asymptotically stationary and ergodic *Markov Chain Monte Carlo* (MCMC) simulated random sequence $\theta^{(t)}$, $\{\underline{z}_j^{(t)}\}_j$ for $t = 1, 2, \ldots,$ generated conditionally given $\underline{Z}_j$. The prior distribution for unknown parameters used in this MCMC simulation is chosen to be 'noninformative' as part of the Bayesian computational details in Section E.2. The posterior distribution for θ and $\underline{z}_j$ given data (for a fixed $j \in \mathcal{J}$) is then estimated as

$$\widehat{Pr}\big((\theta, \underline{z}_j) \in C \,\big|\, \{\underline{Y}_j\}_j\big) \;=\; \frac{1}{T} \sum_{t=B}^{B+T} I_{[(\theta^{(t)}, \underline{z}_j^{(t)}) \in C]}$$

for a moderately large number B (with default value 4000 in `R-STAN`) of *burn-in* iterations, and a very large T (usually $> 100,000$). The posterior expectations of the quantities $\underline{\pi}_j$ are estimated by

$$\hat{\underline{\pi}}_j \;\equiv\; \widehat{E}\Big(\underline{\pi}_j \,\Big|\, \{\underline{Y}_j\}_j\Big) \;=\; \frac{1}{T} \sum_{t=B+1}^{B+T} q(\theta^{(t)}, \underline{X}_j, \underline{z}_j^{(t)}) \tag{30}$$

Estimated Mean Squared Prodiction Errors (MSPEs) in the Bayesian context are obtained (within `R-STAN`) by treating the estimators $\hat{\pi}_j$ in (30) as stationary-time-series variables and from them using time series variance estimation formulas for the conditional (posterior) variance

$$\mathrm{Var}\Big(\hat{\underline{\pi}}_j \,\Big|\, \{\underline{Y}_j\}_j\Big)$$

Formulas for the frequentist conditional-expectation predictors for $\underline{\pi}_j$ are given in the next Section as functions of the unknown parameters defined from integrals that can be numerically evaluated to high accuracy, into which parameter estimates are substituted. The frequentist Variances or MSPEs are estimated by the unconditional quantities

$$E_\vartheta\big(\hat{\pi}_{j,k} \,-\, \pi_{j,k}\big)^2\,\big|_{\vartheta=\hat{\vartheta}} \qquad \text{or} \qquad E_\vartheta\Big(\sum_{b=k+1}^{4}\big[\hat{\pi}_{j,k} \,-\, \pi_{j,k}\big]\Big)^2\,\big|_{\vartheta=\hat{\vartheta}} \tag{31}$$

Predictions of the counts $N_{jg}^A,\ A = V, C, L, I$, (and certain ratios of them) are of primary interest in the Voting Rights analysis. Both in their frequentist and Bayesian versions, the model-based predictions $\hat{\pi}_j$ are translated into model-based predictors $\hat{N}_{jg}^{A,M}$ of these counts by plugging them into the relations (10), where $\hat{N}_j^V$ is the direct survey-weighted estimator given in (5):

$$\hat{N}_j^{C,M} \;=\; \hat{N}_j^V \,\cdot\, (1-\hat{\pi}_{j,1}) \;\;,\quad \hat{N}_j^{L,M} \;=\; \hat{N}_j^V \,\cdot\, (\hat{\pi}_{j,3}+\hat{\pi}_{j,4}) \;\;,\quad \hat{N}_j^{I,M} \;=\; \hat{N}_j^V \,\cdot\, \hat{\pi}_{j,4} \tag{32}$$

Since the estimates $\hat{\pi}_{j,k}$ will all be strictly between 0 and 1, it follows immediately from (32) that all domains (j, g) with positive VOT sample-count n_{jg}^V will have positive VOT total estimate $\hat{N}_{jg}^V$ and positive predicted CIT, LEP, and ILL predictions $\hat{N}_{jg}^A$.

The Variances or MSPEs of interest in the Voting Rights project are estimates of conditional (given data) or unconditional expected squares of $\hat{N}_j^A - N_j^A$ for $A = V, C, L, I$, or of $\hat{N}_j^{A,M} - N_j^A$ for $A = C, L, I$, and corresponding conditional or unconditional expected squares of the differences between certain ratios of these model-based predictions and the same ratios of their targets. The square roots of these Variances or MSPEs are published as 'Margins of Error' in the public data release.

The relations between the different frequentist and Bayesian variance concepts is discussed further in Section D.1 below. Sections D.3 and D.5 present more precisely the formulas used in model-based prediction supporting the VRA determinations. Further details of computation of frequentist and Bayesian predictions and MSPE estimates are described fully in Section E.

D.1 Relations between Different Prediction and Variance Concepts

The preceding two sections have provided high-level formulas (32) for predictors $\hat{N}_j^{A,M}$ and for MSPE's of model-based predictors $\hat{\pi}_{j,k}$. In this subsection, we connect these formulas conceptually through the model-assumption (7) with model-based MSPE formulas applicable to estimate frequentist and Bayesian MSPEs for $\hat{N}_j^{A,M}$, $A = C, L, I$, and their ratios.

Regardless of whether counts are estimated by direct or model-based methods, the error is measured by the conditional or unconditional expected square discrepancy between the estimate and the target. Subtle differences between these notions arise depending on whether we regard the target totals as being pure counts or stochastic, modeled quantities. The estimators or predictors also depend both on the randomly selected samples (and survey weights) and on the modeled quantities $Y_{j,k}$ and underlying random effects $\pi_{j,k}$. The simplest case of estimation and error measurement arises in the case of the survey-weighted direct estimators $\hat{N}_j^A$ defined in (5). In the design-based view, these estimators are unbiased for their nonstochastic target counts N_j^A (defined below (10)), and the variances and MSPEs are the same

$$\text{Var}(\hat{N}_j^A) \;\equiv\; \text{MSPE}(\hat{N}_j^A) \;=\; E\big(\hat{N}_j^A \,-\, N_j^A\big)^2 \quad \text{for} \quad A = V, L, C, I \tag{33}$$

If in this setting we take a model-based point of view as in (10), with $N_j^A = N_j^V \cdot \pi_j^A$, or

$$(N_j^C, N_j^L, N_j^I) \;=\; N_j^V \cdot (1 - \pi_{j,1},\, \pi_{j,3} + \pi_{j,4},\, \pi_{j,4}) \tag{34}$$

then one might imagine taking conditional expectations also given $\underline{Y}_j$ or given the full set of observed data $\mathcal{D}$, in which case the MSPE (33) would be replaced by $E\big((\hat{N}_j^A - N_j^A)^2 \,|\, \mathcal{D}\big)$. Here the main idea is that the models allow the conditional distributions of $\underline{\pi}_j$ given $\underline{Y}_j$ or $\mathcal{D}$ to be used in expressing the conditional moments of $\hat{N}_j^A - N_j^A$ given $\underline{Y}_j$.

When we measure estimation errors using the model-based estimators $\hat{N}_j^{A,M}$, we necessarily view the targets (34) as random, and there are distinct unconditional and conditional MSPEs:

$$MSPE(\hat{N}_j^{A,M}) \;=\; E\big(\hat{N}_j^{A,M} - N_j^A\big)^2 \quad , \qquad MSPE_C(\hat{N}_j^{A,M}) \;=\; E\big((\hat{N}_j^{A,M} - N_j^A)^2 \,\big|\, \mathcal{D}\big) \tag{35}$$

The first of these MSPEs is generally what one estimates in frequentist Small Area Estimation [Rao and Molina, 2015], while the second is what the Bayesian prefers to estimate as posterior MSPE conditional on data. If the conditional MSPE is of direct interest, then the frequentist estimators of MSPE($\hat{N}_j^{A,M}$) developed in Section D.5 below are not useful. If the unconditional MSPE is the operationally useful quantity, then the conditional one is an unbiased estimate of it but has extra variability except when samples are very large.

