

# Essays on the Measurement of Public Opinion

Alissa F. Stollwerk

Submitted in partial fulfillment of the  
requirements for the degree  
of Doctor of Philosophy  
in the Graduate School of Arts and Sciences

**COLUMBIA UNIVERSITY**

2017

©2017

Alissa F. Stollwerk

All Rights Reserved

# ABSTRACT

## Essays on the Measurement of Public Opinion

**Alissa F. Stollwerk**

The study of public opinion has become increasingly central to our understanding of American politics. What the American public believes, why it holds those beliefs, and whether or not those beliefs matter have become essential questions that guide our understanding of how American democracy functions. In order to answer these questions, however, it is important to consider the tools we use to measure public opinion accurately and reliably and to understand the substantive applications and limitations of those tools. This dissertation is composed of three essays that consider important questions in public opinion measurement today. The first considers how the technique of multilevel regression with poststratification (MRP) performs on polling data collected using area-based cluster sampling techniques. While MRP has been a boon to researchers with limited resources, it must still be examined to understand its strengths and shortcomings. The second paper uses two datasets to look at the measurement of scales of political values over time, focusing on both individual and state-level measures, and discusses implications of these results for larger debates around the measurement of partisan sorting and polarization. The third paper turns to the question of social desirability bias in polling. Specifically, it uses list experiments to look at whether survey respondents answer truthfully when asked about support for same-sex rights. These papers all aim to shed light on recent innovations in the measurement of public opinion and illustrate how we can use these innovations to improve our understanding of American public opinion.

# Table of Contents

List of Figures	iv
List of Tables	vii
<b>I Dissertation Chapters</b>	<b>1</b>
<b>1 Introduction</b>	<b>2</b>
<b>2 Estimating Subnational Opinion with Cluster-Sampled Polls: Challenges and Suggestions</b>	<b>5</b>
2.1 Introduction . . . . .	5
2.2 Cluster Sampling . . . . .	8
2.3 MRP and Geography . . . . .	9
2.4 Potential Problems with Cluster Sampling and MRP . . . . .	11
2.4.1 Cluster Sampling in the Real World . . . . .	12
2.4.2 Design Effects and MRP . . . . .	13
2.4.3 Does the Number of Clusters Matter?: GSS, ANES, and the 1980 Sampling Frame . . . . .	16
2.4.4 Simulations . . . . .	21
2.5 Results and Discussion . . . . .	27
2.5.1 Face Validity: Who’s Liberal and Who’s Conservative? . . . . .	27
2.5.2 GSS, ANES, and the National Annenberg Election Study . . . . .	32

2.5.3	Modeling Additional Geographic Information . . . . .	38
2.6	Conclusion . . . . .	41
<b>3</b>	<b>The Polarization of Political Values: An Examination of the Heterogeneity of Support for American Core Values</b>	<b>43</b>
3.1	Introduction . . . . .	43
3.2	Political Values . . . . .	45
3.2.1	What Are Political Values and Why Do They Matter . . . . .	45
3.2.2	Homogeneity and Heterogeneity In American Political Values . . . . .	49
3.3	Individual Political Values . . . . .	55
3.3.1	Data and Measurement . . . . .	55
3.3.2	Individual Political Values Over Time . . . . .	57
3.4	Mapping Political Values in the States . . . . .	68
3.4.1	Data . . . . .	68
3.4.2	Estimating State-Level Opinion using MRP . . . . .	70
3.4.3	State-Level Political Values . . . . .	73
3.5	Discussion . . . . .	83
<b>4</b>	<b>Are Survey Respondents Lying about Their Support for Same-Sex Marriage? Lessons from a List Experiment</b>	<b>86</b>
4.1	Introduction . . . . .	86
4.2	Public Support for Same-sex Marriage . . . . .	88
4.3	Uncovering “True” Opinion . . . . .	91
4.3.1	Using List Experiments to Study Attitudes toward Gays and Lesbians . . . . .	92
4.4	Same-Sex Marriage List Experiment . . . . .	93
4.4.1	Experimental Design . . . . .	93
4.4.2	Evaluating the Design . . . . .	95
4.4.3	Results . . . . .	98
4.4.4	Considering Social Desirability by Subgroups . . . . .	100

4.5	A Further Inquiry . . . . .	107
4.6	Conclusion . . . . .	111
<b>II</b>	<b>Bibliography</b>	<b>112</b>
	<b>Bibliography</b>	<b>113</b>
<b>III</b>	<b>Appendices</b>	<b>120</b>
<b>A</b>	<b>Appendix to “Estimating Subnational Opinion with Cluster-Sampled Polls: Challenges and Suggestions”</b>	<b>121</b>
A.1	GSS, ANES, and Ideology Measures . . . . .	121
<b>B</b>	<b>Appendix to “The Polarization of Political Values: An Examination of the Heterogeneity of Support for American Core Values”</b>	<b>125</b>
B.1	Value Scales and Question Wording . . . . .	125
	B.1.1 ANES Cumulative File . . . . .	125
	B.1.2 2013 CCES Module . . . . .	126
B.2	Reliability of Values Scales . . . . .	127
	B.2.1 ANES . . . . .	127
	B.2.2 CCES . . . . .	128
B.3	Results of Analysis of Individual Values with a Single Equality Scale . . . . .	130
B.4	Analysis of Individual Values and Partisanship Overtime with the Inclusion of the 2012 ANES Internet Supplement Module . . . . .	131
B.5	State-Level Maps for Support for the Liberal Position of Each Values Scale . . . . .	143
B.6	Replication of State-Level Analysis without Including Presidential Vote Share in the Model . . . . .	148

# List of Figures

- 2.1 GSS, ANES, and EWM Correlations for the 1980s Sampling Frame . . . . . 19
- 2.2 GSS, ANES, and EWM Mean Absolute Errors for the 1980s Sampling Frame . . . . . 20
- 2.3 Simulation Results – Correlation Between MRP Estimates and True Values, by Sample Size and Strength of Predictor . . . . . 25
- 2.4 Simulation Results – Mean Absolute Error Between MRP Estimates and True Values, by Sample Size and Strength of Predictor . . . . . 26
- 2.5 MRP GSS Estimates for Democratic and Liberal – 2004 v. 2002-2006 Estimates . . . . . 30
- 2.6 MRP GSS Estimates for Republican and Conservative – 2004 v. 2002-2006 Estimates . . . . . 31
  
- 3.1 Support for Political Values Over Time (1984-2012) . . . . . 58
- 3.2 Standard Deviation of Political Values Over Time (1984-2012) . . . . . 59
- 3.3 Correlation of Political Values and Partisanship and Political Values and Ideology Over Time (1984-2012) . . . . . 61
- 3.4 Coefficient Plot of Estimated Predictors of Support for Moral Traditionalism (1986-2012) . . . . . 63
- 3.5 Coefficient Plot of Estimated Predictors of Opposition to Equality of Opportunity (1984-2012) . . . . . 64
- 3.6 Coefficient Plot of Estimated Predictors of Opposition to Egalitarianism (1984-2012) . . . . . 66
- 3.7 Coefficient Plot of Estimated Predictors of Support for Limited Government (1990-2012) . . . . . 67
- 3.8 Support of Traditional Moral Values, by State and Demographics . . . . . 77

3.9	Support of Limited Government and Individualism, by State and Demographics . . .	79
3.10	Opposition to Egalitarianism, by State and Demographics . . . . .	80
3.11	Opposition to Equality of Opportunity, by State and Demographics . . . . .	82
4.1	Support for Same-Sex Marriage (1994-2014) . . . . .	90
4.2	Multivariate Subgroup Analysis . . . . .	105
4.3	Support for Nondiscrimination Laws (1992-2014) . . . . .	108
B.1	Support for Political Values Over Time (1984-2012) – Single Equality Scale . . . . .	132
B.2	Standard Deviation of Political Values Over Time (1984-2012) – Single Equality Scale	133
B.3	Correlation of Political Values and Partisanship and Political Values and Ideology Over Time (1984-2012) – Single Equality Scale . . . . .	134
B.4	Coefficient Plot of Estimated Predictors of Opposition to Equality (1984-2012) . . .	135
B.5	Support for Political Values Over Time (1984-2012) – Including 2012 Internet Module	136
B.6	Standard Deviation of Political Values Over Time (1984-2012) – Including 2012 Internet Module . . . . .	137
B.7	Correlation of Political Values and Partisanship and Political Values and Ideology Over Time (1984-2012) – Including 2012 Internet Module . . . . .	138
B.8	Coefficient Plot of Moral Traditionalism (1986-2012) – Including 2012 Internet Mod- ule . . . . .	139
B.9	Coefficient Plot of Equality of Opportunity (1984-2012) – Including 2012 Internet Module . . . . .	140
B.10	Coefficient Plot of Egalitarianism (1984-2012) – Including 2012 Internet Module . .	141
B.11	Coefficient Plot of Limited Government (1990-2012) – Including 2012 Internet Module	142
B.12	Opposition to Traditional Moral Values, by State and Demographics . . . . .	144
B.13	Opposition to Limited Government and Individualism, by State and Demographics	145
B.14	Support for Egalitarianism, by State and Demographics . . . . .	146
B.15	Support for Equality of Opportunity, by State and Demographics . . . . .	147
B.16	Support of Traditional Moral Values, by State and Demographics . . . . .	150



B.17 Opposition to Traditional Moral Values, by State and Demographics . . . . . 151

B.18 Support of Limited Government and Individualism, by State and Demographics . . 152

B.19 Opposition to Limited Government and Individualism, by State and Demographics 153

B.20 Opposition to Egalitarianism, by State and Demographics . . . . . 154

B.21 Support for Egalitarianism, by State and Demographics . . . . . 155

B.22 Opposition to Equality of Opportunity, by State and Demographics . . . . . 156

B.23 Support for Equality of Opportunity, by State and Demographics . . . . . 157

# List of Tables

2.1	Design Effects in the ANES . . . . .	15
2.2	Simulation Results – Pearson’s Product-Moment Correlations with True Value . . . .	24
2.3	Simulation Results – Mean Absolute Error Between Estimates and True Value . . . .	24
2.4	Pearson’s Product-Moment Correlations – GSS and NAES . . . . .	33
2.5	Pearson’s Product-Moment Correlations – ANES and NAES . . . . .	35
2.6	Mean Absolute Error – GSS and NAES . . . . .	37
2.7	Mean Absolute Errors – ANES and NAES . . . . .	37
2.8	Comparing State and Congressional Models – Correlation . . . . .	39
2.9	Comparing State and Congressional Models – Mean Absolute Error . . . . .	40
3.1	Value Scale Ranges (CCES Data) . . . . .	74
3.2	Value Scale Correlations (CCES Data) . . . . .	75
4.1	Observed Data – Same-sex Marriage List Experiment . . . . .	96
4.2	Covariate Balance – Same-sex Marriage List Experiment . . . . .	97
4.3	Estimated Support for Same-Sex Marriage . . . . .	98
4.4	Which Way Does Social Desirability Work? . . . . .	101
4.5	Multivariate Analysis of the List Experiment and the Direct Question . . . . .	103
4.6	Observed Data – ENDA List Experiment . . . . .	110
4.7	Estimated Support for Employment Non-Discrimination . . . . .	110
A.1	Pearson’s Product-Moment Correlations – GSS and EWM . . . . .	122

A.2	Mean Absolute Error – GSS and EWM . . . . .	122
A.3	Pearson’s Product-Moment Correlations – ANES and EWM . . . . .	123
A.4	Mean Absolute Error - ANES and EWM . . . . .	124
B.1	Cronbach’s Alphas for Value Scales (ANES) . . . . .	128
B.2	Principle Components Factor Analysis for Value Scales (ANES) . . . . .	129
B.3	Cronbach’s Alphas for Alternate Specifications of Value Scales (ANES) . . . . .	130
B.4	Cronbach’s Alphas for Value Scales and Alternate Specifications (CCES) . . . . .	130
B.5	Principle Components Factor Analysis for Value Scales (CCES) . . . . .	131
B.6	Correlations Between Model Specifications With and Without Presidential Vote Share (CCES Data) . . . . .	148
B.7	Value Scale Range (CCES Data) . . . . .	148
B.8	Value Scale Correlations (CCES Data) . . . . .	149

# Acknowledgments

My dissertation would not have been possible without the support, encouragement, instruction, camaraderie, and wisdom of more people than I can name here.

First, I would like to thank my main advisors throughout my years at Columbia: Bob Shapiro, Bob Erikson, and Justin Phillips.

Bob Shapiro was one of my first instructors at Columbia, and his experience and scholarship in the field of public opinion have been invaluable. His focus on the fundamentals of both political science and statistical scholarship has served as an essential grounding for my own research and approach to our field.

As a leader in the field of state politics and public opinion, Bob Erikson has been an important mentor as I focused on the study of state and local public opinion. I could always count on him to ask questions about my research that reflected both the longer arch of the literature as well as new challenges in the field, and all the while appropriately balancing the two.

During my time at Columbia, I have been lucky to call Justin Phillips an advisor, a co-author, and a friend. I am thankful for the opportunities I have had to work with him and to grow professionally as a researcher and teacher. His advice not only toward this dissertation but also toward the larger undertakings of my academic and professional life has been helpful throughout my graduate education.

More broadly, I'd like to thank all of the professors in the Columbia Political Science Department who have helped shape me into the researcher that I am today. Andy Gelman has provided expert advice in my methodological training, and I'm thankful for the experience to have learned from him, as well as for his serving on my dissertation committee. Don Green, Shigeo Hirano, Ira Katznelson, Jeff Lax, Dorian Warren, and Greg Wawro have all been wonderful teachers and

helpful ears of advice, always ready with an open door to hear more about my research and help me work through problems. I also would like to thank Kay Achar, Betty Howe, and Michael Scott for all they do to make sure the department runs well behind the scenes.

I'd also like to thank several professors at other institutions who have taken time to provide feedback on my research, including Adam Berinsky, Alan Gerber, Jim Gibson, Bill Jacoby, and Chris Warshaw. A special thank you to Pat Egan, who both provided wonderful advice as I was applying to graduate school and then did me the honor of serving on my dissertation committee.

Graduate school is admittedly a long and challenging road, and I would not have learned as much as I did or enjoyed my years at Columbia without the friendship and intellectual community of my peers. It is difficult to name everyone who has shared this journey, but I would like to give particular thanks to Erica Borghard, Alexandra Cirone, Kate Cronin-Furman, Kevin Elliott, Albert Fang, Carolina Ferrerosa-Young, Andy Guess, Sarah Khan, Patricia Kirkland, Lucas Leemann, Michael Schwam-Baird, Amber Spry, Pavithra Suryanarayan, and Steven White.

Outside of graduate school, I would like to thank my colleagues who work in political data and analytics. Through long nights and difficult campaign cycles, they have taught me how to apply the skills I learned in graduate school directly in the arena I had previously studied to help the candidates and causes that we care about. A special thank you to Aaron Strauss, who first suggested I start on this crazy journey, and to other academics who have crossed into the political space and always been there to advise me, especially David Nickerson, Yair Ghitza, and Kym Neck.

My final thanks go to my family, without whom I truly could not have done this. To my parents, Susan and David Stollwerk, thank you for instilling in me a love of learning and knowledge that allowed me to take this path, for always asking how my dissertation was going, and for all the support (and free babysitting) you provided, especially near the end. To my husband, Jason Cohen, who first met me a few months before I started graduate school, thank you for being there every step of the way with encouragement, good humor, and a nudge to keep dissertating, all while you finished medical school and became the "other type" of doctor. I could not have taken this journey without you. And last but not least, to our two daughters, Abigail and Dara,

thank you for bringing me joy and distraction in the present while inspiring me with what the future will bring.

To my husband, Jason

To my daughters, Abigail Cara and Dara Sophie

## **Part I**

# **Dissertation Chapters**



# Chapter 1

## Introduction

The study of public opinion has become an essential component of political science, as scholars work to understand what the public thinks, whether elected officials are responsive to public opinion, and what this means for our democracy. In order to have full faith in our conclusions, we must begin by focusing on the measurement of public opinion and the validity of our data and methods. This dissertation seeks to explore these questions of measurement in three key areas of public opinion research, with a focus on key and advances in survey methodology.

An important recent innovation in the field of public opinion has been the use of modeling and simulation to gain new insights into public opinion data. One specific application of this is a technique known as multilevel regression with poststratification (MRP), which has become widely used in political science to estimate subnational opinion from national polls. This method takes into account both demographic and state-level effects and has greatly expanded knowledge of public opinion at the subnational level. By allowing researchers to estimate subnational opinion from a national poll, researchers can apply old datasets to new questions, addressing research agendas that previously were not possible. While this innovative technique has allowed for important progress in our understanding of public opinion in smaller geographic areas, it has still not been tested in key polling environments. In this paper, I assess the use of MRP on two area-based, cluster-sampled polls, the American National Election Studies (ANES) and the General Social Survey (GSS). I discuss concerns that cluster sampling raises and demonstrate that MRP

can produce problematic results when used on a typical cluster-sampled poll. I evaluate several ways to improve upon MRP with cluster sampling. I find that adding a state-level predictor to the model, pooling surveys across years and sampling frames, and incorporating additional geographic information can all improve MRP's performance on cluster-sampled polls, but that caution should still be used when applying MRP to cluster-sampled data.

I next focus on the question of political values, which have become increasingly important in our understanding of American public opinion. While a growing body of quantitative work has investigated these values, these accounts are problematic because they assume that these political values are shared by most Americans. This paper argues that, like other measures of public opinion such as ideology and various issue-specific attitudes, support for political values varies across demographic and geographic subgroups in meaningful ways. This paper first uses a three-decade time series to examine how individual-level support for values has varied within the population, with a particular focus on how this variation has changed over time. Next, applying advances in the study of state and local public opinion, this paper uses multilevel regression with poststratification (MRP) and an original CCES survey module to create estimates of political values, including moral traditionalism, equality of opportunity, egalitarianism, and limited government, at the state level. This paper shows that Americans disagree on political values in predictable ways along common demographic and geographic lines. Furthermore, this paper shows that while political values previously were more shared across partisan lines, they have increasingly become entwined with the two major political parties, reflecting the increased partisan sorting and political polarization that have characterized the last three decades of American politics.

Last, I consider the question of how to measure public opinion when social desirability bias may prevent respondents from answering questions honestly. I address this question in the context of same-sex rights, an at times controversial political issue that has seen rapid evolution in the public's views on it over the last two decades. Public opinion polls consistently show that a growing majority of Americans support same-sex marriage. Critics, however, raise the possibility that these polls are plagued by social desirability bias, and thereby may overstate public support

for gay and lesbian rights. We test this proposition using a list experiment embedded in the 2013 Cooperative Congressional Election Study. List experiments afford respondents an anonymity that allows them to provide more truthful answers to potentially sensitive survey items. Our experiment finds no evidence that social desirability is affecting overall survey results. If there is social desirability in polling on same-sex marriage, it pushes in both directions. Indeed, our efforts provide new evidence that a national opinion majority favors same-sex marriage. To evaluate the robustness of our findings, we analyze a second list experiment, this one focusing on the inclusion of sexual orientation in employment nondiscrimination laws. Again we find no overall evidence of bias.

## Chapter 2

# Estimating Subnational Opinion with Cluster-Sampled Polls: Challenges and Suggestions

### 2.1 Introduction

Over the last two decades, interest in state public opinion has increased as political scientists have consistently shown that state public opinion is meaningful and that it informs public policy and elections at the state level. Unfortunately, polls that measure public opinion at the state level are few and far between, especially when one is looking for consistency across the fifty states. In response to this paucity of polling data, researchers have used a variety of strategies to derive subnational public opinion estimates from national polls, primarily applying statistical techniques such as disaggregation and simulation. As simulation techniques become more popular, researchers are using simulation methods to generate state-level opinion estimates from a wide range of public opinion polls, many with different sampling methods. This paper examines how one simulation method, multilevel regression with poststratification, performs on cluster-sampled polls, focusing on how a geographically-clustered sample may impact the accuracy and efficiency of sub-national opinion estimates.

As early as the 1960s, political scientists attempted to use simulation, using the demographic characteristics of the survey respondents and of the populations of each state to generate state-level estimates of public opinion from national polls (Pool et al. 1965, Weber et al. 1972, Weber and Shaffer 1972). These early simulations, however, were criticized for ignoring potential geographic variation. In these studies, demographics were the only variables, and they were estimated using simple linear or fixed effects.

Erikson, Wright, and McIver's work on state politics directly criticized this assumption, arguing that geography mattered. In their book *Statehouse Democracy*, they showed that state public opinion was not simply an aggregation of the opinion of demographic categories, but contained distinct information connected to the geography itself. In fact, they argued that "most of the variance in state partisanship and state ideology is due to state-to-state differences that cannot be accounted for by the demographic variables measured here" (61). Rather than using simulation methods, which relied only on demographics, they used a technique known as disaggregation, combining several polls that one could then break down by state, to demonstrate this geographic variation. By emphasizing the importance of geography in forming state-level opinion, they showed that geographic effects could not be ignored.

Recently, however, political scientists have returned to simulation methods and developed more sophisticated techniques that utilize meaningful geographic information as well as the traditional demographic variables. Gelman and Little (1997) developed a multilevel modeling technique, known as multilevel regression with poststratification, or MRP, which creates reliable and valid measures of public opinion at the state level from a single national poll of roughly 1500 respondents. Unlike earlier models of simulation, MRP relies on both demographic and geographic effects in simulating public opinion. MRP also employs partial pooling across subgroups, and Park, Gelman, and Bafumi (2006) showed that partial pooling across states outperformed both running individual models for each state (no pooling) and a model with complete pooling in which only demographic information was used in the modeling phase. In their work, Park et al. test MRP by using it to reproduce the same state-level ideology measures that disaggregation produced in *Statehouse Democracy*. They find that their measures are nearly identical, and show

that in areas where one has seen change over time, such as in a rightward shift in Republican partisan identification during the time series in question, MRP outperforms disaggregation. Lax and Phillips (2009b) further test how different forms of MRP perform against disaggregation and conclude even a simple MRP model often outperforms disaggregation in producing valid state-level estimates.

MRP is quickly becoming an approachable and accepted technique that researchers can use to estimate both state and local public opinion. This growing acceptance, however, warrants some caution. As researchers learn how to conduct MRP and use it on a wide range of polls, it is important that we understand how MRP functions on polls that rely on different sampling methods. Specifically, both Park, Gelman, and Bafumi (2006) as well as Lax and Phillips (2009b) evaluate MRP using polls that rely on a full probability sample. As MRP gains in popularity, however, political scientists will be motivated to use it on two of social science's largest surveys, the American National Election Study (ANES) and the General Social Survey (GSS).<sup>1</sup> These polls, however, are conducted in person and thus rely on area-based cluster sampling rather than random digit dialing or some other completely random sampling procedure. By design, cluster sampling is a variation on standard random sampling, where different individuals in the population have different probabilities of being chosen in certain stages of the design. If all respondents within a state do not have an equal chance of being selected for the survey, as is the case in cluster-sampled polls, it is unclear how accurately MRP will be able to estimate state-level effects and state-level opinion more broadly. Especially given the importance of geographic indicators in predicting state-level opinion discussed above, it is necessary to understand how a nonrandom sample based on geographic location will impact simulation methods such as MRP.

This paper examines how MRP performs on cluster-sampled polls, specifically focusing on the GSS and the ANES. After discussing why cluster sampling may impact state-level estimates produced with MRP, the paper will test MRP on the GSS and ANES and propose some diagnos-

---

<sup>1</sup>Some political scientists have already used multilevel modeling to develop subnational estimates from these surveys, including Berkman and Plutzer (2006), Enns and Koch (2013), and Pacheco (2013). While Enns and Koch as well as Pacheco treat the ANES and GSS similarly to other polls, Berkman and Plutzer anticipate that cluster-sampled polls may be problematic and drop a handful of states with unusual clusters from their analysis. Others, such as Brace et al. (2002), have used disaggregation on the GSS to generate state-level estimates.

tics and corrections to perform when using MRP on cluster-sampled polls.

## 2.2 Cluster Sampling

Two of the longest running and most respected polls in political science, the American National Election Study and the General Social Survey, both rely on cluster sampling rather than random digit dialing or any other full probability sampling method. Given the importance of these two polls in the study of American politics, it is important to understand how cluster sampling works and why it is used for some surveys.

Cluster sampling has been used by major survey houses for several decades. When based on geographic location, this practice may also be known as area sampling. The GSS has used cluster sampling since it fielded its first large poll in 1972, and the ANES has done so since its early polls in the 1950s. At the first level of sampling, large, pre-defined regions are sampled from the country as a whole. Then, in successive rounds, increasingly smaller geographic areas are chosen from within those clusters, eventually narrowing down to the household level and then the individual level within the household. Cluster sampling thus can include certain aspects of random sampling within the chosen clusters, but it is different from simple random sampling because not every respondent in the population of interest has an equal chance of being chosen throughout the design. Rather, those in the selected clusters have a higher probability of being selected and those outside of those clusters have zero probability of being included after the initial clusters have been sampled. While the GSS and ANES employ slightly different sampling techniques and have also both changed their techniques as new technologies and population listings have become available, the general principle remains the same. A national frame of clusters is designed, typically guaranteeing extra or automatic weight to the country's largest population centers. Increasingly smaller clusters nested within these larger clusters are chosen until only certain neighborhoods are left, and it is from these neighborhoods, often as small as 300 households, that respondents will be chosen.

Cluster sampling has many advantages, the most important of which are that it decreases

both the cost and the time of conducting in-person interviews by selecting respondents clustered within specific, smaller geographic areas. Despite the lack of purely random sampling, cluster sampling has long been considered a valid and respected polling technique, and cluster-sampled polls have even been used to develop public opinion estimates for sub-national units. Brace, Sims-Butler, Arceneaux, and Johnson (2002) pool and then disaggregate the GSS over a 25-year period from 1974 to 1998, showing that their estimates hold up to a variety of stability and reliability tests. While their work is promising, they do not directly test their estimates against similar measures that are already viewed as valid, such as state-level polls or national polls designed to be representative at the state level, leaving the accuracy of measures derived through disaggregation of cluster-sampled polls an open question. Furthermore, Brace and his co-authors pool 25 years of the GSS to come to this conclusion, creating such a large sample that it is easy to see why they could defend disaggregation of the GSS, despite problems one might expect because of cluster sampling. As they note, the GSS has changed its sampling frame periodically, and pooling over 25 years would have increased the number of clusters dramatically compared to using a single GSS poll. Given that this pooling over 25 years is not a practical research design for scholars interested in changes over time or questions that were not asked consistently over long periods of time, it is important to investigate the accuracy of subnational estimates derived from a single cluster-sampled poll.

### 2.3 MRP and Geography

To understand how MRP may behave on a cluster-sampled poll, it is important to understand exactly how MRP simulates subnational public opinion and how a nonrandom geographic sample may impact this simulation. Following the model tested by Lax and Phillips (2009b) to estimate public opinion at the state level, I begin by modeling individual responses to public opinion questions as a function of each individual's demographic information and state of residence. Specifically, the model is written out below for each individual  $i$ , with index markers  $j$  for race/gender combination,  $k$  for age,  $l$  for education level, and  $s$  for state.



$$Pr(y_i = 1) = \text{logit}^{-1}(\beta^0 + \alpha_{j[i]}^{\text{race,gender}} + \alpha_{k[i]}^{\text{age}} + \alpha_{l[i]}^{\text{edu}} + \alpha_{s[i]}^{\text{state}}) \quad (2.1)$$

Each of the terms after the intercept is modeled itself based on a normal distribution with a mean of zero and an estimated variance. Race/gender is either a four or six category variable, depending on the model specification and the individual data. Age and education are both four-category variables. State is modeled based on the region the state is in and some state-level variance. In some models, as will be discussed further in the results section, I model the state effect based on both region and an aggregate state level measure, such as state presidential vote for the Democratic candidate, and show that this improves the performance of the model. Finally, region is modeled as a random effect as well.

$$\alpha_j^{\text{race,gender}} \sim N(0, \sigma_{\text{race,gender}}^2), \text{ for } j = 1, \dots, 6 \quad (2.2)$$

$$\alpha_k^{\text{age}} \sim N(0, \sigma_{\text{age}}^2), \text{ for } k = 1, \dots, 4 \quad (2.3)$$

$$\alpha_l^{\text{edu}} \sim N(0, \sigma_{\text{edu}}^2), \text{ for } l = 1, \dots, 4 \quad (2.4)$$

$$\alpha_s^{\text{state}} \sim N(\alpha_{m[s]}^{\text{region}} + \beta^{\text{presvote}} \cdot \text{presvote}, \sigma_{\text{state}}^2), \text{ for } s = 1, \dots, 50 \quad (2.5)$$

$$\alpha_m^{\text{region}} \sim N(0, \sigma_{\text{region}}^2), \text{ for } m = 1, \dots, 4 \quad (2.6)$$

These demographic factors do correlate strongly with ideology, partisanship, and presidential vote, as well as with other attitudes that I estimate in this paper. While this model is quite simple, such a standard model has been shown to perform quite well (Park, Gelman, and Bafumi 2006, Lax and Phillips 2009b, 2012). Thus, it is appropriate to use this model to assess how well MRP performs on cluster-sampled polls.

Once opinion is estimated, the next step is to post-stratify the estimated opinion by state-level population. Specifically, the above model will have generated 4,800 combinations of different demographic and state values.<sup>2</sup> I use the “1-Percent Public Use Microdata Sample” from the US

---

<sup>2</sup>This presumes the model with 6 race/gender categories, 4 age and education categories each, and 50 state cate-

Census to learn the number of people in each state,  $N_c$ , in each demographic cell type,  $c$ . From here, I can generate state-level public opinion estimates by weighting the opinion prediction in each cell,  $\theta_c$ , according to the state's population:

$$y_{state[s]}^{MRP} = \frac{\sum_{c \in s} N_c \theta_c}{\sum_{c \in s} N_c} \quad (2.7)$$

This estimation technique thus relies on geographic information both in the simulation phase and in the poststratification phase. The question I consider, then, is how this geographic information, particularly as used in the simulation phase, may impact the validity of state-level estimates of public opinion produced through MRP on cluster-sampled polls.

## 2.4 Potential Problems with Cluster Sampling and MRP

For MRP to produce valid estimates, it presumes that survey respondents are fairly representative of their geographic-demographic type, an assumption that is reasonable when respondents are selected randomly and a poll has an appropriately large sample size. Cluster sampling, however, alters how sampling occurs within states in a non-random way. If geography matters not only because of interstate variation but also because of intrastate variation, then the accuracy and efficiency of subnational opinion estimates, whether generated through disaggregation or MRP, could be impacted by cluster sampling. Intuitively one would expect that clusters could be unrepresentative of the states they are within, especially in states with diverse populations that are often geographically sorted by ideology and partisanship.

Erikson, Wright, and McIver (1993) note this potential problem in *Statehouse Democracy*, citing it as the reason that they chose to use the CBS/NYT poll as opposed to the ANES and GSS. Unlike disaggregation, however, one might expect MRP's use of poststratification to correct at least partially for any demographic unrepresentativeness of a given cluster (as Lax and Phillips 2009b suggest). On the other hand, MRP might not produce accurate measures of geographic effects in the simulation phase if the sampled respondents are not representative of their state's population,

---

gories. The model that excludes Hispanic as a racial category has 3,200 demographic-geographic types.

and it may even exaggerate that unrepresentativeness. The validity of cluster sampling might also vary among different cluster-sampled polls based on the number of clusters, the size of clusters, and other more idiosyncratic factors of a cluster-sampled poll.

To understand how cluster-sampled polls may be problematic for MRP, I first look at the clusters themselves, showing how clusters often seem unrepresentative of the state within which they are nested. I then measure design effects in some representative cluster-sampled polls, showing that intracluster homogeneity should be cause for further concern. Next, I take advantage of a quirk of the 1980s sampling frame used jointly by the GSS and ANES to examine how the number of clusters might impact the accuracy of MRP estimation. I then conduct simulations to demonstrate in what situations geographic cluster-sampling might be problematic for MRP and what techniques might make such sampling less problematic.

### **2.4.1 Cluster Sampling in the Real World**

It is illustrative to look at the geographic distribution of clusters in an area-sampled poll. While specific geographic information for respondents is typically withheld from the public release of a dataset to ensure the privacy of respondents, some polls do release more information than others. The ANES, for example, releases the state and congressional district for each respondent along with an identification code for the first two levels of sampling for each respondent – the strata and cluster. While not nearly as precise as having the zip code or census tract, this allows us to understand broadly the location of ANES respondents by state and congressional district.

Looking at the 2004 sample, with its almost 1200 respondents, 139 of the country's 435 congressional districts are represented, nested within only 29 of the 50 states, as well as an additional respondent from the District of Columbia. While MRP can handle imputation of missing states, this alone represents a tall order. Further, an inspection of the congressional districts that are represented might cause additional reason for concern. For example, in New Jersey this poll samples 3 people from the state's first congressional district and 33 from the state's second congressional district, with zero sampled from the state's other 11 districts. In 2004, the first congressional district was incredibly Democratic, encompassing the city of Camden. However, the second con-

gressional district, from which almost all of the New Jersey respondents are sampled, was a more Republican area, unusual in a typically Democratic-leaning state. Indeed if one looks at the party identification for the 36 New Jersey residents in this sample, one finds almost a perfectly even split of 18 Democrats, 2 independents, and 16 Republicans – more even than one might expect given the sampling but less Democratic than one would expect for New Jersey.

On the opposite side, one can look at Utah, a small state from which almost as many people – 32 – are sampled, with the majority coming from the more “blue” second congressional district, which includes Salt Lake City, and the rest coming from outside of it. Utah is considered one of the most Republican states in the country, with Bush receiving 71.5% of the vote in 2004. Nonetheless, this sample from Utah oversampled Democrats, including 18 Democrats, 4 independents, and 10 Republicans. An even more extreme example, perhaps, is that of Louisiana. This poll included 46 respondents from Louisiana, but they were all sampled from the conservative 4th congressional district in the northwestern part of the state.

One can show several more examples like these, and they illustrate a problem with any national poll – that one cannot disaggregate at the state level from a single poll to produce valid results. Nevertheless, these results are more worrisome for the use of MRP given that respondents are clustered within certain areas of a state. If these non-random areas are unrepresentative of the state, as they appear to be in these cases, we would expect MRP to produce inaccurate state random effects for these states. We would also expect state random effects to be more problematic in states with fewer clusters.

## 2.4.2 Design Effects and MRP

The above discussion explains why cluster-sampled polls may present a challenge for statistical techniques such as MRP. We can more precisely gauge the challenges posed by cluster-sampled polls by measuring the design effect for our variables of interest.

Specifically, one reason area-sampled polls are potentially problematic for MRP is because one would expect there to be intra-cluster homogeneity within these samples, distinguishing the people in the cluster from the other people in the state. In the case of area-based clusters

specifically, people who live in the same neighborhood are more likely to be similar to each other than to those who live in different neighborhoods on a host of variables, including race, economic class, and even political beliefs. This intracluster homogeneity makes the sample less precise than a simple random sample of the same size would be, increasing the variance of the estimates and decreasing the effective sample size of the poll. The design effect measures this lack of precision and loss of sample size. The design effect can be expressed as a function of the average number of cases selected in each cluster,  $b$ , and the intracluster correlation coefficient,  $\rho$ , which measures how similar respondents are within clusters for a given variable of interest.

$$def f \approx 1 + (b - 1) \times \rho \quad (2.8)$$

We can also define the design effect as the ratio between the variance of a statistic in a clustered sample and the variance of that statistic under the assumption of a simple random sample.

$$def f = \frac{Var_{clustered}(y)}{Var_{SRS}(y)} \quad (2.9)$$

A simple random sample, therefore, has a design effect of 1 by definition. A design effect greater than 1 reflects that there is intracluster homogeneity within the sample. For example, if a variable has a design effect of 2, that means that the sample variance will be twice as big as it would be under a simple random sample. It can also be interpreted to show that half as many respondents would have been needed to produce the same results under a simple random sample. Design effects are not a constant in a survey but rather are measured separately for each variable. For example, in a geography-based cluster sample, we might expect a great deal of intra-cluster homogeneity for a variable such as income or race, given current housing patterns in the United States. However, we might not expect a great deal of intracluster homogeneity in terms of gender or disability.

Design effects can thus be illustrative of where cluster sampling might prove problematic. In their work on large-scale multistage area probability designs, Harter et al. (2010) examine the design effects of several key variables in the 2006 GSS. As one would expect, they found design

effects to be a concern for several variables. The only variable they report as having a design effect of close to 1 is whether or not one has a happy marriage, with a design effect of 0.88. The other variables they report have design effects from 1.51 to 2.85, including those dealing with political attitudes such as legality of abortions and spending on social security.

Likewise, the ANES must also be examined for design effects. In his technical report on how to analyze ANES survey data, Matthew DeBell (2010) advises using a Taylor series approximation, what Harter et al. (2010) call linearization, to take the design effect of the cluster sample into account and produce accurate standard errors given a clustered sampling strategy. While the design effect is different for each variable, DeBell provides the average design effect for recent ANES surveys, which range from 1.21 in the 2004 pre-election study to 1.82 in the 2006 pilot study. In Table 1, I report the design effects for variables of interest in the 2004 pre-election ANES study, calculated using the Taylor series approximation.

Table 2.1: Design Effects in the ANES

ANES Variable	deff
Gender	1.5
Social Class (self-identified)	1.6
Race (6 category)	2.1
Party ID (7 point scale)	2.6
Ideology (7 point scale)	2.1
Abortion self-placement	1.3
Same-sex Couples Allowed to Marry	1.6
Government Spending on Welfare	1.5
Economy Better Since GW Bush Took Office	1.4

Note: This table lists the design effects for several key variables in the 2004 ANES. Design effects greater than 1 indicate a higher amount of intracluster homogeneity that decreases the effective sample size of the survey.

These results add to our concerns about intracluster homogeneity and how it may impact the ability of MRP to produce accurate estimates of public opinion data gathered through multistage area designs. If the presence of a design effect greater than one reflects a decrease in the effective sample size of the model, it might be the case that larger sample sizes are needed to conduct MRP on a cluster-sampled poll than one would need to conduct it on a randomly-sampled

poll. In addition, while knowledge of the design effect is typically used to reflect the lack of precision in a survey and increase the variance of the estimates, the presence of a design effect much greater than 1 might impact the accuracy of the coefficient estimates in addition to the variances when using MRP. If the respondents from a given state are clustered in one location with a high amount of intracluster homogeneity relative to existing heterogeneity at the state level, an MRP model may well produce inaccurate state random effects; the fewer clusters a poll contains in a given state, the more problematic this might be. One might even expect inaccurate demographic coefficients in the model, if the respondents are substantially unrepresentative of their demographic groups as well as their state population. Given these concerns, it is important to investigate how MRP functions when applied to cluster-sampled polls, and how one might improve MRP estimates based on these polls.

### **2.4.3 Does the Number of Clusters Matter?: GSS, ANES, and the 1980 Sampling Frame**

When dealing with cluster-sampled data broadly, we should expect that an increase in the number of clusters increases the efficiency of the sample and helps the poll more closely mimic a random sample (Harter et al. 2010). Therefore, in looking at how we might be able to use MRP on cluster-sampled polls, it is reasonable to ask whether the number of clusters matters, and how an increase in the number of clusters improves the accuracy of MRP estimates. While it can be difficult to tease out the role of an increased overall  $N$  versus the role of an increased number of clusters when pooling multiple surveys, data from the 1980s provides an opportunity to examine the separate role that each factor plays.

National sampling frames can be quite expensive to create, and after the 1980 Census, the GSS and ANES decided to collaborate and use the same national frame (Heeringa 1986, Gibson 1995). The organizations thus agreed on a common sample of primary sampling units (PSUs) but then each did their own individual sampling within those PSUs. The advantage for this study, however, is that while the GSS employed 84 PSUs, the ANES chose only to use 61 of those 84 PSUs. This is a substantive decrease in the number of clusters used in the ANES surveys during

this decade. Thus, with a constant sampling frame, we can compare the estimates that MRP produces on each survey with a “true value” to see if increasing the number of clusters sampled increases the validity of the survey. Given that the 1980 sampling frame overlaps nicely with the original data collected by Erikson, Wright, and McIver, I will use EWM ideology measures of liberal and conservative as a true value for comparison. Note that I drop Hawaii and Alaska from the analysis as the EWM scores do not contain measures for these states. I also drop the District of Columbia, which has consistently performed as an outlier, and Nevada, which Erikson, Wright, and McIver drop from their analyses as well given its unusually high liberal score.

In the following two figures, I summarize the results from a series of models that use MRP to estimate both liberal and conservative ideology from both the ANES and the GSS. I include models that use and do not use Democratic presidential vote share, and for both polls, I apply the models to 5 datasets – 1984, 1986, 1988, a pooled 1984 and 1986 dataset, and a pooled 1984, 1986, and 1988 dataset.<sup>3</sup> Figure 2.1 compares the correlation of each set of MRP state estimates to the “true value” as defined by Erikson, Wright, and McIver’s ideology measures. While these results are mixed, on average one can see that the GSS makes up a higher share of the models with high correlations, whereas the ANES dominates the lower correlations. Models that are conducted on pooled years typically outperform those that are conducted on a single year; even more consistently, presidential vote share almost always increases the correlation with the true value. Figure 2.2 uses the same MRP state estimates and EWM ideology measures to look at mean absolute error. Here one sees again that the GSS does, on average, better than the ANES, and similar patterns as above hold, though presidential vote share is not as consistently a strong performer as it was in predicting higher correlations.

These results confirm the theory that increased clusters do improve the performance of MRP on cluster-sampled polls. However, they also show that the number of clusters is not the only factor, as increasing the  $N$  of the poll and using a state-level predictor also shape the accuracy of the model. From this example, however, it is unclear how to separate the effect of the number of

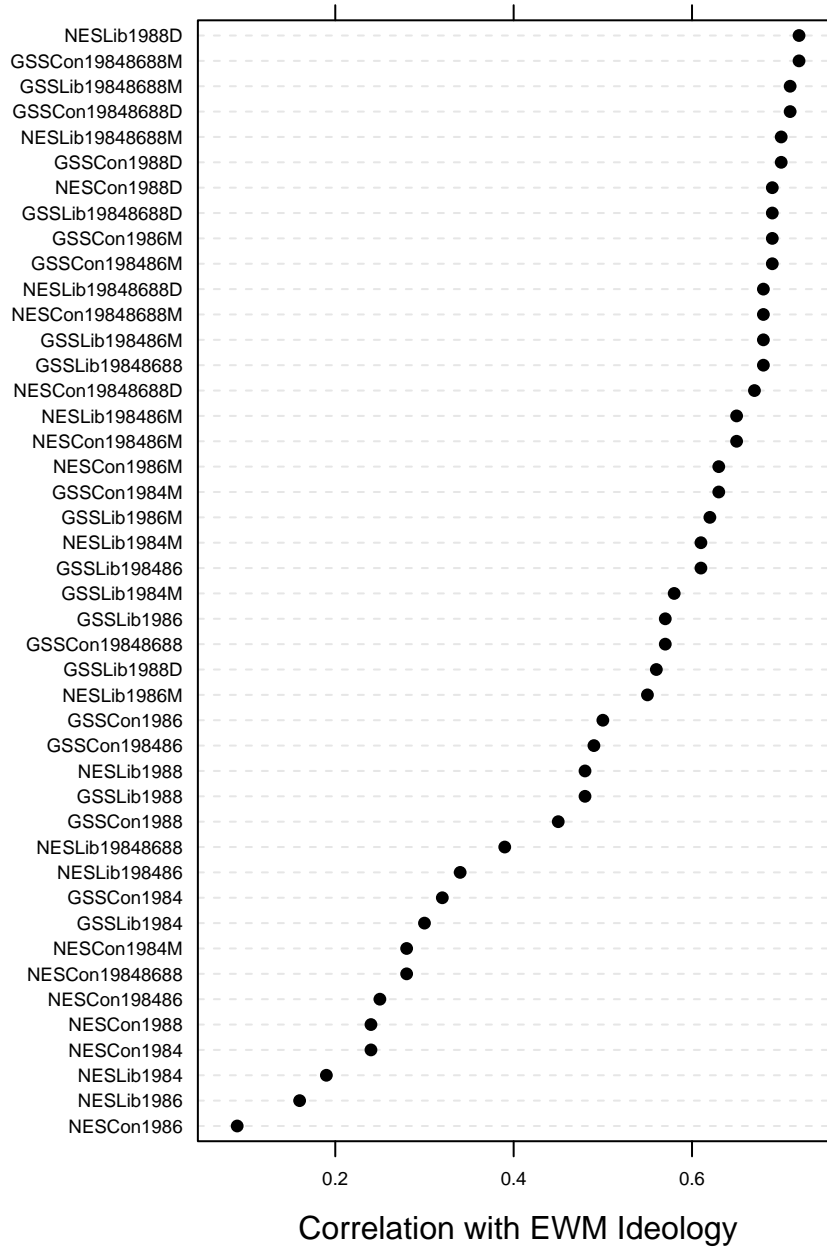
---

<sup>3</sup>While I do not constrict the datasets to have identical  $N$ 's, the size of the polls are quite similar. For the GSS, 1984 has an  $N = 1405$ , 1986 has an  $N = 1394$ , and 1988 has  $N = 1411$ . For the ANES 1984 has an  $N = 1477$ , 1986 has an  $N = 1563$ , and 1988 has an  $N = 1351$ .



