Voter Ideology: Regression Measurement of Position on the Left-Right Spectrum

June 28, 2016

J. Mark Ramseyer and Eric B. Rasmusen

Abstract

For scholars who need a measure of political preferences, a person's position on the ideological spectrum provides a good start. Typically, scholars identify that position through factor analysis on survey questions. In effect, they assume that the calculated synthetic variable marks the person's location on the liberal-conservative spectrum. They then use that ideology variable either as the focus of a study on ideology, or as a control variable in other regressions. The leading attitudinal surveys— the GSS, the CCES, and the ANES— include a variable giving a respondent's self-identified ideology. Factor analysis assigns this variable no special prominence. To treat this self-identification appropriately, we urge scholars to instead measure ideology using the fitted value from a regression of self-identified ideology on other survey responses. In contrast to factor analysis, the regression approach assigns proper priority to self-identification; it lets us test whether voters identify their own ideology through identity-group variables; it avoids the bias introduced in choosing the issue variables to include in the factor analysis; and it identifies the issues that the average voter thinks best define "liberal" and "conservative."

J. Mark Ramseyer: Mitsubishi Professor of Japanese Legal Studies, Harvard Law School; ramseyer@law.harvard.edu.
Eric B. Rasmusen: John M. Olin Faculty Fellow, Olin Center, Harvard Law School; Visiting Professor, Economics Dept., Harvard University; Dan R. & Catherine M. Dalton Professor, Department of Business Economics and Public Policy, Kelley School of Business,Indiana University, Bloomington, Indiana 47405-1701. erasmuse@indiana.edu. http://rasmusen.org. 812-345-8573.

1 Introduction

What does it mean to be "liberal" or "conservative", "leftwing" or "rightwing"? Like members of the general public, scholars use these terms to refer to the ends of the standard uni-dimensional political spectrum. Given that the terms refer to two ends, the questions are obviously tied to each other. Consider three different answers, from a politician, a political theorist, and a journalist:

> "...I believe the very heart and soul of conservatism is libertarianism. I think conservatism is really a misnomer just as liberalism is a misnomer for the liberals— if we were back in the days of the Revolution, so-called conservatives today would be the Liberals and the liberals would be the Tories. The basis of conservatism is a desire for less government interference or less centralized authority or more individual freedom and this is a pretty general description also of what libertarianism is." Ronald Reagan, *Reason Magazine*, Jul. 1, 1975.

> "Conservatives are inclined to use the powers of government to prevent change or to limit its rate to whatever appeals to the more timid mind. In looking forward, they lack the faith in the spontaneous forces of adjustment which makes the liberal accept changes without apprehension, even though he does not know how the necessary adaptations will be brought about." Friedrich Hayek, *Why I Am Not a Conservative.*

> "Liberals and conservatives disagree over what are the most important sins. For conservatives, the sins that matter are personal irresponsibility, the flight from family life, sexual permissiveness, the failure of individuals to work hard. For liberals, the gravest sins are intolerance, a lack of generosity toward the needy, narrow-mindedness toward social and racial minorities." E. J. Dionne, "The War Against Public Life."

People approach the question in several ways. Some take a deductive approach. They start with a definition and explore its implications for the positions a conservative should take. This might seem a subject for political theory. This is the approach most likely to produce a coherent concept of conservatism, but it makes the concept the author's rather than what the world calls conservative. In the quotes above, Ronald Reagan and Friedrich Hayek take this approach.

Others take a synthetic approach. They start by specifying a set of issue positions as conservative and then try to determine what the positions have in common. This, too, is an approach a political theorist might use. It would no doubt appeal to E.J. Dionne. Alternatively, the analyst might specify conservative positions and then rank people by how often they take those positions. That is the method used by online quizzes and politician ratings.

Political science needs a way to operationalize political ideology, a way to rank views numerically from the most liberal to the most conservative. Scholars need a measure for two overlapping but still distinct purposes. Sometimes, they will want to explain the determinants and sources of change in ideological commitment. They will want, in other words, to treat ideology as

a dependent variable and explore its roots. At other times, they will want to use ideology as a control variable in a study of something else. They may try to explain voting patterns through ideological commitment, or to determine how much voting is affected by ideology as opposed to candidate personality. Alternatively, they may hope to explain aspects of popular support for regulation, attitudes toward race relations, or even consumer demand or investment strategies through this ideological commitment.[1]

Within political science there are two contrasting critiques of using the left-right spectrum as a measure. A longstanding view is that common people may say they take a position on that spectrum, but it does not actually affect their issue choices or their behavior. A different view is that people do have political principles, but using a single dimension is insufficient; two or more dimensions such as economic, social, and foreign policy should be used. Carmines & D'Amico (2015) provides a good overview of these debates. Regardless of whether these two views are correct, in ordinary language people do use the terms "liberal" and "conservative" as if they mean something, so we will set about trying to determine that meaning and see if it does correlate with voting for a presidential candidate.

Political scientists commonly take a synthetic approach different from political philosophers or commentators. They avoid defining any set of positions as conservative ex ante, but instead assume that survey respondents take the positions they do because of the degree of their inherent conservatism. Accordingly, they estimate a person's conservatism by calculating the underlying latent variable that best explains his observed positions on political issues. Typically, they do this through factor analysis. The information comes entirely from the data at hand, the positions people take on various issues. It may include the respondents' self-identified degree of conservatism, but that variable receives no special weight. The well-known Aldrich & McKelvey (1977) scaling takes this approach.

We propose a different synthetic approach. As with factor analysis, we do not define what it means to be conservative ex ante. We assume that the respondents take the positions they do because of the their inherent ideology. But where factor analysis either (a) omits self-identified ideology and relies exclusively on issue variables, or (b) includes the self-identified ideology as one variable in a mix, we treat it as the dependent variable in regression analysis. We use the resulting fitted values to estimate each respondent's ideology.