Bayesian predictions for $\underline{\pi}_j$ obtained from MCMC sampling are interpreted as posterior conditional expectations, so the corresponding MSPEs are intrinsically conditional on the data and interpeted as posterior variances. The frequentist unconditional MSPEs are obtained in practice by Balanced Repeated Replication (Successive Difference Replication) techniques that we explain concretely in Section D.5. Are the two approaches definitely estimating the same thing, at least in expectation ? They are, in large samples, when $|\mathcal{J}| \to \infty$. The Bernstein-von Mises theorem and associated 'concentration of measure' results [van der Vaart, 1998, Ch. 10, Bickel and Kleijn, 2012] say that for a large class of smoothly parameterized models including those studied in this report, under diffuse or noninformative prior distributions with positive densities, the posterior distribution $f(\vartheta \,|\mathcal{D})$ of the fixed-effect unknown parameters ϑ given data $\mathcal{D} = \{(\underline{Y}_j, \hat{N}_j^V)\}_{j \in \mathcal{J}}$ is asymptotically normal and centered at the maximum likelihood estimator $\hat{\vartheta}$ with asymptotic variance equal to the inverse of the (limit of the) Fisher information matrix (scaled by $1/|\mathcal{J}|$). Then, in the notation $\underline{\pi}_j = q(\vartheta, \underline{X}_j, \underline{z}_j)$ at the beginning of Section D, and continuing to treat $\underline{X}_j$ data as nonrandom (possibly after conditioning on them), when $\underline{Y}_j = \underline{m}_j$, we have as in (29)

$$\hat{\underline{\pi}}_j = \frac{E\big(q(\vartheta, \underline{X}_j, \underline{z}_j)\, p(\underline{Y}_j = \underline{m}_j \,|\, \underline{\pi}_j = q(\vartheta, \underline{X}_j, \underline{z}_j))) \,\big|\, \mathcal{D}\big)}{E\big(p(\underline{Y}_j = \underline{m}_j \,|\, \underline{\pi}_j = q(\vartheta, \underline{X}_j, \underline{z}_j)) \,\big|\, \mathcal{D}\big)} \Big|_{\vartheta = \hat{\vartheta}} \;\equiv\; Q(\hat{\vartheta},\, \underline{X}_j,\, \underline{m}_j)$$

The main point here is that Q is a well-defined smooth function of its arguments,

$$Q(\vartheta, \underline{X}_j, \underline{Y}_j) \;\equiv\; E_{\vartheta}\big(q(\vartheta, \underline{X}_j, \underline{z}_j) \,\big|\, \underline{X}_j, \underline{Y}_j\big)$$

This implies that in the Bayesian setting the MCMC-based predictor (30) could have been replaced at some computational cost by the asymptotically unbiased predictor

$$\hat{\underline{\pi}}_j^{Bayes} \;\equiv\; \frac{1}{T} \sum_{t=B+1}^{B+T} Q(\vartheta^{(t)}, \underline{X}_j, \underline{Y}_j) \tag{36}$$

The Bayesian predictors (30) and (36) have the same convergent-MCMC limit for large B, T, because that limit is just the posterior conditional expectation of $\underline{\pi}_j$. By the Delta Method applied (with base-point $\hat{\vartheta}$ justified by the Bernstein-von Mises form of the posterior distribution for ϑ) when samples and $|\mathcal{J}|$ are large, we see that in large data-samples governed by a fixed parameter,

$$\hat{\underline{\pi}}_i^{Bayes} - \hat{\underline{\pi}}_j \;=\; \frac{1}{T} \sum_{t=B+1}^{B+T} \big(Q(\vartheta^{(t)}, \underline{X}_j, \underline{Y}_j) \,-\, Q(\hat{\vartheta}, \underline{X}_j, \underline{Y}_j)\big)$$

$$\approx\; \big[\nabla_{\vartheta}\, Q(\hat{\vartheta}, \underline{X}_j, \underline{Y}_j)\big]' \,\frac{1}{T} \sum_{t=B+1}^{B+T} (\vartheta^{(t)} - \hat{\vartheta}) \;=\; O_P\big((T\,|\mathcal{J}|)^{-1/2} \,+\, T^{-1}\big)$$

This argument implies that the order of approximation between the frequentist and Bayesian predictors of $\underline{\pi}_j$ is actually better (for large B, T) than the order of accuracy of $\hat{\underline{\pi}}_j$, and that is what we found in computational comparisons of predictions using Bayesian versus frequentist methods under the MLN2 model on selected large LMGs. The overall impact of this result is that for practical purposes, in large samples the Bayesian and frequentist predictors may be regarded as equivalent, and the Bayesian posterior variances of these predictions can also serve as conditional MSPEs for frequentist estimates.

Despite the difference in outlook between Bayesian and frequentist analysts on reporting variability, the estimated unconditional MSPE quantity (35) is also worth reporting when estimation is done by Bayesian methods. That is true partly to maintain maximal comparability between the estimates of variability reported for LMGs analyzed using Bayesian computations and those using frequentist computations, since the decision to use Bayesian estimates in the VRA project was made for reasons of computational stability more than inferential philosophy.

We must still provide a computational method, in Section D.5, for how to estimate MPSE. But at this point, we explain conceptually how to combine variance estimates for the survey-weighted direct estimates $\hat{N}_j^V$ with the MSPE estimate for Bayesian or frequentist model-based predictors $\hat{\pi}_{j,k}$ to yield estimates for MSPEs (35).

D.2 MSPEs of Counts from MSPEs of Random Probabilities

From either a frequentist or Bayesian point of view, we start now from available predictors and MSPEs for model-based domain-level mixed-effect probabilities $\underline{\hat{\pi}}_j$ to develop MSPE formulas for count-predictors $\hat{N}_j^{A,M}$ and related ratios of counts. The main idea here is the formula

$$E\big(\{\hat{N}_j^V\,\hat{\pi}_j^A\;-\;N_j^V\,\pi_j^A\}^2\big)\;=\;E\big(\{\hat{N}_j^V\,(\hat{\pi}_j^A\;-\;\pi_j^A)\;+\;(\hat{N}_j^V\;-\;N_j^V)\,\pi_j^A\}^2\big)$$

for $A = C, L, I$, where

$$\pi_j^C\;\equiv\;1-\pi_{j,1}\;\;,\qquad \pi_j^L\;\equiv\;\pi_{j,3}+\pi_{j,4}\;\;,\qquad \pi_j^I\;\equiv\;\pi_{j,4} \tag{37}$$

and the analogous notational definitions $\hat{\pi}_j^A$ in terms of $\hat{\pi}_{j,k}$ also hold. Then, by the independence assumption (7′), the MSPE in (35) can be re-expressed in the decomposed form

$$E\big(\{\hat{N}_j^V\}^2\big)\,E\big(\{\hat{\pi}_j^A\;-\;\pi_j^A\}^2\big)\;+\;\mathrm{Var}(\hat{N}_j^V)\,E(\{\pi_j^A\}^2)$$

The variance of $\hat{N}_j^V$ can be estimated by the standard ACS method of SDR covered in Section D.5 below. So the previous discussion of MSPE estimation for $\underline{\hat{\pi}}_j$ completes the derivation of a method for (frequentist) MSPE estimation for $\hat{N}_j^{A,M}$. The high-level formula, specified further in Section D.5, is

$$\widehat{\mathrm{MSPE}}(\hat{N}_j^{A,M})\;=\;\widehat{\mathrm{Var}}(\hat{N}_j^V)\,\cdot\,\hat{E}\big(\{\pi_j^A\}^2\big)\;+\;\big\{\hat{N}_j^V\big\}^2\,\cdot\,\widehat{\mathrm{MSPE}}\big(\hat{\pi}_j^A\big) \tag{38}$$

for $A = C, L, I$. In the Bayesian case, $\{\hat{\pi}_j^A\}^2\;=\;\big\{E\big(\pi_j^A\,\big|\,\mathcal{D}\big)\big\}^2$ itself is an unbiased estimator of $E\big(\{\pi_j^A\}^2\,\big|\,\mathcal{D}\big) - \mathrm{Var}\big(\pi_j^A\,|\,\mathcal{D}\big)$ since the data $\mathcal{D}$ include $\hat{N}_j^V$ while N_j^V is not modeled and considered nonrandom, and

$$MSPE_C(\hat{N}_j^{A,M})\;=\;(\hat{N}_j^V - N_j^V)^2\,E\big(\{\pi_j^A\}^2\,\big|\,\mathcal{D}\big)\;+\;\{\hat{N}_j^V\}^2\,\mathrm{Var}\big(\pi_j^A\,|\,\mathcal{D}\big) \tag{39}$$

Therefore the Bayesian *unconditional* MPSE estimation formula, taking the sampling variability into account and relying on the argument in Section D.1 for the close agreement between the conditional and unconditional variance of π_j^A, becomes

$$\widehat{\mathrm{Var}}(\hat{N}_j^V)\,\cdot\,\big(\{\hat{\pi}_j^A\}^2\;+\;\widehat{\mathrm{Var}}\big(\pi_j^A\,\big|\,\mathcal{D}\big)\big)\;+\;\{\hat{N}_j^V\}^2\,\cdot\,\widehat{\mathrm{Var}}\big(\pi_j^A\,|\,\mathcal{D}\big) \tag{40}$$

D.3 Prediction Formulas

This section provides frequentist prediction formulas for $\hat{\pi}_{j,k}$ for both the models MLN and DM. Unlike the Bayesian predictions in (30), the frequentist predictions $\underline{\hat{\pi}}_j$ are developed from conditional expectations using (14) under MLN and (15) under DM.

Under model MLN, it is easy to check from (14) that

$$\sum_{t=k+1}^{4} \pi_{j,t} \;=\; \prod_{b=1}^{k} \alpha_{j,b} \;=\; \prod_{b=1}^{k} h(\eta_{j,b} + u_{j,b}) \qquad \text{for} \quad k = 1, 2, 3$$

Therefore, the predictors derived from equation (14) are expressed in the form

$$\sum_{t=k+1}^{4} \hat{\pi}_{j,t} \;=\; \prod_{b=1}^{k} \hat{\alpha}_{j,b} \;=\; E\Big(\prod_{b=1}^{k} h(\eta_{j,b} + u_{j,b}) \,\big|\, \underline{Y}_j \Big) \Big|_{\vartheta = \hat{\vartheta}} \tag{41}$$

with unknown parameters ϑ replaced by estimates. The conditional expectation in (41) is expressed as a ratio of 3-dimensional multivariate-normal integrals (or products of 1-dimensional integrals) for the numerator and denominator of (29), to be discussed further in Section E.