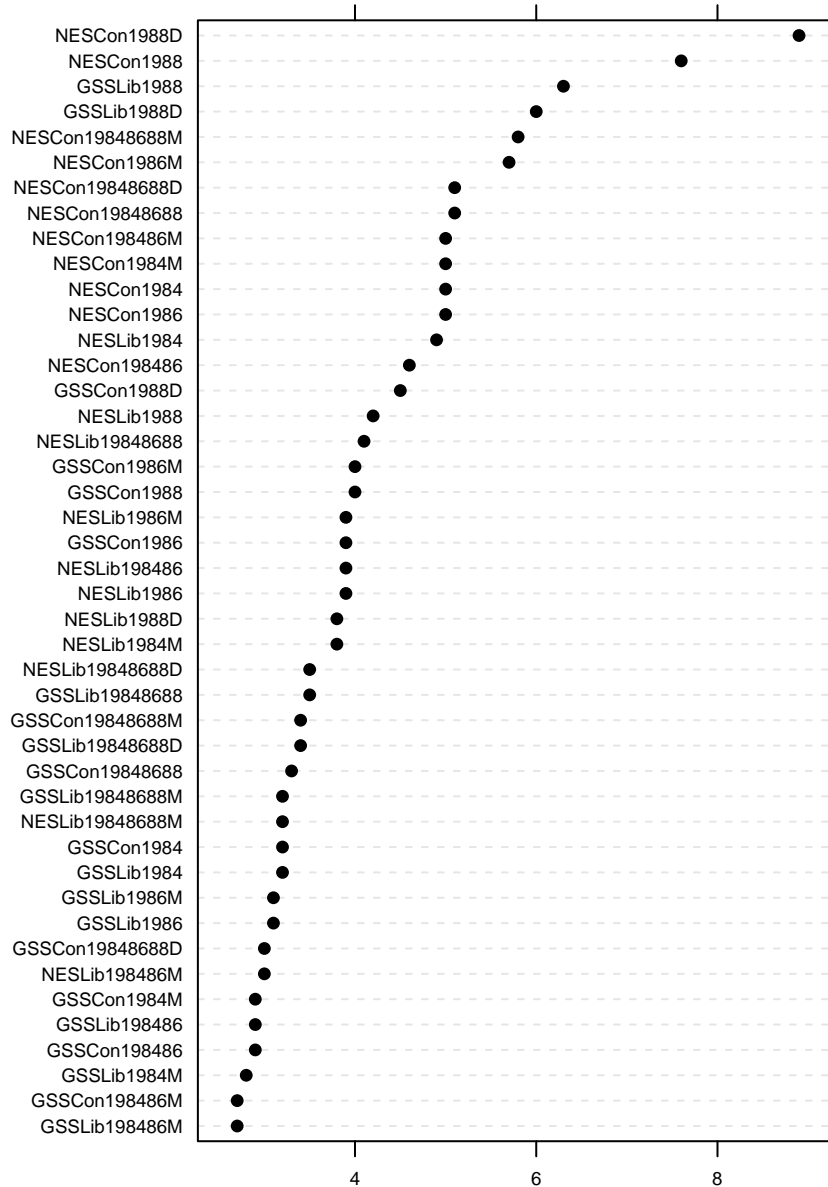
clusters and the effect of the size of the poll.

Figure 2.1: GSS, ANES, and EWM Correlations for the 1980s Sampling Frame



This figure shows the correlation between the original Erikson et al. ideology measures and several MRP estimates for these measures. The correlation is plotted along the x-axis, and each combination of data and model choice along the y-axis. The labels on the y-axis indicate the poll (GSS or ANES), what is being estimated (“Lib” for liberal and “Con” for conservative), the year(s) of the poll(s) included, and whether a state-level predictor is used (D for Dukakis vote-share, M for Mondale vote-share). While the GSS does not always outperform the ANES, the 8 worst performing models are from the ANES and 7 of the top 10 are from the GSS, indicating that the increased number of clusters in the GSS gives it an advantage over the ANES when performing MRP. This comparison also shows that using an aggregate predictor almost always improves upon a basic model, and pooling can be quite helpful as well.

Figure 2.2: GSS, ANES, and EWM Mean Absolute Errors for the 1980s Sampling Frame



Mean Absolute Error Compared with EWM Ideology

This figure shows the mean absolute error between the original Erikson et al. ideology measures and several MRP estimates for these measures. The mean absolute error is plotted along the x-axis, and each combination of data and model choice along the y-axis. As in Figure 3, the labels on the y-axis indicate the poll (GSS or ANES), what is being estimated (“Lib” for liberal and “Con” for conservative), the year(s) of the poll(s) included, and whether a state-level predictor is used (D for Dukakis vote-share, M for Mondale vote-share). Again, the GSS does better overall than the

ANES: 8 of the 10 worst performing models use the ANES and 9 of the top 10 use the GSS, indicating that the increased number of clusters in the GSS gives it an advantage. Including an aggregate predictor and pooling data both seem to be helpful, though neither as consistently as they were in the correlation illustration.

#### 2.4.4 Simulations

To attempt to untangle the potential problems of cluster sampling for MRP in more depth, I next conduct several simulations. By using simulations, we can see how well MRP recovers true opinion under a cluster sample with varying conditions, including the number of clusters in a state, the number of people in the poll, and the use of a state-level predictor. The simulations allow us to examine these problems both individually and jointly to learn more about why area-based cluster sampling might be causing problems for the validity of MRP estimates.

For these simulations, I treat the 2008 Cooperative Congressional Election Study as the population. The 2008 CCES has 32,800 respondents who have responded to a common set of survey questions. For each respondent, the CCES provides the state and congressional district where the respondent lives. I will use this information to approximate geographic clusters in these simulations.<sup>4</sup> I pull samples from this poll and use MRP to attempt to recover the true value as defined by the population estimate, defined by simply disaggregating the poll by state.<sup>5</sup>

For these simulations, I vary three different factors. First, I use three different sampling methods to draw a sample from the larger CCES poll: a random sample, an area-based cluster sample with a small number of clusters (100 congressional districts nestled within 30 states), and an area-based cluster sample with a large number of clusters (200 congressional districts within 45 states). Second, I vary the size of the poll by pulling samples of 1000, 3000, or 5000 respondents, since I hypothesize that larger sample sizes might reduce the problems of cluster sampling. For each of the 9 combinations of sampling method and sample size, I draw 1000 samples, and for each one, I perform MRP on the sample four times, once using a simple MRP model with no state-level aggregate predictor and once each using a weak, a mid-range, and a strong state-level aggregate predictor. The weak predictor is the state percentage of African Americans, which correlates at .28 with the “true” values of the dependent variables in these

---

<sup>4</sup>I use congressional district rather than county since this ensures that each cluster has roughly the same number of people. Furthermore, using county instead of congressional district would create a bias toward rural areas.

<sup>5</sup>Note that I do not use survey weights, as is standard in this type of subsample test (see Erikson, Wright, and McIver (1993), Lax and Phillips (2009b) and Warshaw and Rodden (2012)). Furthermore, the survey weights provided by the CCES are meant to reproduce a national sample, rather than a statewide sample, and thus would be inaccurate for this purpose.

MRP models, state-level ideology (measured both as the percentage who identify as conservative in the state and the percentage who identify as liberal). The medium predictor is Berry, Renquist, Fording, and Hanson's measure of elite state-level ideology, which correlates at .51. The strong predictor is Obama's 2008 presidential vote share, which correlates at .84.

From the theoretical discussion and findings earlier in this paper, I hypothesize that, holding the predictor strength and the sample size constant, random sampling should produce the most accurate results when compared with the true value, followed by a larger cluster sample; the smaller cluster sample would be expected to produce the least accurate results. I also hypothesize that both increased  $N$ -size and increased strength in the predictor used in the model would improve the accuracy of the MRP estimates under all sampling scenarios.

After running these simulations, I calculate the mean absolute error compared to the true state-level estimates and the correlation with the state-level estimates for each set of simulations; these are displayed in Table 2 and 3 below. The general patterns follow the hypotheses above. Larger  $N$ -size and increased strength in the predictor both lead to better results, with higher correlations and lower mean absolute error. Likewise, random sampling performed the best, followed by the cluster sampling with more clusters, with the cluster sampling with fewer clusters performing the worst.

However, while the hypotheses are upheld, the magnitude of the results are somewhat surprising. Overall, sample size and predictor strength matter much more (in the expected directions) than sampling method does. This is quite surprising given other practical findings in this paper. To investigate this further, I plotted the density of these results, looking at the correlations and mean absolute errors for the 1000 runs of each condition of the simulations. Could it be that while the different sampling methods have similar mean correlations and absolute errors, the cluster-based sampling produces a wider range of results, making a single cluster-sample poll less reliable? Overall, Figures 3 and 4 do not prove this point. On each graph, I graph the density of the random sample (solid red line), the large cluster sample (dotted blue line), and the small cluster sample (dashed green line). The overall pattern shows nearly identical densities across the three sampling patterns, though occasionally the small cluster sample starts to deviate from

this pattern as hypothesized.

There are a few ways to interpret these results, especially given that MRP does not always appear to perform as strongly as we would expect on the ANES and the GSS data. The first option is that cluster sampling does not impact MRP results most of the time. There can always be bad “draws” in sampling, and it could be the case that the ANES and GSS samples examined in this paper are examples of this. This interpretation, however, is somewhat strained given the performance of ANES and GSS under MRP. Similarly, it could be the case that simulations reveal that MRP has a more variable performance on randomly sampled polls than previously argued (see Buttice and Highton 2013), and in this context MRP on cluster-sampled polls may not underperform MRP on randomly-sampled polls as much as hypothesized.

The second possibility is that the cluster sampling is too generously defined in these simulations. While the numbers I chose to define the number of clusters do seem to fit past ANES and GSS sample sizes (30 or 45 states, 100 or 200 clusters), the use of congressional district as a proxy for area-based clusters could be cause for concern. While both the GSS and ANES start with similarly larger areas in their sampling, they narrow down to smaller geographic areas before they sample individuals. Perhaps simulations that narrowed geographic areas to smaller units might have shown poorer diagnostics.

The third possibility is that these test statistics do not properly capture the uncertainty involved in using MRP on cluster-sampled polls. While correlation and mean absolute error are standard ways to assess the accuracy of MRP (see Lax and Phillips 2009b), they might not be representing the errors we should be most worried about when using MRP on cluster-sampled polls. The mean absolute error is taking the difference between the true value and the MRP estimate for each state and then taking the mean; likewise, the correlation uses the values for all 50 states. While these are good summary measures, they might not actually capture the set of MRP estimates that researchers would be most worried about – one where 40 of the states are fine but 10 of the states look unusual, because those states have odd clusters. Diagnostic tools relying on average measurements would mask a small number of troubling estimates.

Table 2.2: Simulation Results – Pearson’s Product-Moment Correlations with True Value

Predictor	N=1000			N=3000			N=5000		
	Cluster (100)	Cluster (200)	Random	Cluster (100)	Cluster (200)	Random	Cluster (100)	Cluster (200)	Random
None	0.437	0.452	0.466	0.558	0.592	0.603	0.600	0.651	0.664
Weak	0.442	0.462	0.477	0.563	0.608	0.623	0.606	0.669	0.684
Medium	0.489	0.505	0.515	0.606	0.635	0.645	0.645	0.690	0.703
Strong	0.650	0.657	0.666	0.725	0.738	0.742	0.745	0.764	0.767

Table 2.3: Simulation Results – Mean Absolute Error Between Estimates and True Value

Predictor	N=1000			N=3000			N=5000		
	Cluster (100)	Cluster (200)	Random	Cluster (100)	Cluster (200)	Random	Cluster (100)	Cluster (200)	Random
None	5.40	5.37	5.28	4.81	4.69	4.63	4.60	4.45	4.40
Weak	5.53	5.46	5.37	4.88	4.72	4.64	4.64	4.46	4.40
Medium	5.34	5.32	5.23	4.71	4.60	4.54	4.48	4.35	4.32
Strong	4.90	4.81	4.73	4.20	4.09	4.02	4.01	3.89	3.86

Figure 2.3: Simulation Results – Correlation Between MRP Estimates and True Values, by Sample Size and Strength of Predictor

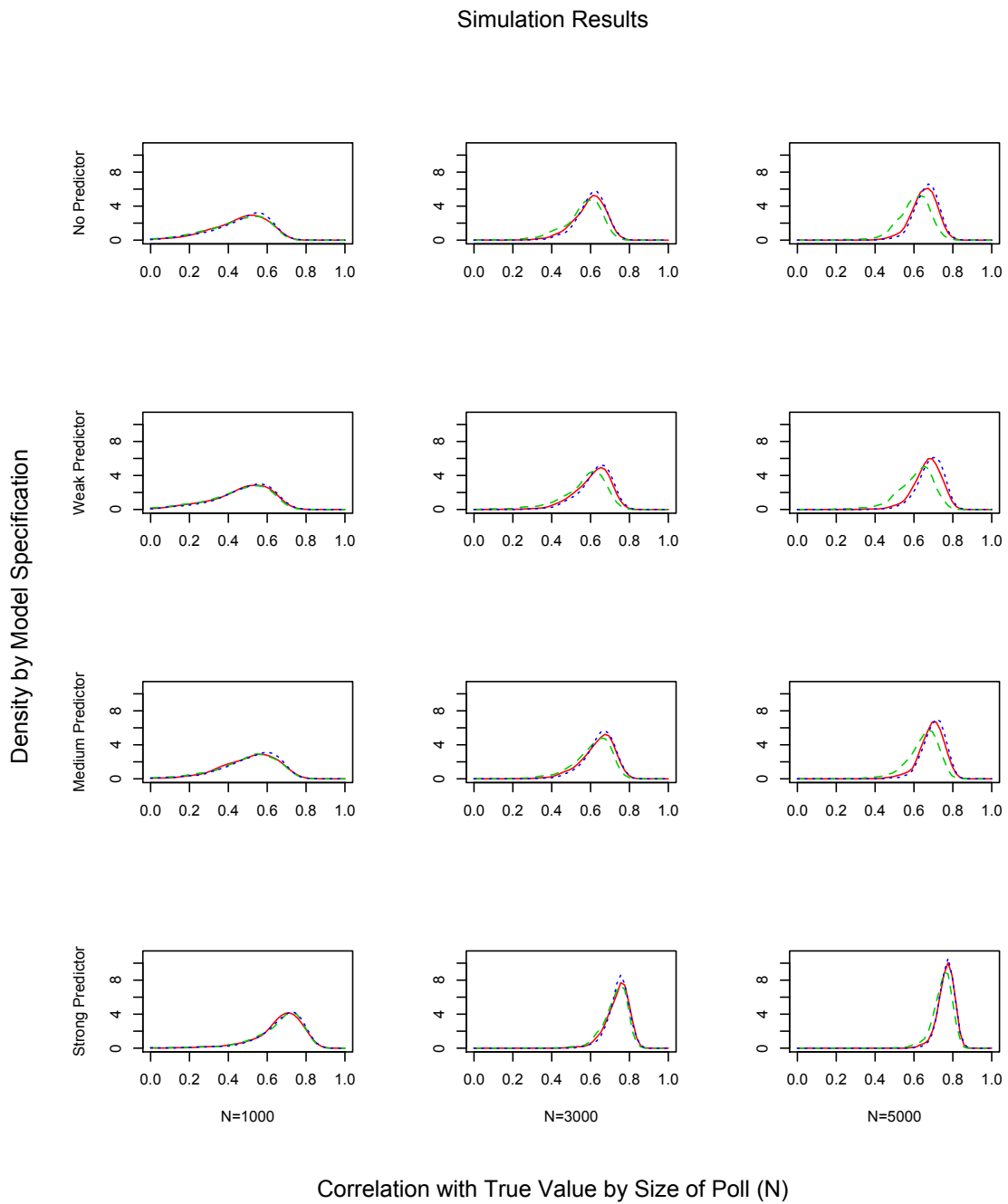
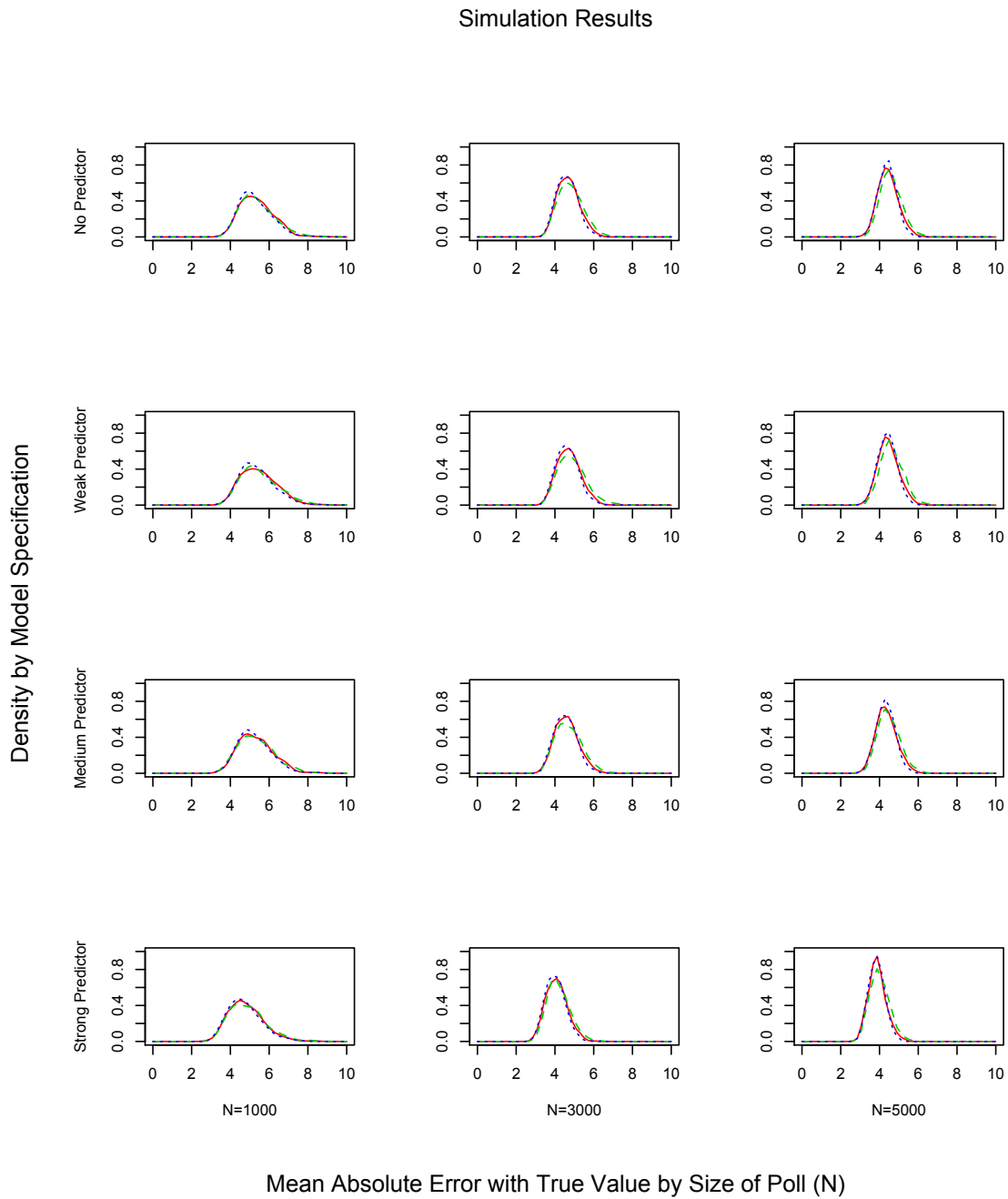




Figure 2.4: Simulation Results – Mean Absolute Error Between MRP Estimates and True Values, by Sample Size and Strength of Predictor



Overall, these simulations confirm some of our hypotheses while leaving others as a somewhat open question. While simulations help us reproduce the real world in more controlled

circumstances, they still cannot perfectly capture the variation present in established cluster sampled polls. Given this, it is important to return to the question of how MRP performs on past cluster-sampled polls so that applied researchers can use these rich data sources for state-level analysis.

## 2.5 Results and Discussion

There are several ways to test the validity of estimates produced by MRP when using cluster-sampled polls. I begin by looking at several sets of state-level estimates generated by MRP from individual GSS polls and pooled sets of GSS polls from several years, considering the face validity of each. I then compare GSS and ANES state-level estimates with state-level results from other polls that are representative at the state level, either by design or after several polls are pooled together. Last, for the ANES opinion estimates, I then investigate how knowledge of a sub-state geographic indicator, in this case congressional district, affects state-level estimates.

### 2.5.1 Face Validity: Who's Liberal and Who's Conservative?

When evaluating the results from multilevel regression with poststratification, it is helpful to look at the face validity of these results. Are liberal states liberal and conservative states conservative in the measures created by MRP? In other words, do the state-level opinion estimates make sense? The short answer, based on an initial analysis of MRP on GSS data, is sometimes.

I use the GSS from recent years to generate state-level estimates for party identification and political ideology. I first create four variables from these: Democratic, Republican, liberal, and conservative, and then I use logistic regression to create estimates for each of these at the state level.<sup>6</sup> I use MRP to model public opinion responses as a function of both geographic factors

---

<sup>6</sup>Both party identification and political ideology are measured along a seven point scale in the GSS. In recoding, I count those who are "extremely liberal," "liberal," and "slightly liberal," as liberal, and likewise for conservative; similarly, I count those who are a "strong Democrat," a "not very strong Democrat," and "Independent, close to Democrat" as Democratic and likewise for Republican. Furthermore, since I focus on estimating percent Democratic and Republican of total respondents, I do include those who responded "other party" in the base category, but I drop the (few) respondents who said "don't know," and "no answer." Likewise, for the ideology variables, I drop those who said "don't know" or "no answer," as well as those who were not asked the question.

(state and region) and demographic factors (age, education, race, and gender), allowing for partial pooling across states. Then, I use Census data for poststratification, weighting the estimates for each type of respondent by the percentage of that type in the actual state population.<sup>7</sup>

Figures 2.5 and 2.6 below graph percent Democratic and percent liberal as generated by MRP in four different models.<sup>8</sup> First, I use a simple MRP model on the 2004 GSS ( $N = 1305$  for ideology,  $N = 2791$  for partisanship). Second, I add an aggregate measure of Democratic presidential vote share in 2004 as a state-level predictor. This statistic highly correlates with ideology and especially with partisan identification at the state level, and, as shown by Lax and Phillips (2009b), adding such an aggregate measure to a logistic estimation should increase its accuracy. Third and fourth, I repeat these graphs on a larger dataset, pooling the 2002, 2004, and 2006 GSS ( $N = 6938$  for ideology,  $N = 9966$  for partisanship). While Lax and Phillips posit that a single national poll of approximately 1400 respondents should produce valid estimates, I hypothesize that a larger  $N$  will be required for a cluster-sampled poll. Given that, as shown earlier, cluster sampling results in large design effects for key political variables, I would expect a corresponding reduction in the poll's effective sample size. Furthermore, the sampling frame used by the GSS changed between the 2002 and 2004 waves of the survey. Thus, pooling GSS surveys from 2002, 2004, and 2006 creates a survey that, in effect, samples from twice the number of clusters as a single GSS; this should produce more valid estimates, since sampling from additional clusters improves the efficiency and accuracy of a cluster-sampled poll (Harter et al. 2010).

These figures help assess the face validity of the estimates in two different manners. First, by looking at the values on the  $x$ - and  $y$ -axes, one can see how accurate each state-level estimate appears to be. For example, a conservative state like Texas should have a high conservative estimate and a low liberal estimate. Second, by looking at both partisanship and ideology simultaneously, one sees how strong the overall relationship is between the estimated measures. While this would not have been a useful example several decades ago, increased polarization had led

---

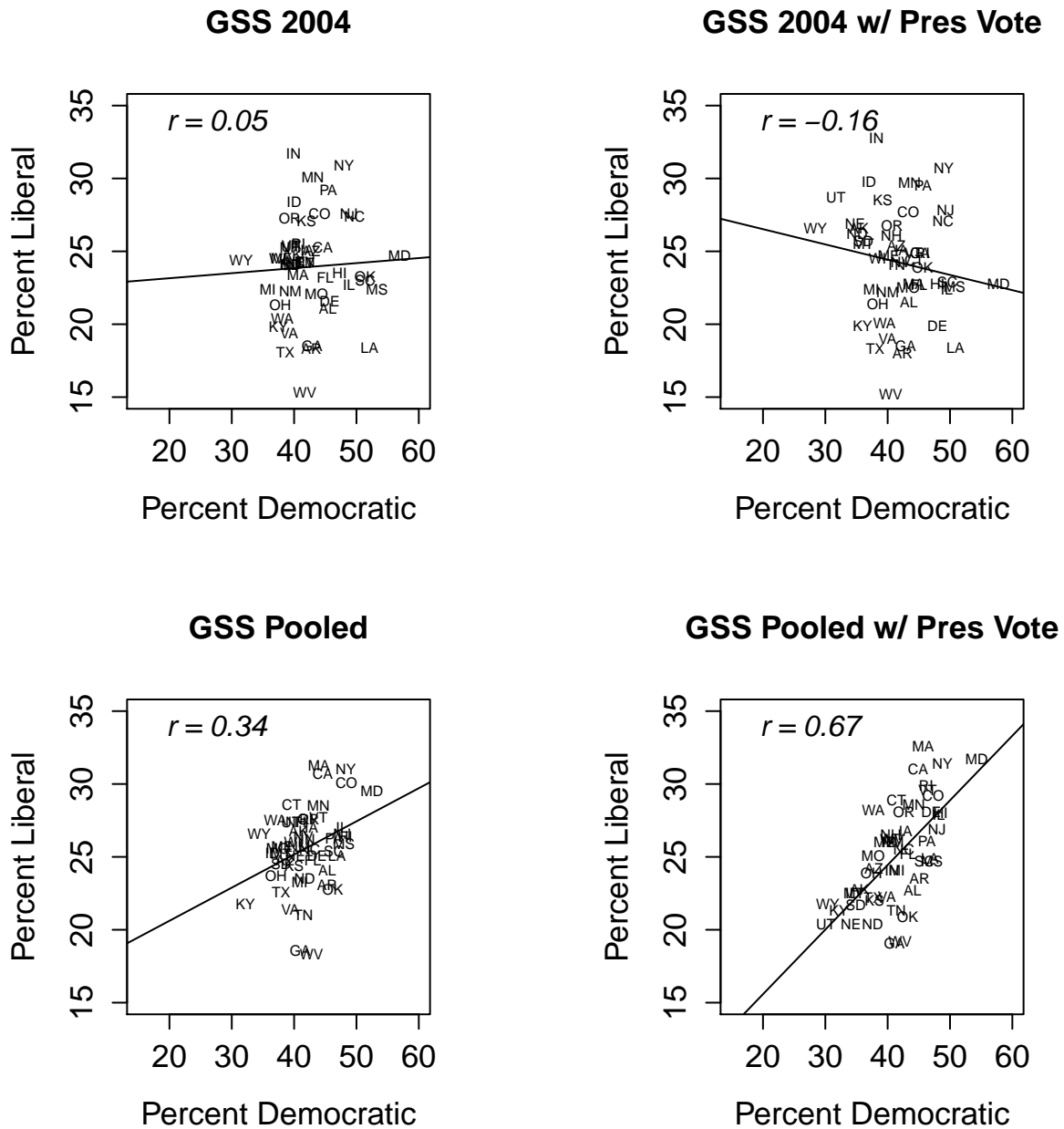
<sup>7</sup>Note that in this and subsequent models in this section, race is a dichotomous white/black variable, given constraints related to how the GSS codes Hispanic respondents. Otherwise, I follow the same categorical specifications described earlier in this paper in defining MRP.

<sup>8</sup>These models and figures also exclude DC, which is an outlier on these measures.

to a strong correlation between state-level ideology and partisanship, especially after 2000. In an updated analysis of their work in *Statehouse Democracy*, Erikson, Wright, and McIver (2006) note that by Bush's first term, this correlation had reached  $r = 0.66$ .

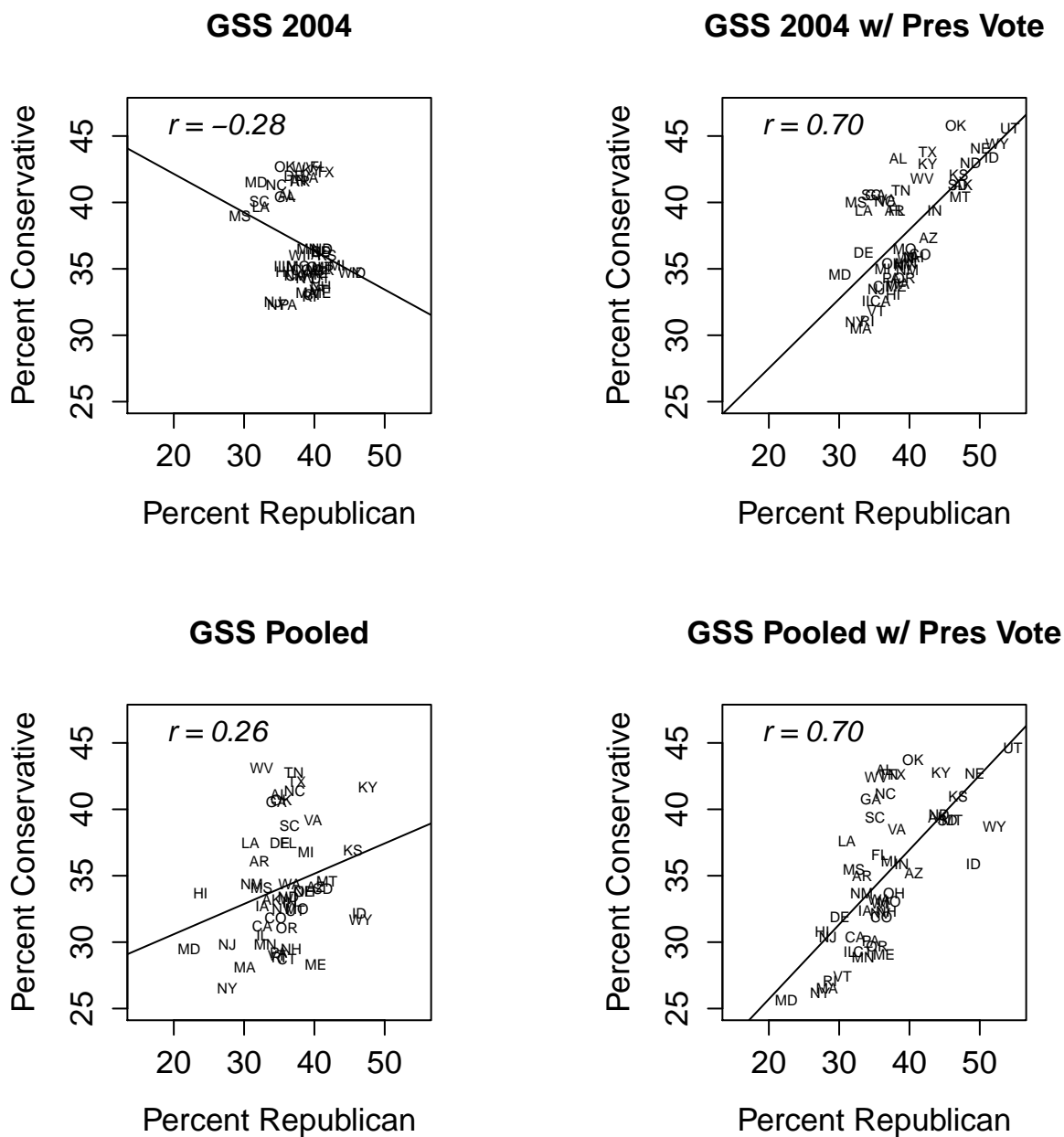
Indeed, looking at the pattern of these four graphs in Figure 2.5, MRP seems to perform as hypothesized, producing problems that would not occur if MRP were performed on a poll that did not use area-based clusters. In the first graph in the top left, when MRP is done on a single cluster-sampled poll without an aggregate state-level measure, the results do not fully meet face validity. Some results make sense: New York is one of the most liberal and Texas is one of the most conservative, for example. However, the correlation between party and ideology is basically zero, and several individual states present surprising results. One would not expect Indiana to be the most liberal state or Idaho and Kansas to be near the top as well. Likewise Massachusetts is in the middle for both partisanship and ideology, and Washington state is more Republican and conservative than one would expect. Adding state-level vote for Kerry in 2004 to the 2004 estimation changes a few results, but keeps the pattern roughly similar, showing that using a more complicated model on a single cluster-sampled poll cannot be counted on as a sufficient solution. Pooling 2004 with the surveys from 2002 and 2006, more than tripling the  $N$  and introducing additional clusters, has a more substantive effect. Massachusetts becomes one of the most liberal states, and Washington improves as well. States that were in a more accurate place, such as Texas and New York, stay where they are relative to the other states. Overall, the correlation between ideology and partisanship strengthens, and this relationship increases further when I add Kerry vote share to the model. The pooled model with an aggregate predictor produces a correlation of 0.67 between ideology and partisanship, similar to that found by Erikson, Wright, and McIver (2006).

Figure 2.5: MRP GSS Estimates for Democratic and Liberal – 2004 v. 2002-2006 Estimates



Note: This figure shows several sets of MRP estimates from the GSS for liberal and Democratic. The first graph uses a simple MRP model on the 2004 GSS. The second (top right) adds presidential vote as an aggregate predictor. The bottom two graphs repeat these two models on a pooled 2002, 2004, and 2006 dataset. To reach the expected correlation between ideology and partisanship and have face validity for individual states, one must both pool clustered data and use a more sophisticated MRP model.

Figure 2.6: MRP GSS Estimates for Republican and Conservative – 2004 v. 2002-2006 Estimates



Note: This figure shows several sets of MRP estimates from the GSS for conservative and Republican. As in Figure 3, the first graph uses a simple MRP model on the 2004 GSS. The second (top right) adds presidential vote as an aggregate predictor. The bottom two graphs repeat these two models on a pooled 2002, 2004, and 2006 dataset. To reach the expected correlation between ideology and partisanship and have face validity for individual states, one must both pool clustered data and use a more sophisticated MRP model.

The Republican and conservative graphs in Figure 2.6 show a similar pattern. In the 2004 data, however, the South has a distinct pattern, with low Republican identification given its conservatism. While this relationship is common in historical data, it is odd it is so strong in 2004. The model using an aggregate predictor corrects for this, as does the pooled data to a lesser degree. Somewhat surprisingly, there are not added benefits of using both the pooled data and the aggregate predictor, compared to using the predictor alone. In both cases, the correlation between partisanship and ideology reaches 0.70.

These initial face validity tests seem to demonstrate that a simple MRP analysis on a small dataset will not produce sufficiently reliable results. Rather, they imply that one must simultaneously add a strong aggregate state-level measure and use a larger dataset to increase the validity of MRP estimates from cluster-sampled polls. These results also seem to indicate that pooling surveys across sampling frames, thus increasing the number of clusters within the sample, further improves the results. In the next section, I evaluate how MRP can be used on cluster-sampled polls by comparing these estimates produced under different specifications to results from a simple random sample, the National Annenberg Election Study.

## 2.5.2 GSS, ANES, and the National Annenberg Election Study

The National Annenberg Election Study (NAES) is a national survey that uses random digit dialing to sample respondents in 48 states and the District of Columbia.<sup>9</sup> In 2004, the Annenberg survey has a sample size of over 80,000 people and is designed to be representative at the state level as well as the national level. Thus, since the NAES is a full probability sample that can be disaggregated into its state components, I can consider its results to be a valid standard against which I can evaluate MRP estimates on cluster-sampled polls. Disaggregation of significantly large datasets is considered to be an ideal method of measuring state-level opinion, and thus the Annenberg dataset provides a useful standard against which to compare MRP estimates (Carsey and Harden 2010). Here, I compare the state-level NAES estimates on partisanship and political

---

<sup>9</sup>Therefore, for this part of the analysis, I will drop Alaska and Hawaii from the analysis. I also drop DC, as I did above, given that it is an outlier on these measures.

ideology to the state estimates produced by MRP of similar questions in the GSS and ANES. While there are minor differences in question wording, these questions are similar enough to assess MRP's performance on the GSS and ANES.

First, I run a series of correlations between the MRP estimates from the GSS data and the disaggregated values from the NAES data, looking at four state-level estimates (percent Democratic, percent Republican, percent liberal, percent conservative) and testing all four specifications for MRP estimation used earlier in this paper against the NAES value. Table 2.4 reports these correlations and shows which MRP estimates, highlighted in bold, reflect an improvement from the simple MRP model on a single survey.

Table 2.4: Pearson's Product-Moment Correlations – GSS and NAES

Proportion	2004 Model	2004 w/ Kerry Vote	Pooled Model	Pooled w/Kerry Vote
Democratic	0.24	<b>0.57</b>	<b>0.41</b>	<b>0.65</b>
Republican	0.18	<b>0.69</b>	<b>0.55</b>	<b>0.79</b>
Liberal	0.29	0.14	<b>0.61</b>	<b>0.80</b>
Conservative	0.51	<b>0.78</b>	<b>0.63</b>	<b>0.77</b>
Average	0.31	<b>0.55</b>	<b>0.55</b>	<b>0.75</b>

*N* (2004) = 1303 (ideology), 2787 (partisanship)

*N* (Pooled) = 6917 (ideology), 9942 (partisanship)

Bold font indicates an improvement over the single-year, basic model.

As the table shows, the simple MRP model using a single poll produces far from desirable results. While the correlations are all correctly signed, they are quite small except for the conservative measure, and they do not inspire much trust in the MRP estimates. Lax and Phillips (2009b) test how a MRP model without an aggregate predictor performs on a sample size of 2800 and find correlations with true values of 0.77. The GSS for 2004 is just under this *N* for the two partisan measures (2787), and half of this for the two ideology measures (1303), and none of these correlations come close to the correlations Lax and Phillips demonstrate from non-clustered polls. Since I follow the standard model specification outlined by Lax and Phillips, this suggests an issue with the underlying data. Specifically, these results suggest that cluster sampling introduces error into MRP estimates because the respondents are non-randomly sampled at the state level.



Including a state-level predictor highly correlated with what one is trying to predict, such as Democratic presidential vote share, somewhat improves the accuracy of the model on the 2004 data, producing an average correlation of 0.55.<sup>10</sup> This improvement reinforces Lax and Phillips' recommendation to include an aggregate predictor when using MRP, especially when using a single national poll. A state level variable is especially important for states without any data, and there are more such states in a cluster-sampled poll than in a randomly-sampled poll. However, while including the aggregate predictor improves the correlation with true values, it does not improve it as much as one would expect given non-clustered data. Lax and Phillips (2009b) find that a standard MRP model applied to a random poll of 1400 typically has a correlation of approximately 0.73 with true values and a correlation of 0.82 for a poll of 2800.

In addition to adding an aggregate predictor, another potential solution to improve MRP's performance is pooling multiple surveys to have a larger  $N$ . In this example, I pool the 2002, 2004, and 2006 GSS surveys ( $N = 9942$  for partisanship identification and  $N = 6917$  for ideological identification), and the results show an increase in correlations both in the simple model and the model with the aggregate predictor, the latter of which reaches correlations similar to those expected by Lax and Phillips mentioned above. Pooling in this case might increase correlations for two reasons. First, an increased  $N$  might be necessary to improve MRP estimates under cluster sampling, given that cluster sampling may reduce the effective sample size of a survey. However, this answer is not completely satisfying. If cluster sampling reduces the efficiency of the sample, including more respondents from within the same cluster alone should not improve it substantially (Harter et al. 2010). Erikson, Wright, and McIver (1993) touched on this concern, as mentioned earlier, when explaining why they use CBS/NYT surveys, rather than GSS and ANES polls, to construct their state-level ideology and partisanship measures. However, the second reason that pooling might improve estimates is that, depending on the sampling design, pooling surveys from several years might increase the number of clusters from which respondents were drawn if the sampling frame has changed over the years pooled. That is the case when pooling

---

<sup>10</sup>The liberal measure, however, presents an unusual outlier, perhaps because of the relatively small number of people who identify as liberal in the survey. Not including the liberal outlier, the average correlation would be 0.68.

GSS surveys from 2002, 2004, and 2006, as the frame changed in 2004. Harter et al. (2010) advise that sampling from additional clusters will improve the quality of one's sample, and that indeed seems to be the case here.

These results indicate that simultaneously pooling data and using an aggregate state-level predictor can produce more accurate measures and make it much less problematic to use MRP on cluster-sampled data. To confirm, I replicate the above section using the American National Election Studies. To analyze the ANES, I use three different sets of the data. First, as with the GSS, I look at only 2004, the same year as the Annenberg survey. The ANES has smaller  $N$ 's than the GSS ( $N = 1175$  for partisanship identification and  $N = 912$  for ideological identification), and I therefore do not expect a strong performance for MRP. Next, I pool the 2004 data with surveys from 2002 and 2000, which I refer to as Pooled A, increasing the sample size to an  $N$  of 4313 for partisanship data and 2755 for ideology.<sup>11</sup> Finally, given that these  $N$ 's are still smaller than what might be needed with a cluster-sampled poll, I also add 2008 data to the pooled model (labeled Pooled B), yielding an  $N$  of 6516 for the partisanship variables and an  $N$  of 4329 for the ideology variables. I generate MRP estimates both using the basic model and adding state-level Kerry vote as an aggregate predictor.<sup>12</sup> Table 3 reports the correlations of the MRP-generated state-level estimates with Anneberg's state-level estimates.

Table 2.5: Pearson's Product-Moment Correlations – ANES and NAES

Proportion	2004	2004 w/ KV	Pooled A	Pooled A w/KV	Pooled B	Pooled B w/KV
Democratic	0.24	<b>0.66</b>	<b>0.38</b>	<b>0.71</b>	<b>0.42</b>	<b>0.68</b>
Republican	0.19	<b>0.61</b>	<b>0.36</b>	<b>0.65</b>	<b>0.41</b>	<b>0.68</b>
Liberal	0.58	<b>0.71</b>	0.54	<b>0.70</b>	0.57	<b>0.71</b>
Conservative	0.53	<b>0.67</b>	0.40	<b>0.65</b>	<b>0.54</b>	<b>0.74</b>
Average	0.39	<b>0.66</b>	<b>0.42</b>	<b>0.68</b>	<b>0.49</b>	<b>0.70</b>

$N$  (2004) = 912 (ideology), 1175 (partisanship)

$N$  (Pooled A) = 2755 (ideology), 4313 (partisanship)

$N$  (Pooled B) = 4329 (ideology), 6516 (partisanship)

Bold font indicates an improvement over the single-year, basic model.

<sup>11</sup>I cannot pool with 2006 data, as I did with the GSS, since the ANES did not conduct a survey in 2006.

<sup>12</sup>Unlike the GSS model, the ANES model uses three categories for race: black, white, and Hispanic.

As with the GSS data, I find that using a strong aggregate predictor at the state-level improves the model considerably, producing stronger correlations with true values than the models that do not include the predictor. Pooling data, however, does not lead to as strong of an improvement as it did with the GSS data. Since both the GSS and the ANES changed sampling frames over this time period, it is unlikely that the number of clusters explains the difference. The pooled datasets here are smaller than those used in the GSS example, so it may be the case that a larger sample size continues to be important for MRP conducted on cluster-sampled polls. It may also be the case that pooling introduces a different bias if the variables that are being estimated change over time. While the GSS example pools 2002, 2004, and 2006 surveys, the ANES data requires pooling from 2000 to 2008 to achieve a similar  $N$ . This larger time frame may create additional tradeoffs, introducing a pooling bias which could harm accuracy over time, rather than help it (Gelman 2013). Despite this, the large pooled model which includes the aggregate predictor does still return strong correlations for all four variables which approach the standard set by Lax and Phillips (2009b).

Another statistic that is helpful to examine is the mean absolute error. This measure takes the average absolute difference between the generated estimates, in this case those generated by MRP from the GSS and ANES, and the baseline measure, which are those measures produced by disaggregating the 2004 Annenberg data. Table 2.6 below shows mean absolute errors calculated by comparing the GSS and the NAES. These mean absolute errors are, on average, slightly larger than one would expect, given that Lax and Phillips (2009b) report mean absolute errors for MRP estimates produced using a full model specification ranging from 4.9 to 4.1 for sample sizes of 1400 and 7000, respectively. Similar to the correlation results, these numbers do improve, on average, when using a state-level predictor, but they actually worsen when pooling the data. This may be another reflection of a potential bias caused by pooling data. This might be especially reflected in the increased error of the Republican measure in the pooled models, given greater instability in partisanship overtime.

I then examine the mean absolute errors for the ANES data as compared to the Annenberg

Table 2.6: Mean Absolute Error – GSS and NAES

Mean Abs. Err.	2004 Model	2004 w/ Kerry Vote	Pooled Model	Pooled w/Kerry Vote
Democratic	6.2	<b>5.2</b>	<b>5.7</b>	<b>5.1</b>
Republican	7.7	<b>5.4</b>	9.0	7.9
Liberal	4.4	4.9	<b>4.3</b>	<b>3.5</b>
Conservative	5.2	<b>3.4</b>	5.7	<b>4.6</b>
Average	5.9	<b>4.7</b>	6.2	<b>5.3</b>

*N* (2004) = 1303 (ideology), 2787 (partisanship)

*N* (Pooled) = 6917 (ideology), 9942 (partisanship)

Bold font indicates an improvement over the single-year, basic model.

data. Table 2.7 below shows similar results to the GSS data, and the mean absolute errors range from 6.5 to 3.7. The ANES estimates do not have the same higher errors for partisanship measures that the GSS estimates showed, but they do further enforce the limited power of pooling alone. By pooling and using a state-level predictor, however, these results do approach the same mean absolute errors reported by Lax and Phillips (2009b).