We believe our approach gives self-identification its proper importance. We treat self-identification as a function of the respondent's personal view of the several issues. This mirrors the way the respondent himself treats his ideology if he identifies himself as conservative or liberal on the basis of the beliefs he holds about politically sensitive issues. Of the many survey

---

[1] Note that scholars most commonly study ideological commitment either of (i) politicians and other government officials (e.g. judges), or of (ii) voters. We focus on voters.

questions, self-identified ideology most clearly reflects the respondent's own sense of what it means to be conservative or liberal. Treating it as just another issue variable miss the way even the respondent himself treats it as a function of those other issue variables.

We are not, however, content with using self-identified conservatism itself as a measure of conservatism. The problem is that it relies on the individual's personal definition. It would be better if we could measure the person's opinions and then calculate how conservative he is based on the way most people think of conservatism, not how he does. We do that with regression analysis. Regression analysis generates something like an average across individuals of the relation between issue positions and self-identified conservatism. The process effectively estimates what the average person thinks is the conservative position on each issue by aggregating all their levels of self-identified conservatism with their issue positions. One can then look at an individual's positions and see how well he corresponds to what other people think "conservative" means.

To demonstrate the procedure, we use a survey (the Cooperative Congressional Election Study, CCES) in which each respondent describes his own degree of conservatism. We use linear regression to determine which positions correlate most closely with that self-identification. The process produces the set of issues, positions on issues, and weights on issues that best match self-identified conservatism. We then take the resulting coefficients, match them to an individual's issue positions, and estimate his position on the ideological scale.

2. SURVEY RESPONSES AND SELF-IDENTIFICATION.

Ideological commitment matters to social scientists for several reasons. Some scholars hope to identify the sources of that commitment itself. They hope to trace the impact on that commitment of upbringing, ethnic status, economic success, education, friendships, geographical location. They hope to identify not just why people hold the commitments that they do, but what events might cause them to change those commitments. In other words, they wish to place ideology on the left-hand side of the regression equation.

Other scholars want to explain something else – attitudes toward government intervention, or perhaps marriage patterns, voting behavior, economic success, geographical mobility, fertility, litigation. They may want to explore the impact of ideological commitment on these phenomena, or they may simply want to hold ideology constant while they explore the impact of yet another variable. In either case, they plan to place ideology on the right-hand side of the regression.

Although surveys routinely ask respondents how they see their own political identity, most scholars try to move beyond this self-identification. After all, respondents do not always answer these questions honestly. They do not always share a common sense of what the various labels mean. They may simply infer their own political status from other attributes (e.g., as a

fifty-year-old white Baptist in a Houston suburb, I must be conservative). They may not call themselves either liberal or conservative. And any time a scholar relies on only one measure, he runs a substantial risk of measurement error. Ansolabehere, Rodden & Snyder (2008) point this out, showing that the average of a person's answers to various survey questions is much more stable over time than his answers to any individual question.

2.1. FACTOR ANALYSIS.

Given these problems with political self-identification, scholars often classify respondents by the positions they take on multiple issues. Rather than treat them as "conservative" if they call themselves conservative, they ask what the respondent actually believes. They then infer a respondent's political status from those survey responses.

For inferring beliefs from surveys, factor analysis has become the tool of choice. Scholars treat a respondent's basic policy position as composed of one or more unobserved variables. They then use the technique to estimate those latent variables from the observed survey responses. Carmines, Ensley & Wagner (2012a: 0), for example, apply factor analysis to American National Election Studies (ANES) data in order to explore the dimensions around which Americans "organize their policy attitudes." In Carmines, Ensley & Wagner (2012b), the same authors use factor analysis on ANES data to study the way voters respond to polarized party leaders.

A wide range of other scholars apply factor analysis to ANES data to estimate belief structures. This includes Conover & Feldman (1981: 617), for instance, who study the "symbolic and nondimensional origins and nature" of idiological self-identification. Feldman (1988) examines the "core beliefs and values" by which people structure their attitudes and beliefs. Feldman & Zaller (1992) ask why people seem to hold contradictory political positions. Feldman & Johnston (2014) explore the dimensional character of ideology. McCann (1997) studies the effect of the choices people make in elections on the values they hold after it. And Layman & Carsey (2002: 791) ask whether attitudes toward "racial, cultural and social welfare issues" constitute three separate attitudes or component parts of a single attitude.

Other scholars apply factor analysis to different survey data – again to estimate a person's underlying (and unobserved) core values. Swedlow & Wyckoff (2009) use a telephone survey to explore the two-dimensional structure of voter ideology. Jacoby (2006) similarly uses a telephone survey, but to test the extent to which "political sophistication" influences the "translation process" from value preferences to issue positions. Conover & Feldman (1984) use student responses to study the "schemas" that people use to understand their political world. Heath, Evans & Martin (1994) use survey data to explore "core beliefs", and Miller (1992) asks whether young people have become more conservative or merely more willing to call themselves conservative. Verhulst, Eaves & Hatemi (2012) study twins to determine whether genetic

endowment might explain political traits. And in more explicitly methodological articles, Alwin & Krosnick (1985) and McCarty & Shrum (2000) use factor analysis to compare the relative usefulness of ranking and rating measures in attitude surveys.

In applying factor analysis, these scholars assume that a respondent's position on an issue is determined by one or more factors, underlying ideological variables that are uncorrelated with each other but that we can interpret as corresponding to such ideas such as "conservatism", "economic conservatism", "populism", and so forth. In effect, the analyst must match the artificially constructed factor to the political idea, a process that requires both interpretation and the assumption that the factors correspond to some ideas we can understand. Without that interpretive process, factor analysis simply leaves us with a linear combination of the survey variables. As Heckman & Snyder (1997) point out, however, because the usual calculations require one to assume that the factors are uncorrelated with each other, they cannot correspond to ideas like "economic conservatism" and "social conservatism" that are usually held by the same group of people.

To facilitate interpretation, scholars usually "rotate the factors." When using two or more factors, one can often construct several sets of factors that yield the same fit to the data. Generally, scholars first assign the most possible fit to factor 1, and then calculate factor 2 as the linear combination (uncorrelated with factor 1) of all variables that explains the greatest variance. "Rotation" is a way in which the scholar can choose an alternative set of factors. There are multiple ways to construct two artificial variables that explain the same total variance, in other words, and rotation can give the analyst a set in which the factors more intuitively correlate with the issue variables.