Under model DM, the Beta distributions in (15) imply conditionally given $\underline{Y}_j$,

$$\begin{aligned}
\alpha_{j,1} \;&=\; 1 - \pi_{j,1} & \sim \quad & \text{Beta}(\tau_1\, v_1 + Y_{j,1},\; \tau_1\,(1 - v_1) + n_j - Y_{j,1}) \\
\alpha_{j,2} \;&=\; 1 - \frac{\pi_{j,2}}{1 - \pi_{j,1}} & \sim \quad & \text{Beta}(\tau_2\, v_1\, v_2 + Y_{j,2},\; \tau_2\, v_1\,(1 - v_2) + Y_{j,1} - Y_{j,2}) \\
\alpha_{j,3} \;&=\; 1 - \frac{\pi_{j,3}}{1 - \pi_{j,1} - \pi_{j,2}} & \sim \quad & \text{Beta}(\tau_3\, v_1\, v_2\, v_3 + Y_{j,3},\; \tau_3\, v_1\, v_2\,(1 - v_3) + Y_{j,2} - Y_{j,3})
\end{aligned} \tag{42}$$

where as before, $v_k = h(\eta_{j,k})$ and the random variables on the left-hand sides of the displayed equations in (42) are independent. Therefore, under the DM model,

$$E(\alpha_{j,k} \,|\, \underline{Y}_j) \;=\; \big(\tau_k \prod_{b=1}^{k} h(\eta_{j,b}) + Y_{j,k}\big) \big/ \big(\tau_k \prod_{b=1}^{k-1} h(\eta_{j,b}) + Y_{j,k-1}\big)$$

Independence of the Beta variables in (15) implies conditional independence given $\underline{Y}_j$, so that

$$\sum_{t=k+1}^{4} \hat{\pi}_{j,t} \;=\; \prod_{b=1}^{k} \hat{\alpha}_{j,b} \;=\; \prod_{b=1}^{k} \Big(\frac{\hat{\tau}_b \prod_{t=1}^{b} h(\hat{\eta}_{j,t}) + Y_{j,b}}{\hat{\tau}_b \prod_{t=1}^{b-1} h(\hat{\eta}_{j,t}) + Y_{j,b-1}} \Big) \tag{43}$$

The final prediction formulas combine (32) with (41) for model MLN-D and with (43) for DM.

D.4 Composing Stagewise Predictions

The prediction formulas described so far under MLN are fully general, encompassing both the frequentist and Bayesian analysis methods under the general MLN-F model in which the random effects $(u_{j,k},\ k = 1, 2, 3)$ are allowed to be dependent. In the more restricted MLN-D model, in which random effects $u_{j,k}$ are independent across k, the estimation and prediction formulas decouple from one k to another, through the successive logistic regression models (19), and the prediction formulas become much more explicit. Similarly, in the DM model the likelihoods factor into stagewise submodels in (26) with separate parameters. In both types of models (MLN-D and DM), the frequentist predictors given by (29) are constructed, according to equations(14) and (15) to have very special structure. First, in the MLN-D model, for $k = 1, 2, 3$,

$$E\Big(\sum_{t=k+1}^{4} \pi_{j,t} \,\Big|\, (Y_{j,b},\ b \le k),\ (u_{j,b}, b < k) \Big) \ = \ \Big[\prod_{b=1}^{k-1} h(\eta_{j,b} + u_{j,b}) \Big] \ \cdot\ E(h(\eta_{j,k} + u_{j,k}) \,|\, Y_{j,k})$$

Using the formulas (29) and (21), we find

$$E\Big(\sum_{t=k+1}^{4} \pi_{j,t} \,\Big|\, (Y_{j,b}, u_{j,b},\ b \le k) \Big) \ = \ \Big[\prod_{b=1}^{k-1} h(\eta_{j,b} + u_{j,b}) \Big] \ \cdot\ g(Y_{j,k-1},\ Y_{j,k},\ \eta_{j,k},\ \sigma_k)$$

Therefore, for $k = 1, 2, 3$, the frequentist predictor in model MLN-D satisfies

$$\sum_{t=k+1}^{4} \hat{\pi}_{j,t} \ = \ \prod_{t=1}^{k} g(Y_{j,t-1},\ Y_{j,t},\ \hat{\eta}_{j,t},\ \hat{\sigma}_b) \tag{44}$$

where $\hat{\eta}_{j,t} \ \equiv \underline{X}_j'\, \hat{\beta}^{(t)}$. The predictor $\hat{\mu}_{j,k} = \prod_{t=1}^{k} h(\hat{\eta}_{j,t})$ in the absence of random effects is the same as the limit of (44) as all $\hat{\sigma}_t$ terms go to 0.

Similarly, in the DM model (where stagewise random effects were already assumed independent), the prediction formulas (43) can be expressed as a product of stagewise predictors in the form

$$\sum_{b=k+1}^{4} \hat{\pi}_{j,b} \ = \ \prod_{t=1}^{k} \hat{\alpha}_{j,t} \ = \ \prod_{t=1}^{k} \frac{\hat{\tau}_t\, \hat{\mu}_{j,t} \ + \ Y_{j,t}}{\hat{\tau}_t\, \hat{\mu}_{j,t-1} \ + \ Y_{j,t-1}} \qquad \text{for} \quad k = 1, 2, 3 \tag{45}$$

where the corresponding predictors $\hat{\mu}_{j,b}$ in the absence of random effects (i.e., in the limit where all $\hat{\tau}_b \to \infty$) are given by

$$\hat{\mu}_{j,b} \ = \ \prod_{t=1}^{b} h(\hat{\eta}_{j,t}) \quad \text{for} \quad b \ge 1, \qquad \hat{\mu}_{j,0} \ = \ 1$$

Recall that in certain extreme cases where the MLN-D or DM models do not converge, we proposed to model $Y_{j,k}$ in terms of $Y_{j,k-1}$ using a beta-binomial model (28). In that case, the beta-binomial posterior-mean predictor for $\alpha_{j,k} = \sum_{t=k+1}^{4} \pi_{j,t} / \sum_{t=k}^{4} \pi_{j,t}$ is

$$\hat{\alpha}_{j,k} \;=\; E\big(Y_{j,k} \,|\, (Y_{j,b}, b < k)\big) \,/\, Y_{j,k-1} \;=\; (Y_{j,k} + 1/2) \,/\, (Y_{j,k-1} + 1)$$

and (in case this beta-binomial form were used only in the stage k model) the overall predictor for $\sum_{t=b+1}^{4} \pi_{j,t}$ is obtained by substituting this stage k predictor whenever $b \geq k$. Under the MLN-D model (with Jeffreys beta-binomial at k'th-stage), the predictor for $\sum_{t=b+1}^{4} \pi_{j,t}$ takes the form

$$\sum_{t=b+1}^{4} \hat{\pi}_{j,t} \;=\; \Big[\prod_{t=1}^{b} g(Y_{j,b-1},\, Y_{j,b},\, \hat{\eta}_{j,b},\, \hat{\sigma}_b)\Big] \cdot \Big[\frac{(Y_{j,k} + 1/2) \,/\, (Y_{j,k-1} + 1)}{g(Y_{j,k-1},\, Y_{j,k},\, \hat{\eta}_{j,k},\, \hat{\sigma}_k)}\Big]^{I[b \geq k]} \tag{46}$$

The analogous formula if the DM model applies to the stages other than the k'th is:

$$\sum_{t=b+1}^{4} \hat{\pi}_{j,t} \;=\; \Big[\prod_{t=1}^{b} \frac{\hat{\tau}_t \,\hat{\mu}_{j,t} \,+\, Y_{j,t}}{\hat{\tau}_t \,\hat{\mu}_{j,t-1} \,+\, Y_{j,t-1}}\Big] \cdot \Big[\frac{(\hat{\tau}_k \,\hat{\mu}_{j,k-1} \,+\, Y_{j,k-1})\,(Y_{j,k} + 1/2)}{(\hat{\tau}_k \,\hat{\mu}_{j,k} \,+\, Y_{j,k})\,(Y_{j,k-1} + 1/2)}\Big]^{I[b \geq k]} \tag{47}$$

D.5 Variance and MSPE Formulas

Conceptual descriptions of variance and MSPE estimation have been given above in Sections D, D.1 and D.2, for the direct estimates and model types under consideration for the VRA project. In this subsection, these descriptions are made more concrete with the mathematical computing formulas used in Variance and MSPE estimation. Further details of the computational implementation of these formulas are given in Section E below.