Table 2.7: Mean Absolute Errors – ANES and NAES

Mean Abs. Err.	2004	2004 w/ KV	Pooled A	Pooled A w/KV	Pooled B	Pooled B w/KV
Democratic	5.9	<b>4.8</b>	<b>5.1</b>	<b>3.7</b>	<b>5.1</b>	<b>4.1</b>
Republican	6.2	<b>5.1</b>	6.5	<b>5.1</b>	6.5	<b>5.1</b>
Liberal	4.5	<b>3.8</b>	5.5	<b>4.4</b>	5.3	<b>4.4</b>
Conservative	5.4	6.4	5.5	5.4	<b>5.2</b>	<b>4.8</b>
Average	5.5	<b>5.0</b>	5.7	<b>4.7</b>	5.5	<b>4.6</b>

*N* (2004) = 912 (ideology), 1175 (partisanship)

*N* (Pooled A) = 2755 (ideology), 4313 (partisanship)

*N* (Pooled B) = 4329 (ideology), 6516 (partisanship)

Bold font indicates an improvement over the single-year, basic model.

To confirm that this is not an artifact of a given dataset, I replicate these results using a different measure of "true" values: the state-level measures of ideology and partisanship that Erikson, Wright, and McIver created using disaggregation of several national level polls. These measures are widely used in political science, and thus provide an important test case. Given that the time frame of the GSS and ANES data used here spans from 2000 to 2008, I use an updated

version of the disaggregated data that was produced using polls from 1996 to 2003 (Erikson, Wright, and McIver 2006).<sup>13</sup> I report tables of these validity tests in Appendix A1. While there are slight differences and more overall error when testing MRP against the EWM scores, we see that the overall principles demonstrated above are reinforced. The accuracy and validity of MRP estimates on cluster-sampled polls appear to improve when adding an aggregate predictor that is correlated with the outcome we are estimating; there are also some smaller gains from pooling data to increase the overall survey size.

### 2.5.3 Modeling Additional Geographic Information

While increasing the number of clusters, increasing the  $N$ , and utilizing an aggregate predictor are all potential ways to improve the accuracy of MRP on cluster-sampled models, another potential tool that researchers could use is to include geographic information about the cluster in the multilevel model itself. If one could account for the distinctiveness of the cluster, it might help to mitigate the impact of an idiosyncratic cluster on state-level estimates. Unfortunately (and unsurprisingly), survey firms do not typically release specific information about the location of their clusters without intensive IRB scrutiny, and thus the average researcher cannot model clusters directly as a level in the multilevel model. While the public data does detail which respondents are in the same PSU, for the purposes of calculating design effects or including sampling weights, it does not say where these PSUs are. Thus, it is not information that could be mapped onto the Census data necessary for MRP.

However, the ANES does provide an opportunity to test this question in its public release data by using another variable as a proxy for cluster. Specifically, the ANES releases the congressional district as well as the state for every respondent in its public data files. Previous work has shown that MRP can be used to produce accurate and efficient estimates of public opinion at the congressional district level (Krimmel, Lax, and Phillips 2016; Warshaw and Rodden 2012). For this example, however, I will include congressional district in the model but continue to poststratify at the state level. Specifically, the only changes in our earlier model can be expressed

---

<sup>13</sup>As before, I drop DC as an outlier from my MRP analysis.

as follows:

$$\alpha_s^{cd} \sim N(\alpha_{m[s]}^{state} + \beta^{presvote} \cdot presvote, \sigma_{cd}^2), \text{ for } s = 1, \dots, 435 \quad (2.10)$$

$$\alpha_m^{state} \sim N(0, \sigma_{state}^2), \text{ for } m = 1, \dots, 50 \quad (2.11)$$

Here, rather than modeling state as a function of region, I instead model congressional district as a function of state and presidential vote at the congressional district level.

I rerun some of my earlier models performed on the ANES, examining how my earlier state-level models compare with models that include congressional district. Below, I look at 2004 and 2008, focusing on models that include presidential vote share as an aggregate predictor. For the partisanship estimates, 2004 has an  $N = 1175$ , and 2008 has an  $N = 2182$ ; for the ideology estimates, 2004 has an  $N = 912$ , and 2008 has an  $N = 1558$ . Since the 2008 survey included more respondents than the 2004 survey, I would also hypothesize that the 2008 results should be slightly stronger on average. Note that because of changes in redistricting, I do not pool data in these instances. As earlier, I treat disaggregated data from the Annenberg survey as "true" values and look at both the correlation and the mean absolute error.<sup>14</sup>

Table 2.8: Comparing State and Congressional Models – Correlation

Proportion	2004 State Model	2004 CD Model	2008 State Model	2008 CD Model
Democratic	0.66	<b>0.71</b>	0.58	<b>0.69</b>
Republican	0.61	<b>0.69</b>	0.64	<b>0.71</b>
Liberal	0.71	0.64	0.77	<b>0.83</b>
Conservative	0.67	<b>0.69</b>	0.76	<b>0.88</b>
Average	0.66	<b>0.68</b>	0.69	<b>0.78</b>

$N$  (2004) = 912 (ideology), 1175 (partisanship)

$N$  (2008) = 1558 (ideology), 2182 (partisanship)

Bold font indicates an improvement over the state model.

<sup>14</sup>Since I am using the NAES data, Alaska and Hawaii are dropped from the analysis of correlations and mean absolute errors as before. Note that as I did earlier, I also drop DC from this analysis. While the 2008 NAES survey was smaller than the 2004 NAES survey, with an  $N$  just under 58,000, this is still large enough to be disaggregated to the state level. Like the 2004 Annenberg data, the 2008 data also relied on random digit dialing rather than cluster

Table 2.9: Comparing State and Congressional Models – Mean Absolute Error

Mean Abs. Err.	2004 State Model	2004 CD Model	2008 State Model	2008 CD Model
Democratic	4.8	<b>4.4</b>	4.7	4.8
Republican	5.1	<b>4.5</b>	6.2	6.3
Liberal	3.8	<b>3.5</b>	3.5	<b>3.2</b>
Conservative	6.4	<b>6.0</b>	3.8	<b>2.7</b>
Average	5.0	<b>4.6</b>	4.6	<b>4.3</b>

*N* (2004) = 912 (ideology), 1175 (partisanship)

*N* (2008) = 1558 (ideology), 2182 (partisanship)

Bold font indicates an improvement over the state model.

In general, this data shows a broad pattern of the congressional district model performing slightly better than the model which includes state and region only. This is especially true when looking at correlations, where the congressional district model performs better than the state model in all but one instance. The mean absolute error analysis reveals less strong results, though the overall pattern indicates modeling the congressional district decreases the mean absolute error. These results also, for the most part, confirm the hypothesis that the 2008 estimates should be more accurate than the 2004 estimates, given that they are generated from a survey with more respondents. It may be that unusual circumstances surrounding Obama’s candidacy decreased the ability of aggregate presidential vote share to predict partisan self-identification.

This data, however, only looks at a small slice of the ANES data available, and a more thorough examination will be necessary to understand to what extent including congressional district in the multilevel model corrects for cluster sampling. While this does seem to produce a potential alternative fix, constant changes in redistricting limit the ability to use pooling if one is modeling the congressional district. Thus, researchers may face a trade-off in choosing which solution, pooling or modeling the congressional district, to use.

---

sampling.

## 2.6 Conclusion

Cluster-sampled polls present specific challenges for analytical tools such as MRP. While cluster-sampled polls are considered representative at the national level, their use of clusters can introduce problems at the state and local level that can complicate simulation-based modeling such as MRP. Given the mixed evidence found in this paper, researchers should use caution when applying subnational opinion estimation techniques to clustered data.

This is not to say, however, that MRP should never be used with cluster-sampled polls. The evidence above shows that the combination of pooling several polls across years and sampling frames, as well as using a state-level aggregate predictor in the MRP model, yields improved state-level MRP estimates on cluster-sampled data. However, pooling may be an unhelpful solution to some researchers if their questions of interest were not asked on several polls or if one expects responses to change over time. In some ways, even researchers who might consider pooling may find these solutions unpalatable, as one of the main advantages of MRP is that it made pooling across several polls unnecessary. For these researchers, modeling the congressional district as an additional level in the multilevel model may be another solution that performs comparably to pooling multiple surveys. Many researchers rely on the rich datasets of the GSS and ANES, and even modified use of MRP could help political scientists to consider several questions concerning state level public opinion.

While using an aggregate predictor and pooling data over a few polls seem to increase the validity of the data, further research will be required to determine to what extent it is appropriate to use MRP estimation techniques on data collected through cluster sampling. Given data restrictions to protect the privacy of respondents, we still know very little about what clusters are chosen, and thus cannot incorporate this information into a model. Further analysis might show that the number, size, and other characteristics of the geographic clusters chosen impact the validity and efficiency of MRP estimates. It may be the case that such estimates are only fruitful in certain instances, or that certain adjustments can be made to increase the accuracy of MRP estimates in this situation, such as by using survey or sampling weights, though the



benefits of using these weights in the context of multilevel models is still an open question. As MRP becomes a more popular tool in political science, it is important that we understand how it operates on polls with different sampling techniques so that we can apply it appropriately in future research.

## Chapter 3

# The Polarization of Political Values: An Examination of the Heterogeneity of Support for American Core Values

### 3.1 Introduction

The study of public opinion has become an increasingly central part of our study of politics. Using both established public opinion surveys and new advances in survey methodology, public opinion researchers have painstakingly studied what the American public thinks and how those opinions influence the political landscape. The breadth and depth of this field of research have provided a rich understanding of American public opinion.

Much of the focus, however, of this school of research has been on either the use of broad measures of ideology and partisan identification or, more recently, more narrow measures of issue-specific opinion. Less empirical energy has gone to a crucial middle ground – the study of the political values that are at the foundation of our political life. These values, however, are an important key to understanding the public's political attitudes. Political values are more specific and meaningful than ideological constructs, and they also are crucial factors in informing issue-specific opinion. While more theoretically- and historically-inclined social scientists have

discussed these values at great length, few empirical studies have considered political values, thus missing a crucial element in the study of American politics.

Those scholars who have studied political values as well as broader strains of political culture and tradition have illustrated the importance of these values in shaping America's political landscape. Political values intuitively resonate as a lens through which to understand political outcomes (Hartz 1955, Smith 1993, Goren 2012) and often prove to have a strong influence on both public opinion and policy outcomes (McClosky and Zaller 1984; Feldman 1988; Gastil, Braman, Kahan, and Slovic 2011). Political values – including egalitarianism, capitalism, democracy, equality of opportunity, individualism, limited government, and traditional moral values – describe large ideas that shape the more specific policy discussions and political campaigns that political scientists study. They also can come into conflict with each other, and it is at these intersections that much of the political debate in the United States occurs.

We still, however, have much that we do not understand about political values and their role in the American political process. First, how do individuals differ in their support for political values, and how have these differences changed over time? How much do Americans actually agree on the political values that they supposedly share, and if there are differences, do these differences follow the patterns we see across other issues in our increasingly polarized country? Is support for political values stable over time? Or does support change in ways that reflect larger patterns in the political landscape?

Second, existing scholarly accounts are somewhat problematic in that most treat America as a single analytical unit. With some exceptions made for a discussion of the unique aspects of the development of the American South (especially in Hartz 1955), scholars have typically studied political values and broader political traditions of "Americans" as a whole, assuming that there is one set of political values that guide political beliefs from one coast to the other and all of the land in between (McClosky and Zaller 1984; Gastil, Braman, Kahan, and Slovic 2011, Goren 2012). While there is little doubt that Americans share certain values and cultural foundations, it is a mistake to assume that these are constant across the fifty states, or within them. No thorough quantitative study of political values at the state level has been conducted, despite the fact that

recent scholarship has repeatedly shown that states do have their own political character that varies independently of the demographic composition of the state (e.g., Erikson, Wright, and McIver 1993; Lax and Phillips 2009b; Lax and Phillips 2012).

This paper focuses on estimating measures of political values at both the individual and the state level. I begin with a discussion of political values and the importance of measuring them empirically. I then create measures of political values at the individual level. Using a three-decade time series in the American National Election Studies, I measure how individual-level support for political values has changed over time and how the relationship between values and other politically relevant factors has changed over time. Next, I create measures of political values at the state level, which I use to demonstrate empirically that there is disagreement across states on “American” political values, arguing that we cannot treat America as a single analytical unit when we discuss support for core political values. This paper concludes with a discussion on the heterogeneity of political values and how they are one important sign of partisan sorting and political polarization in modern politics.

## **3.2 Political Values**

Values differ from other concepts such as ideologies, preferences, and attitudes. Before we can understand the role political values may play in shaping political debate and how they relate to the increased partisan polarization we currently see in American politics, it is important to define political values and understand why they make up a key component of the American political landscape.

### **3.2.1 What Are Political Values and Why Do They Matter**

Rokeach (1973) defines a value as “an enduring belief that a specific mode of conduct or end-state of existence is personally or socially preferable to an opposite or converse mode of conduct or end-state of existence.” Schwartz elaborates on this definition by explaining that a value is a belief concerning “desirable end states or modes of conduct that transcend specific situations,

guide selection or evaluation of behavior, people, and events, and is ordered by importance relative to other values to form a system of value priorities" (1992, 1994). In other words, values are meaningful concepts that cannot simply be collapsed onto an ideological left-right scale but are more general than preferences and attitudes, which tend to be issue-specific and do not apply to multiple situations.

In defining values, Rokeach and Schwartz focus on what would more specifically be called personal values – concepts such as benevolence, conformity, security, and hedonism. Both Rokeach and Schwartz's work have shown relationships between these personal values and support for specific political parties and ideologies in the global context. Similarly, more recent studies on the big five personality traits, which one could argue are connected to or even form the foundation of these values, have shown that personality traits can also be predictive of political beliefs; for example, valuing openness predicts liberal views whereas conscientiousness predicts conservative political views (Jost 2009, Gerber et al. 2010, Gerber et al. 2012).

This line of research has been important in understanding the relationship between values, broadly speaking, and political and ideological attitudes. This paper, however, draws attention to the more specific role of political values, which are more narrow and applied in their scope than personal values in the American landscape. Given their overtly political nature and the high profile role they play in American political debates and rhetoric, focusing on the role of these political values can provide new insights into the broader role of values in our political system. What are political values, and how can we best conceptualize them? McCann defines political values as "overarching normative principles and belief assumptions about government, citizenship, and American society" (1997, p. 565). Goren (2005) similarly notes, "Core political values reflect abstract, prescriptive beliefs about humanity, society and public affairs" (p. 881). Thus, core political values are overarching beliefs that are specific to the realm of government and the public sphere. While scholars of American politics have focused on different values in their own work, a certain core set emerges at the center of American politics, including equality of opportunity, egalitarianism, limited government, and moral traditionalism. These values shape how citizens believe they should interact with their government and how their government

should interact with them.

Well before psychologists and political scientists began to measure values empirically, political philosophers and theorists wrestled with how to explain America's political culture and values. They often focused on America's unique political history that led it to develop a set of political values that they believed were collectively held by the American public and unique to it. Often speaking in terms of the larger political culture that is composed of these values, this literature has long discussed the importance of America's founding principles in shaping our political values, both past and present, often harkening back to Tocqueville's early observations of the American republic. The modern study of political culture finds its grounding in Louis Hartz's classic work, *The Liberal Tradition in America* (1955). Hartz's work describes America as a Lockean liberal community. Focused on explaining why America failed to develop either a socialist or a reactionary tradition, Hartz argues that the liberal values of the Enlightenment have served as a starting point of agreement for American politics. J. David Greenstone (1986) elaborates on Hartz's argument by describing liberalism as a "boundary condition" that sets the framework for American political debate. While Greenstone argues that American history is a struggle between two major types of liberalism, which he terms humanism liberalism and reform liberalism, these disagreements still play out within a shared framework. Political values that we discuss in today's terms with language such as the role of limited government, freedom, and equality were central to both Hartz's and Greenstone's analyses of America's founding and governing principles. Even if we disagree on the details, these authors argue that Americans agree on the same ground rules, and these political values make up our shared political culture.

These theoretical and historical works focus on the guiding nature of political values on developing American institutions and laws. More recent studies have continued in this vein, focusing on the importance of political core values exactly because of their rhetorical place in setting the boundaries of American political discourse. America's political leaders and aspirants for elected office still repeatedly and reliably turn to value-laden rhetoric when they address the American public, just as they have throughout our history, and thus it is worth asking what the American public's political values are. In his work on the role of presidential rhetoric and values, David

Doherty (2008) discusses how the core values of limited government, moral traditionalism, and equality of opportunity all figure prominently in the convention speeches and debate performances given by presidential candidates from 1992 to 2004. Using textual analysis and several specific examples, he shows how candidates from both parties regularly refer to all three of these specific values during high-profile campaign events and that, through this language, are able to educate the public about their political values and positions. One would be hard-pressed to listen to a modern day political address without hearing appeals to the country's foundational political values. In explaining both our historical underpinnings and modern day politics, we see a constant thread of appeals to the political values of equality, limited government, and traditional moral values. Given this defining role, it is important to empirically measure and understand the public's true opinions on these values.

As the study of public opinion has developed, there have been empirical studies of the role of political values and culture in American politics. Several studies have looked at the political values of equality of opportunity, egalitarianism, moral traditionalism, and limited government. These values have even been the focus of a long-standing question battery in the ANES because of their importance in the overarching language of American political culture. Limited government as well as equality are both grounded in the Lockean and Enlightenment values of liberalism that so dominate the founding of the United States. Moral traditionalism, meanwhile, is rooted in the Puritan and Christian underpinnings of that founding story. While how we apply these values has certainly changed over the years as the electorate and its rights have expanded, research has shown that these value scales are still relevant. Empirical studies looking at the ANES value scales have repeatedly shown that these are meaningful concepts: Americans have coherent and consistent views on these values scales and they are distinct from and yet relate to other concepts (such as issue-specific opinion, ideology, partisanship) in rational ways (Feldman 1988; Feldman and Steenbergen 2001; Goren 2005).

It is important to understand political values not just because they are meaningful concepts that define the views of citizens and voters but also because there is increased evidence that these political values shape public policy outcomes. There are two potential avenues through

which one might expect values to influence public policy. First, several studies have indicated that values are prior to attitudes and can predict one's specific attitudes (McClosky and Zaller 1984; Feldman 1988; Feldman and Steenbergen 2001; Gastil, Braman, Kahan, and Slovic 2011). While more work must be done in this area (Feldman 2003), the current evidence would imply that values would influence public policy by shaping issue-specific attitudes, given that there is evidence that public policy is responsive to issue-specific opinion (Lax and Philipps 2012). Thus, studying the responsiveness of policy outcomes to values might help us understand the content and influence of more issue-specific attitudes. It is also the case that representatives could be responsive to values themselves. One would expect this to be the case on less salient and more complicated issues, both of which might be more prevalent at the state level, where news coverage and political information are often quite low. Most individuals cannot be expected to have sophisticated beliefs about the minutia of policy, such as whether one should support a specific environmental regulation or provision in the tax code. Yet representatives would expect their constituents to have strong views on the size of government and preferences for equality and moral regulation. As a variant of the delegate model of representation, one might expect representatives to use values as cues for how to vote on more specific policies, just as they might use partisanship and ideology.

### 3.2.2 Homogeneity and Heterogeneity In American Political Values

Some of these empirical studies on political values explicitly echo the more theoretical literature to argue that Americans fundamentally share the same political values. In *Culture War? The Myth of a Polarized America*, Fiorina, Abrams, and Pope (2006) argue that the political polarization that has fascinated journalists and academics alike is a myth, and that while elite polarization has certainly occurred (see also McCarty, Poole, and Rosenthal 2006), the population is not nearly so divided. Divisions that do occur are more likely to be examples of partisan sorting – individuals simply adjusting their political party to capture their existing views – than the movement toward more extreme views that is implied by polarization. They come to this conclusion both when comparing individuals as well as when comparing geographic units, mainly “red states” and



“blue states.” Benjamin Page and Lawrence Jacobs (2009) make a similar argument, though focusing only on economic values. In *Class War? What Americans Really Think About Economic Inequality*, Page and Jacobs use public opinion data to conclude that Americans fundamentally agree on issues related to economic inequality and fairness, an important subset of political values in a country defined, at least in theory, by economic liberalism.

However, the recent empirical literature on public opinion forces us to question any assumption that Americans are in agreement when it comes to broad political values. In contrast to Fiorina et. al. (2006), several other studies have shown that there is evidence of polarization at the mass level (Abramowitz and Saunders, 2008). Americans are divided across partisanship, ideology, and issue-specific opinion questions (Bafumi and Shapiro 2009), due to a mix of a partisan sorting in the wake of the realignment of the Deep South and other factors as well as actual polarization. These divisions have arguably increased over the last thirty years, and it is clear that Americans are not in as much agreement as some attest (Abramowitz and Saunders 2008, Levendusky 2009). While fundamental political values may hold a symbolic place in the American narrative, it seems naive to assume that Americans would agree in their views of political values, such as equality, moral traditionalism, and limited government, while disagreeing on so many other important questions of the day, especially those that, at least in theory, should be dictated by these values. This reality seems to be heightened in each election year, where the two political parties seem not only to outline two different worldviews but to describe two different realities when describing America’s current strengths and weaknesses. To the casual observer, despite Barack Obama’s well-known assertion to the contrary at the 2004 Democratic Convention, the country does appear to be mainly made up of red and blue states, not purple ones after all.

There have been some studies that have focused on identifying heterogeneity specifically in American political values. These can be divided into two groups. The first focuses on the role of geographic heterogeneity, while the second looks to describe individual heterogeneity.

To consider geographic heterogeneity first, it is logical to presume that support for political values would differ by geographic area, just as we have red and blue states on Election night. This literature traces back to Elazar (1972); in his work *American Federalism: A View from the States*, he

argues that there are three distinct political cultures, which he names individualistic, moralistic, and traditionalistic. He argues, “The national political culture is a synthesis of three major political subcultures that jointly inhabit the country, existing side by side or even overlapping” (93). This argument was in many ways a precursor to the multiple traditions thesis of Rogers Smith (1993). Elazar, however, links these traditions to geography and historical demographic and migratory patterns, classifying the country’s regions and states as each being dominated by one of the three cultures. Elazar was one of the first to advance the concept that political culture could vary by region and state, providing an important complication to the existing literature on American political thought.

While groundbreaking, Elazar’s work is based on historical analysis and impressionistic observations rather than on more quantitative empirical analysis. Given advances in measuring public opinion over the last decades, we can now ask whether the data shows the types of variation Elazar noted. Some work has attempted to quantify Elazar’s work itself, using his cultural classification as a potential explanatory factor for differences in ideology (Erikson, Wright, and McIver 1993) and issue-specific opinion (Lax and Phillips 2012). In *Statehouse Democracy: Public Opinion and Policy in the American States*, Erikson, Wright, and McIver find, “With traditionalist states set aside,<sup>1</sup> the moralist-individualist distinction accounts for a trivial 2 percent of the cultural component of partisanship and 6 percent of the cultural component of ideology” (p. 69). Likewise, Lax and Phillips (2012) do not find that Elazar’s cultural classification has power as a predictor of state-level issue-specific opinion, when controlling for other factors such as state-level ideology and demographic characteristics. While Elazar’s categorization may be intuitive, it does not, on its own, appear to add value to explain heterogeneity in partisanship, ideology, and issue-specific opinion across the United States.

Other recent studies have re-approached the problem that Elazar sought to tackle, using their own quantitative data to classify how support for different political values varies by region. In *Our Patchwork Nation: The Surprising Truth About the ‘Real’ America*, Chinni and Gimpel (2010)

---

<sup>1</sup>These states are set aside since they are definitionally identical to the Southern and border states, which are already accounted for by Erikson, Wright, and McIver as a regional variable.

use factor analysis of Census data to divide the country into 12 regions with distinctive political cultures. They use additional data, including both consumer data and public opinion data, to validate the distinct regions they identify. Their approach, which builds off of earlier work by Gimpel and Schuknecht (2003), has several advantages. The data collection is broad and rigorous, and the regions they describe certainly seem distinct from each other in terms of face validity. Their choice of analytical unit, however, has both benefits and drawbacks. By assigning counties to regions, they can take advantage of a host of Census and polling data available at the county level. Counties also allow them to capture important differences in geography that can be obscured by state-level analysis. However, by ignoring state boundaries (and indeed ignoring the need for geographic contiguity at all), they lose an important application of their work. While states may indeed be heterogeneous themselves, they are still each a political unit whose boundaries are substantively meaningful. In our federalist system, many policy decisions are made at the state level, and many important elections occur at the state-level. Given the importance of states, it is helpful to understand political culture and political values at the state level and how they vary by state.

The second area of research on the heterogeneity of American political values has focused on measuring individual-level heterogeneity. Bafumi and Shapiro (2009) touch on this question during their investigation into partisanship and voting. In their article “A New Partisan Voter,” they show that over the last three decades, partisanship has become increasingly meaningful and important as a way to understand the electorate’s political viewpoints. Rather than viewing partisanship as an allegiance to one team or group (Green, Palmquist, and Schickler, 2002), they argue that partisanship is increasingly connected to a host of issue positions on a left-right ideological continuum. They use longitudinal data from the American National Election Study (ANES) and the General Social Survey (GSS) to show that over time, Democrats and Republicans have diverged sharply on their views on several issues and policies. As part of this discussion, they look at several of the values questions on the ANES, including those dealing with traditional moral values and economic equality. While they document that there has always been some level of disagreement on these values between self-identified liberals and conservatives as well as

between self-identified Democrats and Republicans, these differences have increased markedly from the mid-1980s to the mid-2000s. Furthermore, they view these results as a sign of increased polarization. This paper will, in part, extend this argument by examining how individual support for political values has changed over the last three decades and how we can use these changes to understand partisan sorting and polarization.

William Jacoby (2006, 2014) further examines individual variation in support for political values employing a unique methodological approach. Using an original module on the 2006 CCES, Jacoby asked respondents to rank 7 values in order of importance: economic security, equality, freedom, individualism, morality, patriotism, and social order. He then conducts a host of spatial analyses to show that there is indeed heterogeneity in the way Americans' prioritize value choices: Americans simply don't agree on how to rank the importance of these fundamental values. Furthermore, this heterogeneity is not random but rather is predictable based on factors that have often been identified as connected with the broader idea of a "culture war" in American politics, specifically partisanship, ideology, and to a lesser extent, religion and religiosity. Jacoby finds that political orientations are not only significant in predicting the rank ordering of value choices but also have the largest substantive effects, explaining more than standard demographic predictors of values such as gender and race, though he does find that age and education do have predictive value as well. Jacoby concludes that the alignment of partisanship and ideology with values serves as evidence of a multiple traditions framework theorized by Wildavsky (1987), Smith (1993), and others. He cautions, however, that such evidence of polarization may or may not be evidence of a culture war itself, given the varying level of predictiveness of religion and religiosity in his spatial models.

Jacoby's use of rank ordering to measure values allows us to understand important relationships in the study of American political values and their connection with partisanship. However, while the rank order analysis captures an important reality, it is not the usual way that political scientists measure political values. Traditionally, political scientists have measured these values by allowing survey respondents to say to what extent they agree or disagree with several statements. Rather than forcing people to prioritize their values, it allows them to support values

simultaneously, as is often the case in American politics (Smith 1993).

Another challenge of Jacoby's work is that it examines a snapshot in time. To truly understand the relationship between political values and partisan polarization, it is instructive to see how support for political values has changed over time. Polarization is not a constant of American politics but rather a concept that has slowly emerged over the last four decades (Abramowitz and Saunders 2008). Given this, it is helpful to see to what extent support for values - and who supports which values - has changed over time. One might expect values to be stable over time given their more foundational nature, and this has been shown to be the case in panel studies (Inglehart 2006). However, if the role of values has changed in the American electorate over the last three decades in systematic ways, this would be strong evidence of the emergence of at least increased partisan sorting, if not also polarization and a culture war, either due to change at the individual level or the generational level. The only major data series that asks about Americans' support for values does so by asking survey respondents the degree to which they agree or disagree with several value statements, rather than providing a way to rank order these values, and this is the dataset this paper will use to investigate to what extent political values have changed over time.

This paper examines what values mean in an age of increased partisanship. Are these truly foundational American values, or are they nothing more than talking points to be appropriated by each political party as they see fit? Does support for these core political values vary across individuals and across states, and if so, how does this variation help us understand American politics today? To begin to answer this questions, I first use a 30-year time series to see how individual-level support for values has changed over time. Then I use an original CCES survey module to examine heterogeneity of support for political values across the 50 states.

### 3.3 Individual Political Values

#### 3.3.1 Data and Measurement

The richest source for understanding Americans' political values over time is the American National Election Study (ANES). Fielded every 2 to 4 years since the 1950s, this survey provides one of our greatest sources of longitudinal data on Americans' political beliefs. The ANES began asking a series of questions on political values including moral traditionalism, equality, and limited government in the mid-1980s. These scales, which are listed below, have been the basis of much of the existing research on values in political science (Goren 2005, Feldman 1988, among others) and provide a rich dataset that can be used to study these values.<sup>2</sup>

Following other work in political science, I use the ANES data to measure political values over time. These survey questions ask respondents how much they agree or disagree with several broad statements about political values, including equality, egalitarianism, moral traditionalism, and limited government. When surveys ask about political values, they typically ask several questions about each, allowing researchers to develop latent measures for the value in question, rather than relying on a single survey measure; by aggregating several survey responses up to a single index, we are capturing the respondent's belief in a broader attitude, rather than the idiosyncrasies related to how a single question may be interpreted (Piazza 1980). This analysis focuses on four values measured by the ANES over the last 3 decades – moral traditionalism, equality of opportunity, egalitarianism, and limited government. I summarize the questions composing each value index here:<sup>3</sup>

- **Moral traditionalism** (4 questions, 5-point scale): Newer lifestyles contribute to breakdown of society, should be more emphasis on traditional family, should adjust view of moral behavior to changes, tolerance of different moral standards.
- **Equality of Opportunity** (3 questions, 5-point scale): Society should ensure equal oppor-

---

<sup>2</sup>The series on moral traditionalism has been asked in every year the ANES fielded its main survey between 1986 and 2012 except for 2002. The two series on equality of opportunity and egalitarianism have been asked in every year the ANES fielded its main survey between 1984 and 2012 except for 1998 and 2002. The limited government series has been fielded in 1990, 1992, 2000, 2004, 2008, and 2012.

<sup>3</sup>Full survey questions are included in section B.1 of the Appendix.

tunity to succeed, big problem that not everyone has equal chance, US would have fewer problems if everyone treated equally.

- **Egalitarianism** (3 questions, 5-point scale): We have gone too far in pushing equal rights, not a big problem if some have more of a chance in life, should worry less about how equal people are.
- **Limited Government** (3 questions, 2-point scale): The less government the better, the free market can handle economic problems without government involvement, government is bigger because it has gotten involved in things people should do for themselves.

For each set of values, I use the component questions to create a summary measure of latent support for the value, based on each individual's mean level of support for that value.<sup>4</sup> Averaging the respondents' opinions over several items allows us to capture their latent support for each value (Feldman and Steenbergen 2001; Goren 2005; Gastil, Braman, Kahan, and Slovic 2011). It is worth noting that while value scales should provide increased reliability over any single survey response item, there are also valid concerns that survey item scales can mask differences in the behavior across the individual items in the scale (Piazza 1980). While no scale is perfect, these scales have been well-tested during the thirty years they have appeared on the ANES and are considered standard measures of political values. I reproduce the Cronbach's alphas for these scales by year and overall in section B2 of the appendix. While these vary incrementally over time, they show that the scale items do correlate well and in general are capturing the same latent value.<sup>5</sup>

It is important to note that, unlike personal values, political values are typically not measured through a rank-ordering structure but instead are measured through the same survey instruments used to measure attitudes and ideology. Some may view this as a downside, arguing that voters rank support for political values in meaningful ways, just as people do with personal values (see Jacoby 2014). The refusal to force a ranking system, however, may accurately capture

---

<sup>4</sup>Each question is reordered so that every item in the scale is coded in the same direction. Given that non-response was quite small for these questions, individuals who refused to answer were dropped from the analysis and individuals who volunteered "did not know" were coded as neither agreeing or disagreeing.

<sup>5</sup>There is some debate in the literature on political values regarding whether the equality and egalitarian scales should be combined or separated, as well as if the moral traditional scale should be broken into two components, one on family values and another on tolerance. Based on the alphas of the overall and component scales as well as principle components factor analysis, I choose to treat the equality of opportunity and egalitarian scales separately while leaving moral traditionalism as a single scale. Please see Appendix B.2 for additional discussion.

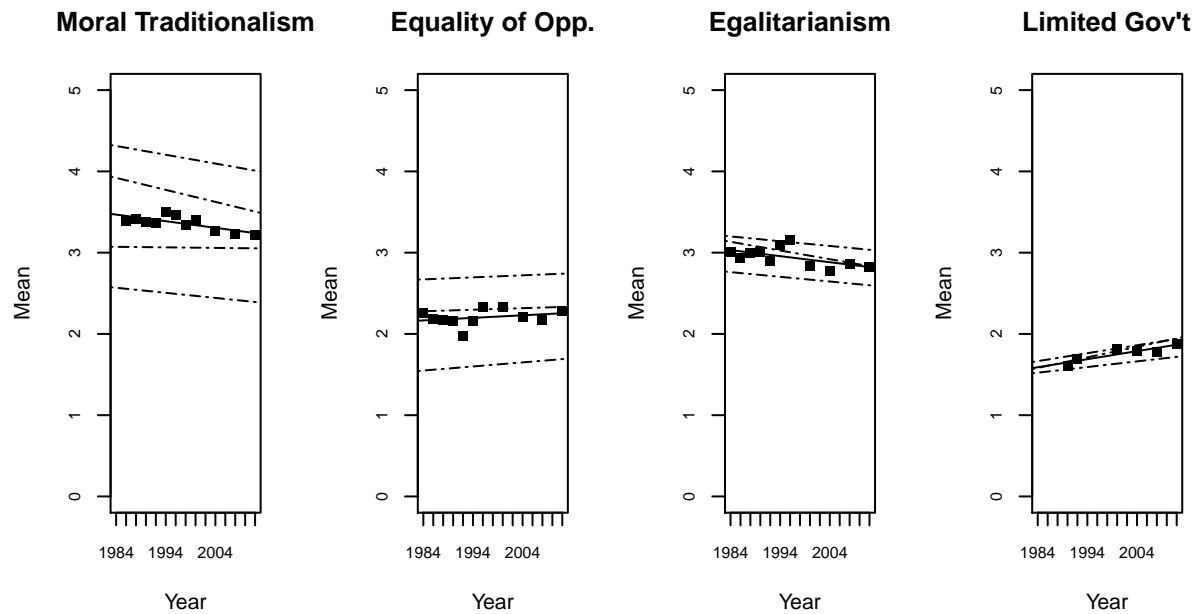
the conflicting nature of many of the core political values that dominate American politics. As Rogers Smith (1993) notes, the multiple traditions that weave through American political culture are often simultaneously held and deeply conflictual, and one can find examples of these contradictory values throughout American history, dating back to the Founding Fathers' insistence on equality while supporting America's "original sin" of slavery. This ambivalence is not something to gloss over or ignore by imposing a forced choice; rather, it is a necessary part of the American political identity that we must allow for in our measurements. Measuring values as individual attitudes or as attitudinal indices rather than as ranked choices allows researchers to capture this ambivalence empirically.

### 3.3.2 Individual Political Values Over Time

Before delving into the data, it is interesting to look at how political values have changed over time. Figure 3.1 shows the mean of each values scale for each year, represented by a black square, as well as a solid black line of best fit showing the trend over time. A higher mean signals a more "conservative" position. The dashed lines show the line of best fit over time for each individual component of each values scale. Overall, we see that Americans have become less conservative on questions of moral traditionalism and egalitarianism while becoming more conservative on questions of limited government and staying roughly constant on their views toward equality of opportunity. As one can see by the slopes of the trend lines, however, none of these values have seen a particularly large shift in aggregate support over the last three decades; support for equality of opportunity has barely moved over time, and the other three values scales see shifts in their means of a few tenths of a percentage point on the 5 point scale, or the equivalent of 4 to 6 percentage points if we were to convert these numbers to a percentage scale. While this might mean that we should expect a fairly static picture of support for values over time, it could also be the case that while the aggregate has not shifted, individuals or specific demographic groups have. These graphs also provide additional face validity to the cohesiveness of the value scales, as we see that the components of each values scale move in the same direction as each other and



Figure 3.1: Support for Political Values Over Time (1984-2012)



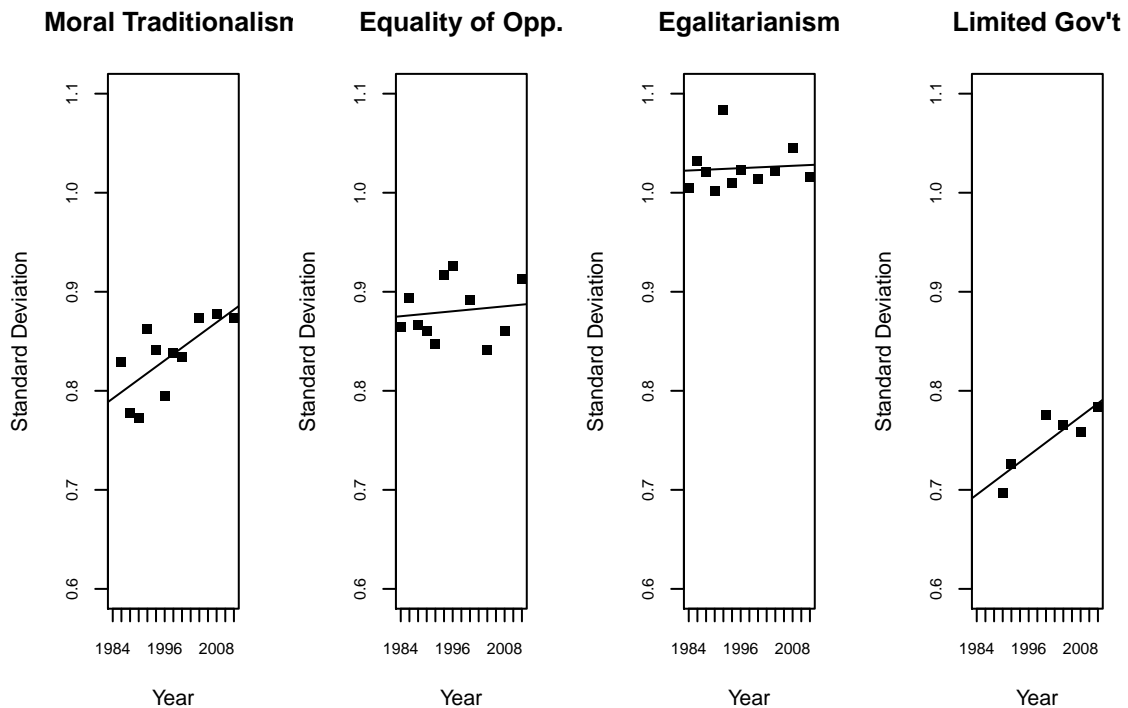
Note: Figure 3.1 shows the mean of each value scale for each year, represented by a black square, as well as a solid black line of best fit showing the trend over time. A higher mean signals a more conservative position. The dashed lines show the line of best fit over time for each individual component of each values scale.

as the aggregate scale over time.

Given the nature of this data series, one simple way to see if people's positions on values have become more polarized is to look at whether or not the standard deviation of each values index has changed over the years. The larger the standard deviation, the more disagreement there is on each value scale and therefore the more polarization there is on each set of political values (Bafumi and Shapiro 2009). As shown in Figure 3.2, the standard deviation for each values scales does indeed increase over time. The increases on the moral traditional scale and on the limited government scale illustrate this clearly; the evidence of this trend on the equality of opportunity and egalitarian scales is smaller but present. Increasing standard deviations over time are thus one hint of the increasing polarization of political values.

Another way to understand the role of political values in American politics over the last

Figure 3.2: Standard Deviation of Political Values Over Time (1984-2012)



Note: These four graphs show the standard deviation of each of the political values scales over the period of time that each question was asked on the ANES. For each political value, its standard deviation has increased over the last three decades.

thirty years is to see to what extent each value correlates with self-reported partisanship across respondents. If these values are truly foundational American values, we might not expect there to be a strong correlation between these values and partisanship – Americans should agree on these values despite deepening partisan fissures. However, if Americans are divided on political values as much as they are on more specific political issues, than we might see a stronger correlation between political values and partisanship. It may also be the case that the strength of the correlation has changed over time; if the strength has increased, this would be another sign that political values are increasingly related to partisanship and polarization.

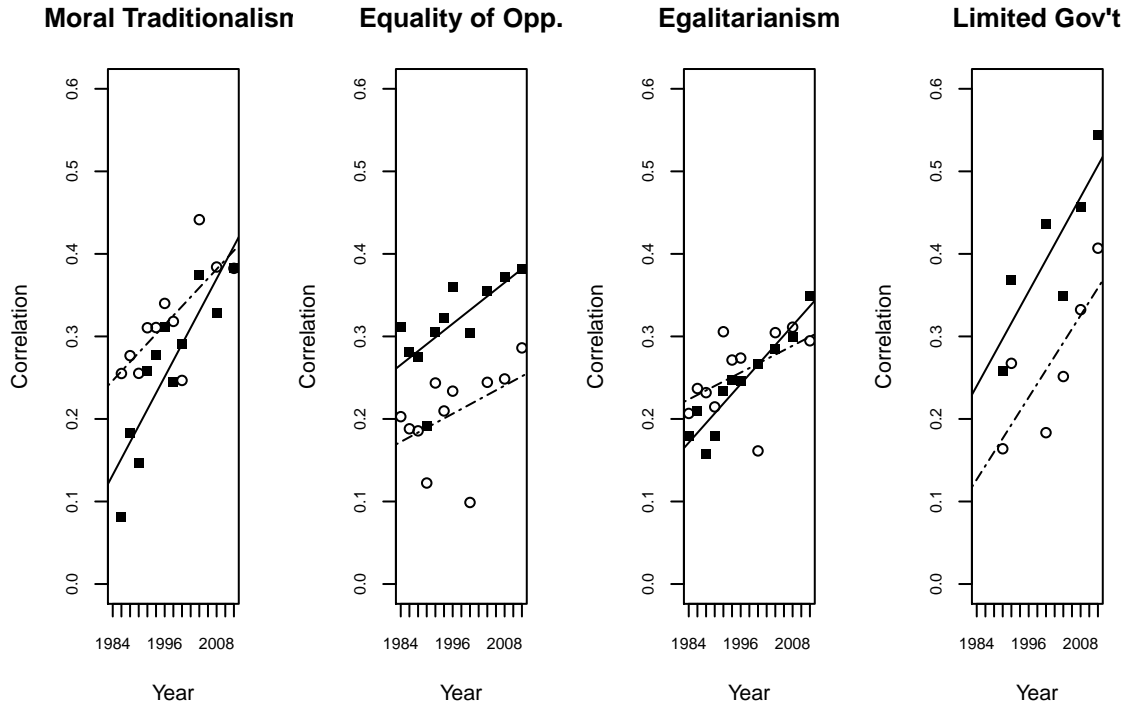
In fact, political values have become increasingly correlated with partisanship over the last three decades. As Figure 3.3 shows, all four value scales show an increasing correlation with

partisanship that trends in a clear pattern (depicted by the solid black line). In the 1980s, none of these values were even moderately correlated with partisanship, but this changed as time passed and as the country became more polarized. Moral traditionalism shows the starkest increase in correlation; in 1986, the correlation between partisanship and moral traditionalism was only 0.08, but by 2012 it had risen to 0.44. This sharp increase makes sense, given that there was little difference between the parties on cultural issues until the “Culture War” of the 1990s and 2000s heated up, especially due to the embrace of Christian conservative issues by the Republican party. The correlation between equality of opportunity and partisanship as well as egalitarianism and partisanship also increased as well. Likewise, limited government showed an increase in correlation, from 0.26 in 1990 to 0.56 in 2012. While the role of government has been a consistent divide between the Democratic and Republican parties, this could reflect the increased sorting that is reflective of patterns noted by polarization scholars. Likewise, the high correlation in 2012 could reflect the rise of the Tea Party and increasing salience of the role of government in today’s politics.

One might ask if this increase is due to increased partisan polarization or only to increased partisan sorting, as people better align their partisanship with their ideological preferences. The dashed lines in Figure 3.3 represent the correlation between each political value and ideology over time. While this correlation does increase over time for each value, mirroring partisanship, the increase is less stark in each case except for the limited government scale, where it is of a similar magnitude. This implies that while ideological sorting is clearly part of the story, partisan polarization appears to capture these changes over time more completely.

These increases in correlations and standard deviations reflect a strengthening relationship between values and partisanship. In order to improve our understanding of how the relationship between values and partisanship has changed over time, it is necessary to look at the relationship in the larger sociopolitical context. Do we still see the same relationship between political values and partisanship over time when we take other factors into account, such as an individual’s sex, age, education, and other demographic variables? Or do more basic demographic factors explain

Figure 3.3: Correlation of Political Values and Partisanship and Political Values and Ideology Over Time (1984-2012)



Note: These four graphs show the correlation between each of the values scales and partisanship (solid line) and ideology (dashed line) over the period of time that each question was asked on the ANES. For each political value, the correlation with partisanship has increased over the last three decades.

these patterns?

To answer these questions, I regress each political value index on partisanship, controlling for a host of other demographic variables that in the past have been shown to be related to either partisanship or political values (or both). I also control for ideology so that we can compare how the relationship with partisanship and each value has changed over time compared to that same relationship between ideology and each value. For each year that the ANES asked political values questions, I run the same ordinary least squares regression. I then plot the coefficients from these regressions with their 95% confidence intervals over time and a linear line of best fit.<sup>6</sup>

<sup>6</sup>Each coefficient has been standardized for ease of comparison across coefficients (Gelman 2008).