Our project depends less on rotation than it does on the implicit assumption that the liberal-conservative dichotomy best explains the variance among respondents. We focus only on one factor – the liberal-conservative spectrum. Because rotation concerns the allocation of variance among multiple factors, it applies only tangentially to this project. The project does depend, however, on the implicit assumption that the liberal-conservative distinction explains the data we have. It might indeed explain it, but a different latent variable might plausibly fit the data better: distrust of elites, for example, or dissatisfaction with the status quo, or some other general attitude. As many political scientsts have observed, Americans may not position themselves along one ideological dimension at all. They might instead take positions determined by, say, five different latent ideological variables (just as scholars in personality psychology generally posit that survey answers are best explained by "the Big Five" variables of approximately equal importance).[2]

---

[2] The Big Five factors are extraversion, agreeableness, conscientiousness, neuroticism, and openness. See e.g. Sanjay Srivastava "Measuring the Big Five Personality Factors," http://psdlab.uoregon.edu/bigfive.html.

2.2. STRUCTURAL STUDIES.

An intriguing alternative to the standard factor analysis is "Bayesian item response theory" (BIRT). To explain the approach, Treir & Hillygus (2009) note that voters tend to hold multidimensional beliefs. As a result, when asked they do not readily catalogue themselves as liberal or conservative. Scholars use factor analysis to tease out these undisclosed basic beliefs from survey questions on specific policy questions.

Treier & Hillygus (2009: 683) urge a Bayesian approach instead. An "additive scale of issues," they observe, would assume "that every issue contributes equally to the underlying preference dimension." Although factor analysis does not make that assumption, it does (id., 684) "assume a multivariate normal distribution for all observed variables." In fact, however, survey responses can be (id., 684) "nominal, binary, ordinal, or continuous." According to Treier & Hillygus (2009: 684), BIRT deals with such variables appropriately:

> "[W]ith the Bayesian IRT model, the latent measures (or factor scores) are estimated directly and simultaneously with the discrimination parameters – rather than as postestimation by-products of the covariance structure, as is the case with conventional factor analysis. Consequently, these traits are subject to inference just like any other model parameter, so we can calculate the uncertainty estimates for the latent measures."

More specifically, Treir & Hillygus (2009) take 23 questions from the ANES, and model issue responses as a function of an unobserved preference dimension. Treier & Jackman (2008: 205) explain the mechanics thus:

> "In a Bayesian analysis, the goal is to characterize the joint posterior density of all parameters in the analysis. This means that the latent variables x are estimable and subject to inference just like any other parameters in the model."

In factor analysis, by contrast (id., 205),

> "The typical implementation of factor analysis is as a model for the covariance matrix of the indicators (and not for the indicators per se), without the identifying restrictions necessary to uniquely recover factor scores, and hence the multiple proposals for obtaining factor scores conditional on estimates of a factor structure ...."

We sympathize. We are no fonder of factor analysis. The Treier, Hillygus and Jackman approach, though, threatens to overwhelm the reader. As Ansolabehere, Rodden & Snyder (2008: 216) put it in their plea for simplicity: "Confronted with complex structural models with many layers and parameters, skeptical readers see an unintelligible black box and are left with the impression that the findings have been manufactured by technique." Simple tools often yield

results close to those from theoretically more rigorous techniques anyway. In the context of legislative voting studies, Heckman & Snyder (1997: S145) note that factor analysis and least squares estimates yield similar results. Ansolabehere, Rodden & Snyder (2008) observe that factor analysis even comes close to the crude index composed of the arithmetic mean of responses on a set of issues.

Although factor analysis does not predict a respondent's self-identified ideology, it does let a scholar estimate a respondent's ideology as an underlying latent variable. The factor loadings, in turn, then help the scholar understand what the estimated factor might mean. If issue positions that we consider conservative are highly correlated with factor 1, we deduce that factor 1 captures the liberal-conservative spectrum.

Heckman and Snyder take this a step further. They show that the factors can be seen as unobservable characteristics of an issue position with coefficients that represent the marginal value of that characteristic to the individual, much like prices of product characteristics in a hedonic pricing model. The factors can be constructed to be uncorrelated with each other, as is standard, but they note that this lack of correlation is purely a convention and there is no real-world reason why characteristics of an issue should be uncorrelated. If liberal-conservative spectrum is one characteristic of an issue position, and benefit-to-richer-citizens is another, there is no reason to expect them to be uncorrelated.

We do not have a structural model, or a model which can be used for inference. Our goal is straightforward: to describe the data in a way that will predict well outside of the sample and whose workings are simple. Like Heckman and Snyder, we wish to avoid the assumption that the most important factor is the liberal-conservative ideology, and we do not want to create a measure of conservatism that by construction is uncorrelated with other characteristics of an issue position.

What we strive for is a measure that bears a meaningful connection to the everyday notion of liberal vs. conservative, but which is simple and is less idiosyncratic than a respondent's answer to the self-identified conservatism question. As mentioned above, the answer to any one question is subject to measurement error, meaning in this context anything from an absent-minded unintended answer to confusion over what the questioner is asking. Self-identified conservatism is also reliant on the respondent's own notion of what it means to be conservative. Our regression approach will avoid both problems by relying on several questions, not just one, and by aggregating the opinions of all the respondents in the sample about what it means to be liberal vs. conservative.

2.3. The Goal of Parsimony

In constructing a summary measure of "conservatism," we prefer simple techniques to

complex. Indeed, simplicity is inherent in trying to measure ideology at all. Were accuracy the only goal, we would retain 100% of the data — the individual's answers to every survey question. We opt instead for a simple technique that sacrifices as little accuracy as possible. Each of us has limited cognitive ability, and lives within time constraints. Between the alternatives of "The height of every American", "The number of Americans in each inch-long interval of height", and "The average height of Americans", we find the average the most useful. We opt for it even though it is the least accurate and the least informative. Financial accounting is based on this principle. An investor who wants to know the financial health of 1,000 firms typically does not want 1,000 annual reports. Usually, he will want only 1,000 numbers — perhaps the return on assets for each firm or the return on equity. He may then delve into how to correct for the error introduced by rigid one-size-fits-all accounting rules, but he starts with simplicity.