The primary technique of design-based variance estimation used in major Census Bureau surveys such as the ACS is Balanced Repeated Replication [Wolter, 2007, Fay and Dippo, 1989] and more specifically Successive Difference Replication (SDR) [Fay and Train, 1995]. This technique estimates the variance of a survey-weighted total estimator $\hat{t}_Y = \sum_{i \in \mathcal{S}} y_i \, w_i$ by re-calculating the survey-total estimate with the survey weights w_i replaced by a system of replicate-weights $w_{i,r} = w_i \, f_{i,r}$ for $r = 1, \ldots R$. The weight-multipliers $f_{i,r}$ are defined systematically from the sequence of sort-ordered respondents $i \in \mathcal{S}$ according to a recipe described in detail in the paper of Fay and Train [1995] and elaborated in the journal paper of Ash [2014], typically with $R = 80$. With $\hat{t}_{Y,(r)} \equiv \sum_{i \in \mathcal{S}} y_i \, w_{i,r}$, the SDR variance estimate for the total $\hat{t}_Y$ is

$$\hat{V}^{SDR}(\hat{t}_Y) \;=\; (4/R) \sum_{r=1}^{R} (\hat{t}_{Y,(r)} \,-\, \hat{t}_Y)^2 \tag{48}$$

It is important to recognize that (48) estimates, in approximately unbiased fashion, only the *sampling* variance due to random sampling with unequal weights the estimated total of constant attribute y_i in a finite population. This estimator makes no attempt to take advantage of relationships between these constants and covariate vectors of constants $\mathbf{X}_i$. Moreover, if the attributes y_i are indicators $I_{[i \in (jg) \cap A]}$, where (jg) indexes geography and LMG and A denotes one of the categorical identifiers CIT, LEP, or ILL, then the random effect incorporated within the model $N_j^A = N_j^V \, \pi_j^A$ under the MLN or DM model must be regarded as fixed with respect to the sampling variability being assessed by (48). Under the multinomial random-effects model MLN or DM,

$$\hat{V}^{SDR}(\hat{N}_j^A) \;\approx\; E\Big((\hat{N}_j^A \;-\; N_j^A)^2 \,\big|\, \underline{u}_l,\; l=1,\ldots,J\Big) \;=\; E\Big((\hat{N}_j^A \;-\; N_j^V \, \pi_j^A)^2 \,\big|\, \underline{u}_l,\; l=1,\ldots,J\Big) \quad (49)$$

is conceptually the same as the conditional Mean Squared Prediction Error for $\hat{N}_j^A$ given the fixed finite population including the full set of population random effects in $\underline{\pi}_l,\; j \in \mathcal{J}$. This measure (49) of variance does not average over alternate possible populations defined by the distribution of random effects. It is the baseline measure of prediction accuracy against which model-based predictors were compared in 2011, 2016, and again in 2021.

Because the SDR Variance (49) does not average over alternative populations, it differs in principle from the model-based Mean-Squared Prediction Error, defined by

$$\texttt{MSPE}(\hat{N}_j^A) \;=\; E\big(\hat{N}_j^A - N_j^V \, \pi_j^A\big)^2$$

More broadly, for any estimator or predictor[8] $\tilde{N}_j^A$ of N_j^A, the *Mean Squared Prediction Error* is defined by

$$\texttt{MSPE}(\tilde{N}_j^A) \;=\; E(\tilde{N}_j^A \;-\; N_j^V \, \pi_j^A\big)^2 \quad (50)$$

Thus, unconditional MSPE represents expected squared discrepancy between an estimator and the random variable N_j^A incorporating the true random effect for unit j and is unconditional in the sense that it is averaged over all potential populations under the model with random effects $\underline{u}_l$.

The method of estimating MSPE in (50) is different in the 2021 VRA cycle than it was in 2016 [Slud et al., 2018]. In 2016, the conditional variance of the predictor $\hat{\pi}_j^A$ was obtained using parametric bootstrap for data derived from each of the set of $R = 80$ replicate weights. In 2021, the primary frequentist method of estimating (50) treats $\tilde{N}_j^A$ as a somewhat complicated nonlinear function of the data $(\hat{N}_l^V, \, \underline{Y}_l, \; l = 1, \ldots, J)$ that can be re-calculated using each of the replicate-weights $w_{j,r}$ (in place of w_j) to yield alternative predictions $\tilde{N}_{j,(r)}^A = \hat{N}_{j,(r)}^V \, \hat{\pi}_{j,(r)}^A$. With this point of view, one obtains an SDR expression for the conditional expected squared discrepancy between $\tilde{N}_j^A$ and its design-based expectation over samples drawn:

$$\hat{V}^{SDR}(\tilde{N}_j^A) \;=\; \frac{4}{R} \sum_{r=1}^{R} \big(\hat{N}_{j,(r)}^V \, \hat{\pi}_{j,(r)}^A \;-\; \tilde{N}_j^A\big)^2 \quad (51)$$

[8] Predictor is the terminology we follow, as Bayesians or frequentists, when an estimated quantity is not a fixed unknown parameter but rather a random variable.

Each of the predictors $\hat{\alpha}_{j,k},\ k = 1,2,3$, is a smooth function of $\underline{X}_j'\,\hat{\beta}^{(k)}$ and either $\hat{\Sigma}$ in MLN or (τ_1, τ_2, τ_3) in model DM. Since n_j^V is viewed as fixed and known, this makes each $\hat{\pi}_j^A$ a smooth nonlinear function of data $\hat{N}_j^A$ for $A = V, C, L, I$ along with roots of a finite-dimensional score equation used to define $\hat{\beta}$ and either $\hat{\Sigma}$ or $\underline{\hat{\tau}}$. By a slight extension of the usual justification [Krewski and Rao, 1981] for BRR estimates of variances of smooth functions of survey-weighted sums, we can base estimates of the variances of $\hat{\pi}_j^A$ on Successive Difference Replication (SDR) as an instance of Balanced Repeated Replication (BRR).

The quantity on the right-hand side of (51) is an estimated conditional MSPE averaged over that part of the data-generating mechanism for $(\underline{Y}_j,\ j \in \mathcal{J})$ associated with the random-response-based assignment of weights to respondents, but with the randomness associated with the random effects $(\underline{\pi}_j,\ j \in \mathcal{J})$ fixed. From this point of view, the variance estimate (51) should be robust to possible dependence between the direct-estimated quantities $\hat{N}_j^V$ and the data $\mathcal{D}$, while independence of these was a necessary assumption in the MSPE estimation method adopted in 2016.

The foregoing discussion of MSPE estimates using SDR applies to frequentist direct and modeled MSPE estimates. The Bayesian estimates, based on MCMC-calculated posterior variance for $\underline{\hat{\pi}}_j$ given data, were already shown in formula (40) to provide unconditional MSPE estimates by making use of SDR variances for $\hat{N}_j^V$. The independence assumption (7) was also used in deriving that Bayesian MSPE formula.

D.6 Variance Calculations for Ratios

We have indicated in several places that the ratio variables `LEPprop` and `ILL/LEP` within LMG for each type of geography are important for Voting Right determinations. The predictors for these ratios are defined straightforwardly as the ratios of the corresponding predictors of numerator and denominator totals appearing in their definitions. Recall the notations $N_{j+}^C,\ \hat{N}_{j+}^C$ respectively for the true and direct-method CIT counts (without regard to LMG) within jurisdiction j, and the analogous notations $N_{a+}^C,\ \hat{N}_{a+}^C$ for counts of AIAN citizens within AIA or ANRC geographies uindexed by a. Also recall the notations $N_{jg}^L, N_{jg}^I, N_{ag}^L, N_{ag}^I$ for true LEP or ILL counts within geographic units j or a, and the corresponding model-based (frequentist or Bayesian) predictors $\tilde{N}_{jg}^L, \tilde{N}_{jg}^I, \tilde{N}_{ag}^L, \tilde{N}_{ag}^I$. Then the ratio variable predictors are defined explicitly as

$$\widetilde{\texttt{LEPprop}}_{jg} = \frac{\tilde{N}_{jg}^L}{\hat{N}_{j+}^C} \cdot I_{[\hat{N}_{j+}^C > 0]} \quad , \qquad \widetilde{\text{ILL/LEP}}_{jg} = \frac{\tilde{N}_{jg}^I}{\tilde{N}_{jg}^L} \cdot I_{[\hat{N}_{jg}^L > 0]} = \frac{\tilde{\pi}_{jg}^I}{\tilde{\pi}_{jg}^L} \cdot I_{[\hat{N}_{jg}^L > 0]} \tag{52}$$

with analogous notations and definitions when Jurisdiction geography-type and indices j are replaced by AIA or ANRC geography-type and indices a. Here π_{jg}^A are defined as in (37), and $\tilde{\pi}_{jg}^A$ analogously with $\tilde{\pi}_{jk}^A$ substituted for π_{jg}^A. In the rest of this section, we describe the formulas for Variance or MSPE estimates for the predictors in (52).