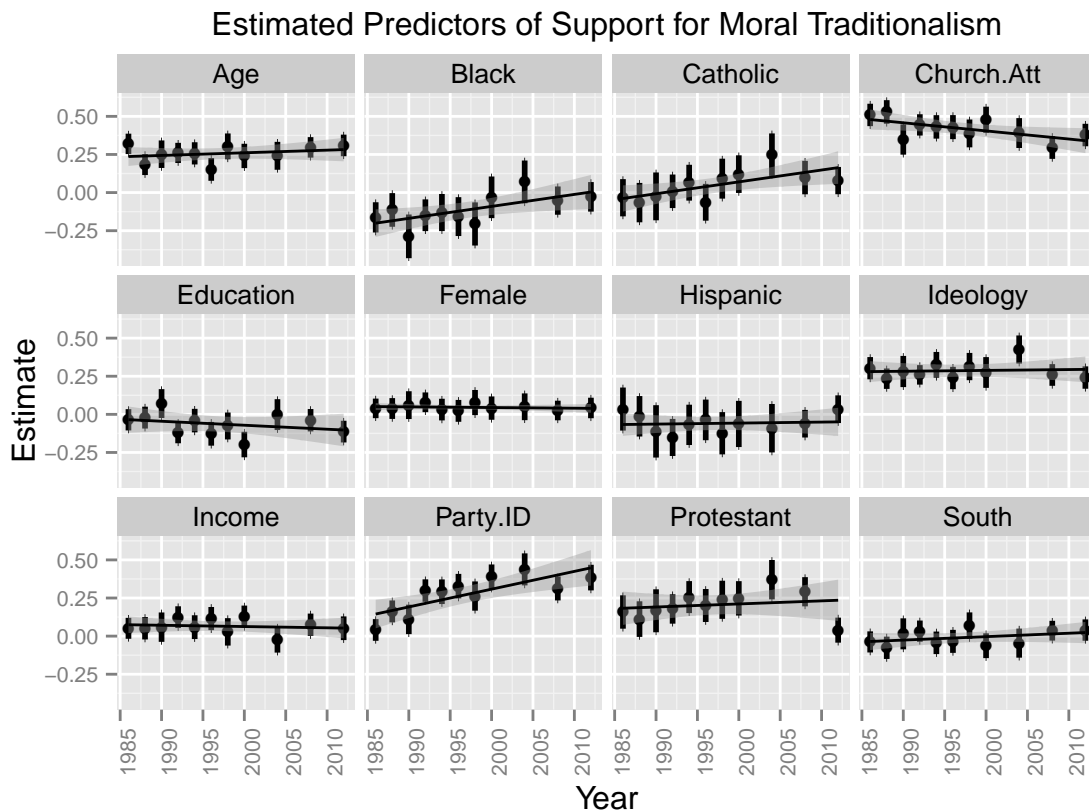
These regressions allow us to begin to answer several questions. Do demographic characteristics predict differences in support for each political value? Furthermore, has each characteristic's role in predicting a respondent's political values changed over time? And does partisanship stand out as a predictor of support for a political value?

I begin by looking at how the relative predictors of moral traditionalism vary and how they have changed over time. Figure 3.4 shows the coefficient plot for moral traditionalism for each variable in the regression equation. Many of the demographic variables behave as we would expect over time, and for many of the demographic variables, there is little variation over time. Increased age significantly predicts increased support for moral traditionalism, and this has not changed much over the last three decades. The same can be said for church attendance and protestantism, though both have seen some movement over time. Blacks and Catholics had been relatively less supportive of moral traditionalism thirty years ago, but support has increased steadily over time. For Catholics, this increased support of traditional moral values may reflect or even help explain their slow electoral shift from the Democratic Party, the home of Catholics since Roosevelt and especially Kennedy, toward the Republican Party, especially given evidence that issues like abortion have motivated individuals to switch parties (Carsey and Layman 2006). This trend is more surprising for African Americans given their steadfast support of the Democratic Party, though perhaps also not surprising given high profile opposition of some African Americans toward issues such as gay rights. In the reverse direction, increased education has gone from having no predictive relationship with moral traditionalism toward leading people to be less supportive of traditional moral values over time. Perhaps unsurprisingly, there is almost no change in the importance of ideology over time, though it is important in predicting support for moral traditionalism throughout this time period.

The relative change and impact of these demographic variables, however, is dwarfed by the role of partisan identification in predicting support for moral traditionalism. While there was minimal predictive value of partisanship on support for moral values in 1986, it has become an important predictor in recent years and indeed is the largest predictor in this model by the end of the time period. The increasing magnitude of the coefficient stands out in comparison to

the other demographic predictors. Given that the predictive value of partisanship has changed much more than the predictive value of any demographic variable, it seems more likely that this reflects increased partisan polarization of the electorate, rather than mass changes in the beliefs of the electorate at large. While such an analysis cannot shed light on causation, it is clear that partisanship and support for traditional moral values have become increasingly and strongly intertwined over the last three decades.

Figure 3.4: Coefficient Plot of Estimated Predictors of Support for Moral Traditionalism (1986-2012)

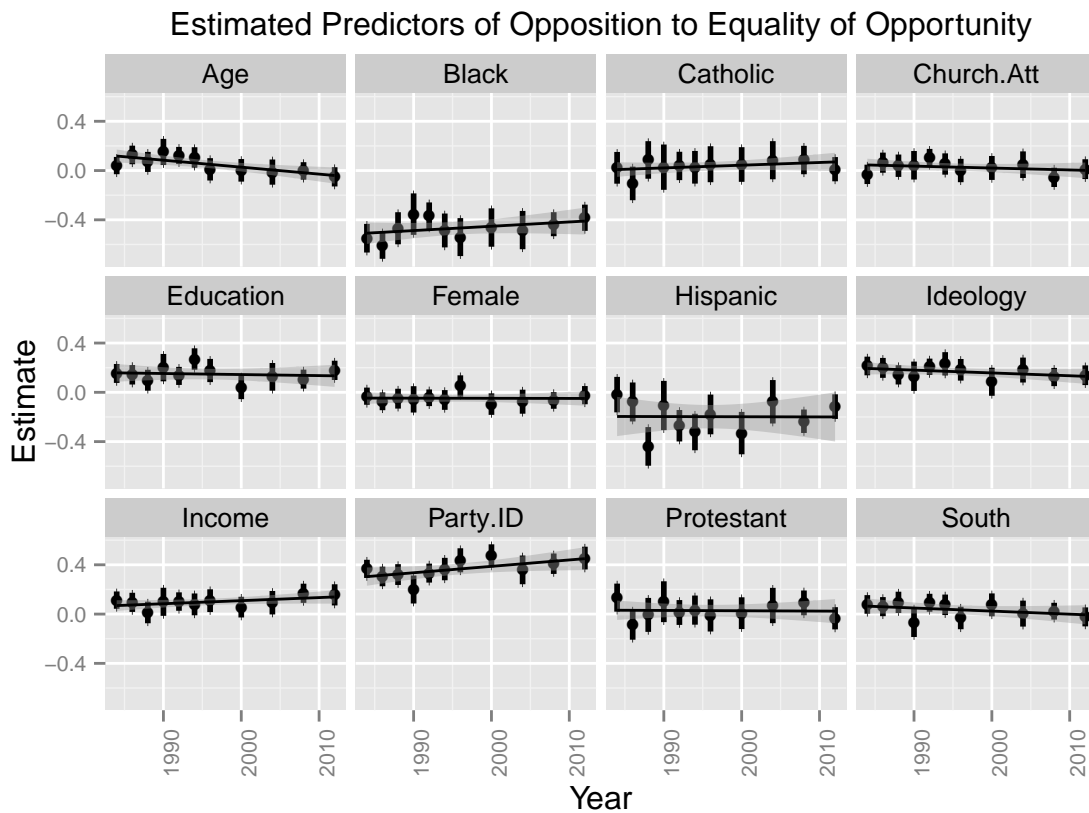


Note: This figure plots the coefficients from regressions over time of a moral traditionalism index on several variables, with their 95% confidence intervals over time and a linear line of best fit. Each coefficient has been standardized for ease of comparison across coefficients.

Figure 3.5 shows a similar coefficient plot for the value scale of equality of opportunity. To

be consistent with the graph above, these plots measure opposition to equality of opportunity rather than support for it, so that in both graphs larger coefficients indicate support for the position deemed more conservative. Again, partisanship stands out as one of the more predictive variables as well as one of the few that has changed the most in its role over time. There is less of a change than there was for moral traditionalism, which is not surprising given that equality of opportunity is the value scale where the American people come closest to consensus.

Figure 3.5: Coefficient Plot of Estimated Predictors of Opposition to Equality of Opportunity (1984-2012)



Note: This figure plots the coefficients from regressions over time of an equality of opportunity index on several variables, with their 95% confidence intervals over time and a linear line of best fit. Each coefficient has been standardized for ease of comparison across coefficients.

Figure 3.6 plots the same coefficient plot for the egalitarianism scale, which comes closest to

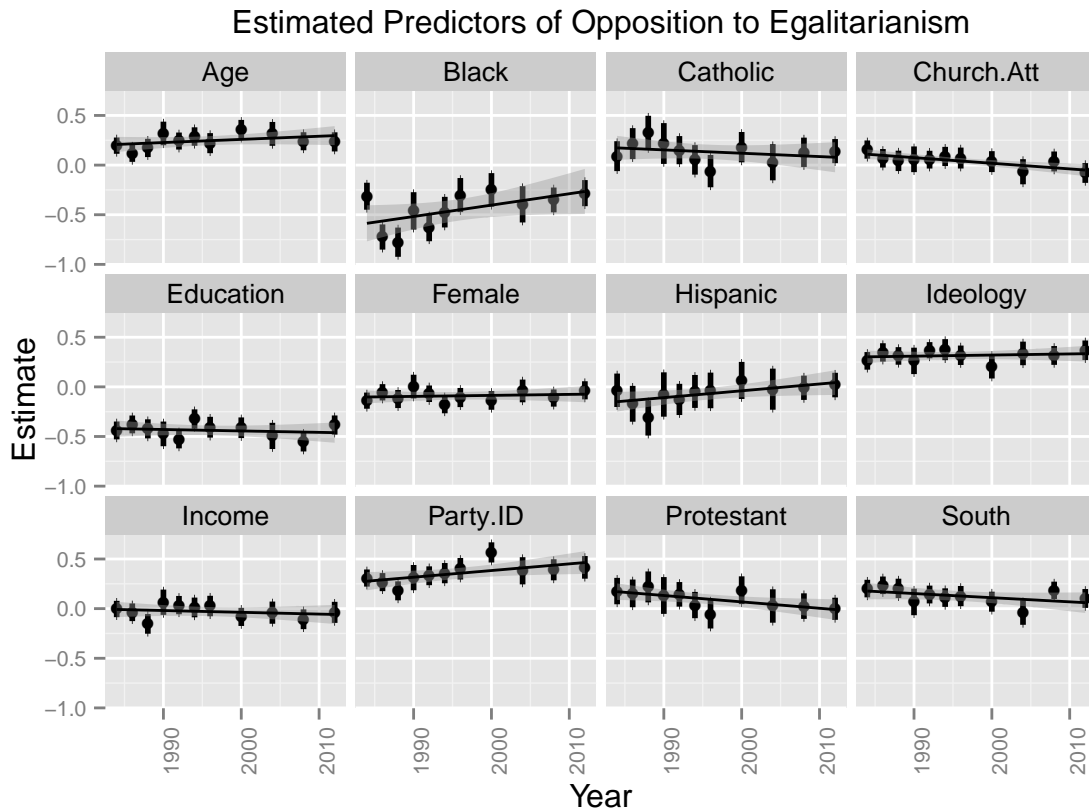
a proxy of what are generally considered civil rights issues. As with equality of opportunity, this is plotted so that more positive coefficients indicate opposition to egalitarianism. As one might expect, we see more movement over time on this scale than we do with the equality of opportunity scale, given less consensus on this issue. Overall those with higher education and minorities are most support of egalitarianism (though less so over time, which is admittedly surprising), whereas more advanced age is one of the stronger predictors of opposition to egalitarianism. As in the other graphs, party identification correlates as we would expect, with self-identified Republicans being more likely to oppose egalitarianism, and this coefficient gets larger over the three decades of the survey, showing its increasing power as an explanatory variable. Ideology plays an important role in predicting opposition to egalitarianism, but that role is constant over time.

Last, Figure 3.7 shows the coefficient plot for limited government. While the data series for limited government is less complete, with only 6 data points, a similar pattern emerges as with the other values scales. Blacks, Hispanics, and women are opposed to limited government, and this opposition has stayed fairly constant over time. Overall, the relationships between these demographic variables and the limited government index are very flat over time. The one exception to the rule is party identification. As before, the standardized coefficient on partisanship has increased substantially over the last two decades, and in the direction we would expect, with Republicans being increasingly in favor of limited government and Democrats increasingly opposed. The increase in partisanship's predictive value between 2008 and 2012 is especially relevant and not surprising, given the rise of the Tea Party beginning in 2010 and its overall (if somewhat controversial) support of the Republican party, though what we can conclude here is obviously limited with only two data points.

Across all four sets of value measurements, we see that partisanship plays a strong and constantly increasing role in predicting the public's support for political values over time. Furthermore, the changing role of partisanship stands in contrast to the fairly constant role of most other predictors over time. While it is unclear whether this is evidence of true polarization or simply



Figure 3.6: Coefficient Plot of Estimated Predictors of Opposition to Egalitarianism (1984-2012)



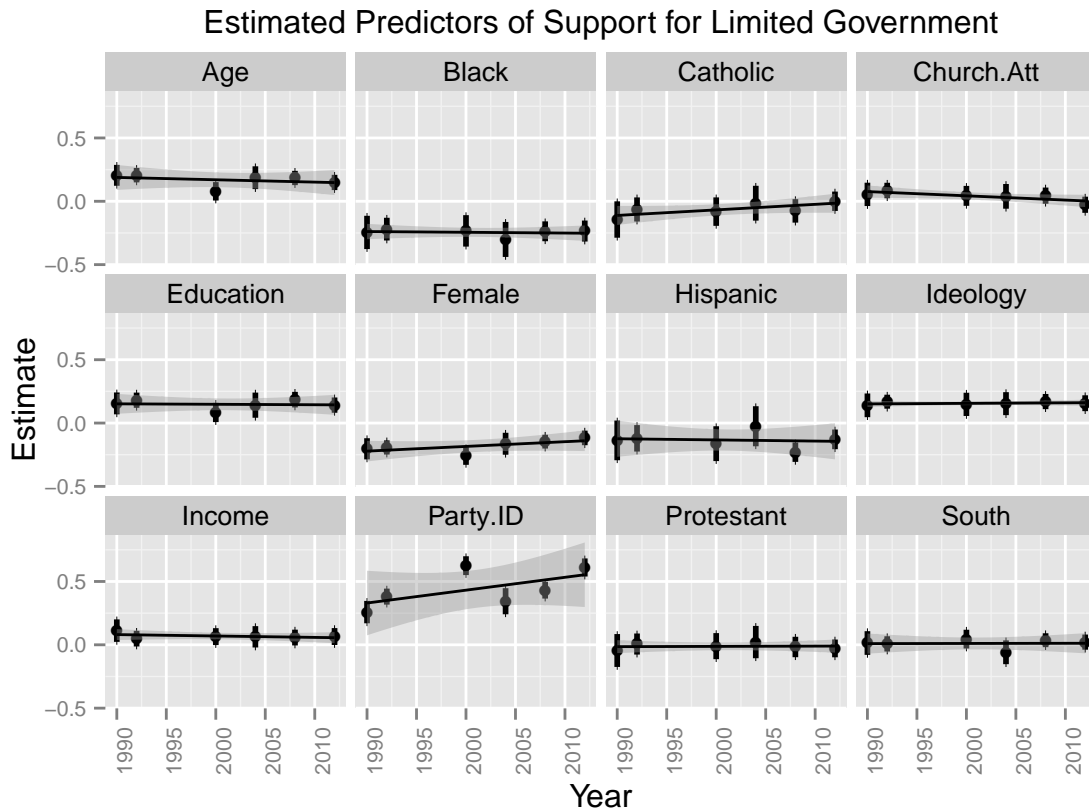
Note: This figure plots the coefficients from regressions of an egalitarianism index on several variables, with their 95% confidence intervals over time and a linear line of best fit. Each coefficient has been standardized for ease of comparison across coefficients.

partisan sorting, it is clear that political values are becoming increasingly partisan over time. Rather than uniting us, political values appear to be another tool that help divide Americans into Democrats and Republicans.

Furthermore, the robustness and consistency of the sampling and interview methods of the ANES over the last three decades allows us to be confident that the changes we see over time are in fact measuring changes in attitudes and their relationship to demographic factors, rather than changes in survey techniques.<sup>7</sup> At the individual level, we see that partisanship has grown to

<sup>7</sup>However, in 2012, the ANES supplemented its traditional face-to-face surveying with a large internet panel, weakening the argument for this consistency. The analyses above exclude this internet component to ensure that

Figure 3.7: Coefficient Plot of Estimated Predictors of Support for Limited Government (1990-2012)



Note: This figure plots the coefficients from regressions over time of a limited government index on several variables, with their 95% confidence intervals over time and a linear line of best fit. Each coefficient has been standardized for ease of comparison across coefficients.

play an increasing role in predicting differences in support for political values at the individual level. No longer shared political values, these values now invoke disagreements that can at least in part be explained by differences in partisanship.

changes over time could be compared reliably. In section B4 of the appendix, I replicate the analyses above including the internet panel data for 2012. In all of these, we see that including the 2012 internet panel does not dramatically change our conclusions, but it does strengthen the relationship between partisanship and values over time. This may not be a surprising effect, given that internet panelists may be more engaged in politics and thus more polarized than the average participant. However, since we do not have internet panel data for prior years, we cannot disentangle the mode effect from other causes of the increased partisanship.

### 3.4 Mapping Political Values in the States

The analysis above shows that there is both heterogeneity in support for political values as well as a growing connection between partisan polarization and support for political values among the American public at the individual level. To understand fully the relationship between political values and partisanship, it is also helpful to look at aggregate measures of support.

An increasing body of work has used an assortment of old and new techniques in public opinion estimation to measure aggregate opinion at the sub-national level in order to understand what differences exist across geographic areas within America and across the American public as a whole. Such techniques have been applied to several political variables of interest, including partisanship and ideology (Erikson, Wright, and McIver 1993), issue-specific opinion (Lax and Phillips 2012, Warshaw and Rodden 2012, Enns and Koch 2014), political culture (Chinni and Gimpel 2010) and personality traits (Rentfrow et al 2009, Mondak and Canache 2014). These works demonstrate the insights that can be gained by understanding patterns of heterogeneity and the meaningfulness of looking at opinion at aggregate levels of geography. Furthermore, aggregate measures of public opinion have important consequences for questions of elections and representation, as they allow us to ask whether or not our democracy is responsive to the opinions of its citizens (Lax and Phillips 2012, Tausanovitch and Warsaw 2014, and Krimmel, Lax, and Phillips 2016).

In this section, I measure support for and opposition to political values at the state level using an original dataset. The analysis shows that there are large differences in support for political values across states and across demographic subgroups within states. Overall, these differences vary predictably across state and demographic lines as one would expect in a political polarized country of red and blue states.

#### 3.4.1 Data

For this state-level analysis, I draw on survey questions that were fielded as part of an original survey module on the 2013 Cooperative Congressional Election Study (CCES). The CCES is a

large-scale academic survey that is fielded annually by a consortium of academic researchers through an online survey conducted by YouGov. Respondents are selected from YouGov's ongoing panel of respondents and the sample is stratified by state and congressional district, making it ideal for the study of state and local public opinion. All respondents answer a set of common questions and are then assigned an additional module.<sup>8</sup>

As part of this module, respondents were asked to what extent they agreed or disagreed with several statements of political values. By design, these questions replicate many values questions that have previously been asked on the American National Election Studies surveys analyzed above, as well as additional questions used by Gastil et. al. (2011). Several studies have shown the utility of these measures, and thus these surveys provide an important foundation for research. Additionally, the Cronbach's alphas and the results of principle components factor analyses for this survey are detailed in section B.2 of the Appendix, confirming the coherence of these scales. The final value indices were informed by the factor analysis and constructed so that each index only has one underlying factor. The value questions are summarized as follows, with the full questions listed in section B.1 of the Appendix:

- **Moral traditionalism** (4 questions, 5-point scale): Newer lifestyles contribute to breakdown of society, should be more emphasis on traditional family, should adjust view of moral behavior to changes, tolerance of different moral standards.
- **Equality of Opportunity** (4 questions, 5-point scale): Society should ensure equal opportunity to succeed, big problem that not everyone has equal chance, US would have fewer problems if everyone treated equally, society would be better off if distribution of wealth were more equal.
- **Egalitarianism** (4 questions, 5-point scale): Not a big problem if some have more of a chance in life, should worry less about how equal people are, we have gone too far in pushing equal rights, it seems like minority groups don't want equal rights but want special rights.
- **Limited Government** (8 questions, 4 on 5-point scale, 4 on 3-point scale): Government interferes too much with everyday lives, society should make sure everyone's basic needs are met, people should be able to rely on government help when needed, it's not the government's business to protect people from themselves; the less government the better, the free market can handle economic problems without government involvement, government is bigger because it has gotten involved in things people should do for themselves.

---

<sup>8</sup>Module size varies based on the needs of the researchers. The CCES module used in this paper had an *N* of 1900 respondents.

Given the replicative nature of these questions, one might ask why I did not use the ANES time series that was analyzed in the above sections. The ANES is designed to produce a nationally representative sample. While national samples typically cannot be subjected to simple disaggregation in order to measure state-level opinion, many can be analyzed through simulation techniques that allow one to generate measures of state-level opinion. However, unlike most polls which rely on a full probability sample or a sampling method designed to mimic a full probability sample (such as the CCES), the ANES relies on a multi-stage area sampling design. Specifically, respondents are only chosen randomly from certain geographic clusters within states. There is reason to believe that area-based cluster sampling will lead MRP to produce inaccurate state-level estimates due to the idiosyncrasies of the chosen clusters, which might lead them to be unrepresentative of the state within which they are nested (Stollwerk 2017). Conversely, MRP has been rigorously tested on data collected through probability samples, as well as through online surveys that mimic a probability sample such as the CCES, and has been shown to be valid on such samples (Lax and Phillips 2009a and 2009b). Given some of the methodological challenges especially present in the ANES for MRP, it was important to replicate these questions on a survey that did not use an area-based cluster design.

### **3.4.2 Estimating State-Level Opinion using MRP**

While this survey provides a rich source of data on the public's political values, it cannot be disaggregated for research on state-level public opinion, as most national polls are designed to be representative solely at the national level. For researchers interested in the American states, there are few polls that are designed to be disaggregated to the state level (both in terms of size and sampling procedures), and those that are, such as the National Annenberg Election Study, do not focus on political values. Multilevel regression with poststratification (MRP), however, can be used to generate state-level estimates of public opinion from a national poll. Developed by Gelman and Little (1997) and tested by Park, Gelman, and Bafumi (2006) and Lax and Phillips (2009b), MRP uses simulation and demographic information to estimate state-level opinion.

In the first stage of MRP, the individual responses to public opinion questions are modeled

as a function of each individual's demographic information and state of residence. Specifically, the model is written out below for each individual  $i$ , with index markers  $j$  for race/gender combination,  $k$  for age,  $l$  for education level, and  $s$  for state.

$$Pr(y_i = 1) = \text{logit}^{-1}(\beta^0 + \alpha_{j[i]}^{\text{race,gender}} + \alpha_{k[i]}^{\text{age}} + \alpha_{l[i]}^{\text{edu}} + \alpha_{s[i]}^{\text{state}}) \quad (3.1)$$

Each of the terms after the intercept is modeled based on a normal distribution with a mean of zero and an estimated variance. Race/gender is a ten-category variable, and age and education are both four-category variables.<sup>9</sup> State is modeled based on the region the state is in, two aggregate state-level measures, and some state-level variance. The first aggregate measure is Democratic presidential vote share in 2012. Partisanship has been shown to structure political values including both moral traditionalism and equality of opportunity (Goren 2005), and state-level presidential vote share serves as a standard proxy for partisanship within the MRP context (Lax and Phillips 2009b).<sup>10</sup> The second aggregate measure I use is a religion measure that captures the percent identifying as Evangelical or Mormon in each state, given the important role religion plays in shaping political values.<sup>11</sup> Finally, region is modeled as a random effect as well.

$$\alpha_j^{\text{race,gender}} \sim N(0, \sigma_{\text{race,gender}}^2), \text{ for } j = 1, \dots, 10 \quad (3.2)$$

$$\alpha_k^{\text{age}} \sim N(0, \sigma_{\text{age}}^2), \text{ for } k = 1, \dots, 4 \quad (3.3)$$

$$\alpha_l^{\text{edu}} \sim N(0, \sigma_{\text{edu}}^2), \text{ for } l = 1, \dots, 4 \quad (3.4)$$

---

<sup>9</sup>Race is divided into 5 categories: White, African-American, Hispanic, Asian, and Other. The age categories are 18-29, 30-44, 45-64, and over 65. The education categories are less than a high school degree, high school degree, some college but no degree, and college degree or graduate degree.

<sup>10</sup>Given that this paper is curious to what extent political values map on to partisanship, it could be problematic to include a measure of partisanship in the MRP estimates. To examine this, I reproduce all the models created in this section without presidential vote share in Appendix B.6. While there is some variation, overall the results are quite similar, showing that the inclusion of this variable is not driving the empirical findings discussed in this section. The overall correlation between the value indices with and without presidential vote share in the model is strikingly high across all of the value indices.

<sup>11</sup>This data is compiled by summing the percent Evangelical Christian and the percent Mormon for each state. Data is from the "2010 U.S. Religion Census: Religious Congregations & Membership Study," collected by the Association of Statisticians of American Religious Bodies and housed online by the Association of Religion Data Archives.

$$\alpha_s^{state} \sim N(\alpha_{m[s]}^{region} + \beta^{presvote} \cdot presvote + \beta^{relig} \cdot relig, \sigma_{state}^2), \text{ for } s = 1, \dots, 50 \quad (3.5)$$

$$\alpha_m^{region} \sim N(0, \sigma_{region}^2), \text{ for } m = 1, \dots, 4 \quad (3.6)$$

In the second stage of MRP, the estimated opinion is post-stratified by state-level population, using the 2010 American Community Survey (ACS) 1-year Public Use Microdata Samples (PUMS) from the U.S. Census to learn the number of people in each state,  $N_c$ , in each demographic cell type,  $c$ . From here, MRP can be used to generate state-level public opinion estimates by weighting the opinion prediction in each cell,  $\theta_c$ , according to the state's population:

$$y_{state[s]}^{MRP} = \frac{\sum_{c \in s} N_c \theta_c}{\sum_{c \in s} N_c} \quad (3.7)$$

Equation 3.7 yields the MRP-estimated aggregate opinion for each state.

In order to capture the complex values scales presented in the CCES data, I present a minor alteration to the traditional MRP model. Rather than create an MRP measure for each question that composes a given value scale and average them after, I instead create two latent measures for each political value scale – one capturing support for the value and one capturing opposition to the value. To do this, I begin with the first question in a particular value scale, and to create the measure of support for the value, I recode that variable to a 1 if the respondent answered the question in support of the broader political value and a 0 if they did not express an opinion or disagreed with the broader value. I then stack the data and repeat this for each question in the scale; separately, I repeat this procedure for each question in the values scale in the reverse direction, measuring opposition to the value scale.<sup>12</sup> After this recoding, I then use MRP to estimate the percentage of people in each state who support the value and also the percentage of people in each state who oppose the value; because of the recoding as dichotomous measures of support of and opposition to the values, I use a logit model as outlined above. In each model, I also model question effects for each question in the value scale, similarly to how other researchers have modeled poll effects (Lax and Phillips 2009a). This strategy allows for the MRP model to

---

<sup>12</sup>The measures of support and opposition to the values scale must be measured separately and are not additive to 100% due to the presence of those who expressed neither support nor opposition for a particular value.

estimate the latent value of the political value in question directly.

A note on terminology is helpful here. In creating the measures and analyzing the results, I will often refer to the two measures of each political value scale as capturing the liberal or conservative side of the value scale, for lack of better terminology. In other words, measures that capture support for equality of opportunity and egalitarianism as well as opposition to limited government and moral traditionalism can all be described as measures of the “liberal” end of the values scale, while those measures that measure opposition to equality of opportunity and egalitarianism as well as support for limited government and moral traditionalism can be described as measuring the “conservative” end of the values scale. While this terminology is not perfect, grouping support and opposition to the values by how they generally fall within the country’s current ideological divisions allows us to examine trends for all four values that we would expect to move in the same direction with increased clarity, as opposed to looking at support for all four values (or opposition to all four values) simultaneously.

### **3.4.3 State-Level Political Values**

To examine these measures of state-level values, it is helpful to begin with some descriptive statistics. The first statistics I examine are the minimum, maximum, and range of each values measure across the 50 states. Table 3.1 shows the range of each of the four values, measured in both the liberal and conservative directions. The range of each also allows us to see where public support lies for each value.

If one were to hypothesize that there were a standard set of American values that all citizens supported and that did not vary substantively by geographic boundaries, we would expect to see this reflected in two ways in these statistics: an incredibly small range of each value scale, and majority (or super-majority) support for each value across the states. Neither of these, however, is reflected in the data. Rather, levels of support and opposition to these four core political values look quite similar to those of a host of other political issues. All of the values measures have a range of at least 15 percentage points, and five of the eight have a difference of at least 20 percentage points. Similarly, support never reaches close to unanimity for any of the values



scales; in fact, none pass a two-thirds majority. From this initial look, it does appear that support for core values varies across the states in substantial ways and does not approach universal support in any of the states.

Table 3.1: Value Scale Ranges (CCES Data)

<i>Liberal Value Scale</i>			
	Minimum	Maximum	Difference
Moral Traditionalism	25.9	46.7	20.8
Equality of Opportunity	43.9	63.8	19.9
Egalitarianism	24.8	47.6	22.8
Limited Government	28.6	47.6	19.0
<i>Conservative Value Scale</i>			
Moral Traditionalism	28.2	58.2	30.0
Equality of Opportunity	17.9	34.6	16.7
Egalitarianism	30.7	51.1	20.4
Limited Government	31.5	52.5	21.0

All minimum and maximum numbers are the average percentage of state-level population support for that value scale in a given state.

Next, it is helpful to look at the correlation matrix for these values. Are values distinct beliefs that are meaningful? Or, do all of these values simply represent the same concept – some latent form of ideology or partisanship? If so, we would expect correlations to be quite high across the scales. This is indeed what we find amongst the different values that involve economic values and the role of government. The correlations among equality of opportunity, egalitarianism, and limited government are all moderate to high whether the scales are measured in the liberal or conservative direction, with a Pearson’s  $r$  above 0.5 in all instances. This is not surprising, especially given how past scholars themselves have debated the composition of these value scales. For example, the equality of opportunity scale and egalitarian scale, as they appeared on the ANES, have been both combined and separated when they were analyzed by past scholars, and some researchers have even changed how they analyzed them across papers (Goren 2005, Goren 2012). The fact that these are all so highly correlated suggests that respondents are quite consistent in their economic value choices.

How moral traditionalism relates to these economic values, however, requires a more nuanced

interpretation. When the values are measured in terms of support for the more liberal end of the value scale, moral traditionalism has no significant relationship to egalitarianism, and its correlations with equality of opportunity and limited government are significant but lower than one might expect. When values are measured conservatively, however, moral traditionalism has a significant and moderate relationship with each value, with correlations around 0.5 for all three other scales. These results may suggest that conservative respondents' values are more ideologically constrained than those of liberal respondents, a conclusion that finds support in earlier literature at the individual level (Feldman and Zaller 1992).

Table 3.2: Value Scale Correlations (CCES Data)

<i>Conservative Value Scale</i>				
	Moral Trad	Equal Opp.	Egal.	Limited Gov.
Moral Traditionalism	1.00***			
Equality of Opportunity	0.51***	1.00***		
Egalitarianism	0.50***	0.53***	1.00***	
Limited Government	0.53***	0.83***	0.60***	1.00***
<i>Liberal Value Scale</i>				
	Moral Trad	Equal Opp.	Egal.	Limited Gov.
Moral Traditionalism	1.00***			
Equality of Opportunity	0.39**	1.00***		
Egalitarianism	0.24	0.54***	1.00***	
Limited Government	0.30*	0.68***	0.64***	1.00***

All numbers are the Pearson's correlation coefficient.

\*  $p < 0.05$ , \*\*  $p < 0.01$ , \*\*\*  $p < 0.001$

It is not enough, however, to understand aggregate variation across states to understand how similar or dissimilar public opinion in the states is. Prior studies have shown that demographic groups can vary across states in both similar and different ways, shedding increased light on geographic and political patterns (Gelman et al 2008). In the next several figures, I map how support and opposition to these four core political values varies across the states by age group, gender, race, and education level.<sup>13</sup> Rather than looking at these values by state alone, a joint

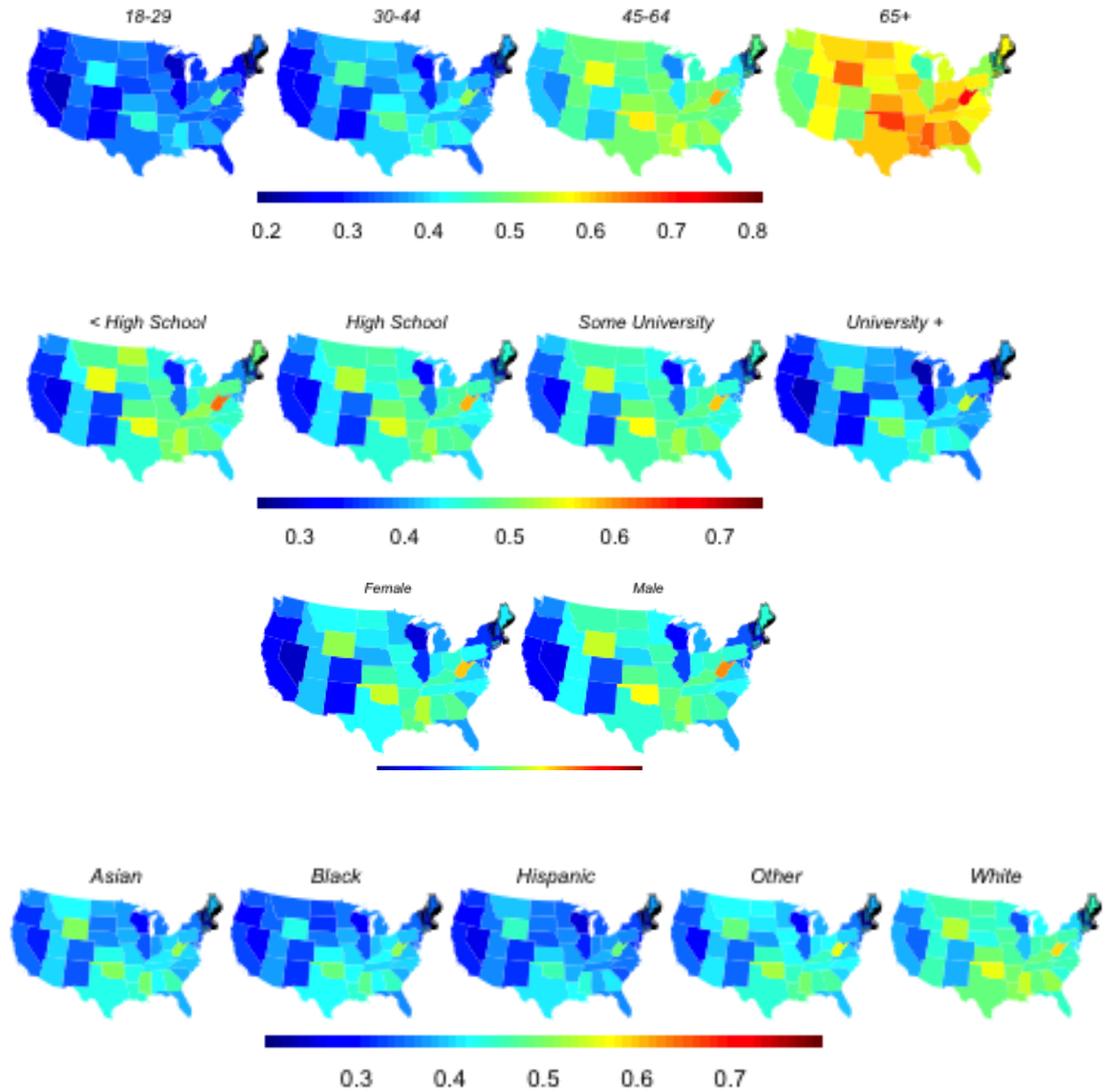
<sup>13</sup>These estimates are created using the *mrp* package in *R*. Just as with producing the overall MRP estimates, the package produces subgroup estimates by using the results of the logit model and then poststratifying those results based on the composition of the predefined population of interest.

examination of state and key demographic variables allows us to see more nuanced patterns that exist in American values. Note that in all of these maps, blue indicates lower percentages and red indicates higher percentages. For all of the graphs here, higher percentages indicate what would be the more conservative value position. We see similar but reverse patterns when looking at the scales in the liberal direction; these are shown in section B.5 of the Appendix.

Figure 3.8 shows how support of traditional moral values maps onto the 50 states across these four demographics. These maps follow the patterns we would expect, with traditionally “blue” Democratic states showing lower levels of support for moral values, and vice versa for Republican states. Additionally, it comes as little surprise that many of the most conservative states are located in the Bible Belt of the South, as one would expect, no matter which demographic segment we look at. While there are almost no differences between men and women within states, we do see differences in the other demographics. The main role of education appears to be to distinguish between those with and without a college degree, with those with a degree being much less supportive of traditional moral values than those without. We see less variation across races, but whites are the most likely to support traditional moral values. The largest distinctions, as one might guess, are those created by age, with increasing age leading to increased support of traditional moral values. One interesting aspect of this graph is how little variation there is in the younger age groups. Not only are younger people more liberal on moral values than older ones, but there are much smaller differences across states than there are within the older age groups. This seems to support current wisdom that culture war issues may be becoming less important for younger voters.

Figure 3.9 shows maps for the support of limited government and individualism. Similar patterns as above hold: Republican and Democratic states look somewhat as we would expect, and younger citizens are considerably more liberal across the country than their older peers. Here, however, we also see a distinct regional pattern emerge. The frontier states of the far west are more conservative in their support for limited government, and this holds true across all age groups, genders, and to a lesser extent, education levels. New England is also more conservative

Figure 3.8: Support of Traditional Moral Values, by State and Demographics



Note: These graphs show the percentage of people who support traditional moral values in each state by age group, education level, gender, and race.

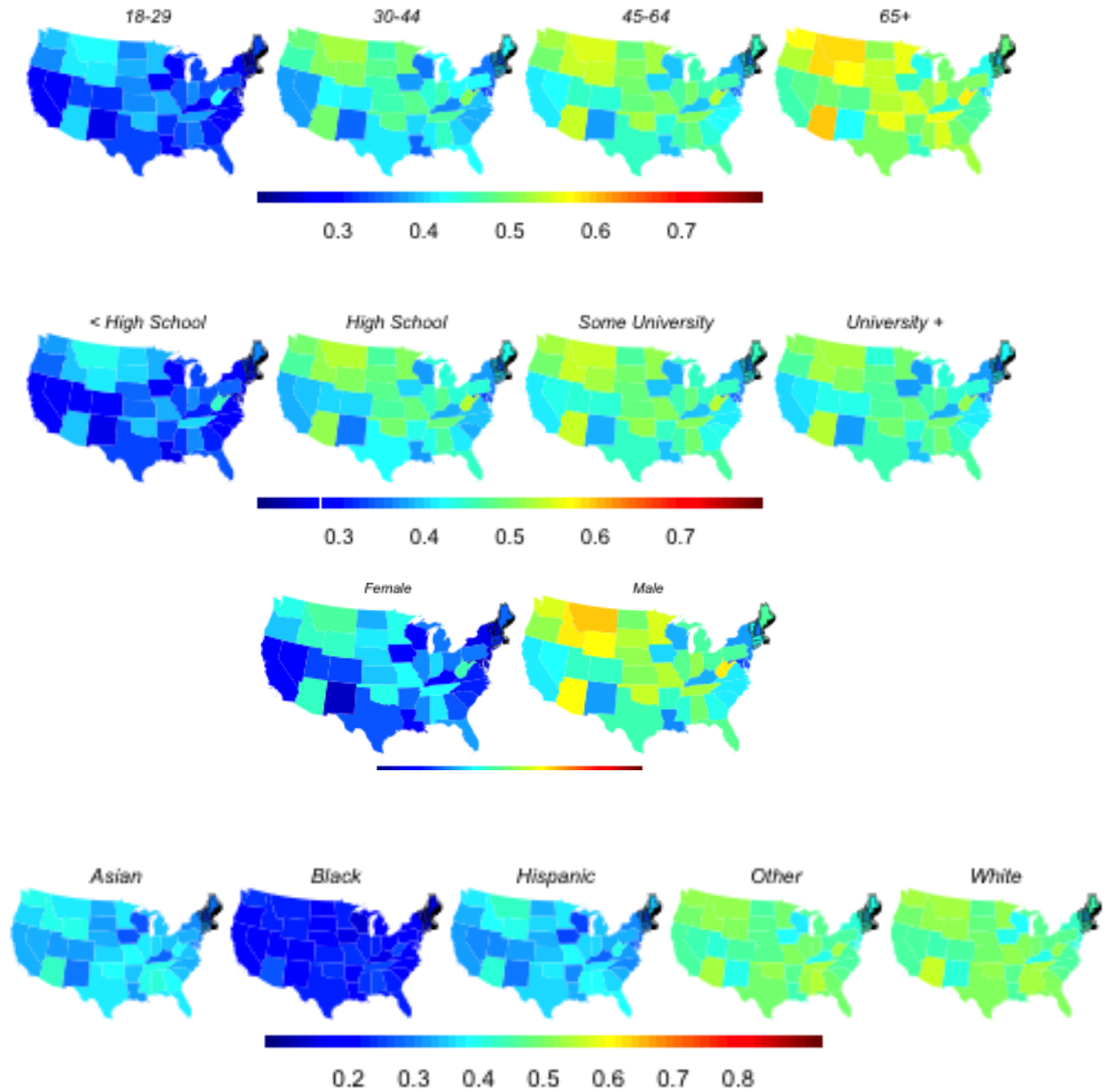
than one might expect given its typical Democratic voting pattern, whereas the South seems less conservative than one might expect; this seems especially true for Louisiana, which may indicate an awareness of the role government can play in times of emergency. When looking at support for limited government by race, however, we see less heterogeneity across the states, with African Americans being uniformly against more limited government and whites showing moderate levels of support across all states.

Figure 3.10 shows opposition to the value of egalitarianism. Across the demographic divisions, we see many patterns similar to those looking at limited government: men are more conservative than women in their views, and African Americans have very low opposition to egalitarianism, while whites are more opposed to it across the states. Unlike with limited government, however, Asian Americans look more like whites in this case than Hispanics, who remain in the middle in both cases. A university education appears to make people more liberal, though the relationship is not entirely linear among the lower education levels. Age again ends up being a strong predictor, with older Americans being more against egalitarianism than younger Americans, who appear to have fairly monolithic opinions.

Figure 3.11 examines opposition to equality of opportunity across the same demographics. These maps break from the more extreme patterns above – this is clearly the value where we see the most unanimity across geographic and demographic divisions. Whites, older Americans, and men appear to be more conservative than others, but there is much less variation overall. Unusually, here a university degree makes one slightly more conservative, but again the differences here are slight compared to those seen in the other values.

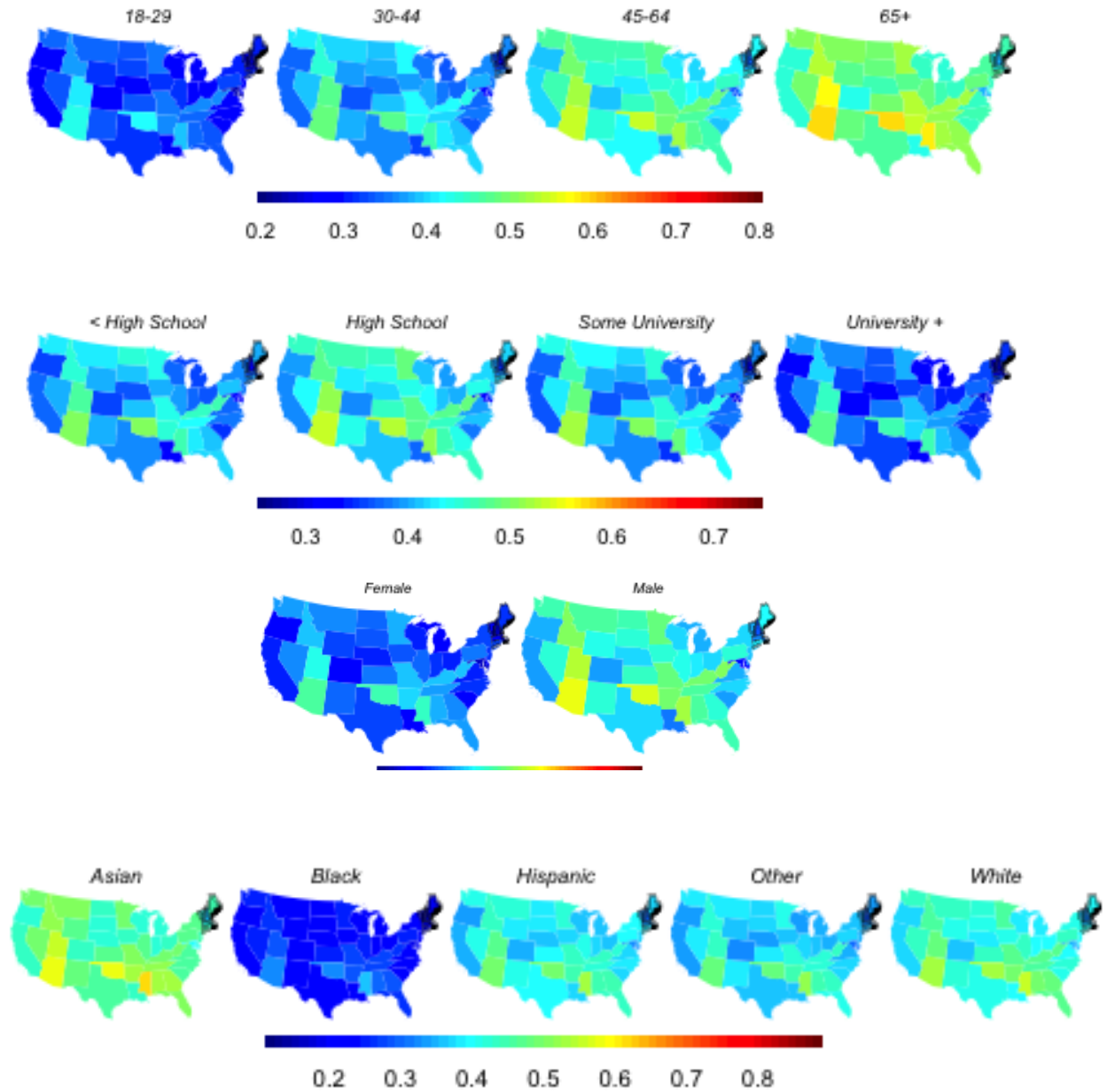
I caution, however, in interpreting too much into this universal low support for opposition to equality of opportunity. When one looks at support for equality of opportunity (where higher values indicate a more liberal position), as in Figure B.15, there is a large range, with younger voters in general being much more supportive than older voters of this supposed cornerstone of American values. Among these younger voters, there is not a truly clear partisan pattern; however, it does seem that states in the west are less supportive than those in the rest of the country.