In designing tests for actual use in decisionmaking, psychologists think hard about the tradeoff between length and informativeness. One paper, for example, is entitled "Measuring Personality in One Minute or Less: A 10-Item Short Version of the Big Five Inventory in English and German" (Rammstedt & John (2007)). When they distribute their preeminent survey, the General Social Survey, sociologists themselves include a 10-question IQ test. In fact, they simply take 10 questions (all verbal) from one of the well-known IQ tests. Nonetheless, the 10-question quiz has a correlation of .71 with more finely measured IQ, compared with .51 for the respondent's educational level, .30 for his father's educational level, or .29 for his father's occupational prestige.[3]

Simplicity lies at the center of our own project. Scholars routinely want a single measure of a respondent's ideological commitments. Some will need it for a dependent variable. Others will want it as a control variable. In either case, they need a single measure that correlates as closely as feasible with a variety of measures relating to the respondent's political ideology. They need a measure that correlates with what the average American thinks is conservatism, that is transparent, and that is easy to measure.

3. THE DATA AND METHOD

We take our data from the 2012 Cooperative Congressional Election Study (CCES). The data in many ways resemble data available from the General Social Survey (GSS) and the American National Election Study (ANES). We choose the CCES because of its large sample size, but we

---

[3] See Wolfle (1980). This is even more remarkable because the short test is so coarse. The 10 questions are graded as right/wrong, so only 10 IQ levels can be measured. See De La Jara, Rodrigo "IQ Percentile and Rarity Chart," *IQ Comparison Site*, http://www.iqcomparisonsite.com/iqtable.aspx (2006) and "A Word about Wordsum," *Half Sigma* blog, http://halfsigma.typepad.com/half_sigma/2011/07/a-word-about-wordsum.html (July 21, 2011). People could game a simple test like this, of course, but since people do not try to game GSS surveys, it serves the purpose well. For our purposes – a person's position on the political spectrum – one similarly need not even worry about strategic behavior by the subjects.

could make the same points with the GSS or ANES.[4] A large sample size is useful in part because it allows us to split the sample into regional or racial subsamples. It is also useful because of the way it lets us use recently developed "machine learning" techniques that replace conventional confidence intervals with a division of the sample between "training" subsamples used for estimation and "testing" subsamples used for verification.
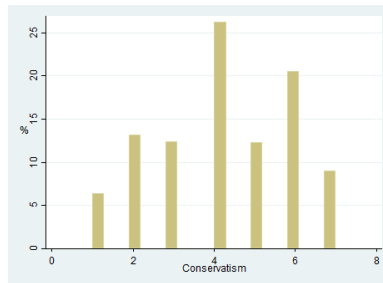
1. The self-identified ideology variable. The CCES asks respondents to locate themselves along an ideological spectrum from 1 (very liberal) to 7 (very conservative). We call this *Conservative-Self*.

Table 1 and Figure 1
Answers to the Self-Identified Conservatism Question, *Conservative-Self*, and the Response Percentages (n = 51,598)

"Thinking about politics these days, how would you describe your own political viewpoint?"

1 Very liberal	6.39
2 Liberal	13.14
3 Somewhat liberal	12.40
4 Moderate	26.24
5 Somewhat conservative	12.25
6 Conservative	20.54
7 Very conservative	9.03

These percentages are adjusted for survey sampling weights.

2. Issue variables. We use the 36 issue variables in Table A1 of the Appendix. These are CCES questions that were more ideological (Was the Iraq invasion a mistake?) than specifically partisan (Is President Obama to blame for the economy?). The questions cover such issues as the Iraq war,

---

[4]The Cooperative Congressional Election Study (CCES) is available at http://projects.iq.harvard.edu/cces/home. The General Social Survey (the GSS) is available at http://www.icpsr.umich.edu/cgi-bin/SDA-ID/ICPSR/hsda?icpsr+31521-0001, which allows downloading as a STATA data set. The GSS codebook is at http://www.icpsr.umich.edu/SDA-ID/ICPSR/31521-0001/CODEBOOK/GSS.htm. The ANES is available at http://www.electionstudies.org/studypages/download/datacenter_all_NoData.php.

gun control, immigration, abortion, environmentalism, gay marriage, affirmative action, tax policy, free trade, the Affordable Care Act ("Obamacare"), and the Keystone pipeline.[5]

3. Identity variables. We use the 17 identity variables in Table A2 of the Data Appendix. They cover such matters as sex, birth year, race, education, marital status, employment, religious affiliation, and income. We include these identity variables for two reasons. First, they might pick up the effect of some omitted political issues. Second, the identity variables might truly be why some people call themselves conservative. As noted earlier, for example, someone might think that he should call himself conservative because he is a male white Southerner, despite his stands on the issues. We want both to untangle that effect from the effect of those issues he lists as important, and to explore whether people call themselves conservative mainly because of issues or mainly because of image. Of course, if an identity variable predicts *Conservative-Self*, we cannot say whether it does so because it is correlated with omitted issue variables or because identity politics gives it a directly causal role. If an identity variable does not predict *Conservative-Self*, however, we can rule out its being important for identity politics

Note that the inclusion of the identity variables distinguishes regression from much of factor analysis. In factor analytic studies, the scholar tries to create a latent variable that approximates the answers respondents give to the issue questions. Thus, he begins the factor analysis by identifying issue questions. Marital status obviously is not itself an issue variable. Potentially, however, it may be more highly correlated with the underlying latent variable than any issue question – either because people take their ideological position from their marital status, or because marital status proxies for important but omitted issue variables.