When the predictors are made from frequentist estimators, the MSPE or variance predictors are constructed fairly painlessly using the same weight-replication idea as the formula (51): the predictor is recalculated using the SDR weight-replicates indexed $r = 1, \ldots, 80$, and the MSPE estimates are calculated directly from these replicate-predictors,

$$\widehat{\text{MSPE}}(\widetilde{\texttt{LEPprop}}_j) = \frac{4}{80} \sum_{r=1}^{80} \big(\widetilde{\texttt{LEPprop}}_{j,(r)} - \widetilde{\texttt{LEPprop}}_j\big)^2 \tag{53}$$

$$\widehat{\text{MSPE}}(\widetilde{\text{ILL/LEP}}_j) = \frac{4}{80} \sum_{r=1}^{80} \big(\widetilde{\text{ILL/LEP}}_{j,(r)} - \widetilde{\text{ILL/LEP}}_j\big)^2 \tag{54}$$

Three special comments must be made here. First, such predictors and MSPE estimates are produced only for geographies with VOT sample, $\hat{N}^V_{jg} > 0$. Second, the indicators in the definitions (52) ensure that these ratio-predictors are well-defined (= 0) when $\hat{N}^L_{jg} = 0$. Third, there are some small geographic units for which the ACS version of the replicate weights $w_{i,r}$ are 0 or even negative, which can result in denominators $\hat{N}^L_{j,(r)}$ being negative or 0. In all such cases, we define the corresponding $\widetilde{\text{ILL/LEP}}_{j,(r)}$ to be 0.

When the predictors $\tilde{N}^A_{jg}$ are derived from Bayesian analysis, the formulas for MSPE of LEPprop and ILL/LEP both are conditioned on the data $\underline{Y}_j$ but have different forms. First, since the sample size from which $\hat{N}^C_{j+}$ is estimated is generally much larger than the sample size n^L_{jg}, it makes sense to use the Bayesian form of linearized variance or MSPE formula for LEPprop:

$$\widetilde{MSPE}\big(\texttt{LEPprop}_j \,|\, \underline{Y}_j\big) \approx \text{Var}\Big(\frac{\tilde{N}^L_j - N^L_j}{N^C_{j+}} - (\hat{N}^C_{j+} - N^C_{j+}) \frac{N^L_j}{(N^C_{j+})^2} \,|\, \underline{Y}_j\Big)$$

and also to disregard the data-conditional covariance between $\tilde{N}^L_j$ and $\hat{N}^C_{j+}$. (This last decision, to ignore covariances in the last expression, is partly motivated by the different scale of observations of the all-geographic-unit CIT versus the LMG-specific LEP, but also because any nonzero covariance is likely to be negative, resulting in a slightly conservative MSPE estimate. This aspect of the joint behavior of assessments should be assessed in future VRA research.) The combination of these two approximations yields

$$\widetilde{MSPE}\big(\texttt{LEPprop}_j \,|\, \underline{Y}_j\big) \approx (\hat{N}^C_{j+})^{-2} \big[\widehat{\text{Var}}(\tilde{N}^L_j \,|\, \underline{Y}_j) + \texttt{LEPprop}^2_j \, \widehat{\text{Var}}(\hat{N}^C_{j+})\big] \tag{55}$$

All terms on the right-hand side of this equation are estimated in ways that have already been described, with the first conditional variance a Bayesian posterior variance obtained from MCMC, and the second (unconditional) variance of a direct estimator calculated by a direct SDR weight-replication formula.

Finally, still in the Bayesian setting, the ratio $\tilde{\pi}_j^I/\tilde{\pi}_j^L$ in the final equality of (52) has posterior MSPE directly obtainable from MCMC in the form

$$\widehat{\mathrm{Var}}\Big(\frac{\pi_j^I}{\pi_j^L}\,|\,\underline{Y}_j\Big) \;+\; \Big(\frac{\hat{E}(\pi_j^I\,|\,\underline{Y}_j)}{\hat{E}(\pi_j^L\,|\,\underline{Y}_j)} \;-\; \hat{E}\Big(\frac{\pi_j^I}{\pi_j^L}\,|\,\underline{Y}_j\Big)\Big)^2$$

E Computational Methods

The multinomial logistic regression model **MLN-F** is an extension of the simpler random-intercept logistic regression model, which has been treated thoroughly both in textbooks [Agresti, 2013, Stroup, 2012] and in software, particularly in `SAS` and in the `R` functions and packages `lme4`, `glmer`, `glmmML`, `GLMMadaptive`, as well as the custom-coded `NRstpGQ` function [Slud, 2000b]. Accurate log-likelihood calculation and maximization is simpler in the lower-dimensional problem than in the mixed-effect multinomial logistic regression MLN-F. It is desirable to make use of the solved, simpler problem wherever possible, either to provide starting values of completely justified parameter estimates or by making use of sub-models.

Whether in the 1- or 3-dimensional integrals over random effects, the numerical quadrature method that has been far and away the most successful and accurate approach to MLN log-likelihood calculation is Adaptive Gaussian Quadrature (AGQ) introduced by Pinheiro and Bates [1995]. Before its advent, even in the 1-dimensional case (random-intercept logistic regression) there had only been approximate methods, the top-order Laplace asymptotic approximation or estimation methods based on penalized likelihood [Breslow and Clayton, 1993]. Since the introduction of AGQ, it has been incorporated into the `R` `glmer` function and other `R` packages and into `SAS` proc's `GLIMMIX` and `NLINMIX`. A study by Slud [2000b] of the accuracy of several competing methods of loglikelihood computation for random-intercept logistic regression showed AGQ's impressive and reliable accuracy.

No AGQ numerical integration was done in the 2016 Dirichlet-Multinomial models [Slud et al., 2018] or in the somewhat different DM models described in Section C.2 in this document. In these models, the loglikelihods can be evaluated directly in terms of Gamma functions and optimized through general-purpose numerical maximization functions `nlm` and `optim` in `R`.

E.1 Adaptive Gaussian Quadrature in MLN Models

The Adaptive Gaussian Quadrature (AGQ) idea in the present context is explained thoroughly in a multilevel GLMM context by Pinheiro and Chao [2006], following the earlier idea of Pinheiro and Bates [1995]. That idea is adapted here in a form similar to one used previously in `R` code for random-intercept logistic regression [Slud, 2000b].

The log-likelihood contribution for domain j (within LMG g that is not explicitly shown in the notations) is given as a function of parameter ϑ :

$$\log L_j(\vartheta) \;=\; \log\Big((2\pi)^{-3/2} \int\!\!\int\!\!\int \exp\Big\{\underline{W}_j \cdot \tag{56}$$

$$\log p(\eta_{j1,} + \sigma_1 z_1,\, \eta_{j2} + a_1 z_1 + \sigma_2 z_2,\, \eta_{j3} + a_2 z_1 + a_3 z_2 + \sigma_3 z_3) \;-\; \frac{1}{2}\,(z_1^2 + z_2^2 + z_3^2)\Big\}\, dz_1\, dz_2\, dz_3\Big)$$

where $\underline{W}_j \;\equiv\; (n_j^V - Y_{j1},\, Y_{j1} - Y_{j2},\, Y_{j2} - Y_{j3},\, Y_{j3})$. For reference later in this section, denote the exponent in the integrand in the last-displayed triple integral as $K_j(\underline{z}\,; \underline{W}_j, \vartheta) \;=$

$$K_j(\underline{z}) \;=\; \underline{W}_j \cdot \log p(\eta_{j1} + \sigma_1 z_1,\, \eta_{j2} + a_1 z_1 + \sigma_2 z_2,\, \eta_{j3} + a_2 z_1 + a_3 z_2 + \sigma_3 z_3) \;-\; \frac{1}{2}\|\underline{z}\|^2 \tag{57}$$

In model **MLN-D** with diagonal covariance matrix Σ, i.e., with the added assumption that $a_1 = a_2 = a_3 = 0$, the log-likelihood factors in a particularly simple way. First, the log-likelihood for data $\underline{Y}_j = \sum_{k=j+1}^4 W_{jk}$ conditionally given $\underline{u}_j$ takes the form (apart from omitted multinomial coefficients, and with cluster-j sample size denoted by $Y_{j0} \equiv n_j^V$) :

$$\sum_{k=1}^{4} W_{jk}\, \log \pi_{jk} \;=\; \sum_{k=1}^{3} Y_{jk}\,(\eta_{jk} + u_{jk}) \;-\; \sum_{k=1}^{3} Y_{j,k-1}\, \log(1 + e^{\eta_{jk} + u_{jk}})$$

When Σ is diagonal, it follows that $u_{jk} = \sigma_k\, z_{jk}$, where $z_{jk} \sim \mathcal{N}(0,1)$ are *i.i.d.*. Then the likelihood $L_j(\vartheta)$ factors into three separate terms with univariate integrals depending on disjoint sets of parameters, in the form

$$\log L_j(\vartheta) \;=\; \sum_{k=1}^{3} \log \int \frac{\exp((\eta_{jk} + \sigma_k z_k) Y_{jk})}{\big(1 + \exp\big(\eta_{jk} + \sigma_k z_k\big)\big)^{Y_{j,k-1}}}\; \phi(z_k)\, dz_k \tag{58}$$

where $\phi(z)$ denotes the standard normal density. Since each integral in (58) is the likelihood for a single random-intercept logistic regression observation, and each integral depends on separate parameters $(\beta^{(k)},\, \sigma_k)$, the ML estimation decouples between separate logistic-regression log-likelihoods (for $k = 1, 2, 3$)

$$\sum_{j \in \mathcal{J}} \log \int \frac{\exp((\eta_{jk} + \sigma_k z_k) Y_{jk})}{\big(1 + \exp\big(\eta_{jk} + \sigma_k z_k\big)\big)^{Y_{j,k-1}}}\; \phi(z_k)\, dz_k \tag{59}$$

and we conclude that the separate maximizations of (59) together provide the joint MLE .