Figure 3.9: Support of Limited Government and Individualism, by State and Demographics



Note: These graphs show the percentage of people who support the values of limited government and individualism in each state by age group, education level, gender, and race.

Figure 3.10: Opposition to Egalitarianism, by State and Demographics



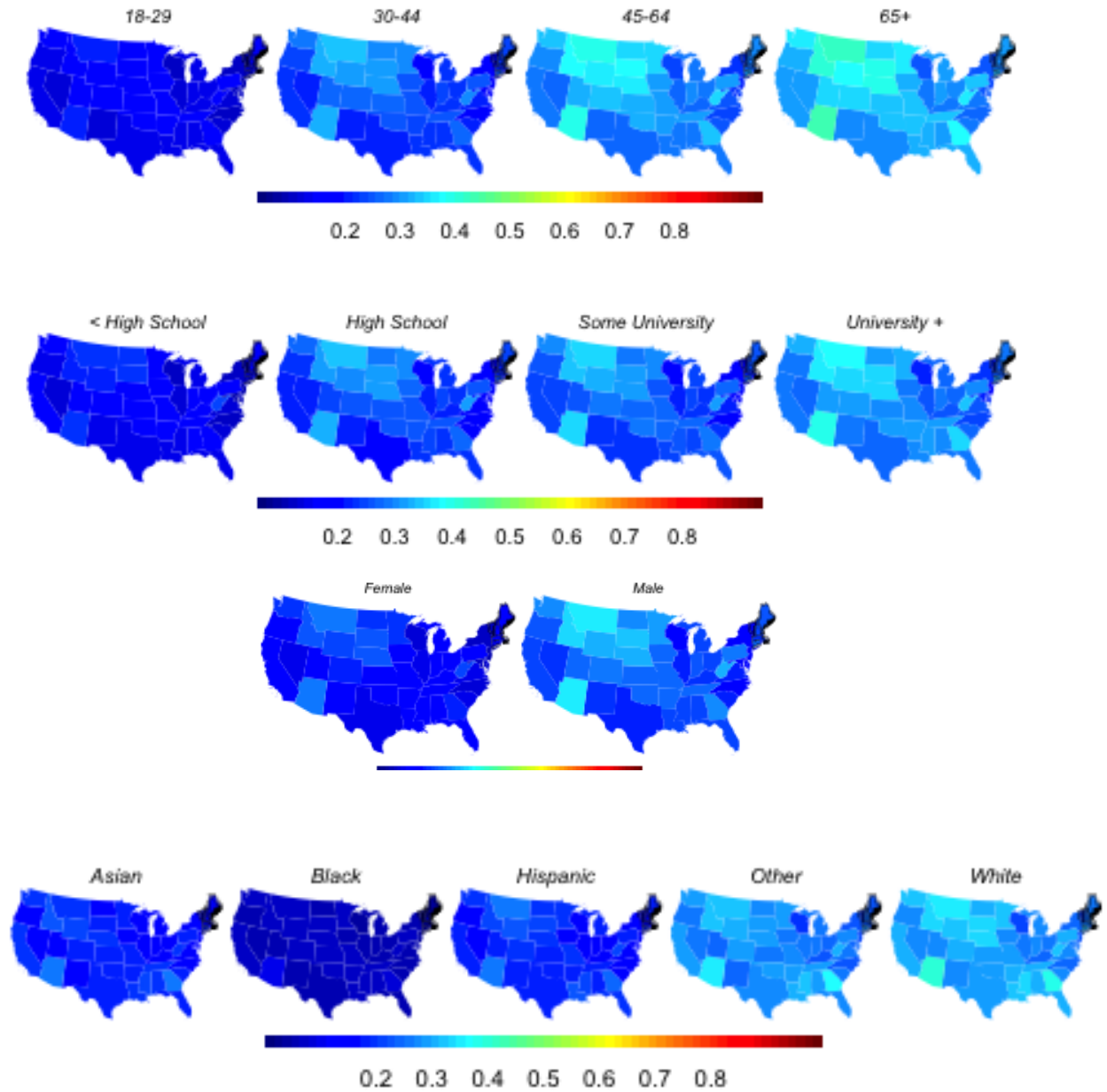
Note: These graphs show the percentage of people who oppose the value of egalitarianism in each state by age group, education level, gender, and race.

We also seem similar variations across races (with African Americans being most supportive), across education levels (with those without a high school degree being most supportive) and across gender, with women being more supportive than men. It may be that equality of opportunity is such a foundational value in how people discuss American politics that those who are opposed are more likely to state a more undecided or neutral opinion, rather than a more unpopular opinion against equality of opportunity (Berinsky 2004).

Overall, these maps show us that states are important units of analysis for the understanding of American political values. While we may share a common language when discussing values, there is strong disagreement on many of these values by state. Furthermore, by breaking the data into different demographic groups, it also becomes clear that states are not the only, or perhaps not even the most important, dimension across which our support for political values varies. Rather, age appears to be a strong predictor of one's political values, with increasing age leading to an increase in more conservative values across every values index studied here. In fact, the aggregate differences across age groups in some states are startling in some instances, with as much as a 30 percentage point gap in many instances. Similarly, we see large differences in ways we would expect across racial groups, gender, and education level of respondent. All of these differences map onto the partisan divisions that predict support for the two major parties within different demographic groups, providing further evidence that values have increasingly become entwined with the larger divisions of polarized partisanship that we see in American politics today.



Figure 3.11: Opposition to Equality of Opportunity, by State and Demographics



Note: These graphs show the percentage of people who oppose the value of equality of opportunity in each state by age group, education level, gender, and race.

### 3.5 Discussion

Political values are often at the center of the political narrative in the United States, but there is still much to understand about what our political values truly mean and how they shape the opinions of everyday Americans. In this paper, I use advances in survey research and an original dataset to understand what patterns exist and shed light on what it truly means to be guided by “American” values.

Looking at 30 years of longitudinal data on public opinion, this paper shows that individuals increasingly do not share political values. These disagreements and differences can primarily be explained by the increasingly large role of party identification. As the country has seen an increase in political polarization, disagreements over what our core political values mean increasingly fall over partisan lines. It may be the case that the increasing role of partisanship does not actually reflect an increasing role of polarization but rather a more precise sorting of people into the appropriate political party to reflect their attitudes and values. Nonetheless, the role of the two major political parties in predicting disagreement over political values is incredibly revealing. Rather than viewing these values as shared beliefs that unite us, they should be viewed as political beliefs on which we may, often vehemently, disagree with our fellow Americans.

These findings are echoed in the analysis of aggregate support for political values at the state level. When measuring support for political values across the states, we see strong differences in the average level of support and opposition to each key value. Furthermore, most of these differences can be predicted according to the standard red and blue states that the country divides into each election night. This is not to say that the states are monolithic. There are important differences within states across demographic groups. Even these, however, fall along the typical partisan and demographic patterns that we would expect in quite predictable ways. For example, young people living in more conservative states look more similar to their age cohort in liberal states than they do to older residents of their states when it comes to measuring the value of moral traditionalism.

This is not to say that these political values are not meaningful, or that they do not represent

distinct concepts. We see this in the occasions where they do depart from the typical partisan patterns, such as when the frontier states in the west along with New England show more support for limited government, a typically conservative opinion, than we might expect given their politics, broadly speaking. But to those who have studied the history of these regions, such observations are not surprising.

There is still much to learn in understanding both how we measure political values and how we understand their role in our political process. Surveys designed to be representative at the state level would allow us to make additional insights into state and regional differences in support for political values without relying on modeling tools such as MRP, which do require their own assumptions.

The study of political values would also benefit from the use of experiments to learn how values might be activated to gain support for or rally opposition to a particular cause. A special case of this is in the role of shifting the value frame. Advocates for same-sex marriage succeeded in achieving this over the last decade, switching from a discussion of moral values, which they were losing, to a discussion of equality under the law. Even if there is not complete agreement on values across the political spectrum, the power of values as a rhetorical tool to unite Americans should not be brushed aside.

Future research should also ask what role, if any, political values have in shaping public policy. Prior research has shown that at the individual level, political values often shape issue-specific opinion on public policy. Given these developed measures of aggregate support for political values, researchers can use them to examine if these political values shape state-level policies, either directly or through an intermediate force, such as by shaping state-wide issue specific opinion on specific policies. Given the partisan nature of these political values and their salience, understanding if and how they shape state-level policy is key to understanding heterogeneity in policy across the states.

Whether looking at measurements of political values at the aggregate or the individual level, the same pattern emerges. Support for political values varies along traditional partisan lines. Rather than serving as a way to unite individuals, political values are becoming another dimen-

sion upon which individuals disagree in predictable partisan ways. In an era of polarization, political values now provide another fault line across which we can express our differences.

## Chapter 4

# Are Survey Respondents Lying about Their Support for Same-Sex Marriage? Lessons from a List Experiment

with Jeffrey R. Lax and Justin H. Phillips

### 4.1 Introduction

Public support for gay and lesbian rights has risen dramatically over the past two decades. Nowhere is this more apparent than in responses to survey questions about same-sex marriage. While in the mid-1990s, fewer than one-third of Americans thought that it should be legal for same-sex couples to marry, surveys now indicate that a growing majority support marriage equality. This sea change in public attitudes has received a great deal of attention in the academic literature and has been widely reported in the media. But is it real? How certain can we be that a majority of the public now supports same-sex marriage?

Concerns arise because survey responses to potentially sensitive questions (e.g., questions about prejudice, religious attendance, drug use, etc.) are subject to a social desirability bias. That is, respondents may lie to pollsters when they believe that their true opinion runs counter to per-

ceived societal norms. As messages from cultural and political elites have becoming increasingly supportive of gay and lesbian rights, survey respondents who oppose same-sex marriage may now feel psychological pressure to conceal from pollsters their true preferences. This possibility has been raised in academic work (cf., Egan 2008; Powell 2013; Emerson and Essenburg 2013) as well as in some media outlets (cf., Regnerus 2013). The presence of social desirability bias may be particularly plausible given that opponents of same-sex marriage are now sometimes portrayed as being on the “wrong side of history.” If such a bias is present in polling, scholars, the media, the courts, and elected officials may have a false sense of the public’s opinion on marriage equality.

While traditional public opinion polls are ill-equipped to tease out the presence of these effects, a technique known as a list experiment can do so. List experiments, by design, afford survey respondents anonymity that allows them to provide truthful answers to sensitive questions. Indeed, this technique is commonly employed in the social sciences to study views or behaviors that may be difficult to measure with direct questions (c.f., Gilens, Sniderman, and Kuklinski 1998; Streb et al. 2008; Lyall, Blair, and Imai 2013). In our case, we embedded a list experiment in the 2013 Cooperative Congressional Election Study (CCES), a large online academic survey that is nationally representative. The design of our list experiment enables us to test not only whether social desirability bias is skewing overall measures of public support for same-sex marriage, but also to consider the possibility that this bias is not unidirectional—i.e., that it may lead some subgroups of the population to overreport their support for same-sex marriage, while leading others to underreport their support.

The results of our list experiment have important implications. First, they contribute to our understanding of the state of American public opinion on same-sex marriage. Because battles over legal recognition for such marriages have dominated public debate, knowing where the public stands is of crucial importance, especially given the long-established link between public preferences and policymaking (e.g., Page and Shapiro 1983; Burstein 2003; Brooks and Manza 2007; Lewis and Oh 2008). Second, our study speaks to the prevalence of social desirability bias in computer surveys. Research suggests that such surveys, because they are completed in

private, are likely to elicit truthful answers (Holbrook and Krosnick 2010).<sup>1</sup> As computer surveys become more common, it is important to evaluate the accuracy of the data they generate. Indeed, we find no evidence that social desirability bias is skewing overall survey results on same-sex marriage. If there is such bias in polling on this issue, it pushes in both directions. Furthermore, our efforts provide new evidence that a national opinion majority favors same-sex marriage and should increase confidence in the ability of computer surveys to produce opinion estimates free of social desirability bias.

We begin by documenting the dramatic rise in public support for same-sex marriage as well as the existing evidence which suggests that a social desirability bias may exist in polling on this issue. We then discuss list experiments as an approach for generating estimates of public preferences that avoid this bias. Next, we present and evaluate our experimental design. After confirming that the assumptions for a successful list experiment have been met, we present our findings, looking for the presence of social desirability bias at the aggregate level and across specific subgroups of the population. To evaluate the robustness of our results, we analyze a second list experiment, one focusing on the inclusion of sexual orientation in employment nondiscrimination laws. We conclude by more fully discussing the implications of our findings.

## 4.2 Public Support for Same-sex Marriage

In May of 1993, the Hawaii Supreme Court issued a landmark decision in *Baehr v. Lewin*. The Court held that by denying marriage licenses to same-sex couples, Hawaii was discriminating on the basis of sex, and thereby violating the Equal Rights Amendment in the state's constitution. This made the Hawaii Supreme Court the first court of last resort in the United States to issue a ruling in favor of same-sex marriage. While the legislature and voters overturned *Baehr* via a constitutional amendment, the Hawaii decision placed the issue of same-sex marriage on the national political agenda. It also engendered a backlash, resulting in the passage of Defense of Marriage Acts (DOMAs) by Congress and more than 30 state legislatures (Pinello 2006).

---

<sup>1</sup>Though a recent study by Ansolabehere and Schaffner (2014) suggests that there is no difference in social desirability bias between telephone and computer surveys.

In the years following *Baehr*, national polling firms began to sporadically measure public support for legalizing marriages between same-sex couples. The first Gallup poll on the subject was conducted in March of 1996 and asked respondents, “Do you think marriages between homosexuals should or should not be recognized by the law as valid, with the same rights as traditional marriages?”<sup>2</sup> The results were not positive for advocates of gay and lesbian rights: only 27 percent thought that such marriages should be valid. The results of this early Gallup poll were consistent with those of other reputable polling firms conducted around the same time.

As the years passed, however, support for legalizing same-sex marriage began to rise.<sup>3</sup> Figure 4.1 plots the results of nearly 100 national surveys over a 20-year period, along with the results from our own online survey conducted as part of the 2013 CCES. Average opinion rose steadily from the mid-1990s through the present, with recent polls showing a clear national majority in favor of legal recognition. Indeed, among those polls conducted in 2014, average support for same sex-marriage is just over 56 percent, with a Pew survey (from February 2014) placing support for marriage at 59%. This change in public opinion is correlated (though imperfectly) with changes in public policy (Lax and Phillips 2009, Krimmel, Lax and Phillips 2016).

---

<sup>2</sup>The wording of this question has changed very little over time. The most notable change is that polling firms have replaced the word “homosexuals” with “same-sex couples” or “gay and lesbian couples.”

<sup>3</sup>For a detailed discussion of the potential causes of changing public opinion on gay and lesbian rights see Brewer (2008).



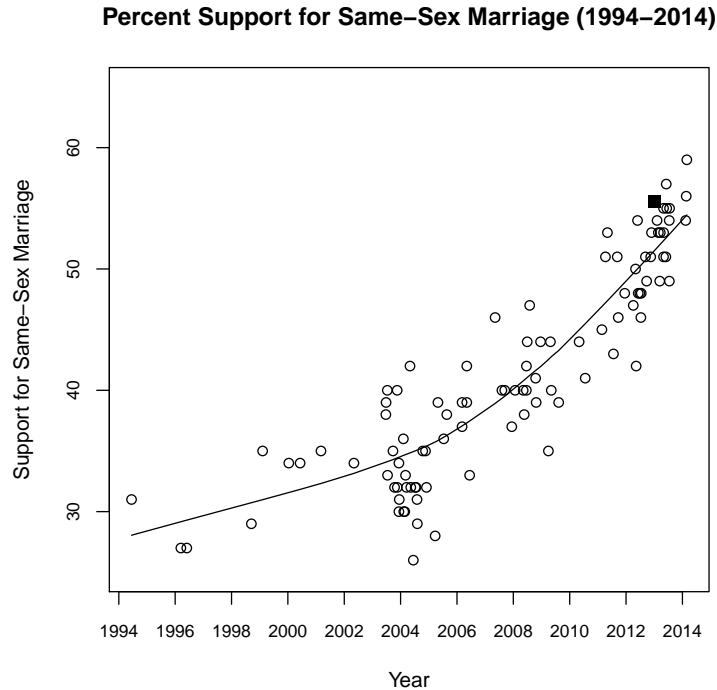


Figure 4.1: *Support for Same-Sex Marriage (1994-2014)*. Each plotted circle represents a single poll result. The x-axis is the year in which a poll was conducted and the y-axis is the percentage of respondents who report (under direct questioning) support for the legalization of marriages between same-sex couples. The time trend is measured using a lowess curve. Polling data were obtained from iPoll, which is housed at the Roper Center for Public Opinion Research. The graph includes all available polls from reputable polling firms conducted between January 1994 and March 2014. The solid square is the weighted percentage of untreated respondents from our CCES module who directly report supporting same-sex marriage.

However, not everyone is convinced that public opinion on this matter has changed as much as polling suggests. A study by Powell (2013), for example, compares the accuracy of pre-election polling on same-sex marriage ballot measures to similar polling on other statewide ballot issues, including taxes, bonds, and term limits. He finds that opposition to same-sex marriage is 5 to 7 percentage points higher on election day than in pre-election polling, but that corresponding inaccuracies are absent for other issues. For Powell, these results are consistent with the presence of social desirability bias in polling on same-sex marriage (p. 14). In a similar study, Egan (2008) compares the outcomes of 33 state-level ballot measures on same-sex marriage to opinion polls, finding that pre-election polls consistently underestimated opposition to same-sex marriage (i.e.,

support for constitutional bans) by an average of 7 percentage points. Finally, a recent longitudinal study by Emerson and Essenburg (2013) tracks the opinions of nearly 1,300 Americans from 2006 to 2012. They find little evidence that opinion has changed between 2006 and 2012. The red flags raised by the research of Egan, Powell, and Emerson and Essenburg indicate that a careful, individual-level inquiry into the existence of social desirability bias in same-sex marriage polling is warranted.

### 4.3 Uncovering “True” Opinion

There is plenty of evidence that respondents sometimes provide socially acceptable (as opposed to true) responses to direct survey questions. For example, respondents have been shown to overreport voting in the most recent election (Silver, Anderson, and Abramson 1986; Presser 1990), church attendance (Hadaway, Marler, and Chaves 1993; Smith 1998), and their willingness to vote for black and female candidates (Finkel, Guterback, and Borg 1991; Berinsky 1999; Streb et al. 2008). While traditional public opinion polls are ill-equipped to tease out the presence of these effects, a list experiment is one accepted method for doing so. This technique affords survey respondents an additional layer of anonymity that has been shown to elicit more truthful reports of behaviors and beliefs that are perceived as socially undesirable (Dalton, Wimbush, and Daily 1994; LaBrie and Earlywine 2000; Tsuchiya, Hirai, and Ono 2007; Tourangeau and Yan 2007).

In a list experiment, subjects are randomized into control and treatment groups. In the control group, subjects are given a list of  $J$  non-sensitive items and asked to report *how many*, not which ones, they support. Members of the treatment group are assigned the same task, but receive a list of  $J+1$  items. This list includes the same non-sensitive items given to the control group plus the sensitive item of interest to researchers. With a sufficiently large sample, researchers can estimate the population proportion that supports the sensitive item by taking the difference between the average response of the treatment group and the control group. By not directly asking respondents their views on the sensitive question, it is impossible for the researcher to

infer a specific individual's response to the sensitive item. This veiled approach all but eliminates pressure to mislead the researcher.

List experiments, however, are not as straightforward to implement in practice as they may seem in theory. Problems can arise from a lack of statistical power, the construction of lists in which the sensitive item is obviously distinct from the control items, uneven implementation by enumerators, and the construction of lists for which a larger number of respondents support or oppose all of the non-sensitive items. In designing our experiment, we take great care to avoid these potential pitfalls, adopting best practices recommend by Glynn (2013) and others. In particular, we use a large sample size, employ nonsensitive list items that (like our sensitive item) are political in nature, and use pairs of nonsensitive items for which respondents' answers are likely to be negatively correlated.<sup>4</sup> For a more detailed discussion of the logic behind our design choices, see Section 4.4.1.

### 4.3.1 Using List Experiments to Study Attitudes toward Gays and Lesbians

To the best of our knowledge, list experiments have been used in two prior instances to study the public's evolving attitudes toward gay and lesbian rights. In an unpublished paper, Goldman (2008) employs two list experiments embedded in the 2006 Cooperative Congressional Election Study to measure the public's anti-gay attitudes. Specifically, Goldman measures what proportion of the population is angered by "a gay or lesbian family moving next door" as well as "the growing acceptance of homosexuality." Goldman also asks these questions directly to measure social desirability bias, but finds no evidence of such bias in the aggregate. This paper unfortunately only includes respondents from 16 states, and thus is not nationally representative. It also does not ask about the specific policies (e.g., same-sex marriage) that are at the center of debates

---

<sup>4</sup>Also, a list experiment assumes that people are willing to tell the truth when provided increased anonymity. If the norm to lie is so strong that respondents continue to do so even when afforded anonymity, the experiment will not reveal true opinion (for more discussion of the no-liars assumption see Blair and Imai (2012)). We do not, however, expect the urge to lie to be so strong in our case; the sort of bias we study here is indeed the usual target for a list experiment.

over gay and lesbian rights and that have been the subject of most existing polling.<sup>5</sup>

In an NBER working paper, Coffman et al. (2013) find evidence of social desirability bias. Using a series of list experiments and direct questions, they show that respondents, when given the anonymity of a list experiment, are more likely to self-identify as gay, express disapproval of an openly-gay manager at work, and support discrimination against, gay, lesbian, or bisexual individuals. However, Coffman et al. found only a small (and statistically insignificant) difference in responses to direct and indirect questioning about support for same-sex marriage. The notable exception is self-identified Democrats, who are significantly more willing to admit to opposing same-sex marriage on a list experiment. While this study suggests the presence of social desirability bias on policy questions involving LGB rights, some caution is warranted. In particular, Coffman et al. conduct their investigation using Amazon's Mechanical Turk (an online labor market) and therefore do not have a representative sample of the American public.<sup>6</sup>

## 4.4 Same-Sex Marriage List Experiment

### 4.4.1 Experimental Design

Our data come from a survey experiment embedded in the 2013 Cooperative Congressional Election Study (CCES).<sup>7</sup> Our sample, unlike those used in prior list experiments on gay and lesbian rights, is nationally representative. The list experiment was part of a survey module that asks respondents a series of direct questions about state government and their opinions on a variety of public policy matters, some of which are traditionally set at the state level and others at the national level. Given the content of the survey module, the list experiment was unlikely

---

<sup>5</sup>Goldman does show that those with higher levels of education are somewhat more likely to censor expressions of anger, but only when questioned about the growing acceptance of homosexuality and not about a gay or lesbian family moving next door.

<sup>6</sup>Mechanical Turk (MT) is an on-line labor market in which employers post solicitations for paid work (sometimes this work involves participating in social science surveys). Pay is typically substantially less than the minimum wage.

<sup>7</sup>The CCES is a large-scale academic survey that is fielded annually by a consortium of academic researchers through an online survey conducted by YouGov. Respondents are selected from YouGov's ongoing panel of respondents, and the sample is stratified by state and congressional district. All respondents answer a set of common questions and are then assigned a specific module created by one team of researchers.

to strike respondents as odd. All that differentiated the list experiment was that respondents were asked to report the number of policies they support from a list instead of responding to individual direct questions. The full text of the experiment is listed below (with the sensitive item last).

Please take your time and tell us **how many** of the following you support. We do not need to know which ones, just **how many**.

- President Obama’s health care reform (“ObamaCare”)
- Making birth control illegal
- Cutting spending on food stamps
- Laws that make drunk driving illegal
- **Allowing gays and lesbians to marry legally**

In designing our list experiment, we were careful to avoid potential pitfalls. First, list experiments require larger sample sizes than are typically necessary for a direct question, given the larger standard errors they produce. Corstange (2009) recommends researchers use samples of at least 1,000 respondents but suggests closer to 2,000 if possible. We follow this advice, obtaining a sample of 1,900. Second, to draw less attention to our sensitive item, each non-sensitive item in our list is also political in nature (Kuklinski, Cobb, and Gilens 1997; Glynn 2013; Aronow, Coppock, Crawford, and Green 2015). Doing so also ensures that the list experiment blends with the rest of the survey. Third, we were careful to avoid the presence of ceiling and floor effects. That is, in selecting nonsensitive items for our lists, we ensure a low probability that any respondent would answer either “yes” or “no” to all non-sensitive items, since doing so can remove the anonymity that is essential to a list experiment. To guard against these effects, we do not include too many high or low prevalence items (Kuklinski, Cobb, and Gilens 1997; Tsuchiya, Hirai, and Ono 2007; Glynn 2013).

We also attempted to design a set of non-sensitive items for which the mean number of items supported is two (out of a possible four). To achieve this, we included one statement that, based on existing public opinion data, we expected almost all respondents to support—laws that make drunk driving illegal. Likewise, we included one statement that we expected almost everyone

to reject—making birth control illegal. For our last two items, we chose statements that we thought would be negatively correlated—President Obama’s health care reform (“ObamaCare”) and cutting spending on food stamps. Glynn (2013) demonstrates that negative correlation within the list items and a modal response of support for 2 out of 4 control items will reduce variance. Given the inherent noisiness of list experiments (since they are indirect measures of preferences), it is important to use the design to lower variance wherever possible.

Participants were randomized into two groups. The control group received a list that included the first four items, while the treatment group received the full list (i.e., the control items plus the sensitive item). The order in which items appear in the lists was randomized across respondents. All respondents were also directly asked whether they support same-sex marriage.<sup>8</sup> The sensitive question was asked near the end of the survey module, well after respondents had completed the list experiment. The inclusion of the direct question provides us with the baseline estimate of public support for same-sex marriage in the survey and allows us to detect whether social desirability bias is present among some groups of respondents but not others. Of course, in asking the direct question after the list experiment, we assume that the presence of the list experiment does not change answers to the direct question. Fortunately, this assumption can be tested by comparing the mean response to the direct question among those in the control group and those in the treatment group.

#### 4.4.2 Evaluating the Design

Before we analyze our results, we first evaluate the design of the list experiment. Glancing at the data presented in Table 4.1, it appears we were able to avoid potential problems of list experiment design. First, the potential for floor and ceiling effects appears to be quite small. Out of the 899 survey respondents in the control group, only 4% said they supported zero policies and only 3% said they supported all four policies. Additionally, the modal response to our set of nonsensitive items was two, with just over 52% of the respondents in the control group

---

<sup>8</sup>“Do you favor or oppose legally sanctioned marriages between gay and lesbian couples?”

providing that number.<sup>9</sup> This suggests that we achieved the desired negative correlation among our nonsensitive items.

Table 4.1: Observed Data – Same-sex Marriage List Experiment

Response Value	Control Group		Treatment Group	
	Frequency	Percentage	Frequency	Percentage
0	39	4.34%	22	2.3%
1	243	27.03	137	14.0
2	474	52.73	338	34.5
3	114	12.68	385	39.3
4	29	3.23	68	6.9
5			29	3.0
Total	899		979	

Note: This table displays the number and percentage of respondents for each value of  $Y$ , the number of items that the respondent supports in the list experiment, for both the control and treatment group. Percentages may not add to 100% due to rounding.

Next, we consider whether design effects are present. A list experiment has a design effect when an individual’s response to the non-sensitive items changes whether or not the sensitive item is present. For the list experiment difference-in-means estimate to be valid, the mean for support for the nonsensitive items must be the same on average across treatment and control (Imai 2011). Following Blair and Imai (2012), we use the *List* package in R to test for design effects. The p-value on this test is 0.63, so we cannot reject the null hypothesis of no design effect. Given this result, we move forward in the analysis of our list experiment under the assumption of no design effect.

It is also important to determine whether there is balance among demographic covariates across the treatment and control groups. Table 4.2 suggests that there is balance—the difference in nearly all demographic categories is quite small, with the exception of ideological conservatives (the control group is 6% more conservative than the treatment group). When we regress assignment to treatment on these covariates, we find that none are statistically significant. This

---

<sup>9</sup>Of the 1,900 respondents who were presented with either the control or treatment list, the nonresponse rate was 0.6%. We observe the same non-response rate for the direct question.

is true even for ideological conservatives.<sup>10</sup>

Table 4.2: Covariate Balance – Same-sex Marriage List Experiment

	Treatment Mean	Control Mean	Difference
Male	0.45	0.47	-0.02
Female	0.55	0.53	0.02
< High School	0.03	0.04	-0.00
Graduated College	0.22	0.25	-0.02
High School	0.31	0.29	0.02
Post-Grad	0.12	0.09	0.03
Some College	0.31	0.34	-0.03
Asian	0.02	0.01	0.00
Black	0.10	0.11	-0.01
Hispanic	0.06	0.06	0.00
Other Race	0.04	0.03	0.01
White	0.78	0.78	-0.00
18-29	0.17	0.18	-0.01
30-44	0.17	0.18	-0.01
45-64	0.44	0.43	0.01
65+	0.22	0.20	0.02
Democrat	0.39	0.35	0.04
Independent	0.31	0.31	-0.00
Other Party	0.07	0.07	-0.00
Republican	0.23	0.26	-0.03
Conservative	0.31	0.37	-0.06
Liberal	0.26	0.25	0.01
Moderate	0.34	0.30	0.04
Not Sure (ideology)	0.09	0.08	0.01

Note: This table displays the proportion of respondents by demographic group in the control and treatment groups and the difference between them. Percentages may not add to 100% due to rounding.

<sup>10</sup>As an additional test, we then use randomization inference to simulate the random assignment procedure 100,000 times and calculate the F-statistic of this regression for each hypothetical sample. This collection of F-statistics can be thought of as the sampling distribution of the F-statistic under the null hypothesis that none of the covariates have any effect on assignment to treatment (Gerber and Green 2012, p. 107-08). We find the p-value of the F-statistic by finding its location within the simulated sampling distribution. Since we obtain a p-value of 0.20, we cannot reject the null hypothesis that the covariates are unrelated to treatment assignment, further confirming the validity of the randomization procedure.



### 4.4.3 Results

We present results in Table 4.3, which shows the mean number of items supported by the control and treatment groups, the difference between the two, and the mean response to our direct question. Since one should avoid making inferences about the general population using unweighted survey data, we focus our discussion in the text on the results of our weighted sample. (Note that none of our substantive findings would differ if we relied primarily on our unweighted survey data.) By subtracting the mean number of items the control group supports from the mean number of items the treatment group supports, we obtain the list experiment estimate of support for same-sex marriage. The weighted estimate is 58.6%, with a 95% confidence interval bounded by 46.7% and 70.4%.<sup>11</sup> Since the confidence interval is fairly wide, we cannot conclude with 95% certainty that a majority supports marriage equality. However, we can conclude majority support with a confidence level of 85%.<sup>12</sup>

Table 4.3: Estimated Support for Same-Sex Marriage

	Control Mean	Treatment Mean	Difference-in-Means	Direct Question
Sample	1.83 (.03)	2.44 (.03)	0.60 (.04)***	0.57 (.01)
Weighted Sample	1.83 (.05)	2.42 (.04)	0.59 (.06)***	0.56 (.02)
<i>N</i>	899	979		1878

Note: The numbers in the parentheses are the standard errors. The reported difference-in-means may not equal the difference between the control mean and the treatment mean due to rounding.

\*\*\*  $p < .001$ .

Do our estimates suggest that there is an overall social desirability bias? To address this question, we compare our list experiment estimate of opinion to our direct question estimate. In doing so, however, we must keep in mind that the direct question is a post-treatment covariate—that is, the direct question was asked after respondents had completed the list experiment (albeit with many questions in between the two). To test whether this interfered with our direct estimate,

<sup>11</sup>For the analysis applying weights, linearized standard errors were calculated using the *svy* command in Stata.

<sup>12</sup>For our sample itself, we can conclude majority support for marriage equality. Our list experiment indicates that support for same-sex marriage among our respondents (before adjusting for population weights) is 60.2%, with a 95% confidence interval bounded by 50.4% and 66.8%.

we consider whether respondents who received the marriage treatment in the list experiment answered the direct question systematically differently from those in the control group. Since respondents were randomly assigned to control or treatment, in expectation the two groups should answer the direct question the same way if the list experiment has no effect on the direct response. Using a well-powered difference-of-proportions test, we find no significant difference across these two groups (55.6% of respondents in the control group say that they support same-sex marriage, while 56.9% report doing so in the treatment group).<sup>13</sup>

Given that there do not appear to be any order effects, we compare our indirect and direct estimates of public opinion. We find that the weighted direct estimate of 56.3% is similar to our indirect estimate (indeed, the former is actually 3 percentage points lower than the latter). If we only use answers to the direct question from the control group, the estimate of weighted direct support for same-sex marriage is virtually unchanged at 55.6%. This indicates that respondents are *not* systematically lying about their support for same-sex marriage, at least not in the way that we would anticipate if social desirability bias were inflating direct estimates of opinion. Furthermore, when we compare our list experiment estimate of opinion to recent estimates of support for same-sex marriage obtained by polling firms that employ a direct question approach, we find them to be very similar (as can be seen using the data presented in Figure 4.1).<sup>14</sup> This serves as additional evidence against the presence of social desirability bias.

The results we present here obviously stand in contrast to those of Powell (2013), who finds evidence consistent with the presence of social desirability effects. As noted previously, Powell looks for social desirability by comparing pre-election surveys on same-sex marriage to the results of ballot measure elections. That our results differ, then, is not necessarily surprising. The tests employed by Powell are indirect, using aggregate-level data to study what is fundamentally

---

<sup>13</sup>The difference of 1.3% is small with a p-value of 0.66. The power of the test for a difference of proportions of size 1.3 percentage points, which is what we found, with our sample sizes of 899 and 979, is roughly 0.9, making this a well-powered test. With these size samples and aiming for a standard power level of 0.8, we could detect a difference of proportion as small as a percentage point.

<sup>14</sup>Mean public support for same-sex marriage across all of the phone-based surveys reported in iPoll for 2013 is 53%.

an individual-level phenomenon.<sup>15</sup> We believe that individual-level evaluations are ultimately stronger tests of the truthfulness of respondents on surveys.

#### 4.4.4 Considering Social Desirability by Subgroups

Until now, we have assumed that social pressure may lead opponents of same-sex marriage to conceal their true preferences. However, it may also be the case that for some types of respondents, social pressure works in the opposite direction. That is, they may feel pressured to state that they oppose same-sex marriage (due to norms or pressures of their community or reference groups), when, in fact, they actually support marriage equality. Indeed, if both types of social pressure exist, they may be offsetting at the aggregate level and, therefore, not appear in our overall analysis. Fortunately, the inclusion of the direct question allows us to conduct the nuanced investigation that is necessary to test for conflicting forms of bias.

We begin with Table 4.4, which compares the list experiment difference-in-means for two subgroups—those who said they support same-sex marriage when asked directly and those who did not. If there is no lying when answering the direct question, the list-experiment estimate should be 1 among those who directly report supporting same-sex marriage and 0 among those who directly report opposition.<sup>16</sup> This is not, however, what we find. Among those who report that they oppose same-sex marriage, the difference of means is 0.15, and zero is not included in a 95% confidence interval around the estimate. The opposite pattern emerges when we look at those who report, under questioning, that they support marriage equality. Among this group of respondents, the difference of means estimate is 0.93. (While 1 falls within the 95% confidence interval, it does not fall within a 90% confidence interval.) These results raise the possibility that a social desirability bias exists in polling on same-sex marriage, but that it pushes some respondents into overstating their support for marriage equality and others into underreporting

---

<sup>15</sup>Of particular concern is that the sample in pre-election surveys may not be representative of the individuals who turn out to vote. Furthermore, as Egan (2008) notes, ballot measures on same-sex marriage are often written in a manner that is intentionally confusing, whereas the questions asked by pollsters tend to be much clearer. Both of these factors can lead to differences between pre-election polls and election results.

<sup>16</sup>For examples of scholars who also utilize a direct question in the context of a list experiment see Aronow, Coppock, Crawford, and Green (2015) and Gonzales-Ocantos et al. (2012).

their support. Indeed, the point estimates produced in the table indicate that, if anything, more respondents are *underreporting* than over-reporting their support for same-sex marriage.

Table 4.4: Which Way Does Social Desirability Work?

<i>Sample</i>					
Direct Question	Control Mean	Treatment Mean	Expected Diff.-in-Means	Actual Diff.-in-Means	95% Conf. Interval
For Gay Marriage	1.78 (.03)	2.71 (.04)	1.00	.93 (.05)	[.83, 1.03]
N	490	575			
Against Gay Marriage	1.90 (.04)	2.04 (.05)	0.00	.15 (.07)	[-.02, .28]
N	409	404			

Note: The numbers in the parentheses are the standard errors. The reported difference-in-means may not equal the difference between the control mean and the treatment mean due to rounding.

\*  $p < .05$ , \*\*\*  $p < .001$ .

These results prompt us to further explore the possibility that social desirability operates in unique ways across subgroups. It may be the case that the direction in which social desirability bias works is predicted by a respondent’s key reference groups. For example, a religious conservative who personally favors same-sex marriage but whose religious community is against marriage equality may conclude that the socially desirable answer is to say that one is against same-sex marriage; we might expect the reverse among someone who is a Democrat or who is not religious.

In keeping with these examples, our analysis considers whether social desirability effects differ by a respondent’s partisan identification and religious affiliation. Indeed, prior work has found some evidence that self-identified Democrats are more willing to admit opposition to same-sex marriage in a list experiment than under direct questioning (Coffman et al. 2013). In addition, we also consider the possibility that social desirability effects vary by geography (comparing respondents from the South to those from other regions of the country) and by educational attainment. These distinctions have been used in prior studies that explore attitudes toward minority groups, and Goldman (2008) found that highly-educated respondents were more likely to censor expressions of anger about the growing acceptance of homosexuality than were the less educated.

Previous researchers who have explored cross-group differences in social desirability bias have typically done so by limiting their data only to the group of interest and then employing the difference-of-means estimator used earlier to produce the comparisons shown in Table 4.3 (cf., Kuklinski, Cobb, and Gilens 1997). But conducting analyses in this fashion is less than ideal. When subgroups are small, tests will be underpowered. Furthermore, the basic difference-in-means approach does not allow researchers to adjust for multiple covariates at the same time.

Fortunately, Imai (2011) has proposed a regression methodology for list experiments as a solution to these problems (see also Blair and Imai 2012). Imai's approach first estimates a multivariate model of support for same-sex marriage that uses only answers from the list experiment. To do so, he has developed a maximum likelihood estimator to estimate the joint distribution of  $(Y_i(0), Z_{i,J+1}^*)$ , where  $Y_i(0)$  is the number of control items supported by the  $i$ 'th respondent and  $Z_{i,J+1}^*$  is the  $i$ 'th respondent's truthful answer to the sensitive item. This yields coefficients for predicting the count of nonsensitive items a respondent supports as well as the likelihood that a respondent will support the sensitive item. The multivariate regression allows the researcher to model the relationships between several respondent characteristics and their answers to the sensitive question (Blair and Imai 2012). By analyzing the treatment and control group together, this approach relies on the fact that identical control items were asked to both groups to improve statistical efficiency.

Using the results of this regression, one then generates predicted probabilities of support for the sensitive item (in this case same-sex marriage) by the respondent characteristics of interest. This first set of predicted probabilities can be thought of as being devoid of social desirability bias. The next step is to use a standard binary logistical regression and responses to only the direct question to generate a second multivariate model of support for the sensitive item. Again, results of the model are used to generate predicted probabilities by respondent characteristics. These estimates, because they do not rely on responses to the list experiment, can be thought of as being contaminated by social desirability bias, assuming such bias exists. The difference between the first and second set of estimates is the size of the bias, controlling for other demographic characteristics. These can be easily plotted with corresponding confidence intervals to evaluate

statistical significance.

We employ the Imai approach here. The respondent characteristics that are included in our models are gender, education, race, age, partisan affiliation, geography (a dummy variable indicating whether the respondent lives in a southern state), and a dummy variable for identifying as a religious conservative.<sup>17</sup> We use the *R* package designed by Blair and Imai called *list* to estimate both models.

Table 4.5: Multivariate Analysis of the List Experiment and the Direct Question

Variables	Sensitive Item	Control Items	Direct Question
	Est. (SE)	Est. (SE)	Est. (SE)
(Intercept)	1.78 (1.24)	-0.17 (0.17)	2.74 (0.37)***
Male	0.28 (0.62)	0.035 (0.06)	-0.41 (0.12)***
High School	2.81 (1.51)	-0.16 (0.17)	-0.08 (0.34)
Some College	4.18 (1.51)**	-0.19 (0.16)	0.08 (0.34)
Graduated College	3.72 (1.48)*	-0.08 (0.16)	0.08 (0.35)
Post-Grad	5.70 (2.24)*	-0.21 (0.19)	0.59 (0.38)
Black	0.24 (0.88)	0.16 (0.10)	-0.99 (.20)***
Hispanic	-2.18 (0.97)*	0.21 (0.11)	-0.42 (0.24)
Age 30-44	-1.69 (1.27)	0.13 (0.10)	-0.43 (0.21)*
Age 45-64	-3.13 (1.35)*	0.22 (0.10)*	-0.76 (0.17)***
Age 65+	-2.02 (1.27)	0.22 (0.10)*	-1.25 (0.20)***
Republican	-2.91 (0.90)**	-0.10 (0.09)	-2.14 (0.16)***
Independent	-1.27 (0.80)	-0.18 (0.07)*	-0.95 (0.14)***
South	-0.93 (0.59)	0.04 (0.06)	-0.27 (0.12)*
Relig. Conservative	-2.67 (0.69)***	0.07 (0.08)	-1.62 (0.13)***

\* p < .05, \*\* p < .01, \*\*\* p < .001.

The results of the first model (using only responses from the list experiment) are presented in the first two columns of Table 4.5. The coefficients in the sensitive item column are coefficients that predict whether someone will answer yes to the sensitive item in the list experiment (i.e., support same-sex marriage). Again, these can be thought of as showing the demographic correlates of support for same-sex marriage absent any potential social desirability effects. The

<sup>17</sup>We use the Census definition of the south and define a religious conservative as someone who self-identifies as either Mormon or as a “born-again” Christian. We measure education using four dummy variables (the reference category is “less than a high school education”). To generate the results reported in Figure 2, we re-estimate the model, replacing these four education dummy variables with a single dichotomous measure indicating whether a respondent has a college degree. This alternative specification is employed for ease of presentation, but its use does not change our findings.

coefficients in the control items column predict the count of non-sensitive items a respondent will answer in the affirmative. (Note the coefficients in columns one and two are on different scales.) The last column shows the results of the standard logit model that uses only data from the direct question. The results in the table largely confirm findings in the existing public opinion literature (see Brewer 2008).<sup>18</sup> Both younger respondents and respondents with higher levels of education are more likely to support marriage equality, while Republicans and religious conservative are much less likely to support same-sex marriage, as are those residing in the South (though the coefficient on “South” fails to reach statistical significance).

Figure 4.2 displays the results of our subgroup analysis graphically. The subgroups of interest are along the x-axis, and proportion support for same-sex marriage is along the y-axis. The bars around each estimate depict 95% confidence intervals. We use 10,000 Monte Carlo draws to estimate confidence intervals on effects and differences in effects.<sup>19</sup> The top two panels report estimates of social desirability by party and region (South vs. non-South). The next set of panels report estimates by religion (religious conservative vs. other) and education level (a college degree or higher vs. less than a college degree). The final set unpacks the education results a bit further.

---

<sup>18</sup>The only odd result is that the coefficients on our measures of education fail to reach statistical significance in the direct question regression. That being said, they all have the anticipated sign and “Post-Grad” would reach statistical significance at the 90 percent level, using a one-tailed test. We are not overly concerned by this result, since the education variables perform as expected in column one.

<sup>19</sup>As in Blair and Imai (2012), our confidence intervals are calculated by first “sampling parameters from the multivariate normal distribution with mean set to the vector of parameter estimates and the variance set to the estimated covariance matrices.” Next, we “calculate each quantity of interest...and average over the empirical distribution of covariates for the entire data (p. 61).”