4. Constructing the ideology measure. To construct our measures of conservativism, we need first to know which issue variables best predict political ideology. Note that we seek to explain the data parsimoniously, not to find the correct structural model. We want to discover which variable best predicts *Cons-Self*, which two variables best predict it, which three variables, and so forth. In this exercise, we have no need for measures of statistical significance. Instead, we can be boldly ad hoc – even opportunist – and consider such observations as "R^2 hardly goes up at all once we have included 3 variables instead of 2."

A scholar could envision the "best predictors of *Conservative-Self*" in several different ways.

---

[5]Because we compare regressions using different explanatory variables, missing values present a special problem. Starting with a given regression with a particular R^2, if we add an explanatory variable the R^2 may fall. This is arithmetically impossible when the dataset stays unchanged, but can occur if the new explanatory variable has many missing values. The sample size will then fall and the remaining observations may be the hardest to explain. To address this problem, we impute values to the missing observations through "mean imputation" — that is, we insert the mean value of the non-missing observations. This technique leaves the point estimates unchanged, although it biases the standard errors (see Little [1992]). Crucially, the mean value that we impute will not help explain the variation. Hence, any increase in the R^2 results from the actual values for the variables.

He could, for example, simply look at the unconditional correlation between *Conservative-Self* and the issue variables. He could then identify the five variables with the highest correlations. In doing so, he would answer the question: "If you could use one variable to predict *Conservative-Self*, which would be your top five choices?" Alternatively, the scholar could find the five variables that best predict *Conservative-Self* through linear regression. Here, he would be looking to conditional correlations, and answering the question: "If you could choose a set of five variables to predict *Conservative-Self*, which set would be your top choice?"

4.1 Factor Analysis

To explore different ways in which scholars could estimate ideological commitment, we start with the most commonly used technique, factor analysis. Because we have 36 issue variables plus *Conservative-Self*, we could – hypothetically – generate as many as 37 factors, each of which is by definition uncorrelated with the others. Scholars always stop with fewer, however, since most use the technique primarily to reduce the total number of variables.

The idea behind factor analysis is that there is some latent variable explaining someone's position on issues. Thus, we will exclude the identity variables, as is conventional in the literature. They are not variables we think are caused by conservatism. As explained earlier, we might think that the causality goes the other way, and identity causes a person to be conservative, but that idea does not fit into the framework of factor analysis. One might try calculating a conservatism variable for each person and then use regression to see if identity explains that variable, but that would be mixing techniques in a way that would have dubious statistical underpinnings.

The creation of the factors results in an eigenvalue for each factor, and it is conventional to discard any factor with an eigenvalue of less than one. Here, factor analysis of the 51,598 observations yields 3 factors with eigenvalues over one. Factor 1 explains 71% of the variance, factor 2 explains 15%, and factor 3 explains 9%, a total of 95%.

The factors in this first step of factor analysis are created so that the first factor explains as much of the variance of the 37 variables as possible (roughly speaking, it is the single artificial variable most correlated with those 37 variables). The second factor is constructed to explain as much of the remaining variance as possible (it is the single artificial variable most correlated with what's left over of the 37 variables after we remove the values of them as predicted by the first factor). The third factor explains what's left over after the first and second factors are used, and so forth.

We could stop here and take Factor 1, known as the "first principal component", to be "conservatism". It is conventional, however, in factor analysis to "rotate" the factors. This is because the 95% of the variance explained by Factors 1, 2, and 3 could be explained by many

other combinations of three artificial variables. The first step uses a combination in which Factor 1 is constructed to explain as much as possible, 71%. An alternative would be to construct three factors each of which explain 32%, so no one factor gets primacy. There are actually an infinite number of ways to construct three factors.

The most common rotation method is the "orthogonal" rotation known as "varimax". An orthogonal rotation is one that keeps the factors constructed so they remain uncorrelated with each other. A varimax rotation is one that is orthogonal and, roughly speaking, drives the values of the factor loadings as far away from .5 and -.5 as possible. The motivation is to construct three factors that each are either strongly correlated or strongly uncorrelated with the underlying issue variables, rather than having a mediocre correlation with all of them. This has the effect of pushing some issues out of Factor 1 and into Factors 2 and 3, so each factor specializes in a particular set of issues. Varimax rotation yields three factors explaining 62%, 18%, and 15% of the variance. We might want to say that varimax factor 1 is conservatism. Or, we could "normalize". That yields three factors explaining 58%, 21%, and 16% of the variance. We might want to say that normalized varimax factor 1 is conservatism. The most important other kind of orthogonal rotation is "quartimax". This is opposite of varimax. Instead of making each factor specialize in issues, it finds the factors such that each issue is explained by as few factors as possible, which generally results in one big factor, just as we have without rotation at all. That yields three factors explaining 70%, 15%, and 10% of the variance. We might want to say that quartimax factor 1 is conservatism. Again, we could "normalize" the factors.

So far we have discussed orthogonal rotation. The other class of rotations is "oblique" rotations. These result in 3 factors that may be correlated with each other, but that explain the same amount of variance in total. As with orthogonal rotation, there are many ways to do oblique rotations. The most common kind is "oblimin", which minimizes the squared loading covariances between factors under the same kind of motivation as varimax: to generate specialized factors. That yields three factors explains 66%, 50%, and 16% of the variance, which adds up to more than 95% because now the three factors are correlated, with overlapping explanatory power. Together, they explains 95% of the variance, but, for example, Factor 2 would explain 50% of the variance if you used it by itself. We might want to say that quartimax factor 1 is conservatism. Again, we could "normalize" the factors.

Factors are interpreted using their "factor loadings," which are equivalent to the Pearson correlation coefficient between the estimated factor and each variable. These will be affected by the rotation method used. The top five factor loadings here for the unrotated first factor are for *Conservative-Self, Global Warming, ACA Health Plan, Repeal ACA, Affirmative Action,* and *Black Favors* (blacks should not get special favors), ranging in magnitude from .67 to .78. (Recall that definitions of the questions are in Appendices I and III.) As this shows, similar issues have

similar factor loadings— *ACA Health Plan* and *Repeal ACA* are similar, and so are *Affirmative Action* and *Black Favors.* If rotation is used, different sets of variables have the highest loadings. Using normalized varimax, for example, the top five factor loadings are for *Conservative-Self, Global Warming, ACA Health Plan, Repeal ACA, Mand. Birth Cntrl. Ins.*, and *Abortion*.