The idea of Adaptive Gaussian Quadrature (AGQ) is implemented in a few clear steps, which we describe for each $\log L_j$ log-likelihood term in parallel, to achieve efficiencies in `R`.

Step 1. First the unique maximizer $\underline{z}^*_{j,k}$ of the curly-bracketed exponent in $\log L_j(\vartheta)$ given in (56) must be found. This can be accomplished in vectorized fashion, i.e., in parallel across domains j, via the Newton-Raphson (NR) algorithm. The derivatives of the integrand terms are explicit, because the derivatives for p are explicit. The checking of the NR stopping-criterion at each step can also be vectorized, and the algorithm continues vectorized for those domains where it has not stopped, up to a maximum number of steps, say 10 or 20. In the special random-intercept logistic regression sub-case, 5 NR iterates were generally sufficient.

Step 2. Using the same explicit derivatives as in Step 1, we calculate for each j the 3×3 matrix equal to the negative Hessian of $K_j(\underline{z}\,; \underline{W}_j, \vartheta)$ in $\underline{z}$ at $\underline{z}^*_j$:

$$D(j) \;=\; I_{3\times 3} - \sum_{k=1}^{4} W_{jk}\, \nabla^{\otimes 2}_{\underline{z}} \, \log p_k(\eta_{j1} + \sigma_1 z_1,\, \eta_{j2} + a_1 z_1 + \sigma_2 z_2,\, \eta_{j3} + a_2 z_1 + a_3 z_2 + \sigma_3 z_3) \Big|_{\underline{z}=\underline{z}^*_j}$$

The corresponding gradient $\nabla_{\underline{z}} K_j(\underline{z}^*_j, \underline{W}_j, \vartheta)$ evaluated at $\underline{z}^*_j$ is identically $\mathbf{0}$, and the matrix $D(j)$ is positive definite: both assertions hold because $\underline{z}^*_j$ is the calculus maximizer of K_j.

Step 3. Using Steps 1 and 2, we re-write

$$\log L_j(\vartheta) \;=\; K_j(\underline{z}^*_j\,; \underline{W}_j, \vartheta) \;+\; \log\Big((2\pi)^{-3/2} \int \exp\big(- (\underline{z} - \underline{z}^*_j)^{tr}\, D(j)\, (\underline{z} - \underline{z}^*_j)/2\big) \cdot$$
$$\exp\Big(K_j(\underline{z}) - K_j(\underline{z}^*_j) \;+\; (\underline{z} - \underline{z}^*_j)^{tr}\, D(j)\, (\underline{z} - \underline{z}^*_j)/2\Big)\, d\underline{z}\Big) \tag{60}$$

The exponent in the second line of (60) is the remainder in the 2-term Taylor expansion of $K_j(\underline{z})$ around $\underline{z}^*_j$, and thus of order $\|\underline{z} - \underline{z}^*_j\|^3$, and is a smooth function of parameters ϑ. Moreover, $K_j(\underline{z}) - K_j(\underline{z}^*_j) \le 0$, and in domains with large n^V_j, the matrix $D(j)$ is also roughly of order n^V_j, so that almost all the contribution to the integral occurs in the near neighborhood of $\underline{z}^*_j$.

Step 4. In the final (triple) integral of (60), change variables by $\underline{x} \;=\; D(j)^{1/2}\, (\underline{z} - \underline{z}^*_j)$. Then the argument of the logarithm in (60) becomes

$$\det(D(j))^{-1/2} \int\!\!\int\!\!\int \exp\Big(K_j(\underline{z}) - K_j(\underline{z}^*_j) \;+\; \|\underline{x}\|^2/2\Big)\, \phi(x_1)\, \phi(x_2)\, \phi(x_3)\, dx_1\, dx_2\, dx_3 \tag{61}$$

Step 5. The final step of AGQ evaluates the triple integral (61) by doing each univariate integral numerically using Gaussian quadrature. For a fixed choice of the number m of quadrature points, a set of carefully defined weights w_s and real values v_s (satisfying symmetry conditions $w_{m-s+1} \;=\; w_s,\;\; v_{m-s+1} \;=\; -v_s$) is determined so that $\int f(y)\, \exp(-y^2)\, dy \;\approx\; \sum_{s=1}^m w_s\, f(v_s)$ or $\int g(x)\, \phi(x)\, dx \;\approx\; \pi^{-1/2} \sum_{s=1}^m w_s\, g(\sqrt{2}\, v_s)$ to a high order of approximation, and this formula is exact for polynomials f of degree up to $2m$. The final AGQ approximation of $\log L_j(\vartheta)$ becomes

$$\log L_j(\vartheta) \;\;\approx\;\; -\frac{1}{2}\, \log\det(D(j)) \;+\; K_j(\underline{z}^*_j\,; \underline{W}_j, \theta) \;+\; \log\Big[\frac{1}{\pi^{3/2}} \sum_{s_1=1}^{m} \sum_{s_2=1}^{m} \sum_{s_3=1}^{m} w_{s_1} w_{s_2} w_{s_3} \cdot$$

$$\exp\Big(K_j(\underline{z}_j^* + \sqrt{2}\,D(j)^{-1/2}(v_{s_1}, v_{s_2}, v_{s_3})^{tr}) - K_j(\underline{z}_j^*) \;+\; v_{s_1}^2 + v_{s_2}^2 + v_{s_3}^2\Big)\Big] \tag{62}$$

E.1.1 Gradients and Prediction formulas for MLN Models

Because the MLN models can be high-dimensional ($3d+3$ or $3d+6$ parameters if all covariates are used for each category of outcome), off-the-shelf numerical maximization software can be somewhat slow in maximizing even a well-approximated log-likelihood for them. However such programs are designed to run faster when analytical functions computing or closely approximating log-likelihood gradients (scores) are supplied. In fact, this can be done readily in the MLN models, where the form of the gradients is closely related to the computation of empirical-Bayes predictors of random-effect scores $\beta^{(k)tr} X_i + u_{k,i}$ from data $\underline{W}_i$ and estimated parameters. Under the **MLN-F** model, we are particularly interested in the probabilities, for $k = 1, 2, 3$,

$$\sum_{t=k+1}^{4} \pi_{j,k} \;\equiv\; E(Y_{jk}/n_j^V \,|\, \underline{X}_j,\, \underline{u}_j) \;=\; \prod_{t=1}^{k} h(\eta_{jt} + u_{jt}) \tag{63}$$

and their conditional expectations given the observed data are

$$E(\sum_{t=k+1}^{4} \pi_{j,k} \,|\, \underline{X}_j,\, \underline{Y}_j,\, \vartheta) \;=\; E\big(\prod_{t=1}^{k} h(\eta_{jt} + u_{jt}) \,\big|\, \underline{X}_j,\, \underline{Y}_j,\, \vartheta\big) \tag{64}$$

Under model MLN-D, when the random-effects covariance matrix is diagonal, the last conditional expectation becomes a product of three univariate conditional expectations

$$\alpha_{jk} \;\equiv\; E\big(h(\eta_{jk} + u_{jk}) \,\big|\, \underline{X}_j,\, \underline{Y}_j\big) \;=\; E\big(\texttt{plogis}(\eta_{jk} + u_{jk}) \,|\, \underline{X}_i,\, \underline{Y}_j\big)$$

$$= \int \frac{\exp\big((\eta_{jk} + \sigma_k z)(Y_{jk} + 1)\big)}{(1 + \exp(\eta_{jk} + \sigma_k z))^{Y_{j,k-1}+1}}\, \phi(z)\, dz \Big/ \int \frac{\exp\big((\eta_{jk} + \sigma_k z) Y_{jk}\big)}{(1 + \exp(\eta_{jk} + \sigma_k z))^{Y_{j,k-1}}}\, \phi(z)\, dz$$

equal to $g(Y_{j,k-1} + 1, Y_{jk} + 1, \eta_{jk}, \sigma_k)/g(Y_{j,k-1}, Y_{jk}, \eta_{jk}, \sigma_k)$ where g was defined in formula (21) in Section (C.1.1).