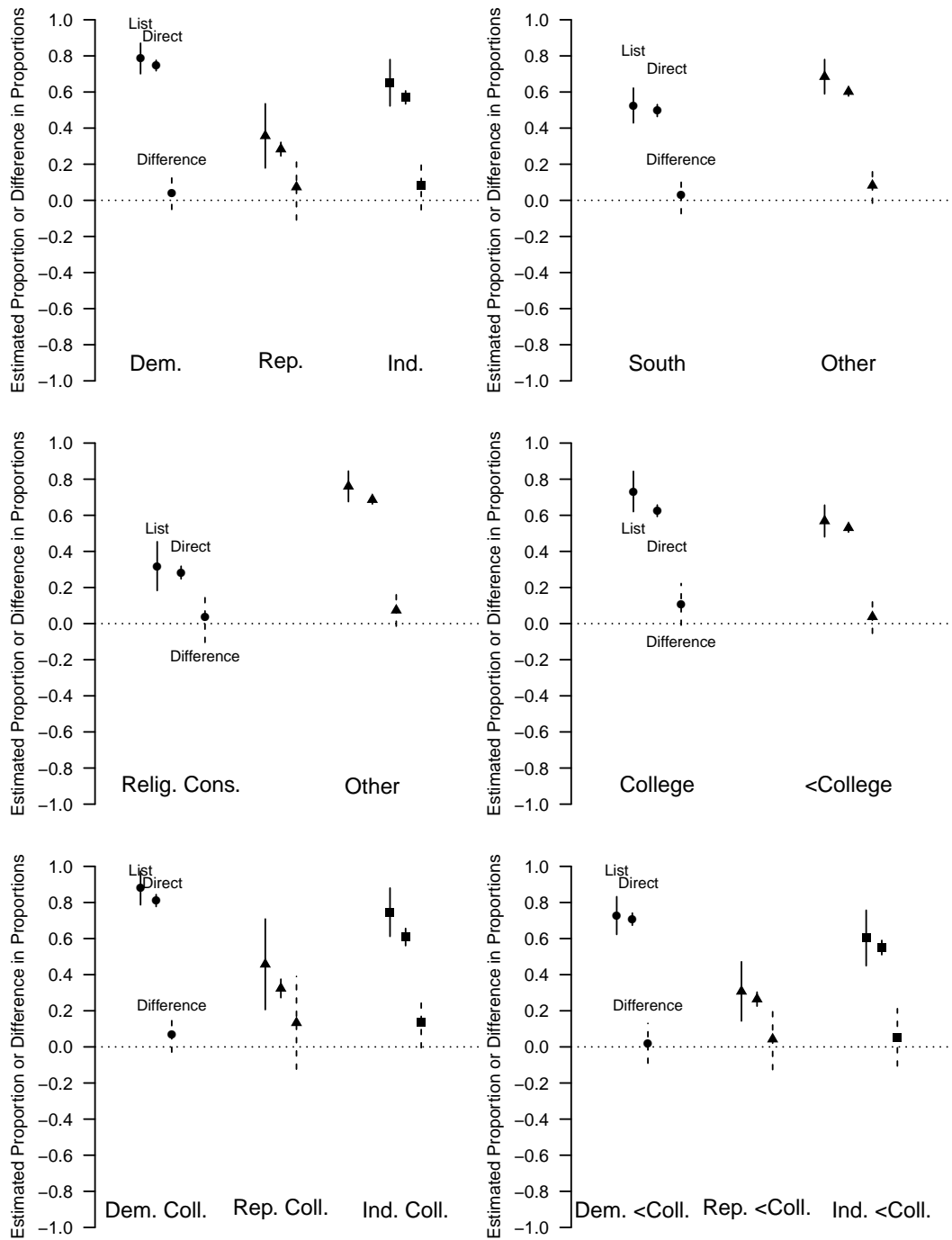


Figure 4.2: *Multivariate Subgroup Analysis* The lines are the 95% confidence interval generated by Monte Carlo simulations.

In the top two panels, we see that in the case of all subgroups, the difference between the list



experiment estimate and the direct question estimate of support for same-sex marriage is greater than zero. This implies that in each of the key subgroups shown above, there is some understating of support for same-sex marriage when respondents are asked directly. Though such a result is somewhat surprising, it is consistent with the population-level point predictions in Table 4.4. However, in all of these cases, given the 95% confidence intervals, none are statistically significant.

We come closest to finding statistical significance among respondents with at least a college education. These individuals are, on average, 10 percentage points more likely to report that they support same-sex marriage in a list experiment than on a direct question. The direction and magnitude of this result is not what we would have expected. Because support for marriage equality is strongly correlated to education levels, we would predict (if anything) that among educated respondents there would be social pressure to state that they support same-sex marriage when asked directly, but that when given the anonymity of a list experiment, that pressure would disappear (potentially leading to evidence of greater opposition).

To explore this further we break the education results into smaller subgroups, considering both education and party, again using predictions based on the results in Table 4.5. This further exploration indicates that the result in the second panel is largely (though not exclusively) being driven by Republican respondents—it is well-educated Republicans who are more likely to report on a list experiment that they support marriage equality. This is a more predictable result. Still, however, the finding is not statistically significant.

Given this, what (if anything) does our subgroup analysis tell us about social desirability bias in polling on same-sex marriage? At best, the analysis presented in this section indicates that social desirability bias exists, but is not unidirectional. The results in Table 4.4 are consistent with a world in which some respondents feel pressured to overreport their support for marriage equality while others feel pressure to underreport their support. In our data, these competing pressures are largely offsetting, and have little effect on national-level estimates of opinion.

However, it is also possible that Table 4.4 is simply picking up noise in the data. List experiments are computationally more demanding than direct questions, which may lead some respondents to provide seemingly inconsistent answers when confronted with both types of

questions. Furthermore, in neither the list experiment nor the direct question were respondents given the opportunity to provide a “don’t know” answer. This means that respondents with weak or unclear preferences may be switching answers across questions.<sup>20</sup> Finally, that we do not uncover statistically meaningful evidence of social desirability bias in a more nuanced analysis of subgroups provides additionally evidence that such a bias is simply not a factor in polling on same-sex marriage.

## 4.5 A Further Inquiry

While we find little to no evidence of a social desirability bias in polling on same-sex marriage, one might argue that it is too soon for such an effect to have emerged. Might we find evidence of social desirability in areas where opposition to gay rights may more clearly go against perceived societal norms of tolerance?

To test for this possibility, we analyze a second list experiment, this one focusing on employment nondiscrimination, which has been on the policy agendas of LGBT rights organizations for decades (much longer than same-sex marriage) and appears to be significantly less controversial with the American public. Figure 4.3 plots polls on this topic for the prior 20 years: support has been quite high throughout, with the most recent surveys indicating that a large supermajority—over 70%—favors such laws. The high level of support suggests greater social pressure to conform to the pro-gay policy position. This is reflected in the rhetoric of elites—mainstream elected officials and candidates for office rarely suggest that individuals should be fired on the basis of their sexual orientation. Indeed, as Brewer suggests in his book about public opinion and gay rights, “On some policies, such as employment nondiscrimination and gays in the military, support for gay rights has approached the near-consensus levels attained by support for the principle of racial equality” (2008, p. 37). This makes employment nondiscrimination an ideal area for evaluating the robustness of our findings.

---

<sup>20</sup>We do not believe that noise from the absence of a “don’t know” option would bias our results in any particular direction. Indeed, there are reasons not to provide respondents with such an option when trying to get opinion devoid of social desirability bias. Berinsky (2004) finds that individuals who hold socially unacceptable opinions may hide their opinions behind a “don’t know” response.

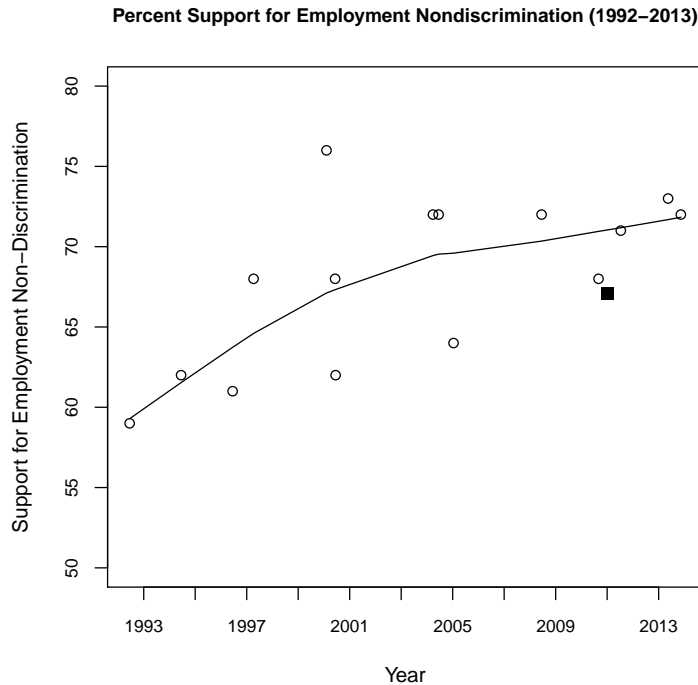


Figure 4.3: *Support for Nondiscrimination Laws (1992-2014)*. Each plotted circle represents a single poll result. The x-axis is the year in which a poll was conducted and the y-axis is the percentage of respondents who report (under direct questioning) supporting laws that protect gays and lesbians against employment discrimination. The time trend is measured using a loess curve. Polling data were obtained from iPoll. The graph includes all available polls from reputable polling firms conducted between January 1992 and March 2014. The solid square is the weighted percentage of untreated respondents from our CCES module who directly report supporting employment nondiscrimination.

The employment list experiment was embedded in the 2011 CCES. Unlike our same-sex marriage experiment, we employ a design in which participants were randomly divided into three (as opposed to two) groups: (1) the control group, consisting of 592 respondents, each of whom received a list that included only the first four (i.e., the non-sensitive) items; (2) the treatment group, consisting of 595 individuals, each of whom received the full list; and (3) a group of 608 respondents who were not given either list but were simply asked directly whether they favor or oppose such laws.<sup>21</sup> The three-group approach is similar to that of Gilens, Sniderman, and Kuklinsky (1998) and was the convention at the time our survey went into the field. Unfortunately

<sup>21</sup>The non-response rate was 0.03%.

this design limits our ability to directly test for the presence of social desirability bias among subgroups of respondents (as we did above) and reduces our overall sample size. That being said, the experiment still provides us with the necessary leverage to test for the presence of social desirability in the overall population.

Once again we use a nationally representative sample and unobtrusively embed our list experiment within a larger survey module that asks respondents their opinions on a variety of public policy matters. To draw less attention to our sensitive item, we include some non-sensitive items that are also political in nature. We were careful not to include too many high or low prevalence items to avoid ceiling and floor effects. The full text is:

**Employment Non-discrimination List:**

Please take your time and tell us **how many** of the following you support. We do not need to know which ones, just **how many**.

- A law requiring seat belts while driving
- Professional athletes getting million dollar-plus salaries
- Keeping a large number of troops in Afghanistan
- An amendment to the federal constitution requiring a balanced budget
- **A law protecting gays and lesbian against employment discrimination**

Table 4.6 summarizes the data. The distribution of responses indicates that there is unlikely to be much in the way of either large ceiling or floor effects. The modal response in the control group is 2 (as desired), with just over 2% of respondents supporting all of the items and approximately 15% supporting none. The Blair and Imai (2012) test for design effects produces a p-value of 0.18, meaning that we cannot reject the null hypothesis of no design effect. Finally, there is balance among demographic covariates across treatment and control.

Table 4.7 presents the mean number of items supported by the control and treatment groups, the difference between the two, and the mean response to our direct question. (Note that the direct question was not asked to individuals in either the control or treatment group.) The weighted estimate is 69.5% with a 95% confidence interval bounded by 51.9% and 87.1%.<sup>22</sup> Even

---

<sup>22</sup>For the analysis applying weights, linearized standard errors were calculated using the *svy* command in Stata.

Table 4.6: Observed Data – ENDA List Experiment

Response Value	Control Group		Treatment Group	
	Frequency	Percentage	Frequency	Percentage
0	89	15.03%	23	3.87%
1	197	33.28	128	21.51
2	217	36.66	222	37.31
3	77	13.01	159	26.72
4	12	2.03	46	7.73
5			17	2.86
Total	592		595	

Note: This table displays the number and percentage of respondents for each value of  $Y$ , the number of items that the respondent supports in the list experiment, for both the control and treatment group. Percentages may not add to 100% due to rounding.

with the large confidence interval, we can safely conclude that a majority supports employment nondiscrimination laws for gays and lesbians.<sup>23</sup>

Table 4.7: Estimated Support for Employment Non-Discrimination

	Control Mean	Treatment Mean	Difference-in-Means	Direct Question
Sample	1.54 (.04)	2.22 (.04)	0.678 (.06)***	0.655 (.02)
Weighted Sample	1.51 (.07)	2.21 (.06)	0.695 (.09)***	0.671 (.03)
$N$	592	595		608

Note: The numbers in the parentheses are the standard errors. The reported difference-in-means may not equal the difference between the control mean and the treatment mean due to rounding.

\*\*\* < .001.

Our estimates show no overall social desirability bias. The list experiment estimate of opinion is similar to that obtained from our direct question, and both our direct and indirect estimates of support are consistent with those from recent national polls on the topic (see Figure 4.3). This serves as additional evidence against the presence of social desirability bias in polling on issues related to same-sex rights.

<sup>23</sup>The unweighted list experiment estimate of support for laws that protect gays and lesbians against employment discrimination is 67.8%.

## 4.6 Conclusion

It is natural to ask whether some of the purported increase in support for same sex-marriage is due to the presence of social desirability bias. In our list experiment in the 2013 Cooperative Congressional Study, we find no evidence of social desirability bias at the population level. Our list experiment measure of support for same-sex marriage (59%) is almost identical to the estimate we obtain from direct survey response. These estimates also match those returned in other recent national surveys. If there is social desirability in polling on same-sex marriage, our results indicate that it pushes in both directions. This experiment also provides new evidence that there exists majority national support for extending marriage rights to same-sex couples.

In our second list experiment, on the adoption of laws that make employment discrimination against gays and lesbians illegal, we again find no evidence of bias at the population level. In tandem, these list experiments cast serious doubt on claims that social desirability bias is plaguing estimates of support of gay rights policies. This finding should strengthen social science work that has relied upon national surveys to study the opinion-policy relationship in this issue area (Lax and Phillips 2009a; Krimmel, Lax, and Phillips 2016).

It is important to note that our results only directly speak to one mode of polling—surveys that are taken over the computer in the privacy of a respondent’s home or office, where one might expect to observe the lowest levels of social desirability bias. Indeed, our results are consistent with work suggesting that the prevalence of social desirability bias in computer surveys should be low.<sup>24</sup> Furthermore, because there do not appear to be many differences between measures of public support for gay rights obtained via internet and telephone polling,<sup>25</sup> we suspect that there is little if any social desirability bias in telephone polling on same sex-marriage as well.<sup>26</sup> This comports with recent research that suggests that differences in survey mode may not produce different levels of social desirability bias (Ansolabehere and Schaffner 2014).

---

<sup>24</sup>Though such bias has previously been uncovered in these types of surveys (Goldman 2008; Coffman et al. 2013).

<sup>25</sup>Note that our CCES estimate of support for same-sex marriage (obtained via a direct question) falls nicely in line with the results of recent phone surveys (see Figure 4.1).

<sup>26</sup>If there is social desirability bias in telephone polling, but not internet polling, we would expect the levels of support for same-sex marriage and employment nondiscrimination to be notably higher in telephone polls.

## **Part II**

# **Bibliography**

# Bibliography

- Abramowitz, Alan I and Kyle L. Saunders. 2008. "Is Polarization A Myth?" *The Journal of Politics* 70(2):542–555.
- Ansolahehere, Stephen and Brian F. Schaffner. 2014. "Does Survey Mode Still Matter? Findings from a 2010 Multi-Mode Comparison." *Political Analysis* 22(3):285–303.
- Aronow, Peter M., Alexander Coppock, Forrest W. Crawford and Donald P. Green. 2015. "Combining List Experiment and Direct Question Estimates of Sensitive Behavior Prevalence." *Journal of Survey Statistics and Methodology*. 3(1):43–66.
- Bafumi, Joseph and Robert Y. Shapiro. 2009. "A New Partisan Voter." *The Journal of Politics* 71(1):1–24.
- Berinsky, Adam J. 1999. "The Two Faces of Public Opinion." *American Journal of Political Science* 43(4):1209–1230.
- Berinsky, Adam J. 2004. *Silent Voices: Opinion Polls and Political Participation in America*. Princeton: Princeton University Press.
- Berkman, Michael B. and Eric Plutzer. 2005. *Ten Thousand Democracies: Politics and Public Opinion in America's School Districts*. Georgetown University Press.
- Blair, Graeme and Kosuke Imai. 2010. "list: Statistical Methods for the Item Count Technique and List Experiment." Available at The Comprehensive R Archive Network (CRAN).
- Blair, Graeme and Kosuke Imai. 2012. "Statistical Analysis of List Experiments." *Political Analysis* 20(1):47–77.
- Brace, Paul, Kellie Sims-Butler, Kevin Arceneaux and Martin Johnson. 2002. "Public Opinion in the American States: New Perspectives Using National Survey Data." *American Journal of Political Science* 46:173–89.
- Brewer, Paul R. 2008. *Value War: Public Opinion and the Politics of Gay Rights*. Rowman and Littlefield Publishers, Inc.
- Brooks, Clem and Jeff Manza. 2007. *Why Welfare States Persist: The Importance of Public Opinion in Democracies*. The University of Chicago Press.
- Burstein, Paul. 2003. "The Impact of Public Opinion on Public Policy: A Review and an Agenda." *Political Research Quarterly* 56(1):29–40.



- Buttice, Matthew K. and Benjamin Highton. 2013. "How Does Multilevel Regression and Post-stratification Perform with Conventional National Surveys?" *Political Analysis* 21(4):449–467.
- Carsey, Thomas M. and Geoffrey C. Layman. 2006. "Changing Sides or Changing Minds? Party Identification and Policy Preferences in the American Electorate." *American Journal of Political Science* 50(2):464–477.
- Carsey, Thomas M. and Jeffrey J. Harden. 2010. "New Measures of Partisanship, Ideology, and Policy Mood in the American States." *State Politics and Policy Quarterly* 10(2):136–156.
- Chinni, Dante and James G. Gimpel. 2010. *Our Patchwork Nation: The Surprising Truth About the "Real" America*. New York: Gotham Books.
- Ciuk, David J. and William G. Jacoby. 2014. "Checking for Systematic Value Preferences Using the Method of Triads." *Political Psychology* .
- Coffman, Katherine B., Lucas C. Coffman and Keith M. Marzilli Ericson. 2013. "The Size of the LGBT Population and the Magnitude of Anti-Gay Sentiment are Substantially Underestimated." NBER Working Paper No. 19508.
- Corstange, Daniel. 2009. "Sensitive Questions, Truthful Answers? Modeling the List Experiment with LISTIT." *Political Analysis* 17:45–63.
- Dalton, Dan R., James C. Wimbush and Catherine M. Daily. 1994. "Using the Unmatched Count Technique (UCT) to Estimate Base Rates for Sensitive Behavior." *Personnel Psychology* 47(4):817–829.
- de Toqueville, Alexis. 1969. *Democracy in America*. HarperCollins Publishers, Inc.
- DeBell, Matthew. 2010. How To Analyze ANES Survey Data. ANES Technical Report Series. nes012492. Stanford University and the University of Michigan. Palo Alto, CA, and Ann Arbor, MI.: .
- Doherty, David. 2008. "Presidential Rhetoric, Candidate Evaluations, and Party Identification: Can Parties "Own" Values?" *Political Research Quarterly* 61(3):419–433.
- Egan, Patrick. 2008. "Findings from a Decade of Polling on Ballot Measures Regarding the Legal Status of Same-Sex Couples." Working Paper, New York University.
- Elazar, Daniel. 1972. *American Federalism: A View From the States*. Thomas Y. Crowell Company.
- Emerson, Michael O. and Laura J. Essenburg. 2013. "What is Marriage? Americans Dividing." Kinder Institute for Urban Research.
- Enns, Peter K. and Julianna Koch. 2013. "Public Opinion in the U.S. States: 1956 to 2010." *State Politics and Policy Quarterly* 13(3):349–372.
- Erikson, Robert S., Gerald C. Wright and John P. McIver. 1993. *Statehouse Democracy: Public Opinion and Policy in the American States*. Cambridge University Press.

- Erikson, Robert S., Gerald C. Wright and John P. McIver. 2006. Public Opinion in the States: A Quarter Century of Change and Stability. In *Public Opinion in State Politics*, ed. Jeffrey E. Cohen. Stanford University Press.
- Feldman, Stanley. 1988. "Structure and Consistency in Public Opinion: The Role of Core Beliefs and Values." *American Journal of Political Science* 32(2):416–440.
- Feldman, Stanley. 2003. Values, Ideology, and the Structure of Political Attitudes. In *Oxford Handbook of Political Psychology*, ed. David O. Sears, Leonie Huddy and Robert Jervis. Oxford University Press chapter 14, pp. 477–508.
- Feldman, Stanley and John Zaller. 1992. "The Political Culture of Ambivalence: Ideological Responses to the Welfare State." *American Journal of Political Science* 36(1):268–307.
- Feldman, Stanley and Marco R. Steenbergen. 2001. "The Humanitarian Foundation of Public Support for Social Welfare." *American Journal of Political Science* 45(3):658–677.
- Finkel, Steven E., Thomas M. Guterbock and Marian J. Borg. 1991. "Race-of-Interviewer Effects in Preelection Polls: Virginia 1989." *Public Opinion Quarterly* 55(3):313–330.
- Fiorina, Morris P., Samuel J. Abrams and Jeremy C. Pope. 2006. *Culture War?: The Myth of a Polarized America*. Pearson Education, Inc.
- Gastil, John, Don Braman, Dan Kahan and Paul Slovic. 2011. "The Cultural Orientation of Mass Political Opinion." *PS: Political Science and Politics* 44(4):711–714.
- Gelman, Andrew. 2008. "Scaling regression inputs by dividing by two standard deviations." *Statistics in Medicine* 27:2865–2873.
- Gelman, Andrew. 2013. "Everyone's Trading Bias for Variance at Some Point. It's Just Done at Different Places in the Analyses."  
**URL:** <http://andrewgelman.com/2013/03/14/everyones-trading-bias-for-variance-at-some-point-its-just-done-at-different-places-in-the-analyses/>
- Gelman, Andrew, Boris Shor, Joseph Bafumi and Jeronimo Cortina. 2008. *Red State, Blue State, Rich State, Poor State: Why Americans Vote the Way They Do*. Princeton University Press.
- Gelman, Andrew and Thomas C. Little. 1997. "Poststratification into Many Categories Using Hierarchical Logistic Regression." *Survey Methodology* 23(2):127–35.
- Gerber, Alan S. and Donald P. Green. 2012. *Field Experiments: Design, Analysis, and Interpretation*. W. W. Norton and Company, Inc.
- Gerber, Alan S., Gregory A. Huber, David Doherty and Conor M. Dowling. 2012. "Personality and the Strength and Direction of Partisan Identification." *Political Behavior* 34:653–688.
- Gerber, Alan S., Gregory A. Huber, David Doherty, Conor M. Dowling and Shang E. Ha. 2010. "Personality and Political Attitudes: Relationships Across Issue Domains and Political Contexts." *American Political Science Review* 104:111–133.

- Gibson, James L. 1995. "The Political Freedom of African-Americans: A Contextual Analysis of Racial Attitudes, Political Tolerance, and Individual Liberty." *Political Geography* 14(6/7):571–599.
- Gilens, Martin, Paul M. Sniderman and James H. Kuklinski. 1998. "Affirmative Action and the Politics of Realignment." *British Journal of Political Science* 28(1):159–183.
- Gimpel, James G. and Jason E. Schuknecht. 2003. *Patchwork Nation: Sectionalism and Political Change in American Politics*. Ann Arbor: The University of Michigan Press.
- Glynn, Adam N. 2013. "What Can We Learn with Statistical Truth Serum?: Design and Analysis of the List Experiment." *Public Opinion Quarterly* 77(S1):159–172.
- Goldman, Seth. 2008. "Effects of Social Desirability on Expressions of Anti-Gay Attitudes: Revealing the Role of Education in Reducing Prejudice." Presented at the Annual Meeting of the Midwest Political Science Association, Chicago, IL.
- Gonzalez-Ocantos, Ezequiel, Chad Kiewiet de Jonge, Carlos Meléndez, Javier Osorio and David W. Nickerson. 2012. "Vote Buying and Social Desirability Bias: Experimental Evidence from Nicaragua." *American Journal of Political Science* 56(1):202–217.
- Goren, Paul. 2005. "Party Identification and Core Political Values." *American Journal of Political Science* 49(4):881–896.
- Goren, Paul. 2012. *On Voter Competence*. Oxford University Press.
- Goren, Paul, Christopher M. Federico and Miki Caul Kittilson. 2009. "Source Cues, Partisan Identities, and Political Value Expression." *American Journal of Political Science* 53(4):805–820.
- Green, Donald, Bradley Palmquist and Eric Schickler. 2002. *Partisan Hearts and Minds: Political Parties and the Social Identities of Voters*. Yale University Press.
- Greenstone, J. David. 1986. "Political Culture and American Political Development: Liberty, Union, and the Liberal Bipolarity." *Studies in American Political Development* 1(1):1–49.
- Hadaway, C. Kirk, Penny Long Marler and Mark Chavez. 1993. "What the Polls Don't Show: A Closer Look at U.S. Church Attendance." *American Sociological Review* 58(6):741–752.
- Harter, Rachel, Stephanie Eckman, Ned English and Colm O'Muircheartaigh. 2010. Applied Sampling for Large-Scale Multistage Area Probability Designs. In *Handbook of Survey Research*, ed. Peter V. Marsden and James D. Wright. Second ed. Emerald Group Publishing Limited pp. 169–197.
- Hartz, Louis. 1955. *The Liberal Tradition in America*. Harcourt, Inc.
- Heeringa, Steven G., Judith H. Connor and Doris C. Darrah. 1986. 1980 SRC National Sample: Design and Development. Technical report Institute for Social Research, University of Michigan. Ann Arbor, MI: .
- Holbrook, Allyson L. and Jon A. Krosnick. 2010. "Social Desirability Bias in Voter Turnout Reports: Tests Using the Item Count Technique." *Public Opinion Quarterly* 74(1):37–67.

- Imai, Kosuke. 2011. "Multivariate Regression Analysis for the Item Count Technique." *Journal of the American Statistical Association* 106(494):407–416.
- Inglehart, Ronald. 2006. "Mapping Global Values." *Comparative Sociology* 5(2-3):115–136.
- Jacoby, William G. 2006. "Value Choices and American Public Opinion." *American Journal of Political Science* 50(3):706–723.
- Jacoby, William G. 2014. "Is There a Culture War? Conflicting Value Structures in American Public Opinion." *American Political Science Review* 108(4):754–771.
- Jost, John T. 2009. "Elective Affinities: On the Psychological Bases of Left-Right Differences." *Psychological Inquiry* 20:129–141.
- Kluegel, J. R. and E. R. Smith. 1986. *Beliefs About Inequality*. de Gruyter.
- Krimmel, Katherine L., Jeffrey R. Lax and Justin H. Phillips. 2016. "Gay Rights in Congress: Public Opinion and (Mis)representation." *Public Opinion Quarterly* 80(4):888–913.
- Kuklinski, James H., Michael D. Cobb and Martin Gilens. 1997. "Racial Attitudes and the 'New South'." *The Journal of Politics* 59(2):323–349.
- LaBrie, Joseph W. and Mitchell Earleywine. 2000. "Sexual Risk Behaviors and Alcohol: Higher Base Rates Revealed Using the Unmatched-Count Technique." *The Journal of Sex Research* 37(4):321–326.
- Lax, Jeffrey R. and Justin H. Phillips. 2009a. "Gay Rights in the States: Public Opinion and Policy Responsiveness." *American Political Science Review* 103(3):367–386.
- Lax, Jeffrey R. and Justin H. Phillips. 2009b. "How Should We Estimate Public Opinion in the States?" *American Journal of Political Science* 53(1):107–121.
- Lax, Jeffrey R. and Justin H. Phillips. 2012. "The Democratic Deficit in the States." *American Journal of Political Science* 56(1):148–166.
- Levendusky, Matthew. 2009. *The Partisan Sort: How Liberals Became Democrats and Conservatives Became Republicans*. Chicago: University of Chicago Press.
- Lewis, Gregory B. and Seong Soo Oh. 2008. "Public Opinion and State Action on Same-Sex Marriage." *State and Local Government Review* 40(1):42–53.
- Lyall, Jason, Graeme Blair and Kosuke Imai. 2013. "Explaining Support for Combatants during Wartime: A Survey Experiment in Afghanistan." *American Political Science Review* 107(4):679–705.
- Malecki, Michael, Jeffrey R. Lax, Andrew Gelman and Wei Wang. 2013. "mrp: Multilevel Regression and Poststratification." R package version 1.0-1.
- McCann, James A. 1997. "Electoral Choices and Core Value Change: The 1992 Presidential Campaign." *American Journal of Political Science* 41(2):564–583.

- McCarty, Nolan, Keith T. Poole and Howard Rosenthal. 2006. *Polarized America: The Dance of Ideology and Unequal Riches*. The MIT Press.
- McClosky, Herbert and John Zaller. 1984. *The American Ethos: Public Attitudes Toward Capitalism and Democracy*. Harvard University Press.
- Mondak, Jeffrey J. and Damarys Canache. 2014. "Personality and Political Culture in the American States." *Political Research Quarterly* 67(1):26–41.
- Pacheco, Julianna. 2011. "Using National Surveys to Measure Dynamic U.S. State Public Opinion: A Guideline for Scholars and an Application." *State Politics and Policy Quarterly* 11(4):415–439.
- Pacheco, Julianna. 2013. "The Thermostatic Model of Responsiveness in the American States." *State Politics and Policy Quarterly* 13(3):306–332.
- Page, Benjamin I. and Lawrence R. Jacobs. 2009. *Class War?: What Americans Really Think About Economic Inequality*. The University of Chicago Press.
- Page, Benjamin I. and Robert Y. Shapiro. 1983. "Effects of Public Opinion on Policy." *American Political Science Review* 77(1):175–190.
- Park, David K., Andrew Gelman and Joseph Bafumi. 2006. State-Level Opinions from National Surveys: Poststratification Using Multilevel Logistic Regression. In *Public Opinion in State Politics*, ed. Jeffrey E. Cohen. Stanford University Press chapter 11, pp. 209–228.
- Pedlow, Steven. 2013. "Calculating Design-Corrected Standard Errors for the General Social Survey, 1975-2012."   
**URL:** [http://publicdata.norc.org:41000/gss/documents//OTHR/GSS design variables.pdf](http://publicdata.norc.org:41000/gss/documents//OTHR/GSS%20design%20variables.pdf)
- Piazza, Thomas. 1980. "The Analysis of Attitude Items." *American Journal of Sociology* 86(3):584–603.
- Pinello, Daniel R. 2006. *America's Struggle for Same-Sex Marriage*. Cambridge University Press.
- Pool, I. d. S., R. P. Abelson and S. L. Popkin. 1965. *Candidates, Issues, and Strategies*. MIT Press.
- Powell, Richard J. 2013. "Social Desirability Bias in Polling on Same-Sex Marriage Ballot Measures." *American Politics Research* 41(6):1052–1070.
- Presser, Stanley. 1990. "Can Changes in Context Reduce Vote Overreporting in Surveys?" *Public Opinion Quarterly* 54(4):586–593.
- Regnerus, Mark. 2013. "Right Side of History, or Primed to Say Yes? When Asking about Same-Sex Marriage, Polling Methods Matter." National Review Online.
- Rentfrow, Peter, John T. Jost, Samuel D. Gosling and Jeffrey Potter. 2009. Statewide Differences in Personality Predict Voting Patterns in 1996-2004 U.S. Presidential Elections. In *Social and Psychological Bases of Ideology and System Justification*, ed. John T. Jost, Aaron C. Kay and Hulda Thorisdottir. Oxford University Press pp. 314–347.
- Rokeach, Milton. 1973. *The Nature of Human Values*. Free Press.

- Schwartz, Shalom H. 1992. Universals in the Content and Structure of Values: Theoretical Advances and Empirical Tests in 20 Countries. In *Advances in Experimental Social Psychology*, ed. M. P. Zanna. Vol. 25 Academic Press chapter 1, pp. 1–65.
- Schwartz, Shalom H. 1994. "Are There Universal Aspects in the Structure and Contents of Human Values?" *Journal of Social Issues* 50(4):19–45.
- Silver, Brian D., Barbara A. Anderson and Paul R. Abramson. 1986. "Who Overreports Voting?" *American Political Science Review* 80(2):613–624.
- Smith, Rogers M. 1993. "Beyond Tocqueville, Myrdal, and Hartz: The Multiple Traditions in America." *American Political Science Review* 87(3):549–566.
- Smith, Tom W. 1998. "A Review of Church Attendance Measures." *American Sociological Review* 63:131–136.
- Status of State Action on the Medicaid Expansion Decision, 2014*. 2014.  
**URL:** <http://kff.org/health-reform/state-indicator/state-activity-around-expanding-medicaid-under-the-affordable-care-act/>
- Stollwerk, Alissa F. 2013. "Estimating Subnational Opinion with Cluster-Sampled Polls: Challenges and Suggestions." Working Paper.
- Streb, Matthew J., Barbara Burrell, Brian Frederick and Michael A. Genovese. 2008. "Social Desirability Effects and Support for a Female American President." *Public Opinion Quarterly* 72(1):76–89.
- Tausanovitch, Chris and Christopher Warshaw. 2013. "Measuring Constituent Policy Preferences in Congress, State Legislatures, and Cities." *The Journal of Politics* 75(2):330–342.
- Tourangeau, Roger and Ting Yan. 2007. "Sensitive Questions in Surveys." *Psychological Bulletin* 133(5):859–883.
- Tsuchiya, Takahiro, Yoko Hirai and Shigeru Ono. 2007. "A Study of the Properties of the Item Count Technique." *Public Opinion Quarterly* 71(2):253–272.
- Warshaw, Christopher and Jonathan Rodden. 2012. "How Should We Measure District-Level Public Opinion on Individual Issues?" *The Journal of Politics* 74(1):203–219.
- Weber, Ronald E., Anne H. Hopkins, Michael L. Mezey and Frank J. Munger. 1972. "Computer Simulation of State Electorates." *Public Opinion Quarterly* 36:549–65.
- Weber, Ronald E. and William R. Shaffer. 1972. "Public Opinion and American State Policy-Making." *Midwest Journal of Political Science* 16(4):683–699.
- Wildavsky, Aaron. 1987. "Choosing Preferences by Constructing Institutions: A Cultural Theory of Preference Formation." *American Political Science Review* 81(1):3–22.

**Part III**  
**Appendices**

## Appendix A

# Appendix to “Estimating Subnational Opinion with Cluster-Sampled Polls: Challenges and Suggestions”

### A.1 GSS, ANES, and Ideology Measures

I replicate the comparisons of the GSS and MRP estimates against the Annenberg survey estimates using another source of “true” public opinion: the state-level measures of ideology and partisanship that Erikson, Wright, and McIver created using disaggregation of several national level polls. These measures are widely used in political science, and thus it seems appropriate to measure MRP estimates against them. Given that the time frame of the GSS and ANES data used here spans from 2000 to 2008, I use an updated version of the disaggregated data that was produced using polls from 1996 to 2003 (Erikson, Wright, and McIver 2006). As before, I drop DC as an outlier from my MRP analysis. Beginning with the GSS, this produces the following correlations, as shown in Table A.1.

In general, these correlation results are similar to those produced with the Annenberg data. The single-year, simple model produces poor to moderate correlations, and while the pooled model



Table A.1: Pearson's Product-Moment Correlations – GSS and EWM

Proportion	2004 Model	2004 w/ Kerry Vote	Pooled Model	Pooled w/Kerry Vote
Democratic	0.55	<b>0.61</b>	0.47	<b>0.58</b>
Republican	0.26	<b>0.69</b>	<b>0.54</b>	<b>0.78</b>
Liberal	0.22	0.06	<b>0.57</b>	<b>0.77</b>
Conservative	0.46	<b>0.82</b>	<b>0.58</b>	<b>0.79</b>
Average	0.37	<b>0.55</b>	<b>0.54</b>	<b>0.73</b>

*N* (2004) = 1303 (ideology), 2787 (partisanship)

*N* (Pooled) = 6917 (ideology), 9942 (partisanship)

Bold font indicates an improvement over the single-year, basic model.

improves on them slightly, the correlations are still smaller than one would expect compared to a non-clustered poll. Adding a state-level predictor to the single year model produces strong correlations for all variables but percent liberal, which is the measure I attempt to estimate with the fewest respondents. As in the earlier tables that compared MRP estimates to the Annenberg data, the model using pooled data with the state-level predictor performs the best, reaching an average correlation of 0.73 among the four variables tested. This pattern is somewhat maintained in the mean absolute errors reported below in Table A.2.

Table A.2: Mean Absolute Error – GSS and EWM

Mean Abs. Err.	2004 Model	2004 w/ Kerry Vote	Pooled Model	Pooled w/Kerry Vote
Democratic	8.4	<b>8.0</b>	<b>8.0</b>	<b>7.5</b>
Republican	7.2	8.0	<b>4.9</b>	<b>5.3</b>
Liberal	4.5	5.0	5.5	4.8
Conservative	4.3	<b>4.2</b>	<b>3.6</b>	<b>3.0</b>
Average	6.1	6.3	<b>5.5</b>	<b>5.2</b>

*N* (2004) = 1303 (ideology), 2787 (partisanship)

*N* (Pooled) = 6917 (ideology), 9942 (partisanship)

Bold font indicates an improvement over the single-year, basic model.

I repeat this analysis using data from the American National Election Studies. Again, I consider six models rather than four, given the lower *N*. Initially, there is a similar pattern as in the GSS results above, with correlation becoming stronger when an aggregate predictor, vote share for Kerry, is added to the model, as shown in Table A.3. Even more so than with the GSS, the ANES

shows that pooling alone adds little additional value; as discussed earlier, this may be because the longer time frame of the ANES data included here. These results are further reflected in the mean absolute error analysis in Table A.4.

Table A.3: Pearson's Product-Moment Correlations – ANES and EWM

Proportion	2004	2004 w/ KV	Pooled A	Pooled A w/KV	Pooled B	Pooled B w/KV
Democratic	0.38	<b>0.50</b>	<b>0.48</b>	<b>0.58</b>	0.36	<b>0.50</b>
Republican	0.20	<b>0.61</b>	<b>0.26</b>	<b>0.57</b>	<b>0.34</b>	<b>0.62</b>
Liberal	0.61	<b>0.75</b>	0.57	<b>0.74</b>	0.59	<b>0.74</b>
Conservative	0.53	<b>0.74</b>	0.41	<b>0.71</b>	<b>0.55</b>	<b>0.78</b>
Average	0.43	<b>0.65</b>	0.43	<b>0.65</b>	<b>0.46</b>	<b>0.66</b>

*N* (2004) = 912 (ideology), 1175 (partisanship)

*N* (Pooled A) = 2755 (ideology), 4313 (partisanship)

*N* (Pooled B) = 4329 (ideology), 6516 (partisanship)

Bold font indicates an improvement over the single-year, basic model.

However, this ANES analysis, both in the correlations and especially the mean absolute errors, amplifies a trend also seen in the GSS results above: In these validity tests using the disaggregated Erikson, Wright, and McIver data, the results for ideology are more comforting to those who wish to use MRP on cluster-sampled data than those for party identification, reflected in both the correlations and the mean absolute errors. This at first seems surprising, given that there was no such difference in the validity testing using the Annenberg data. One explanation may be that, as Erikson, Wright, and McIver argue, partisanship is less stable than ideology; given that the time periods of the various measures do not completely overlap, especially with the ANES data (which spans up to 2008), this might explain the lower correlations and the higher mean absolute errors in this series of validity tests. These results, therefore, may serve as another caution against pooling when the variable of interest changes over time. It is also worth remembering that in both the individual and pooled models, the MRP estimates generated from the GSS rely on a larger *N* than those generated from the ANES, which could further explain discrepancies in the results.

Table A.4: Mean Absolute Error - ANES and EWM

Proportion	2004	2004 w/ KV	Pooled A	Pooled A w/KV	Pooled B	Pooled B w/KV
Democratic	10.7	<b>9.8</b>	12.2	11.6	13.2	12.8
Republican	12.5	13.7	<b>11.0</b>	<b>11.0</b>	<b>9.0</b>	<b>9.3</b>
Liberal	5.3	<b>4.4</b>	6.4	5.6	6.4	5.7
Conservative	8.8	10.3	<b>8.0</b>	9.2	<b>7.8</b>	<b>8.6</b>
Average	9.3	9.6	9.4	9.4	<b>9.1</b>	<b>9.1</b>

N (2004) = 912 (ideology), 1175 (partisanship)

N (Pooled A) = 2755 (ideology), 4313 (partisanship)

N (Pooled B) = 4329 (ideology), 6516 (partisanship)

Bold font indicates an improvement over the single-year, basic model.

## Appendix B

# Appendix to “The Polarization of Political Values: An Examination of the Heterogeneity of Support for American Core Values”

### B.1 Value Scales and Question Wording

#### B.1.1 ANES Cumulative File

- **Moral traditionalism** (5-point scale):
  - The newer lifestyles are contributing to the breakdown of our society (VCF0851).
  - This country would have many fewer problems if there were more emphasis on traditional family ties (VCF0853).
  - The world is always changing and we should adjust our view of moral behavior to those changes (VCF0852).
  - We should be more tolerant of people who choose to live according to their own moral standards, even if they are very different from our own (VCF0854).
- **Equality of Opportunity** (5-point scale):

- Our society should do whatever is necessary to make sure that everyone has an equal opportunity to succeed (VCF9013).
- One of the big problems in this country is that we don't give everyone an equal chance (VCF9015).
- If people were treated more equally in this country we would have many fewer problems (VCF9018).
- **Egalitarianism** (5-point scale):
  - We have gone too far in pushing equal rights in this country. (VCF9014)
  - It is not really that big a problem if some people have more of a chance in life than others (VCF9016).
  - This country would be better off if we worried less about how equal people are (VCF9017).
- **Limited Government** (2-point scale)
  - The less government the better, OR There are more things that government should be doing (VCF9131).
  - We need a stronger government to handle today's complex economic problems, OR The free market can handle these problems without government being involved (VCF9132).
  - The main reason government has become bigger over the years is because it has gotten involved in things that people should do for themselves, OR Government has become bigger because the problems we face have become bigger (VCF9133).

### B.1.2 2013 CCES Module

- **Moral traditionalism** (5-point scale):
  - The newer lifestyles are contributing to the breakdown of society (COL326).
  - The world is changing, and we should adjust our view of moral behavior to those changes (COL327).
  - We should be more tolerant of people who choose to live according to their own moral standards, even if they are very different from our own (COL328).
  - This country would have fewer problems if there were more emphasis on traditional family ties (COL329).
- **Equality of Opportunity** (5-point scale):
  - Our society should do whatever is necessary to make sure that everyone has an equal opportunity to succeed (COL330).
  - One of the big problems in this country is that we don't give everyone an equal chance (COL332).

- If people were treated more equally in this country, we would have many fewer problems (COL334).
- Our society would be better off if the distribution of wealth were more equal (COL337).
- **Egalitarianism** (5-point scale):
  - It is not really that big a problem if some people have more of a chance in life than others (COL331).
  - This country would be better off if we worried less about how equal people are (COL333).
  - We have gone too far in pushing equal rights in this country (COL335).
  - It seems like blacks, women, homosexuals and other groups don't want equal rights, they want special rights just for them (COL336).
- **Limited Government** (5-point scale, 3-point scale):
  - Government interferes far too much in our everyday lives (COL338).
  - Society should make sure everyone's basic needs are met (COL339).
  - People should be able to rely on the government for help when they need it (COL340).
  - It's not the government's business to try to protect people from themselves (COL341).
  - Do you favor a smaller government with fewer services OR a larger government with many services (COL348).
  - We need a stronger government to handle today's complex economic problems, OR The free market can handle these problems without government being involved (COL349).
  - The main reason government has become bigger over the years is because it has gotten involved in things that people should do for themselves, OR Government has become bigger because the problems we face have become bigger (COL350).
  - The less government the better, OR There are more things that government should be doing (COL351).

## B.2 Reliability of Values Scales

### B.2.1 ANES

Table B.1 below shows the Cronbach's alphas for the three overall NES values scales for each year asked – moral traditionalism, equality, and limited government, both overall and by year. As prior research has noted, while there is some variation over time, the scale reliability coefficients are sufficiently high to give us confidence in the latent values that these scales are measuring.

I then conducted principle components factor analysis on each item scale, looking both at the overall pooled data and at each year of data individually. The equality scale clearly shows two separate factors, one that could be more clearly labeled equality of opportunity and one that could be referred to as egalitarianism (or equality of outcome). The analysis retains two factors

Table B.1: Cronbach's Alphas for Value Scales

	Morality	Equality	Limited Gov.
1984	–	0.5859	–
1986	0.6379	0.6606	–
1988	0.5859	0.6757	–
1990	0.5688	0.6120	0.6939
1992	0.6487	0.7167	0.7099
1994	0.6714	0.6625	–
1996	0.6312	0.706	–
1998	0.5953	–	–
2000	0.6331	0.683	0.7357
2004	0.6722	0.7185	0.7031
2008	0.5451	0.6589	0.6498
2012	0.5502	0.6844	0.7046
Average	0.6195	0.6735	0.7017

Note: These calculations exclude the 2012 Internet panel.

when the data is looked at both pooled and in every year of the data set. Thus, this paper divides this scale into two separate scales, following Goren 2005.