Factor analysis also yields a predicted value of the factor for each respondent in the sample. This, using the unrotated first factor, is our measure of conservatism, which we will call *Conservative-Factor*. When *Conservative-Self* is regressed on *Conservative-Factor*, it yields an R^2 of .4841. This will be useful for comparisons later.

We could also include the identity variables in the factor analysis, which reduces the proportion of variance explained by the first factor. This was because, roughly speaking, the average identity variable was less correlated with the latent variable than the average issue variable. Ideally, a newly added variable would be exactly correlated with the latent variable. This would give it a "factor loading" of 1, and (obviously) increase the proportion of variance explained.

4.2 Regression Methods

Turn now to our alternative to factor analysis: a regression of *Conservative-Self* on a set of issue variables, and the use of the fitted values to estimate a conservatism score for each survey respondent. We shall explore several ways to select the appropriate issue variables.

Although we use ordinary least squares, ordered probit is what would ordinarily be appropriate, since conservatism is a categorical variable with only seven possible values. Ordered probit would measure how an underlying conservatism variable plus random error would show up as those seven values when observed. It would take into account the fact that the value could not be less than 1 or greater than 7, no matter what the value of the error. It also would account for the fact that intermediate values such as 4.5 cannot be observed, and that the true difference between the values 2 and 3 is not necessarily the same as the difference between 4 and 5 (that is, that the choice of linear scaling may not be correct). OLS is inconsistent, and its standard errors cannot be trusted. On the other hand, ordered probit requires that we assume normality for the error distribution, would be computationally intensive, and less transparent than least squares. Ordered probit would generate better estimates of the standard errors, but we are not using those. We aim not to test hypotheses but to describe the data, to predict, and to create an index variable. We aim to replace self-identified conservatism and factor analysis, and toward that end to identify useful variables. OLS works well as a way to find conditional correlations. In the interests of retaining a computationally tractable and analytically transparent way of measuring conservatism, we thus use least squares. It is best to think of what we are doing as finding a best linear projection of *Conservative-Self* on different sets of variables.

One kind of predictive equation is to include every variable, in a "universal regression". We have 36 issue variables. Regressing *Conservative-Self* on all of them for the 51,598 observations generates an R^2 of .52. The variables with t-statistics over 2 are:

Issue variables: Abortion, Gay Marriage, ACA Health Plan, Global Warming, Taxes v. Spending, Iraq Mistake, Gun Control, Immigpatrol, Immigpolice, Immigservices, Jobsenvironment, Affirmative Action, Balanced Budget, Ryan Budget, Tax Cut, Tax Hike Act, Birth Control, Repeal ACA, Gay Military, Keystone Pipeline, Troops–Allies, Troops-UN, Income v Sales Tax, Black Favors, Black Class

Although this method is more transparent than factor analysis, it is cumbersome. Moreover, in a regression with this many variables, interpretation of t-statistics is problematic. The t-test asks whether a variable's conditional correlation significantly differs from zero. If we examine the t-statistics of all coefficients at once with 36 variables some variables will likely appear significant by chance. What is more, we risk overfitting the data. To maximize R^2, we should not omit any variable, no matter how low its t-statistic. Even with a sample of more than 50,000, however, doing that will result in overfitting. Some variables will help explain the data in our particular sample even though they are unimportant in the true population. Thus, if we try to use the regression result on a different sample, the coefficients of those variables will just be adding random noise.

Reducing the Number of Issues

For parsimony, we should use fewer variables than in the universal regression. Recall that our goal is not statistical inference, but prediction. One possibility is to see how each variable performs individually in predicting *Conservative-Self*. That is measured by regressing *Conservative-Self* on each variable in a simple regression (which is also equivalent to find the five variables with the top pairwise correlations with *Conservative-Self*). Those five variables (with the R^2 of the simple regressions) are *ACA Health Care* (.27), *Gay Marriage* (.27), *Climate Change* (.27), *Repeal ACA* (.21), and *Mandatory Birth Control Insurance* (.21). A regression of *Conservative-Self* on these five variables yields an R^2 of .46, which is close to the .48 of *Conservative-Factor*, the latent variable from factor analysis. Another simple method would be to use the five variables that in the universal regression have the highest t-statistic: *ACA Health Care, Gay Marriage, Climate Change, Abortion,* and *Taxes v. Spending.* That yields an R^2 of .47.

A third method is "best subsets" regression, finding the five variables that generate the highest R^2 when *Conservative-Self* is regressed upon them. Maximizing the Akaike Information Criterion is, with minimal assumptions, asymptotically efficient as a way of finding the true set of explanatory variables (see, Cavanaugh & Neath [2011]). The Akaike is log(estimate of variance of the error term) + penalty-function-for-adding-RHS-variables. This is similar to maximizing

14

adjusted R^2, which is consistent but not efficient. See Castle, Qin & Reed (2013). In our case, the Akaike and adjusted R^2 criteria are optimized with 34 variables, which defeats the goal of parsimony. We will instead fix k (the number of explanatory variables) in the best-k regression, in which case maximizing the Akaike is equivalent to maximizing R^2. That is the "best subsets" approach.

With a small number of variables, best subsets regression can be done by exhaustive search. For 37 variables it can be done using a leaps-and-bounds algorithm. We used Stata's *vselect* command. Table 2 shows the resulting sets of size one to ten variables that were selected. The last column shows the R^2 for a simple regression of *Conservative-Self* on each variable individually (which is the squared correlation).