When Σ is not diagonal, the formulas for the terms $E(\sum_{t=k+1}^{4} \pi_{j,t} \,|\, \underline{Y}_j)$ are slightly more complicated (i.e., are 3-dimensional integrals in z_1, z_2, z_3). For example, in the general-variance case,

$$E\big(\pi_{j3} + \pi_{j4} \,\big|\, \underline{X}_j,\, \underline{Y}_j,\, \theta\big) \;=\; \int\!\!\int\!\!\int \exp\big(K_j(\underline{z}; \underline{W}_j, \vartheta)\big)\, h(\eta_{j1} + \sigma_1 z_1) \,\cdot$$

$$\cdot\, h(\eta_{j2} + a_1 z_1 + \sigma_2 z)\, dz_1\, dz_2\, dz_3 \Big/ \int\!\!\int\!\!\int \exp\big(K_j(\underline{z}; \underline{W}_j, \theta)\big)\, dz_1\, dz_2\, dz_3 \tag{65}$$

The 'empirical BLUP' predictors for $E(Y_{j,k} \,|\, \underline{X}_j,\, \underline{u}_j) \;=\; n_j^V \;\cdot\; E(\sum_{t=k+1}^4 \pi_{jt} \,|\, \underline{X}_j,\, \underline{Y}_j)$ are then given by ratios of integrals as in the last displayed equation, with parameters ϑ replaced by their maximum likelihood estimators. Similarly, the Empirical Best Linear Unbiased (EBLUP) predictor for the population ratio modeled by $\alpha_{jk} = \sum_{t=k+1}^4 \pi_{jt} / \sum_{t=k}^4 \pi_{jt} = h(\eta_{jk} + u_{jk})$ is given by substituting $\hat{\vartheta}$ into

$$E\big(h(\eta_{jk} + u_{jk}) \,\big|\, \underline{X}_j, \underline{Y}_j\big) = \iiint \exp\big(K_j(\underline{z}; \underline{W}_j, \vartheta)\big)\, h(\eta_{jk} + u_{jk}) d\underline{z} \Big/ \iiint \exp\big(K_j(\underline{z}; \underline{W}_j, \theta)\big) d\underline{z}$$

E.2 Bayesian Computation in MLN Nodels

The Bayesian analysis of the MLN model was conducted via Markov Chain Monte Carlo (MCMC) using the STAN programming language [STAN Development Team, 2019]. The covariance matrix of the random effects was parametrized as above in terms of the three stagewise marginal variances and the Cholesky decomposition of the correlation matrix, using the LKJ prior recommended in STAN Development Team [2019, Sec. 1.13] and also in Koster and McElreath [2017] and McElreath [2020]. For the marginal priors, we used the recommended Cauchy distribution as suggested in the STAN User's guide, comparing it in sensitivity analysis with other priors including the unit Exponential prior suggested by McElreath [2020]. The regression coefficients were given the 'weakly informative' $\mathcal{N}(0, 10)$ priors for regression coefficients, following recommendations of `STAN`'s author Andrew Gelman.

For each MCMC run, we used four chains, as recommended in STAN Development Team [2019]. We checked that all the chains converged for all the LMGs that used the Bayesian code, which as described on page 4 were numbered 1, 3, 4, 5, 6, 7, 8, 9, 10, 12, 14, 15, 16, 27, 44, 45, 56, 63, 64, 72, and 73. (The Hispanic LMG that we call LMG 73 was analyzed in this way separately in four regional pieces.) We checked that the `Rhat` statistics never exceeded the recommended level 1.05, and examined several sample trace plots and posterior densities for anomalies. Because the model development was done under a different version of JAGS than the productions runs, we also checked whether the posterior expectations and variances under the two versions yielded similar results, and checked for a number of LMGs that increasing the number of iterations by 50% did not much change the results.

There is a very fast experimental STAN program called `vb` for *Variational Bayes* that can quickly produce posterior distributions under the specified model. Because `vb` is currently experimental, we did not use it for the actual production of the estimates, but it proved very useful for rapidly generating initial values for the posterior distribution. In particular, we use certain percentiles $(2.5\%, 25\%, 75\%, 97.5\%)$ of `vb` estimated posteriors and then proceeded with usual `STAN` MCMC. With these initializations and the 'weakly informative' priors mentioned above, we found that the Bayesian and frequentists estimates were substantially equivalent, as discussed in the next Section.

E.3 Computational Comparison of Bayesian and Frequentist Estimates

In the 2021 cycle of production of the LMG-level statistical estimates used in VRA Section 203(b) Determinations, the decision of whether to use Bayesian or frequentist methods of analysis were made for primarily computational reasons. To explain why, and precisely which LMGs and models were analysed by which methods, we present various checks on the relative accuracy of the methods and comparisons of run-times, convergence and reliability of the computed estimates. A summary of how the results of these comparisons were used in the production decisions can be found in Section 3.5 of the main text of this report.

The Bayesian analyses were run, in individual LMGs, only for the MLN-F model. The MCMC analysis of these models required four parallel computations (in order to check convergence of the estimates by the current standards of Bayesian analysis), at the end of which both the parameter estimates and MSPE estimates were rapidly collected. By contract, the frequentist analyses required iterative numerical maximization of the likelihood for the MLN model in each LMG, resulting in LMG geographic-unit-level predictions of CIT, LEP and ILL, and then a further computation of the analogous estimates and predictions for the analogous survey data $\{\underline{Y}_{jg}\}_{j \in \mathcal{J}_g}$ re-generated with 80 separate sets of 'SDR replicate' weights in order to produce margin-of-error MSPE estimates by the methods described in this Appendix.

The frequentist estimates and predictions based on the simpler MLN-D model (the same model with random-effect variance matrix Σ constrained to be diagonal) were so much simpler to compute numerically that in all LMGs this was the method of choice for exploratory data analyses with alternative models (based on different sets of covariates $\underline{X}_{jg}$ and different thresholds `minsamp` for a (Geo, LMG) domain to be included in the parameter estimation step. To get an idea of the relative run-times for the two types analysis, see Table 21 for results on LMGs 3–10 and 73 (NE Region) analyzed in this way. (The respective numbers of jurisdictions with ACS 2015-2019 sample in these LMGs were: $989, 3309, 4004, 742, 845, 3029, 3126, 1052$ and 1045.) In the first row of Table 21, run-times in hours are given for the *average* of the 4 MCMC chains (because the 4 chain computations are run in parallel), although this way of doing the comparison advantages the Bayes methods. The third row displays the *total* run-time for the 81 replicate-weight frequentist MLN-D estimation and prediction runs, and the second row gives the additional times for each LMG beyond the MLN-D intializations for a single fit using the original ACS weights.

Table 21 shows that the time required to produce estimates in the MLN-F models are definitely faster for the frequentist AGQ calculations than for the Bayesian MCMC. However, the Bayesian MCMC runs produce posterior variances as MSPEs as part of the same computation, while according to the current method, MSPEs from frequentist estimates must be produced from 81 separate model-fitting calculations using SDR weight-replicates. The time required for 81 frequentist MLN-F calculations would not have been faster than the Bayesian MCMC calculations. The Bayesian approach was ultimately employed only for the most data-rich LMGs with Jurisdiction geography. In less data-rich LMGs, it was found that the MLN-F fitted models were not significantly different

Table 21: Run-times (hrs) for Bayesian and frequentist MLN runs computing parameter estimates, outcome (Geo, LMG) predictions and MSPEs in Jurisdiction Geography for selected LMGs.

		Language Minority Group Number								
Method	Model	3	4	5	6	7	8	9	10	73
Bayes	MLN-F	.25	3.50	3.20	.12	.21	1.60	2.00	.20	1.6
Freq	MLN-F	.10	.43	.41	.07	.12	.67	.65	.11	.17
Freq	MLN-D	.02	.14	.12	.01	.01	.12	.18	.02	.04

Notes. Bayes MLN-F: average of 4 MCMC runs; Freq MLN-D: total for 81 replicate runs; Freq MLN-F: from MLN-D initialization to convergence. LMG numbering 1–73 described on page 4.

from the MLN-D fits, and the tremendously greater speed of MLN-D frequentist computations resulted in a preference for that method on all but the largest LMGs.

All but one of the frequentist MLN-F fits in Table 21 used the standard `R` nonlinear minimization by function `nlm` of the AGQ numerically integrated log-likelihood and analytical (numerically integrated) gradient. The one exception was LMG 6, where `nlm` and similarly `optim` did not run properly and had to be restarted and completed with a different method, optimal-steplength gradient descent. This kind of unaccountable non-convergence requiring a re-start with modified optimization method did occur multiple times in our testing of frequentist software for MLN-F model fitting, and this was another reason we chose a Bayesian method of MLN-F model-fitting in the LMGs that were sufficiently large to require it. The run-times for frequentist MLN-F model fits could be made smaller, perhaps by 10-50% without much change in predictive performance, by tinkering with convergence criteria. However, we are not sure that the optimization could be made fully automatic without the need for re-starts involving intermediate convergence failures. But probably the Bayesian MCMC runs could also be made quicker using Metropolis-Hastings MCMC with normal posterior proposal distributions starting from MLN-D fits.

These purely computational considerations are relevant because the estimates and predictions from the same MLN-F models by the two analytical approaches were substantially the same. The estimates are not identical, but typically differ only in the second or third decimal place, occasionally in a very few cases in the first decimal place only when the estimate SEs are large (> 0.7). The sense of estimation accuracy is considerably different in the Bayesian and frequentist computations. In Bayesian MCMC, 'convergence' represents a confirmation of indicators of approximate stationarity tested by approximately equal distributions for the 4 parallel simulated Markov chains. The sense of convergence in the frequentist AGQ likelihood maximization is based on checks for smallness of the calculated gradients of log-likelihood components, which typically must be $< 10^{-6}$ of the log-likelihood for the calculation to stop. In the example LMGs of Table 21, with MLN-F parameter dimension from 13 to 19, fully half the computation time is spent in iterations reducing the gradient sizes by the last factor of 0.01 while the log-likelihoods change by at most 0.01. So the standard

Table 22: Differences between MLN-F frequentist and Bayesian predictions of outcome counts CIT, LEP, ILL by jurisdiction within selected LMGs. Columns 1-3 are mean absolute differences of predictions within jurisdictions with MLN-D predicted counts ≤ 30. Columns 4-6 are mean absolute differences of log counts for jurisdictions with MLN-D predicted count > 30.