The moral traditional scale is slightly less clear. When pooling the data, only one factor is retained. When looking at the data year by year, there is a mix: of the 11 years where the index was fielded, one factor is retained in 6 of those years and two factors in the other 5. There is not a clear pattern based on time period, and thus there is some room for discretion in the interpretation of this scale. From a substantive perspective, the two potential scales, one on upholding traditional moral values and one on tolerance of differences in moral values, seem less distinct than the two sub-scales in the equality scale. For this reason, I will analyze the moral traditional scale as a single scale in this paper.

The limited government scale, with only three items, is clearly more compact. With the highest scale reliability coefficients, as shown in Table B.1, it is clearly the most reliable scale. In analysis, only one factor is retained both when the data is pooled and when analyzed year by year.

The results of these factor analyses are shown below in Table B.2. While the below table shows the results of this analysis on all years pooled, the year-by-year results are available upon request. Given the findings and discussion above, Table B.3 below shows the Cronbach's alphas for the overall NES values scales of moral traditionalism and equality as well as the potential component scales of each for reference.

### B.2.2 CCES

Table B.4 below shows the Cronbach's alphas for the three overall CCES values scales for each year asked – moral traditionalism, equality and limited government. For each scale, I also show the alphas for subcomponents of the scale, as suggested by theory as well as prior research (Goren

Table B.2: Principle Components Factor Analysis for Value Scales

<b>Morality</b>				
<i>Retained Factors=1</i>				<i>N=18036</i>
Factor	Eigenvalue	Difference	Proportion	Cumulative
Factor 1	1.88928	0.91222	0.4723	0.4723
Factor 2	0.97706	0.36914	0.2443	0.7166
Factor 3	0.60792	0.08217	0.1520	0.8686
Factor 4	0.52575	–	0.1314	1.0000
<b>Equality</b>				
<i>Retained Factors=2</i>				<i>N=18970</i>
Factor	Eigenvalue	Difference	Proportion	Cumulative
Factor 1	2.28637	1.02546	0.3811	0.3811
Factor 2	1.26092	0.51449	0.2102	0.5912
Factor 3	0.74642	0.07394	0.1244	0.7156
Factor 4	0.67248	0.14344	0.1121	0.8277
Factor 5	0.52904	0.02427	0.0882	0.9159
Factor 6	0.50477	–	0.0841	1.0000
<b>Limited Government</b>				
<i>Retained Factors=1</i>				<i>N=9625</i>
Factor	Eigenvalue	Difference	Proportion	Cumulative
Factor 1	1.88054	1.27187	0.6268	0.6268
Factor 2	0.60866	0.09786	0.2029	0.8297
Factor 3	0.51080	–	0.1703	1.0000

Note: Cronbach's alphas exclude the 2012 Internet panel.

2005, Gastil et. al. 2011). While there is some variation, what is most striking is the high alpha each scale has, especially when compared to the ANES above. It may be the case that internet survey respondents have more consistent and ideologically constrained worldviews, producing these higher reliability coefficients for similar (though admittedly not always identical) scales. It could also be that the mode of survey somehow facilitates more consistent answers. Since these are two independent surveys rather than a random experiment, it is difficult to pinpoint the cause of these differences, even while it is interesting to note them.

Table B.5 shows the principle components factor analysis for each scale. As with Table B.2, we see that the moral traditionalism scale loads on 1 factor while the equality scale loads on two, which can be broadly broken down into equality of opportunity and egalitarianism. Despite the additional items added to the survey, the limited government scale continues to load on one factor as well.



Table B.3: Cronbach's Alphas for Alternate Specifications of Value Scales

	Morality			Equality		
	Full	Family	Tolerance	Full	Equal Opp.	Egalitarian
1984	–	–	–	0.5859	0.5418	0.5735
1986	0.6379	0.6127	0.5601	0.6606	0.6121	0.6361
1988	0.5859	0.6105	0.5515	0.6757	0.6566	0.6700
1990	0.5688	0.5010	0.5353	0.6120	0.6127	0.6173
1992	0.6487	0.6862	0.6388	0.7167	0.6582	0.7308
1994	0.6714	0.6532	0.6274	0.6625	0.6348	0.6330
1996	0.6312	0.5916	0.6001	0.7060	0.6789	0.6909
1998	0.5953	0.6786	0.4833	–	–	–
2000	0.6331	0.5994	0.5673	0.6830	0.5815	0.6529
2004	0.6722	0.7317	0.4998	0.7185	0.6439	0.7241
2008	0.5451	0.5860	0.5269	0.6589	0.6421	0.7221
2012	0.5502	0.6052	0.4974	0.6844	0.6288	0.7030
Average	0.6195	0.6313	0.5628	0.6735	0.6289	0.6737

Note: Cronbach Alphas are calculated in Stata and exclude the 2012 Internet panel.

Table B.4: Cronbach's Alphas for Value Scales and Alternate Specifications

<b>Morality</b>			
	Full	Family	Tolerance
CCES	0.8113	0.8470	0.7179
<b>Equality</b>			
	Full	Equal Opp.	Egalitarian
CCES	0.9013	0.8976	0.8515
<b>Limited Government</b>			
	Full	Lim. Gov.	Individualism
CCES	0.8890	0.8896	0.7630

### B.3 Results of Analysis of Individual Values with a Single Equality Scale

This section shows the figures for the individual values section with equality represented as a single values scale, rather than broken into its two component parts of equality of opportunity and egalitarianism. Overall the patterns are quite similar to those shown in the main paper, but often the picture of change over time of the equality scale is somewhat less clean than when it is broken down into its component parts. This is not surprising given that the two scales often show contradictory patterns, as discussed in the paper.

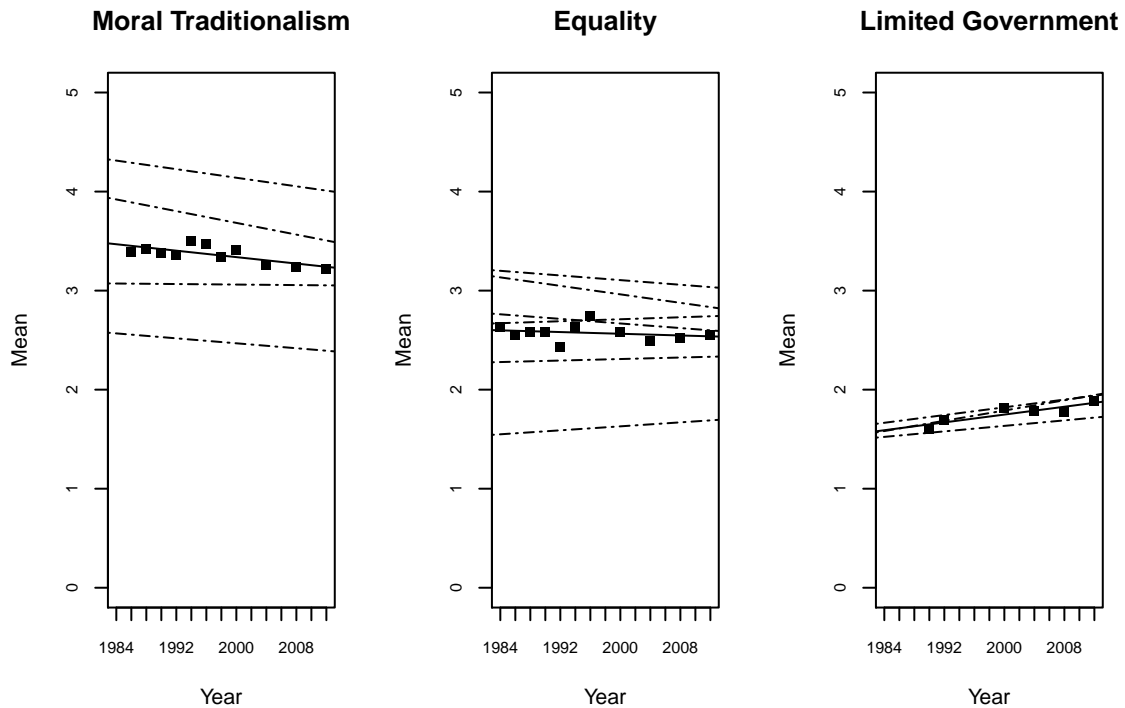
Table B.5: Principle Components Factor Analysis for Value Scales

<b>Morality</b>				
<i>Retained Factors=1</i>				<i>N=1900</i>
Factor	Eigenvalue	Difference	Proportion	Cumulative
Factor 1	2.55550	1.78182	0.6389	0.6389
Factor 2	0.77368	0.36180	0.1934	0.8323
Factor 3	0.41187	0.15292	0.1030	0.9353
Factor 4	0.25895	–	0.0647	1.0000
<b>Equality</b>				
<i>Retained Factors=2</i>				<i>N=1884</i>
Factor	Eigenvalue	Difference	Proportion	Cumulative
Factor 1	4.77600	3.66298	0.5970	0.5970
Factor 2	1.11302	0.46394	0.1391	0.7361
Factor 3	0.64907	0.12278	0.0811	0.8173
Factor 4	0.52630	0.17655	0.0658	0.8830
Factor 5	0.34975	0.00995	0.0437	0.9268
Factor 6	0.33980	0.09373	0.0425	0.9692
Factor 7	0.24607	0.24607	0.0308	1.0000
Factor 8	-0.00000	–	-0.0000	1.0000
<b>Limited Government</b>				
<i>Retained Factors=1</i>				<i>N=1866</i>
Factor	Eigenvalue	Difference	Proportion	Cumulative
Factor 1	4.74325	3.82781	0.5929	0.5929
Factor 2	0.91545	0.30205	0.1144	0.7073
Factor 3	0.61339	0.21216	0.0767	0.7840
Factor 4	0.40123	0.02317	0.0502	0.8342
Factor 5	0.37806	0.02087	0.0473	0.8814
Factor 6	0.35719	0.01773	0.0446	0.9261
Factor 7	0.33946	0.08751	0.0424	0.9685
Factor 8	0.25196	–	0.0315	1.0000

#### B.4 Analysis of Individual Values and Partisanship Overtime with the Inclusion of the 2012 ANES Internet Supplement Module

This section shows how the results presented in this paper change when including the internet supplement module from the 2012 ANES. While the direction of the story is the same across the analysis, the magnitude of the narrative does change slightly when the internet data is included. Overall, including this data appears to indicate an even stronger relationship between partisanship and political values over time, though the differences are minor and not always

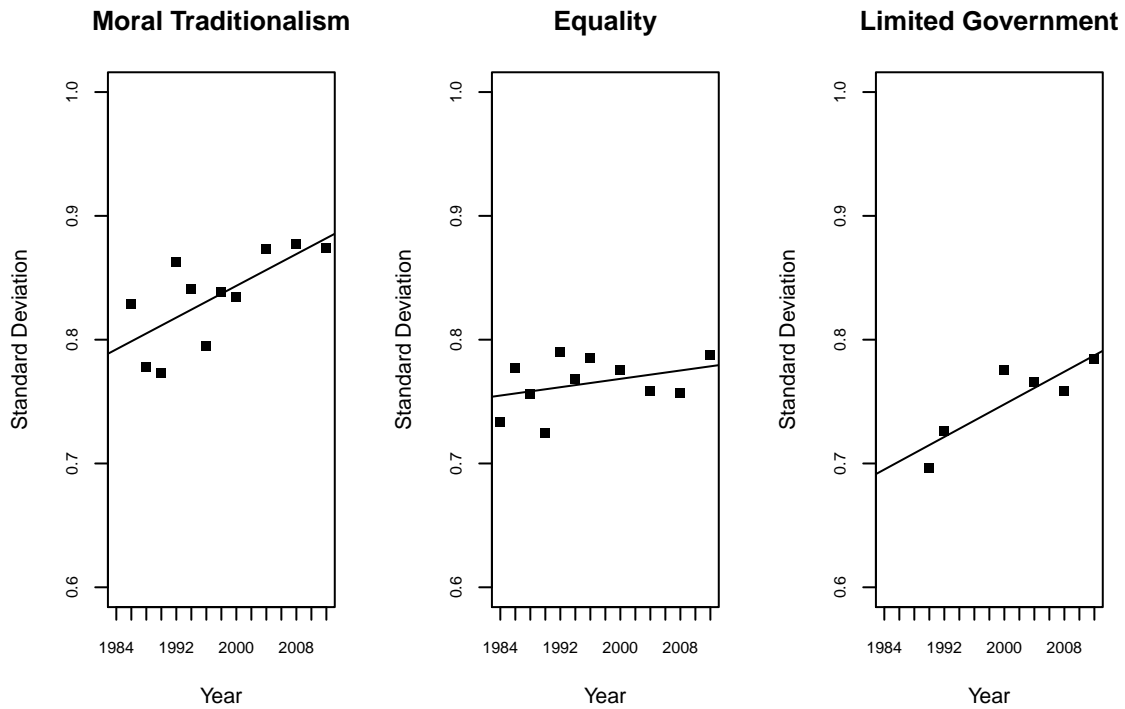
Figure B.1: Support for Political Values Over Time (1984-2012) – Single Equality Scale



Note: These three graphs show the mean of each of the values scales for each year, represented by a black square, as well as a solid line of best fit showing the trend over time. A higher mean signals a more conservative position. The dashed lines show the line of best fit over time for each individual component of each values scale. Note that for the equality scale, not all of its component parts move in the same direction, obscuring movement over time. This issue is remedied when we shift to looking at equality in its two component parts.

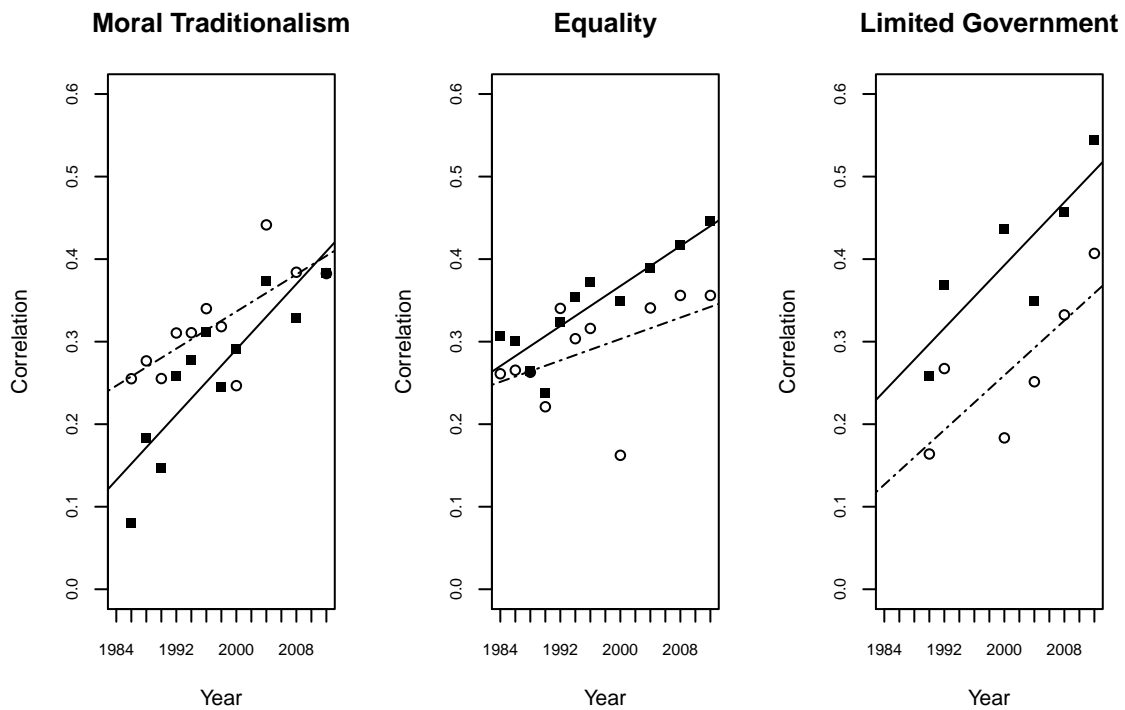
consistent.

Figure B.2: Standard Deviation of Political Values Over Time (1984-2012) – Single Equality Scale



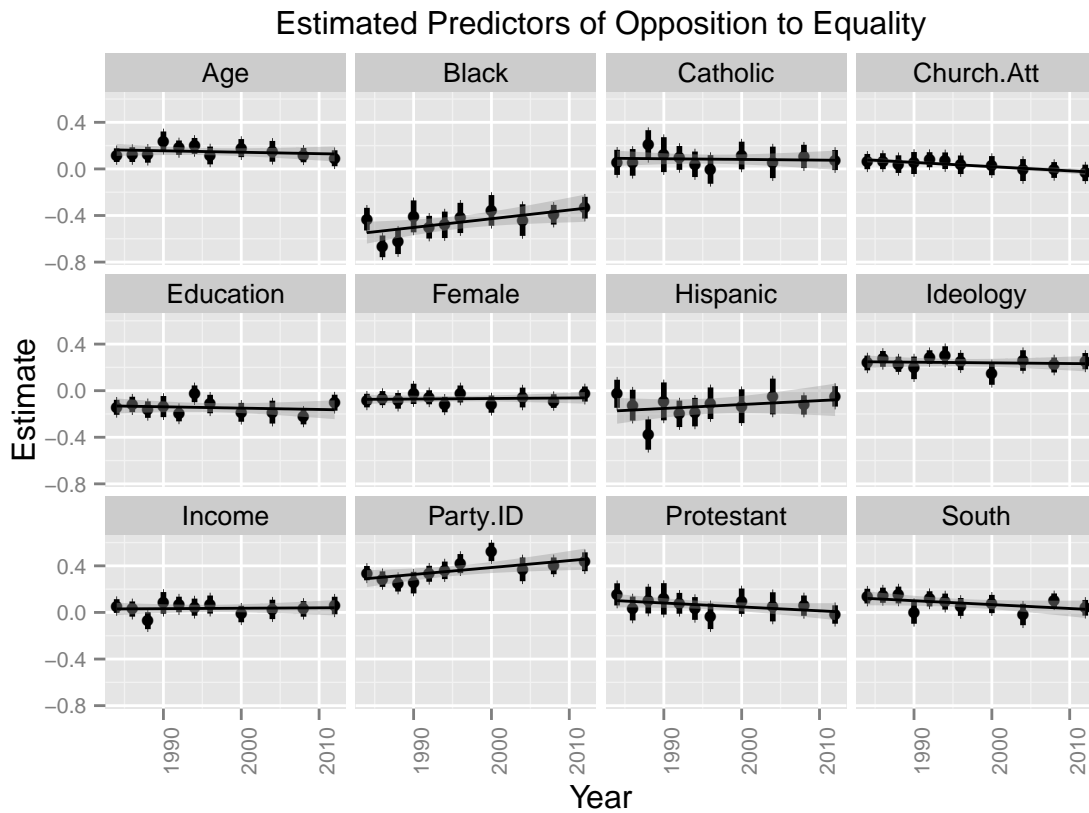
Note: These three graphs show the standard deviation of each of the political values scales over the period of time that each question was asked on the ANES. For each political value, its standard deviation has increased over the last three decades.

Figure B.3: Correlation of Political Values and Partisanship and Political Values and Ideology Over Time (1984-2012) – Single Equality Scale



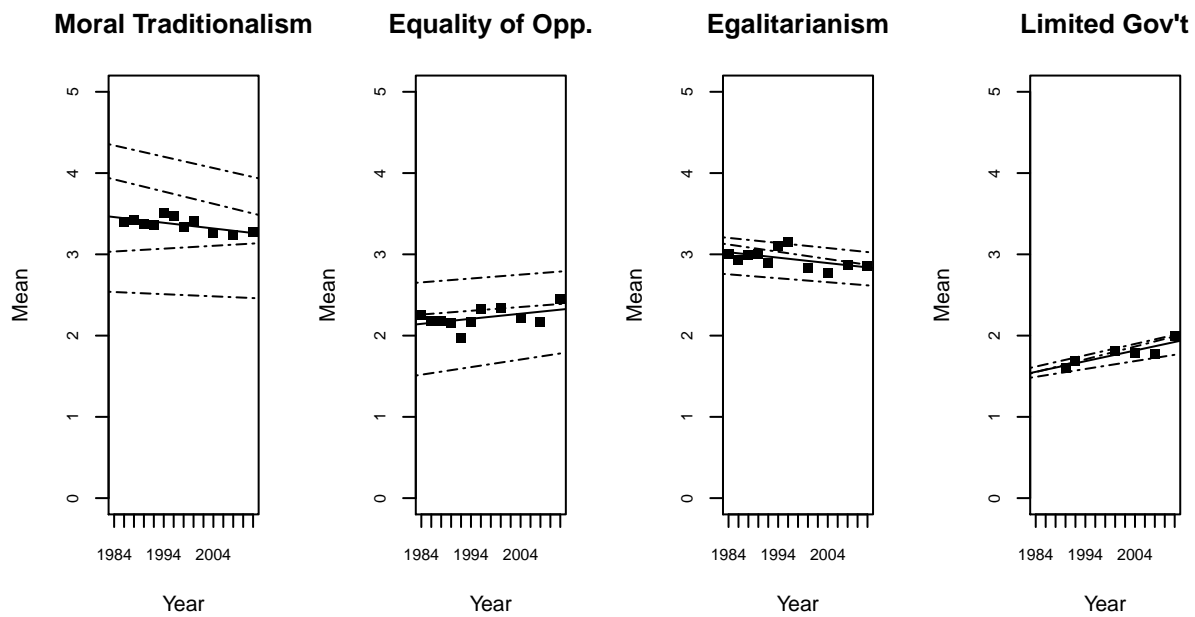
Note: These three graphs show the correlation between each of the values scales and partisanship (solid line) and ideology (dashed line) over the period of time that each question was asked on the ANES. For each political value, the correlation with partisanship has increased over the last three decades.

Figure B.4: Coefficient Plot of Estimated Predictors of Opposition to Equality (1984-2012)



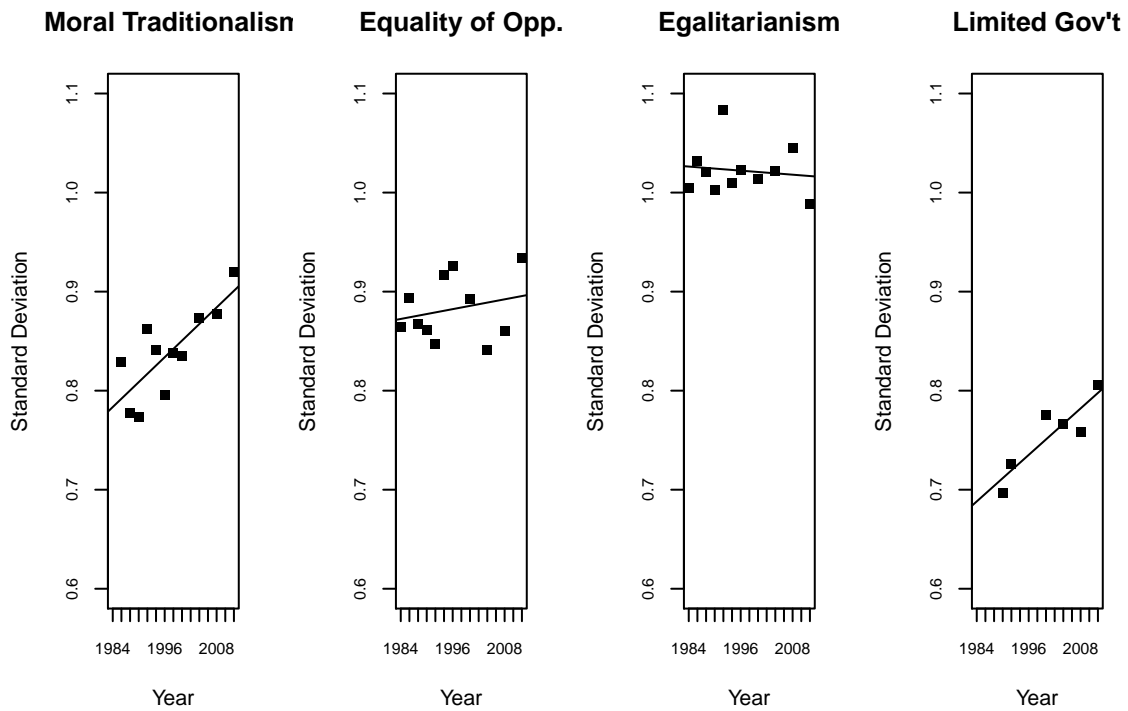
Note: This figure plots the coefficients from regressions of an equality index on several variables, with their 95% confidence intervals over time and a linear line of best fit. Each coefficient has been standardized for ease of comparison across coefficients.

Figure B.5: Support for Political Values Over Time (1984-2012) – Including 2012 Internet Module



Note: Figure B.5 shows the mean of each value scale for each year, represented by a black square, as well as a solid black line of best fit showing the trend over time. A higher mean signals a more conservative position. The dashed lines show the line of best fit over time for each individual component of each values scale. The results in this figure are quite similar to those that exclude the data from the internet panel.

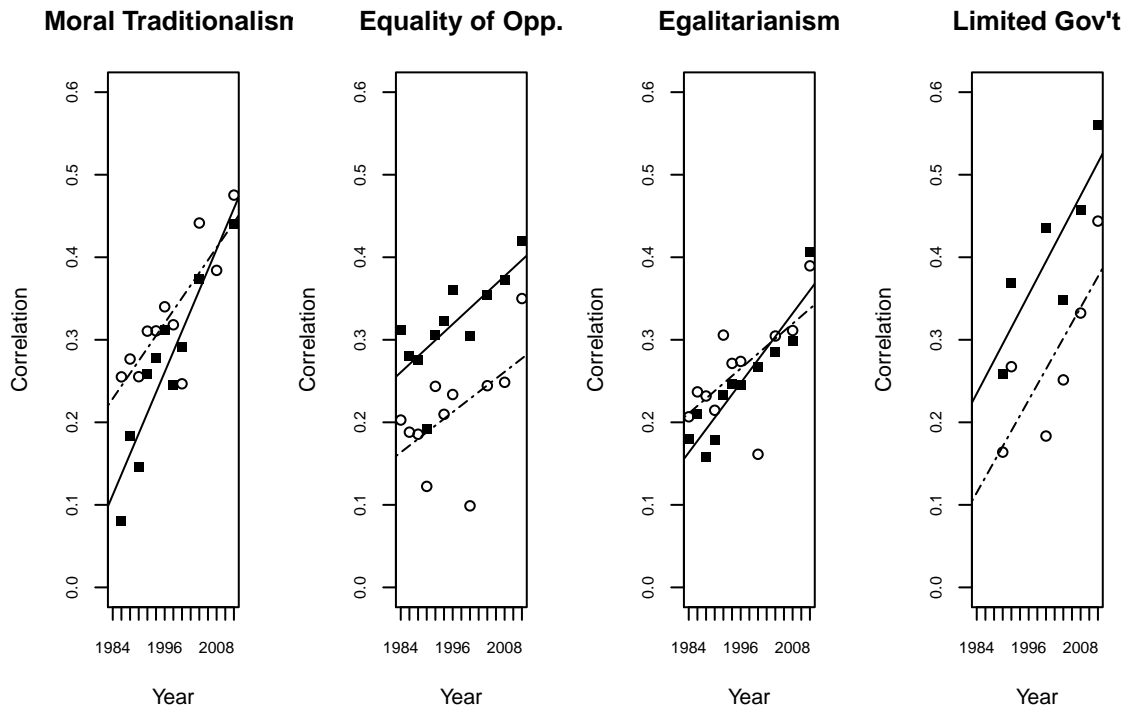
Figure B.6: Standard Deviation of Political Values Over Time (1984-2012) – Including 2012 Internet Module



Note: These four graphs show the standard deviation of each of the political values scales over the period of time that each question was asked on the ANES. For each political value other than egalitarianism, its standard deviation has increased over the last three decades; when the internet data is included, egalitarianism's standard deviation is fairly constant.

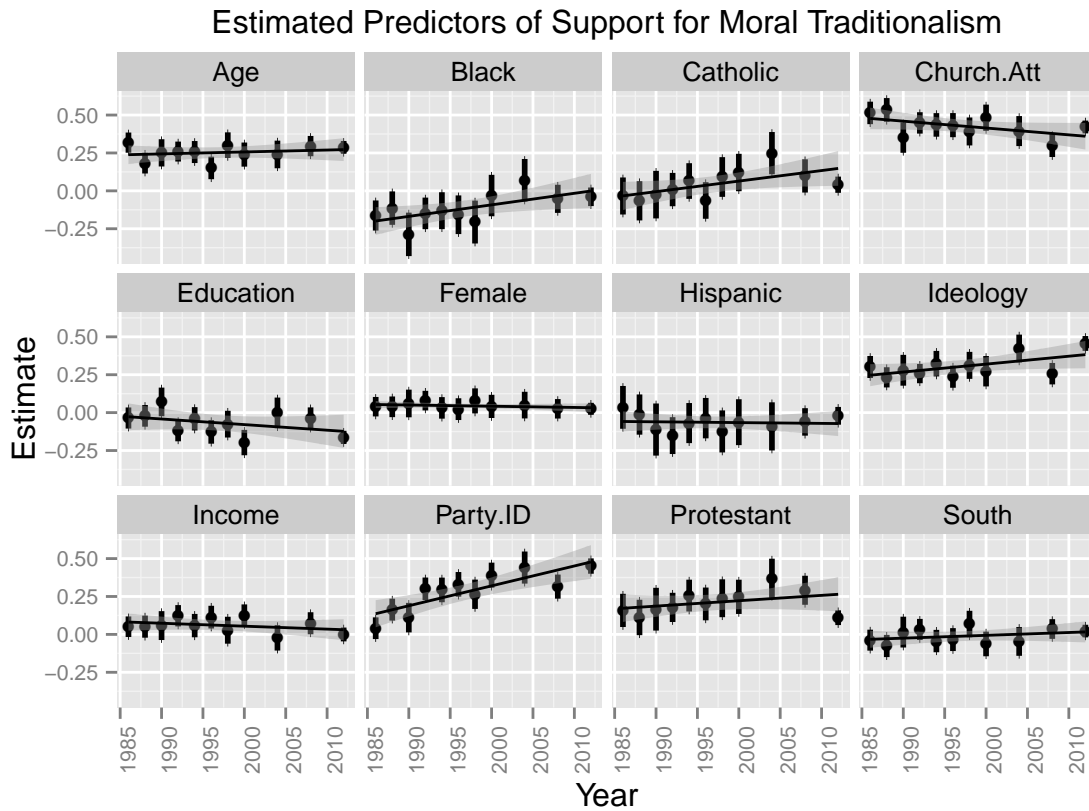


Figure B.7: Correlation of Political Values and Partisanship and Political Values and Ideology Over Time (1984-2012) – Including 2012 Internet Module



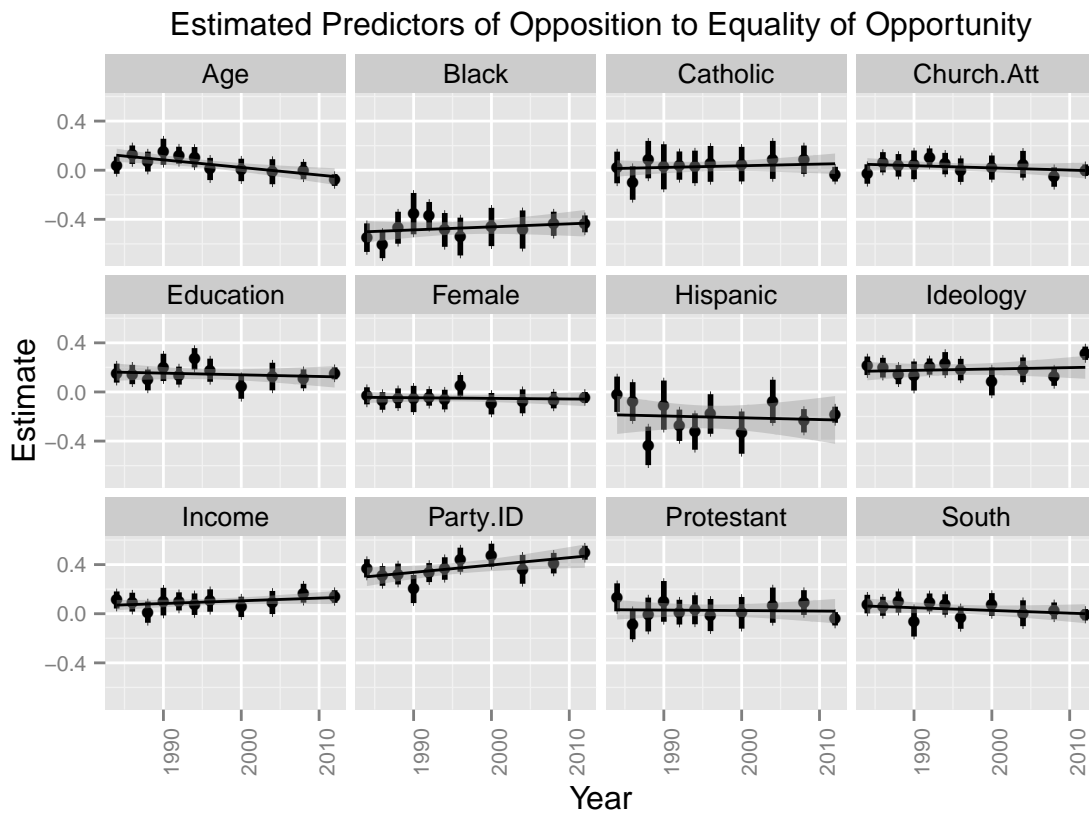
Note: These four graphs show the correlation between each of the values scales and partisanship (solid line) and ideology (dashed line) over the period of time that each question was asked on the ANES. For each political value, the correlation with partisanship has increased over the last three decades. While this figure looks quite similar to the figure without the internet data, it's worth noting that the correlations between ideology and political values are slightly stronger in this data, indicating that respondents on the internet panel may be better sorted on ideological grounds than those taking the more traditional survey.

Figure B.8: Coefficient Plot of Moral Traditionalism (1986-2012) – Including 2012 Internet Module



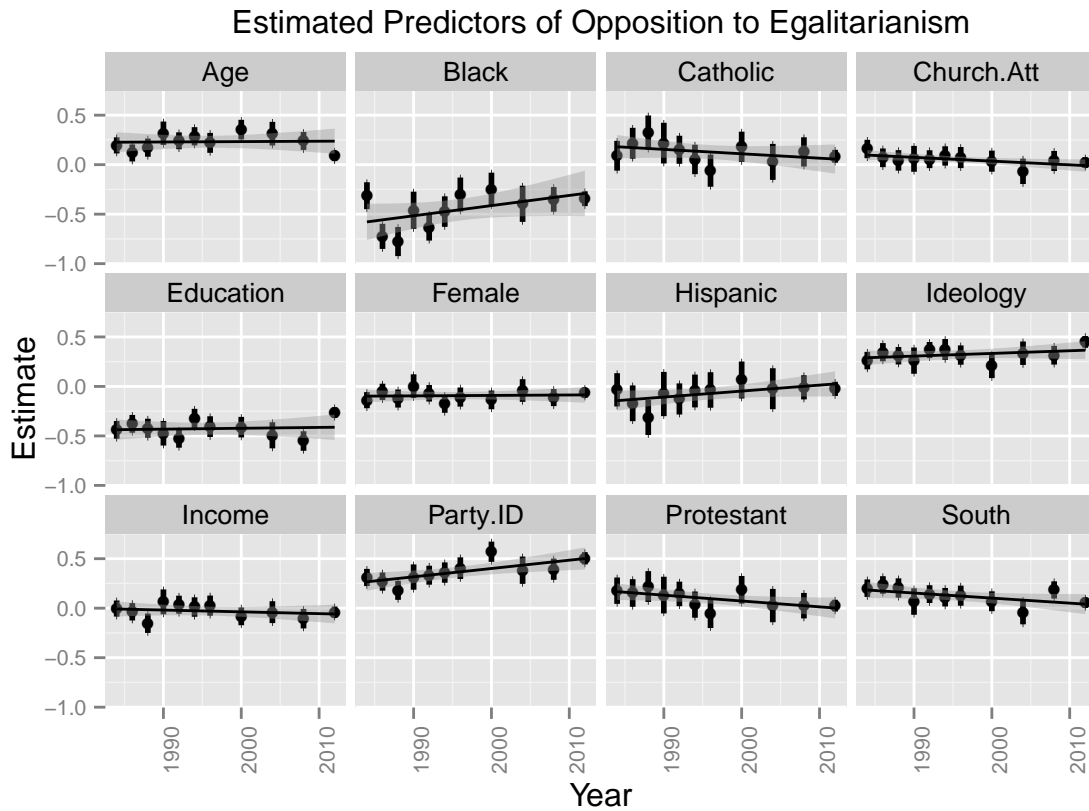
Note: This figure plots the coefficients from regressions of a moral traditionalism index on several variables, with their 95% confidence intervals over time and a linear line of best fit. Each coefficient has been standardized for ease of comparison across coefficients.

Figure B.9: Coefficient Plot of Equality of Opportunity (1984-2012) – Including 2012 Internet Module



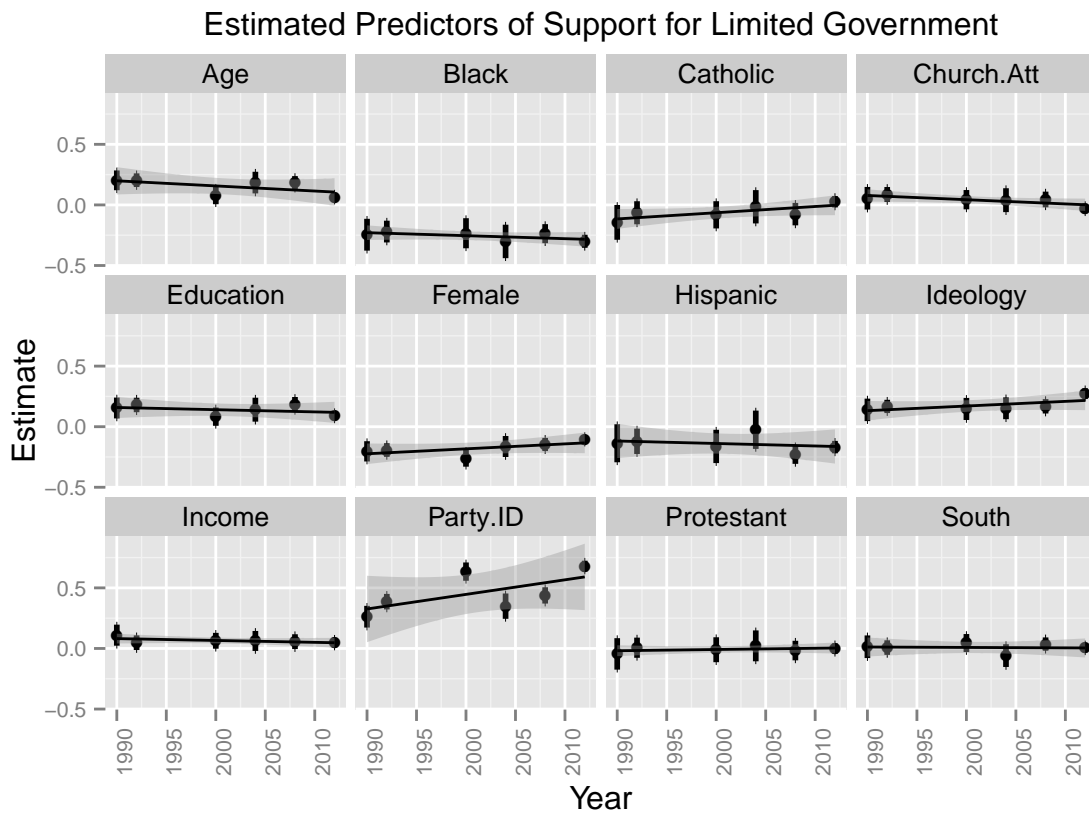
Note: This figure plots the coefficients from regressions of an equality of opportunity index on several variables, with their 95% confidence intervals over time and a linear line of best fit. Each coefficient has been standardized for ease of comparison across coefficients.

Figure B.10: Coefficient Plot of Egalitarianism (1984-2012) – Including 2012 Internet Module



Note: This figure plots the coefficients from regressions of an equality of opportunity index on several variables, with their 95% confidence intervals over time and a linear line of best fit. Each coefficient has been standardized for ease of comparison across coefficients.

Figure B.11: Coefficient Plot of Limited Government (1990-2012) – Including 2012 Internet Module

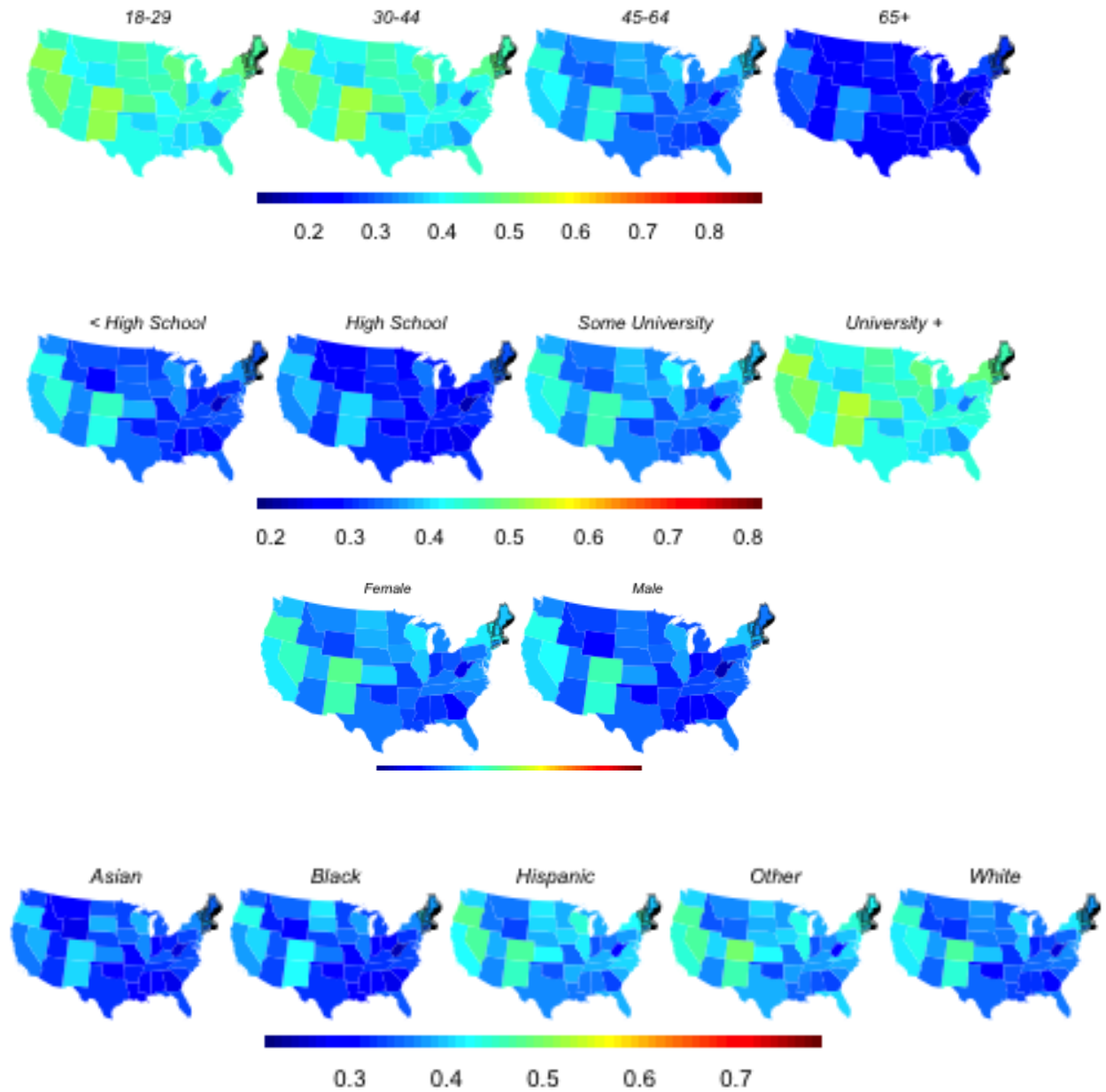


Note: This figure plots the coefficients from regressions of a limited government index on several variables, with their 95% confidence intervals over time and a linear line of best fit. Each coefficient has been standardized for ease of comparison across coefficients.