TABLE 2
R^2 FOR THE BEST-k REGRESSIONS FOR *Conservative-Self*

Best-k Predictors	b1	b2	b3	b4	b5	b6	b7	b8	b9	b10	Alone
1. Global warming is not a problem *(Global Warming)*	.2702		.4328	.4547	.4708	.4821	.4885	.4939	.4988	.5022	.2702
2. Against gay marriage *(Gay Marriage)*		.3809	.4328	.4547	.4708	.4821	.4885	.4939	.4988	.5022	.2650
Against ACA health plan *(ACA Health Plan)*		.3809	.4328	.4547	.4708	.4821	.4885	.4939	.4988	.5022	.5181
4. Blacks should not get special favors *(Black Favors)*				.4547	.4708	.4821		.4939	.4988	.5022	.2684
5. Abortion should be legal *(Abortion)*					.4708	.4821	.4885	.4939	.4988	.5022	.1868
6. Spending cuts better than tax increases *(Tax v. Spending)*						.4821	.4885	.4939	.4988	.5022	.2003
7. Invading Iraq was not a mistake *(Iraq Mistake)*							.4885	.4939	.4988	.5022	.0949
8. Oppose affirmative action *(Affirmative Action)*								.4885		.4988	.5022	.1714
9. Mandatory birth control insurance *(Birth Control)*									.4939	.4988	.5022	-.2136
10. Increase border patrol *Border Patrol)*										.5022	.1045

Notes. $n = 51,598$. For the exact wording of the questions, see Appendix 3. the last column shows how well the variable performs when it is the only regressor.

Table 2 gives the best-k predictors of conservatism: the k independent variables that when regressed on *Conservative-Self* yield the highest R^2. Note that the best-1 regression picks *Global Warming*, but the best-2 regression drops it and uses *ACA Health Plan* and *Gay Marriage.* The variable *Global Warming* correlates highly with both *ACA Health Plan* and *Gay Marriage*

16

(correlations of .50 and .41, as Table 6 shows below). It is the best single variable to use if only one explanatory variable is allowed. *ACA Health Plan* and *Gay Marriage*, however, each explain different aspects of *Conservative-Self*, and they therefore perform better in combination than either one does with *Global Warming*. Two other gaps in the table similarly indicate where a variable temporarily dropped out of the best-k as k increased. Observe also that R^2 generally increases at a decreasing rate.

None of the top 5 variables directly involves taxes, though of course, both *ACA Health Plan* and *Global Warming* do implicate taxation and government regulation. It may be of interest that in the best-10 regression, the t-statistics range from 10.4 to 22.7, though they do not have their usual meaning because we have selected for the variables with the largest coefficients and smallest standard errors.

We will take the best-5 regression as our benchmark. Table 3 shows the coefficients, and Table 4 shows how the variables correlate with each other.

TABLE 3
THE BEST-5 REGRESSION FOR *Conservative-Self*

Regressor	Coefficient	Possible values of the variable
Global warming is not a problem (*Global Warming*)	.31	1,2,3,4,5
Gay marriage should not be legal (*Gay Marriage*)	.74	1,2
Favor ACA health plan (*ACA Health Plan*)	.77	1,2
Blacks should not get special favors (*Black Favors*)	-.23	1,2,3,4,5
Abortion should be legal (*Abortion*)	-.24	1,2,3,4
Constant	2.53	1

Notes: $n = 51,598$. $R^2 = .51$. The descriptions in this table are summaries; for the precise questions see Appendix 3.

TABLE 4

THE BEST-5 CORRELATION MATRIX

	Conservative-Self	Warming	Gay	ACA	Blk Fav	Abortion
Global Warming	.52	1.00				
Gay Marriage	.51	.41	1.00			
ACA Health Plan	.52	.50	.39	1.00		
Black Favors	-.43	-.38	-.32	-.38	1.00	
Abortion	-.46	-.36	-.51	-.34	.28	1.00

4.3 LASSO

LASSO is a relatively recent technique for choosing among variables to get the best predictive set. It is well known in statistics but is just now entering the toolkit of researchers in economics and political science. LASSO finds the regression with the highest R^2 subject to the constraint that the sum of the absolute values of the coefficients not exceed a threshold size penalty. This drives down the coefficients of some variables to zero. It also reduces the coefficients of the variables that remain in the regression; LASSO is a "shrinkage estimator". Shrinkage estimators do not maximize R^2 and they are biased, but they may nonetheless be better predictors than the conventional multiple regression coefficients in terms of mean squared error.

Consider the issues involved. "Bias" is the expected value of the difference between the estimator's value and the true population value ($E\hat{\theta} - \theta$)). "Sample variance" is the expected value of the square of the difference between the estimator and the estimator's expected value ($E(\hat{\theta} - E\hat{\theta})^2$). Mean squared error is the expected value of the square of the difference between the estimator and the true population value ($E(\hat{\theta} - \theta)^2$), which happens to equal the sum of the square of the bias plus the variance (*bias² + sample variance*). If an estimator is unbiased, then with an infinite amount of data the mean squared error goes to zero, since the sample variance (the error arising from just having a sample instead of the entire population) goes to zero. With a small amount of data, however, the sampling error will be so big that a biased estimate could well do better.

Shrinkage estimators represent a tradeoff. They accept some bias in return for reducing sampling error. The fact that they do not maximize R^2 is a feature, not a bug. Rather, it means they depend less heavily on the particular sample at hand. For the normal distribution, dividing by $n+1$ has lower mean squared error in finite samples even though with an infinite amount of data dividing by $n-1$ is better. The intuition is, we speculate, that if the average size of the sample estimate's error is zero, then since the underestimates are limited to the range $[0, \sigma^2)$ but the overestimates are in the much larger (σ^2, ∞), squaring an overestimate will on average give a

larger number than squaring an underestimate. For estimating means (or regression coefficients), the standard example is the James-Stein estimator, which for three or more variables with normal distributions and identical variances has lower mean squared error than the sample mean. See the original James & Stein (1961) and the non-technical *Scientific American* article by Efron & Morris (1977).