	Predicted count ≤ 30			Predicted count > 30		
	CIT	LEP	ILL	CIT	LEP	ILL
LMG 3	0.024	0.075	0.046	0.002	0.009	0.012
LMG 4	0.022	0.032	0.020	0.002	0.006	0.014
LMG 5	0.014	0.026	0.011	0.001	0.006	0.020
LMG 6	0.028	0.098	0.059	0.002	0.012	0.016
LMG 7	0.058	0.125	0.060	0.006	0.033	0.199
LMG 8	0.020	0.026	0.025	0.002	0.010	0.047
LMG 9	0.015	0.033	0.025	0.001	0.007	0.022
LMG 10	0.029	0.056	0.060	0.003	0.007	0.015
LMG 73	0.011	0.040	0.016	0.001	0.015	0.014

Table 23: Measures of difference between MLN-F and MLN-D frequentist predictions of outcome counts CIT, LEP, ILL by jurisdiction within selected LMGs. Columns 1-3 are mean absolute differences of predictions within jurisdictions with MLN-D predicted counts ≤ 30. Columns 4-6 are mean absolute differences of log counts for jurisdictions with MLN-D predicted count > 30.

	Predicted count ≤ 30			Predicted count > 30		
	CIT	LEP	ILL	CIT	LEP	ILL
LMG 3	0.066	0.276	0.275	0.005	0.029	0.061
LMG 4	0.050	0.144	0.087	0.004	0.021	0.044
LMG 5	0.042	0.205	0.032	0.004	0.037	0.044
LMG 6	0.059	0.102	0.312	0.005	0.011	0.049
LMG 7	0.173	0.408	0.109	0.015	0.119	0.422
LMG 8	0.049	0.081	0.039	0.004	0.024	0.070
LMG 9	0.111	0.133	0.132	0.009	0.021	0.104
LMG 10	0.115	0.276	0.313	0.009	0.031	0.079
LMG 73	0.037	0.327	0.097	0.003	0.059	0.059

of accuracy is considerably more stringent for the frequentist analysis. In any case, it is the outcome predictions [CIT, LEP, ILL counts within (Geo, LMG)] that must be compared to establish practical equivalence between the Bayesian and frequentist methods of analysis in this project. This comparison is done in Tables 22 and 23.

Tables 22 and 23 respectively quantify the differences between frequentist versus Bayes MLN-F predictions and between frequentist MLN-F versus MLN-D predictions, for (Geo, LMG) outcome counts in the categories CIT, LEP, and ILL. Both tables quantify these prediction differences by separate measures for small and large counts. For jurisdictions in which counts for a specific outcome category are predicted (by frequentist MLN-D) to be ≤ 30, mean absolute differences of predicted counts are shown in the first 3 columns of both tables. For jurisdictions with predicted category counts > 30, relative errors of prediction are quantified by mean absolute differences of (natural) logarithms of predicted counts, in the last 3 columns of both tables. The discrepancies in MLN-F frequentist versus Bayes predictions are very small, both in their own right and by comparison with the corresponding differences in MLN-F versus MLN-D frequentist predictions. We argue in some detail in Section 3.5 of the main text of this report that the MLN-D model is aleady quite effective for estimation and prediction of domain outcome-category population counts in the main text. Yet a comparison between the corresponding entries of Tables 22 and 23 shows that the MLN-F outcome-count predictions of the same model by the radically different frequentist-AGQ and Bayesian-MCMC computational techniques are closer together by a factor of 3 or more than the frequentist predictions between the slightly different MLN-F and MLN-D models.

F Prediction Diagnostics

This section describes a simple systematic approach to the assessment of quality of predictions made by different models and different variants of the same model in order to inform modeling decisions and validate that the models chosen for VRA prediction fit adequately. The quantities that we are interested in predicting are: N_{jg}^A where $j = 1, \ldots J$ indexes geographic unit (jurisdiction, AIA or ANRC), g indexes LMG, and $A = V, C, L, I$ respectively denote the nested decreasing categories of VOT, CIT, LEP, ILL.

Our data come in the form of observed sample sizes n_{jg}^A (of which we regard n_{jg}^V as fixed and known and disregard the rest) and survey-weighted estimates $\hat{N}_{jg}^A$. These survey-weighted estimates are noisy, but they are supposed to be unbiased estimates of their targets N_{jg}^A based on ACS weighting, and this is not a bad assumption. We recode the survey-weighted data into scaled sample-sizes $Y_{jg}^A = n_{jg}^V \cdot \hat{N}_{jg}^A / \hat{N}_{jg}^V$ that we assume to satisfy (jointly in $A = V, C, L, I$) a mixed-effects multinomial generalized linear model (MLN or DM) in terms of covariate vectors X_{jg}. As in Section C, we focus attention on a single LMG g and suppress the notation g.

Corresponding to a fitted generalized linear model, with a particular covariate specification and set of model estimates $\hat{\vartheta}$ consisting of coefficients $\hat{\beta}^{(k)},\ k = 1, 2, 3$ and 3×3 mixed-effect variance $\hat{\Sigma}$ or dispersion parameters $\hat{\underline{\tau}}$, resulting in a set of model-based predictions $\tilde{N}_j^A = \tilde{N}_j^A(\hat{\theta})$.

We are interested in comparing many model specifications and associated predictions $\tilde{N}_j^A$. It is not helpful to compare these directly with the survey-weighted estimates $\hat{N}_j^A$ because the latter are so noisy. On the other hand, aggregated versions of the discrepancies $\tilde{N}_j^A - \hat{N}_j^A$ across indices j are supposed to have mean approximately 0 for a properly specified model. (That is, the mean would be exactly 0 if the predictions $\tilde{N}_j(\theta)$ were based on the true θ in a correct model.) So we propose to use diagnostic statistics of the form

$$\Delta^A(\hat{\theta}) \equiv \sum_{j=1}^{J} x_j \left(\tilde{N}_j^A - \hat{N}_j^A\right) , \qquad A = C, L, I$$

where the quantities x_j may be functions or components of the covariates X_j or indicators of the indices j falling into designated subsets defined by sample sizes n_j^V. If we are comparing two models with different specifications, estimators $\hat{\theta}$ and predictions $\tilde{N}_j^A(\hat{\theta})$, we will tend to prefer the model with systematically smaller absolute values of the statistic $\Delta^A = \Delta^A(\hat{\theta}, \mathbf{x})$.

One systematic set of diagnostic comparisons would use a number of different sequences x_j defined from model covariates, other potential covariates, and geographic size measures for which n_j^V may be a proxy. As implemented in Section 3.6, the quantities $\Delta^A(\hat{\theta}, \mathbf{x})$ are usefully tabulated within VOT sample-size classes since somee important differences in the performance of model predictions were found to relate strongly to domain sample size. However, we found it useful also to tabulate $\Delta^A(\hat{\theta}, \mathbf{x})$ for jurisdictions aggregated by intervals of values of certain key covariates.

In applying this diagnostic approach to data from the 2021 VRA Section 203 project, the survey-weighted estimates $\hat{N}_j^A$ are drawn directly from arrays tabulated from ACS 2015-2019 data, and the predictors $\tilde{N}_j^A(\hat{\theta})$ are the model-based results from MLN or DM models, with specified rules for choosing the covariates (the $\beta^{(k)}$ components chosen not to be structural zeroes, with indices $\mathcal{I}_k$, and further rules defining the minimum sample-size threshold for including geographic units in the LMG-specific fitting of unknown statistical parameter. The results of these diagnostic comparisons, and the consequent modeling decisions made for the VRA 2021 production of estimates used in Section 203(b) determinations, are described in Sections 3.4 and 3.6 of this report.

A somewhat different diagnostic is applied to the overall assessment of prediction quality in Section 3.6.1. In that Section, the objective is to document the sense in which the model-based predictions with covariates improve on corresponding model-based predictions without covariates. Both the predicted counts $\tilde{N}_{jg}^A$ and their direct-estimated comparators $\hat{N}_{jg}^A$ increase with geographic-unit VOT population size $\hat{N}_{jg}^V$, and are for that reason highly correlated, and the predictions $\tilde{N}_{jg}^A$ also make use of the direct counts $\hat{N}_{jg}^A$ through the empirical-Bayes conditioning on $\hat{Y}_{jg}^A$. To learn whether the model made effective use of covariates, another approach is to compare *synthetic* model-based predictions $\tilde{\pi}_{jg}^A$ — defined from the model as though no there were no sampled data in domain (j, g) — in predicting the observed (direct-estimated) ratios $\hat{N}_{jg}^A/\hat{N}_{jg}^V$. The diagnostic used in Section 3.6.1 is the LMG-specific correlations between these predicted and observed rates, computed across weight-replicates.