## **B.5 State-Level Maps for Support for the Liberal Position of Each Values Scale**

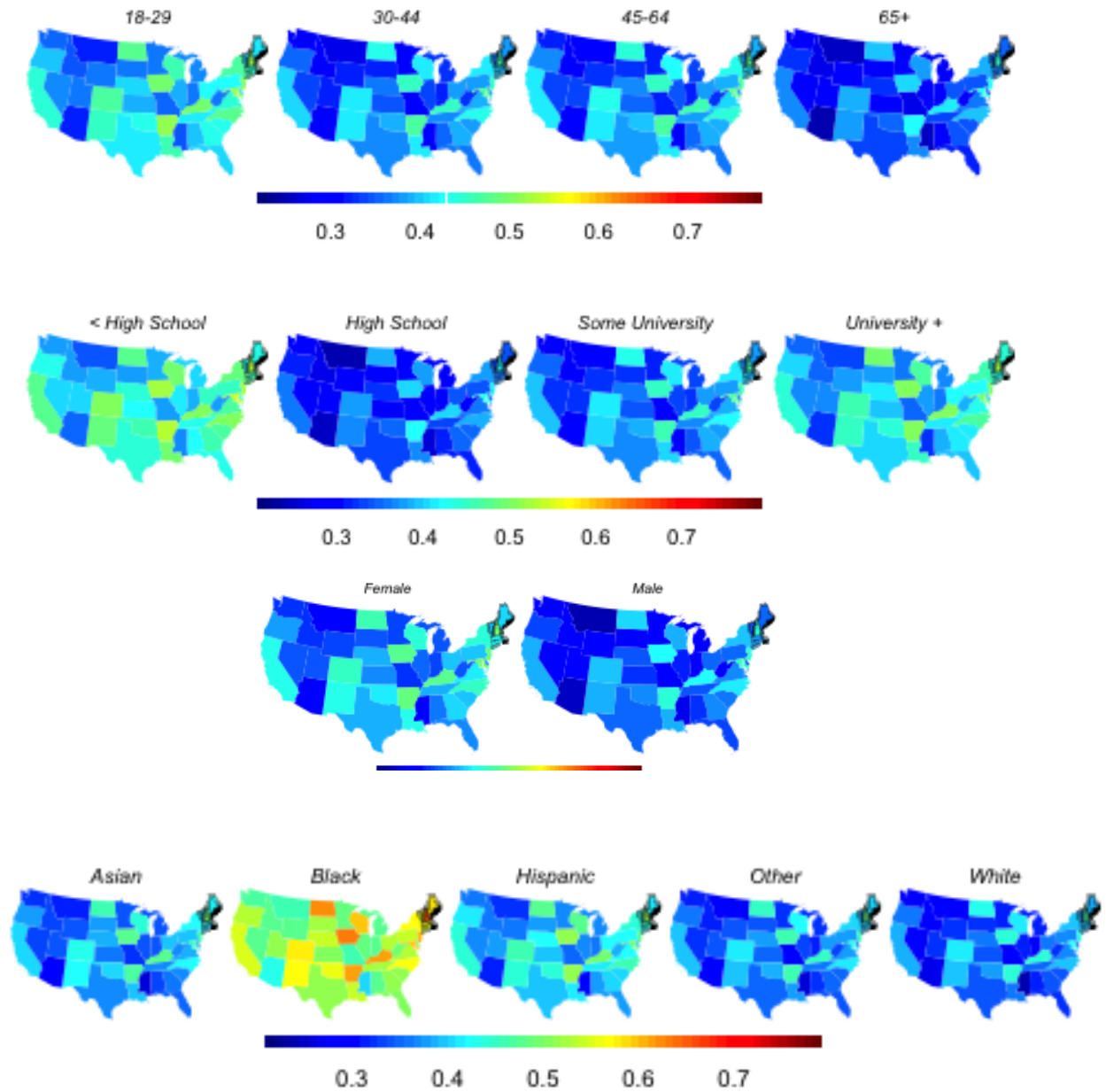
This section shows the demographic maps for the four value scales when measuring the level of support for the traditionally-liberal end of the values scale. While the patterns are quite similar for three of the value scales, we see much more variation across demographic and geographic groups when looking at support for equality of opportunity than we did when looking at opposition to it.

Figure B.12: Opposition to Traditional Moral Values, by State and Demographics



Note: These graphs show the percentage of people who oppose traditional moral values in each state by age group, education level, gender, and race.

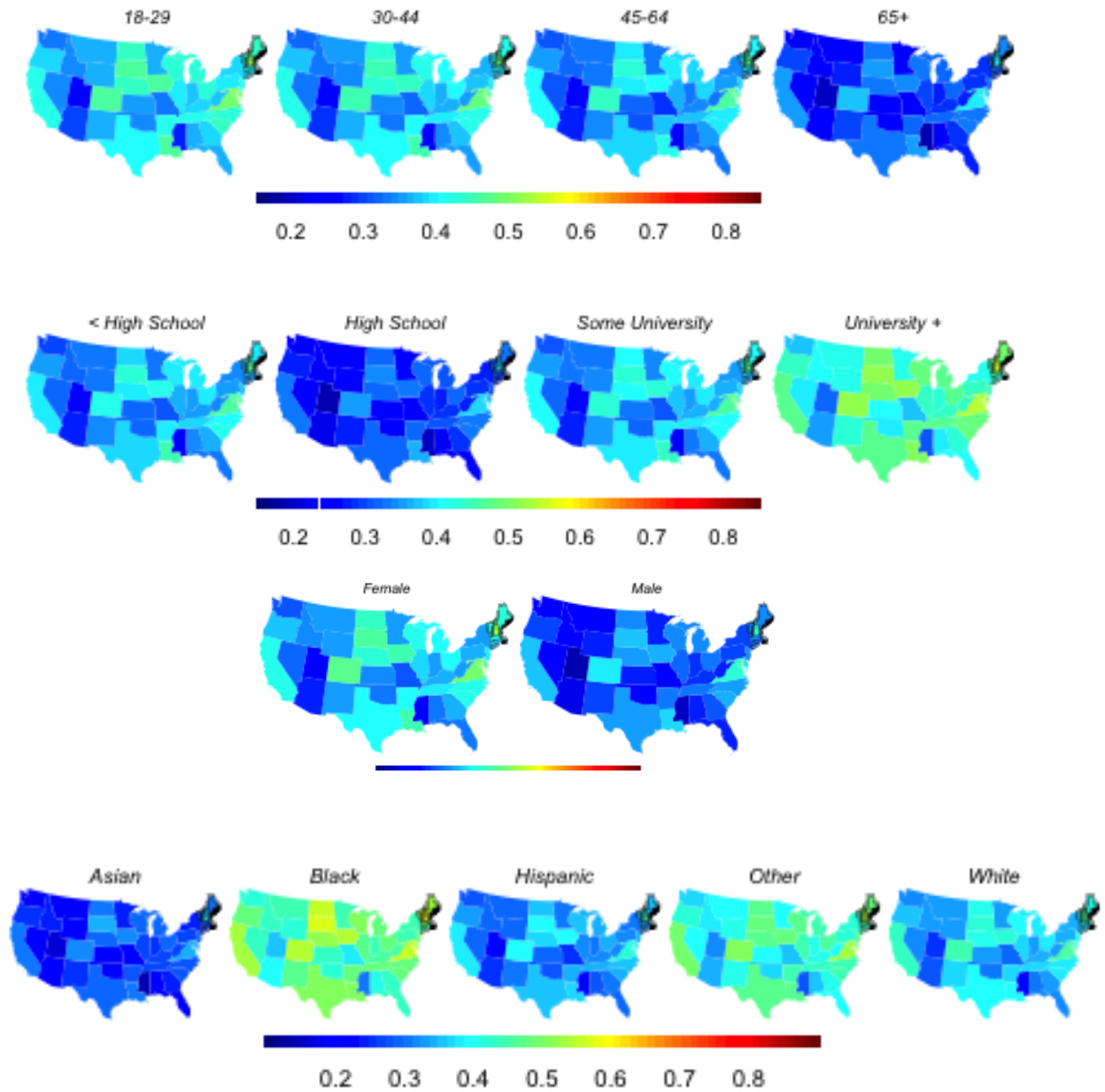
Figure B.13: Opposition to Limited Government and Individualism, by State and Demographics



Note: These graphs show the percentage of people who oppose the values of limited government and individualism in each state by age group, education level, gender, and race.

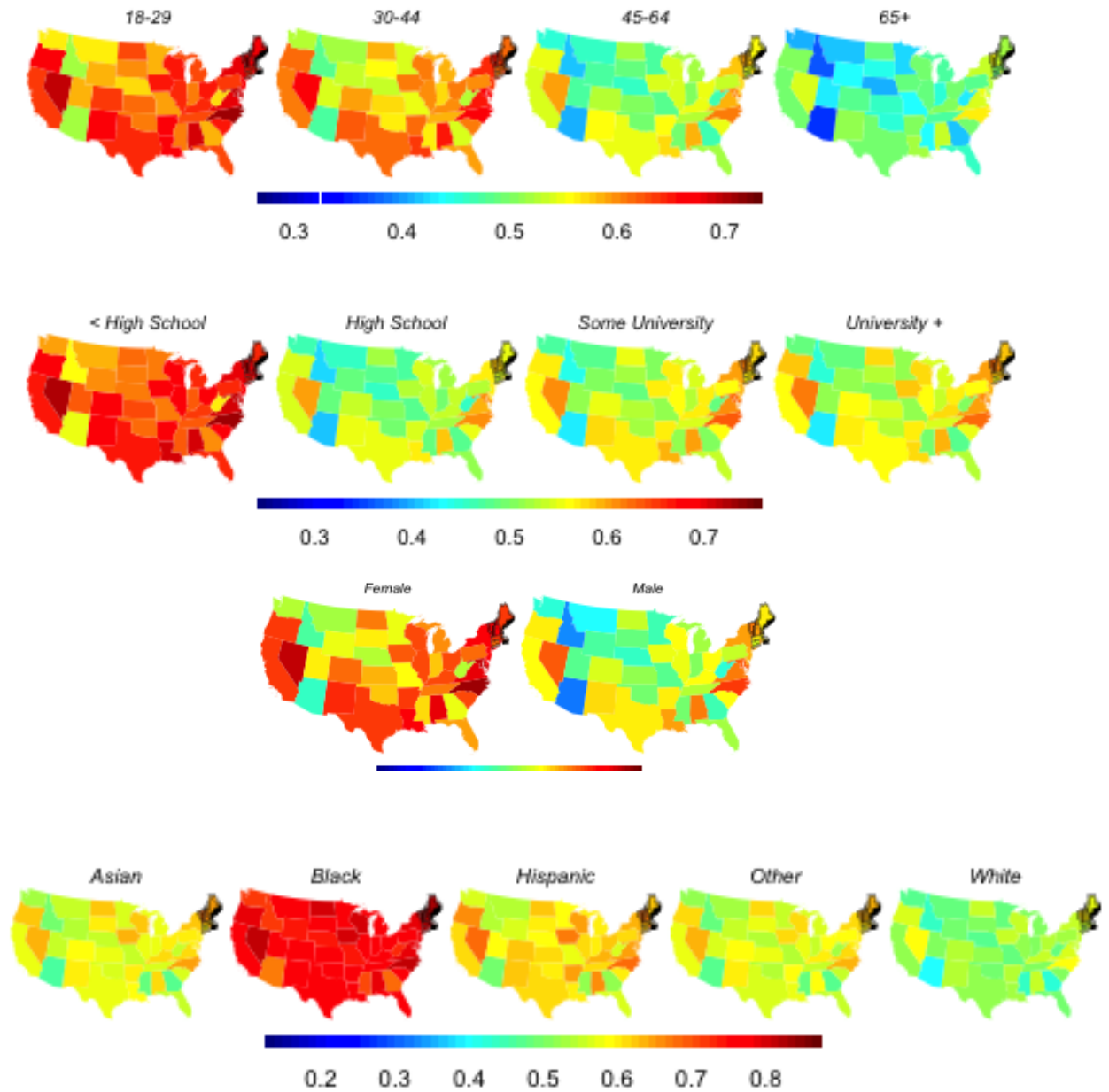


Figure B.14: Support for Egalitarianism, by State and Demographics



Note: These graphs show the percentage of people who support the value of egalitarianism in each state by age group, education level, gender, and race.

Figure B.15: Support for Equality of Opportunity, by State and Demographics



Note: These graphs show the percentage of people who support the value of equality of opportunity in each state by age group, education level, gender, and race.

## B.6 Replication of State-Level Analysis without Including Presidential Vote Share in the Model

This section replicates the models and subsequent analysis created in the state-level section without the use of the partisanship measure, state-level presidential vote share in 2012, in the multilevel models. While partisanship has been shown to be an important tool in the MRP toolkit, there are reasons to be concerned that it would lead to endogeneity in this analysis, given that the analysis focuses on the increased role that partisanship may play in predicting and shaping political values. Thus, it could be problematic to include a measure of partisanship in the MRP estimates. The models still model state effects using percent religious as well as the Census region within which the states are located, as well as the same demographic categories. While there is variation across the two model specifications, overall the results are quite similar, showing that the inclusion of this variable is not driving the empirical findings discussed in this section. Table B.6 shows that the correlations between the two model specifications are quite high across all value scale measurements. Table B.7 and B.8 replicate Tables 3.1 and 3.2 with very similar results. The subsequent figures reproduce the maps for support for the conservative and liberal direction of each scale, again with quite similar patterns across all value scales.

Table B.6: Correlations Between Model Specifications With and Without Presidential Vote Share (CCES Data)

	Conservative Scale	Liberal Scale
Moral Traditionalism	0.94***	0.96***
Equality of Opportunity	0.96***	0.99***
Egalitarianism	1.00***	1.00***
Limited Government	0.99***	1.00***

All numbers are the Pearson's correlation coefficient.

\*  $p < 0.05$ , \*\*  $p < 0.01$ , \*\*\*  $p < 0.001$

Table B.7: Value Scale Range (CCES Data)

<i>Liberal Value Scale</i>			
	Minimum	Maximum	Difference
Moral Traditionalism	28.9	46.7	17.8
Equality of Opportunity	43.7	63.9	20.2
Egalitarianism	24.8	48.0	23.2
Limited Government	28.9	47.7	18.8
<i>Conservative Value Scale</i>			
Moral Traditionalism	30.8	54.9	24.1
Equality of Opportunity	17.9	33.7	15.8
Egalitarianism	30.7	51.2	20.5
Limited Government	33.3	50.9	17.6

All numbers are the average percentage of state-level population support for that value scale.

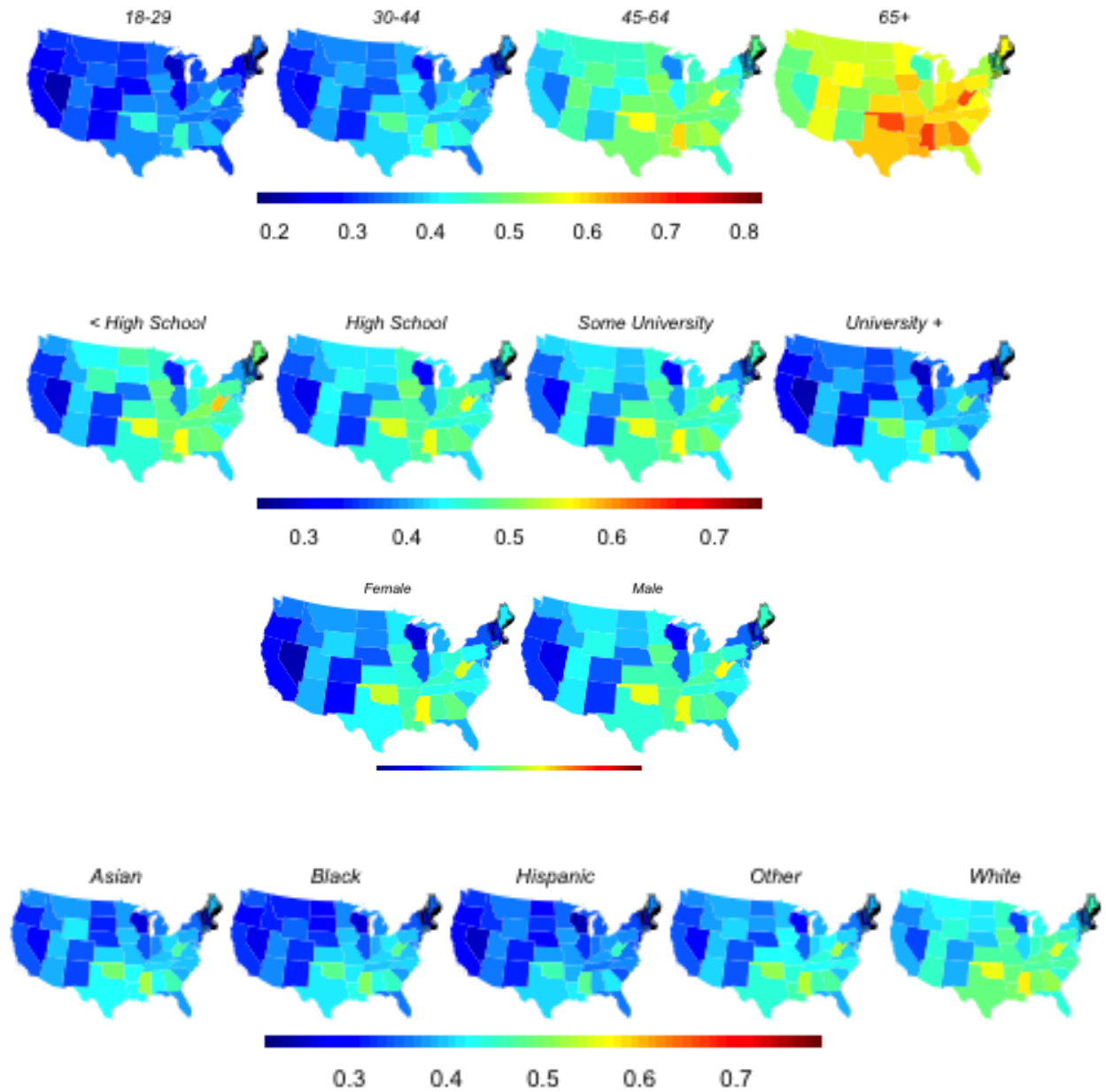
Table B.8: Value Scale Correlations (CCES Data)

<i>Conservative Value Scale</i>				
	Moral Trad	Equal Opp.	Egal.	Limited Gov.
Moral Traditionalism	1.00***			
Equality of Opportunity	0.31*	1.00***		
Egalitarianism	0.51***	0.57***	1.00***	
Limited Government	0.38**	0.79***	0.61***	1.00***
<i>Liberal Value Scale</i>				
	Moral Trad	Equal Opp.	Egal.	Limited Gov.
Moral Traditionalism	1.00***			
Equality of Opportunity	0.28	1.00***		
Egalitarianism	0.26	0.55***	1.00***	
Limited Government	0.24	0.67***	0.65***	1.00***

All numbers are the Pearson's correlation coefficient.

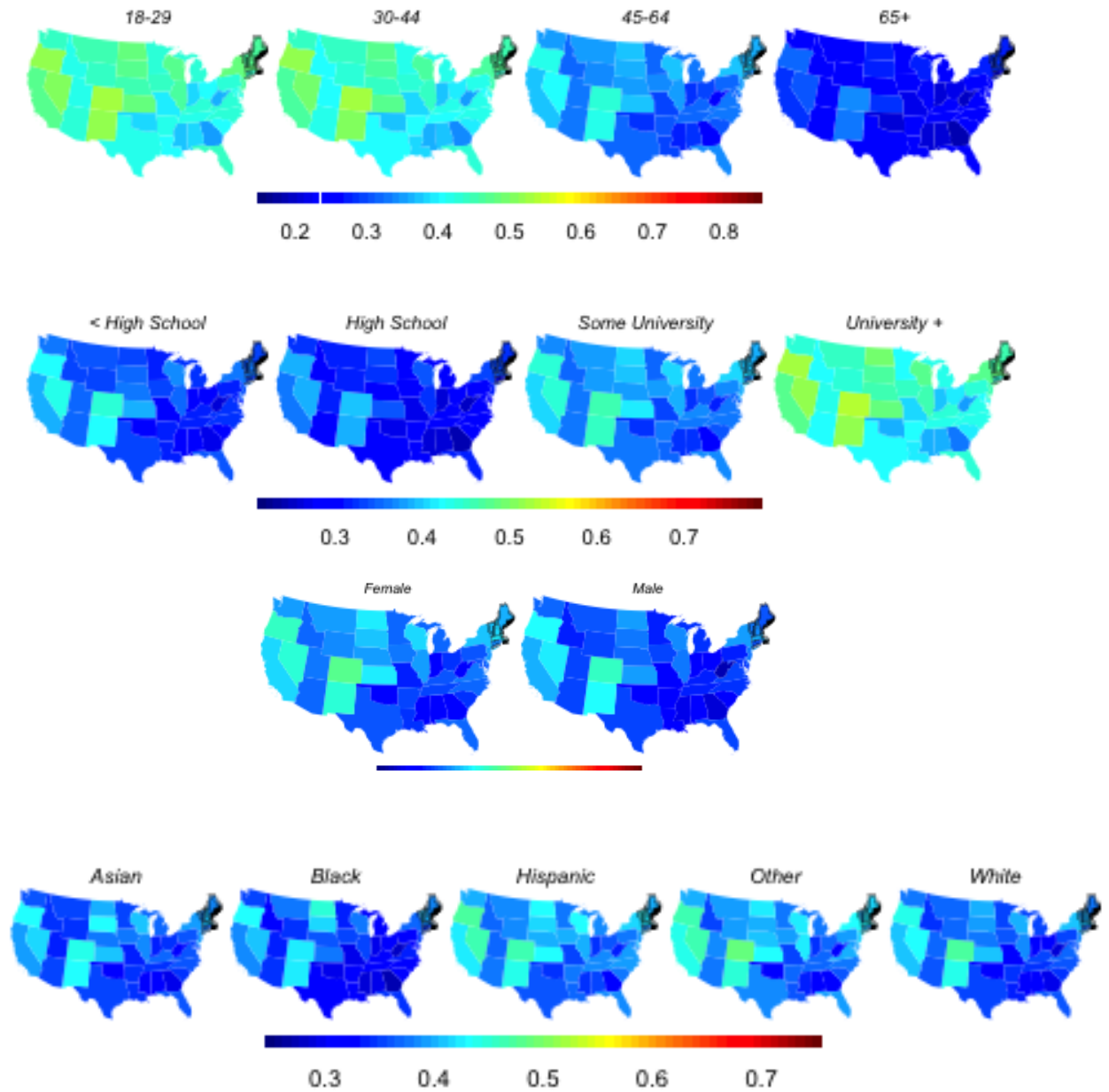
\*  $p < 0.05$ , \*\*  $p < 0.01$ , \*\*\*  $p < 0.001$

Figure B.16: Support of Traditional Moral Values, by State and Demographics



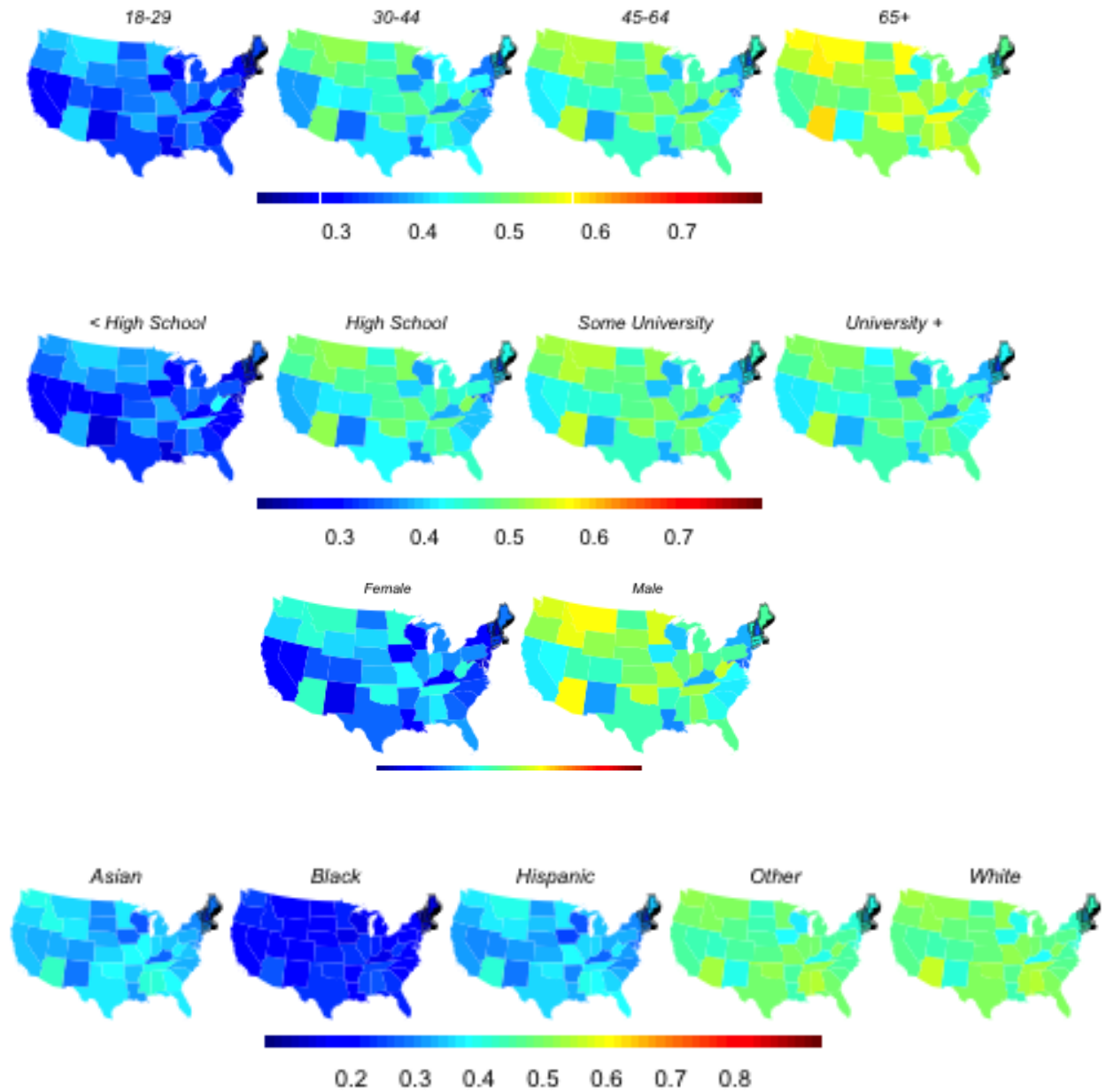
Note: These graphs show the percentage of people who support traditional moral values in each state by age group, education level, gender, and race.

Figure B.17: Opposition to Traditional Moral Values, by State and Demographics



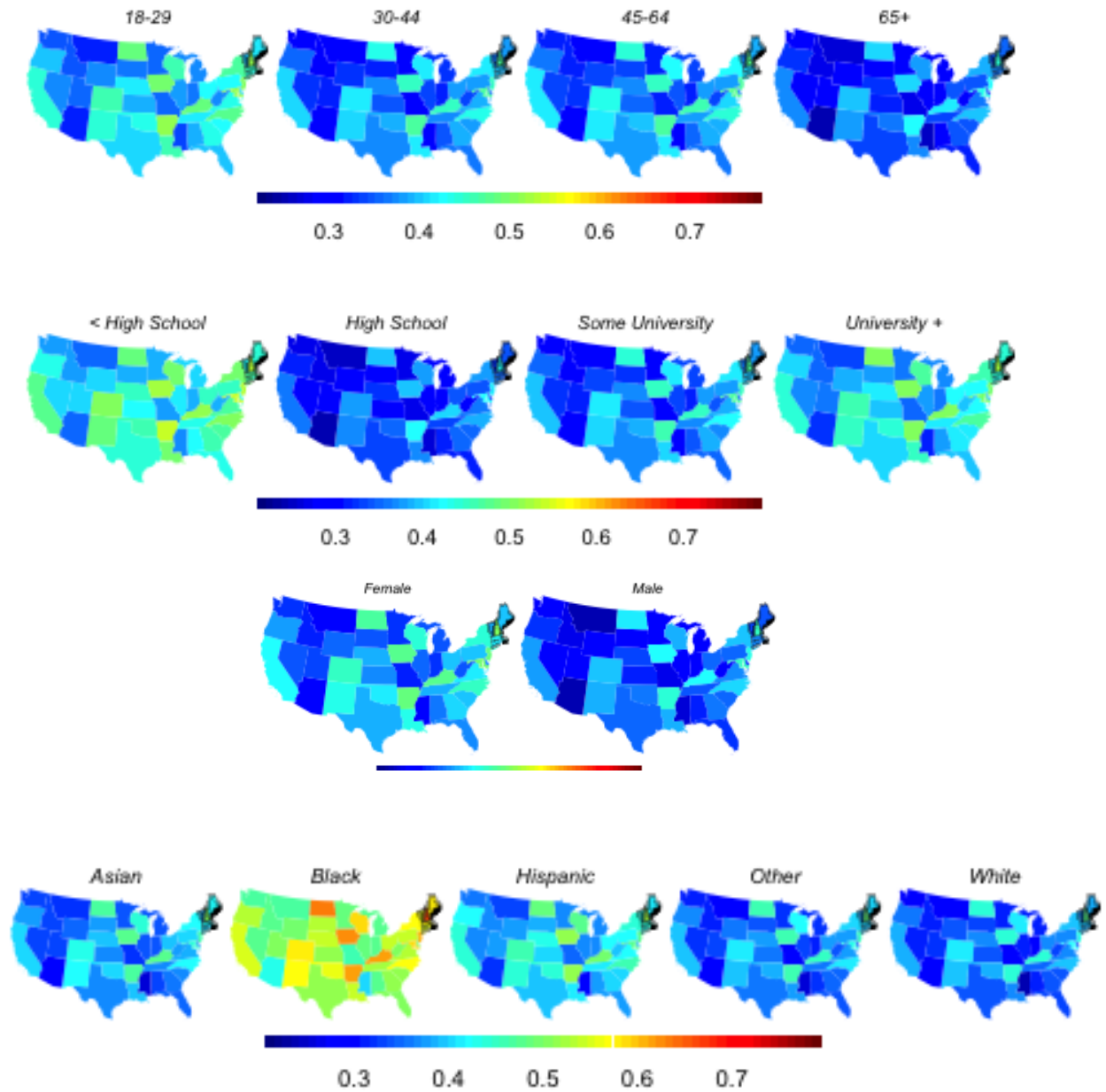
Note: These graphs show the percentage of people who oppose traditional moral values in each state by age group, education level, gender, and race.

Figure B.18: Support of Limited Government and Individualism, by State and Demographics



Note: These graphs show the percentage of people who support the values of limited government and individualism in each state by age group, education level, gender, and race.

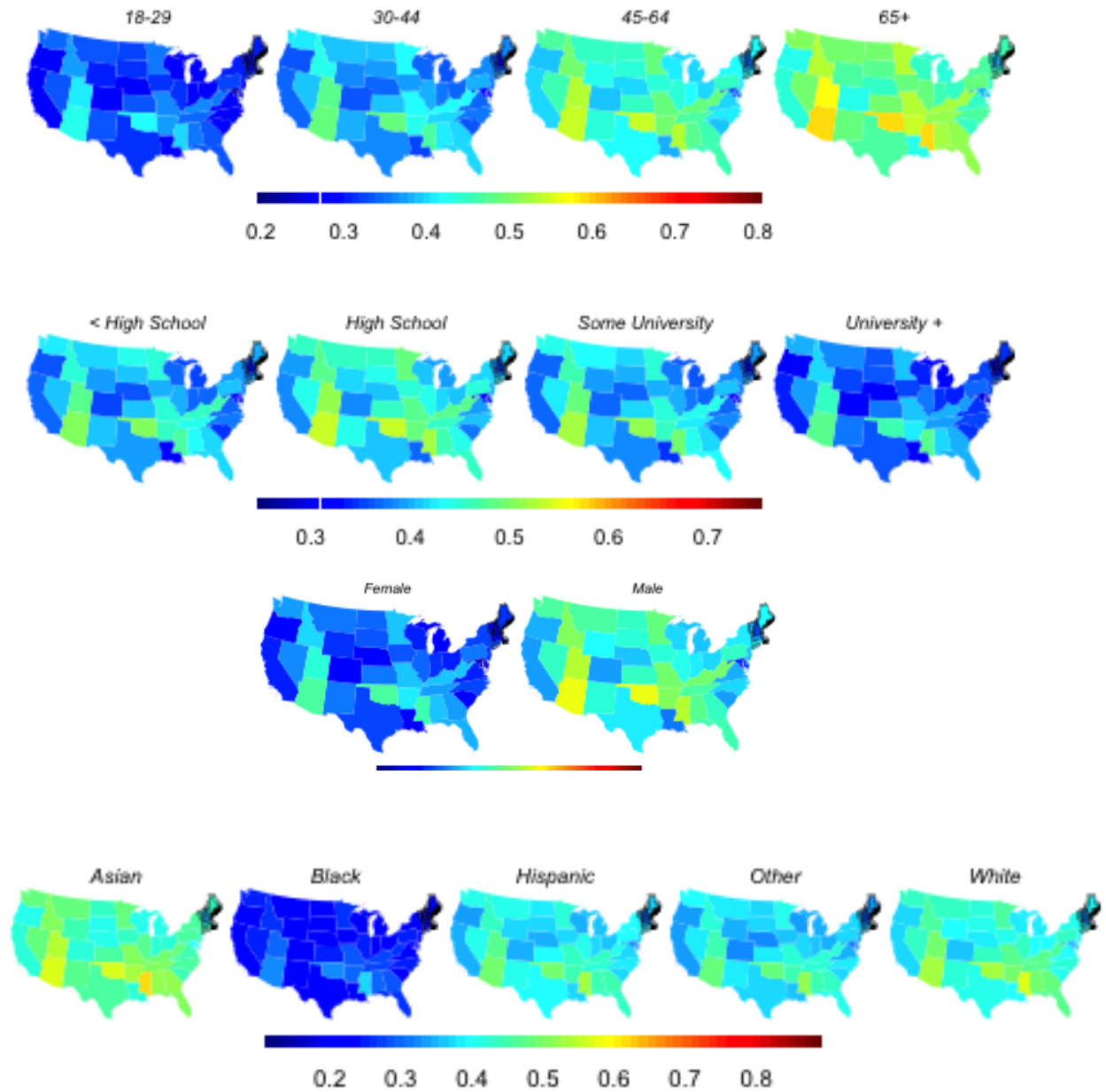
Figure B.19: Opposition to Limited Government and Individualism, by State and Demographics



Note: These graphs show the percentage of people who oppose the values of limited government and individualism in each state by age group, education level, gender, and race.

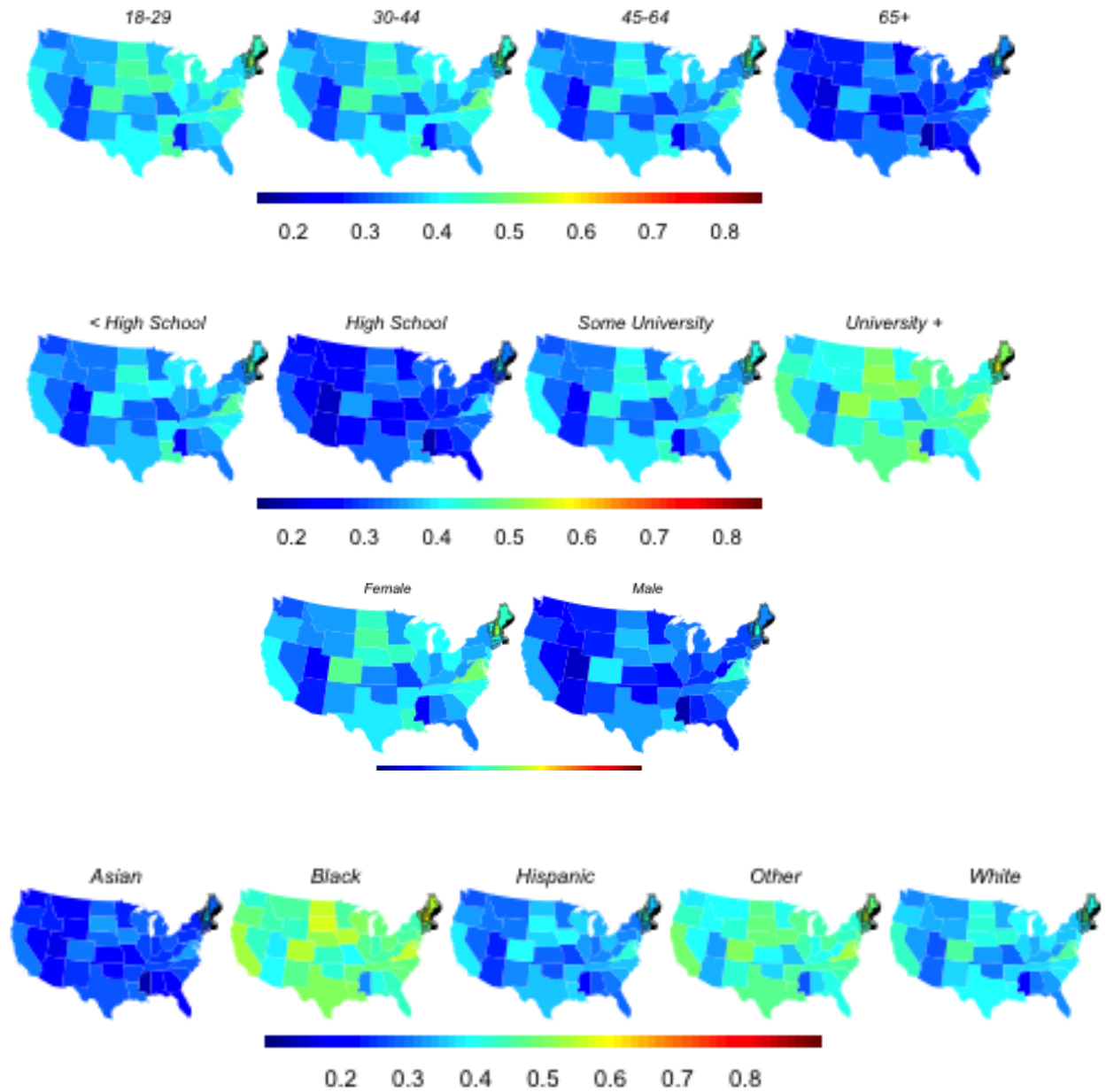


Figure B.20: Opposition to Egalitarianism, by State and Demographics



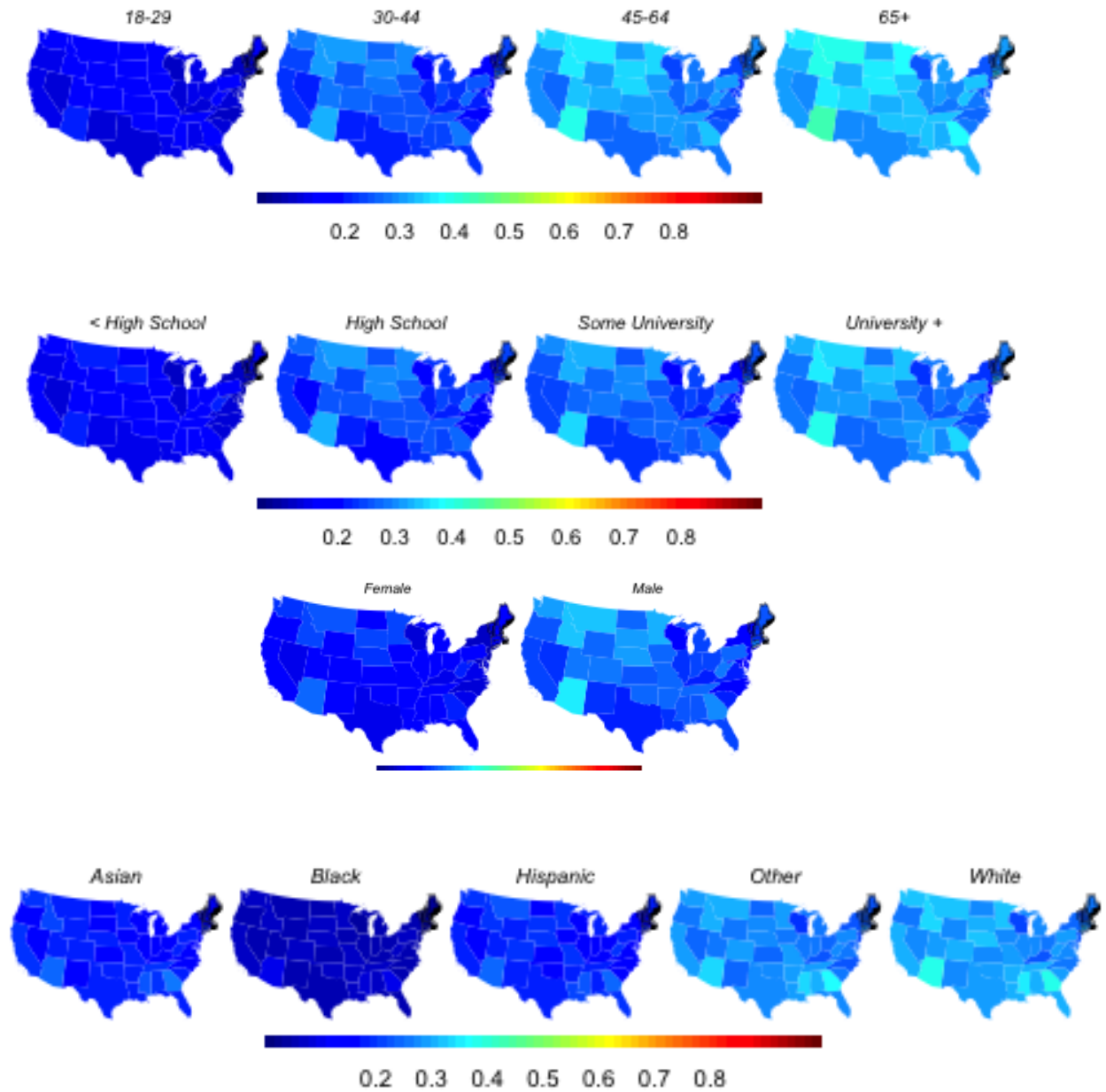
Note: These graphs show the percentage of people who oppose the value of egalitarianism in each state by age group, education level, gender, and race.

Figure B.21: Support for Egalitarianism, by State and Demographics



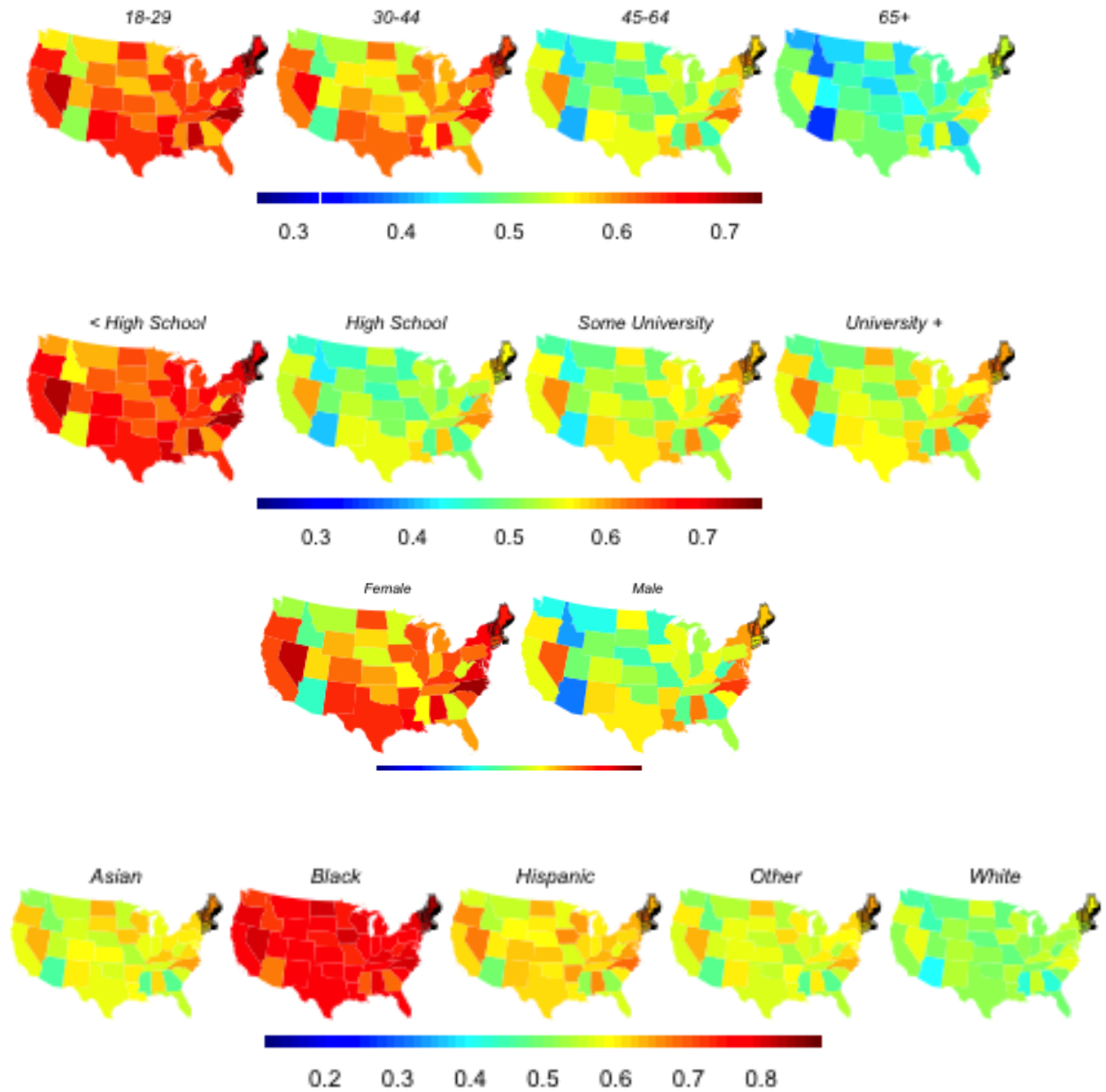
Note: These graphs show the percentage of people who support the value of egalitarianism in each state by age group, education level, gender, and race.

Figure B.22: Opposition to Equality of Opportunity, by State and Demographics



Note: These graphs show the percentage of people who oppose the value of equality of opportunity in each state by age group, education level, gender, and race.

Figure B.23: Support for Equality of Opportunity, by State and Demographics



Note: These graphs show the percentage of people who support the value of equality of opportunity in each state by age group, education level, gender, and race.