The idea of shrinkage estimators is similar to the idea of variable selection itself. Recall that if we want to maximize R^2 in our prediction equation for $Conservative - Self$ we should use the universal regression with all 54 variables. Such an approach is unbiased, because if our sample were the entire population the estimated coefficients for irrelevant variables would equal zero. With our limited sample, however, some irrelevant variables will accidentally look important. If we tried using our estimated regression equation on a different sample, it would no longer give the highest R^2. By assigning importance to irrelevant variables the universal regression adds random noise to the prediction— random, because the irrelevant variable's mistaken effect might be either negative or positive.

LASSO combines variable selection with shrinkage. One thus could use LASSO for the variable selection (selecting the best-k variables), and then run a final regression with OLS on the selected variables to get the coefficient estimates and a higher R^2. This technique is given formal theoretical support in Belloni & Chernozhukov (2013).[6]

---

[6]See chapter 3 of Hastie, Tibshirani & Friedman (2003) for an explanation and comparison with stepwise and best subsets regression. We used STATA's $lars, a(lasso)$ command. Note that this command does not allow sample weights, unlike the other methods.

TABLE 5

LASSO COEFFICIENTS AS THE SIZE PENALTY IS RELAXED

Variable	Number of Variables									
	1	2	3	4	5	6	7	8	9	10
Global Warming	9.43	15.62	44.03	46.83	49.81	50.44	51.28	52.73	58.48	58.53
ACA Health Plan		6.19	36.27	39.19	41.69	42.26	42.80	43.83	48.10	48.13
Gay Marriage			34.77	38.42	42.68	43.39	44.63	47.03	60.57	60.72
Mand. B. C. Ins.				-3.07	-6.20	-6.78	-7.73	-9.87	-21.12	-21.25
Repeal ACA					-3.13	-3.87	-5.07	-7.35	-19.09	-19.24
Abortion						-.95	-2.74	-6.54	-28.83	-29.14
Affirmative Action							1.94	4.99	22.26	22.44
Black Favors							-3.31	-18.94	-19.12	-21.44
Tax v. Spend									16.78	16.99
Immigration Patrol										-.27
R^2	.03	.06	.27	.29	.32	.33	.34	.37	.51	.52

Table 5 shows how LASSO adds variables and increases coefficient sizes as the size penalty is relaxed. If the size penalty is set high enough, *Global Warming* is the only variable with a positive coefficient. As the penalty is relaxed, the coefficient on *Global Warming* rises. When it reaches 9.43, LASSO starts increasing the coefficient on a second variable, *ACA Health Plan*, to above zero. Relaxing the penalty still further, increasing both variables' coefficients is the best way to increase R^2 until they reach 15.62 and 6.19, at which point raising *Gay Marriage*'s coefficient above zero becomes worthwhile. If the size penalty were completely relaxed, the coefficients would take the OLS values and all variables would be used. Note that the reductions in the size penalty are not the same between columns: this table shows the size of coefficients when a new variable is introduced, not the size as the size penalty is reduced by a given amount. That is why the R^2 does not show the typical diminishing returns as variables are added: since *ACA Health Plan* is almost as useful a variable as *Global Warming*, the coefficient on *Global Warming* is still small (9.43) when *ACA Health Plan* enters the regression and so the one-variable estimate has a poor goodness of fit. The shrinkage feature of LASSO is also why the one-variable R^2 is low compared to other methods. The advantage of shrinkage does not show up in regressions which use the same data for estimation and prediction. LASSO's advantage is that by reducing the importance of a variable by reducing its coefficient size, it avoids overemphasizing variables which by chance are better predictors than they would be in other samples. We will next separate the samples used for estimation and prediction to allow a fair comparison between estimators.

COMPARING THE METHODS: FIVEFOLD CROSS-VALIDATION

One way to compare the various measures of conservatism is to see how well they predict *Conservative-Self.* Table 6 summarizes the various measures we have used. It contains two measures of R^2. One is what we have been talking of till now: the R^2 from applying the method to the full dataset. As noted earlier, the universal regression must mathematically have the highest R^2. Factor analysis is also advantaged because a factor is a linear combination of all 37 variables. Also, a problem running throughout our analysis is that significance measures such as the F test or t tests do not have their usual statistical interpretations, so it is hard to know whether one method really is better than another.

To get at this, we will use fivefold cross-validation. This technique creates parameter estimates using one part of the data and tests them on the remaining data. We randomly divided the data into five groups, and performed the estimation for each method five times. The first estimation used groups 2,3,4, and 5 to form the conservatism measure, after which *Conservative-Self* was regressed on it using just group 1. This was repeated five times using the five distinct partitions, each with 4/5 of the data for estimation and 1/5 for the prediction.

TABLE 6
R^2 ACROSS METHODS OF CONSTRUCTING A CONSERVATISM MEASURE

	Factor Analysis	Universal Regression	Universal Top 5	Correlations Top 5	Best Subset Top 5	LASSO Top 5
(1) 5-Fold cross-validation regression on *Conservative-Self*	.475	.509	.463	.455	.470	.406
(2) Full sample regression on *Conservative-Self*	.484	.517	.470	.455	.471	.403

Notes. Row (1) shows the average R^2 from the five cross-validation prediction regressions. Row (2) is the R^2 when the full sample is used for both estimation and prediction.

In comparison with the R^2 using all the data, factor analysis and the universal regression have the biggest decline in fivefold cross-validation. These are the two methods that use all available variables and so are the most likely to fit the data accidentally. Picking the five top variables from the universal regression and using them also shows a decline. The regression with the top 5 simple correlations and using *Conservative-5* show very little loss of R^2, and LASSO actually performs better in fivefold cross-validation than when it uses all the data. This is as one would expect. Parsimony reduces accidental fit, and LASSO's shrinkage feature prevents large

coefficients that suit only the particular sample and not the population.

In terms of absolute performance in fivefold cross-validation, LASSO does worst; the loss from biasedness seems to outweigh the gain from shrinkage. The universal regression still achieves the highest average R^2. Factor analysis and *Conservative-5* perform similarly, with a small edge to factor analysis that we think is outweighed by *Conservative-5*'s simplicity and transparency.

ANOTHER PREDICTION: VOTING FOR PRESIDENT OBAMA

Another way to compare the measures of conservatism is to see how well they predict whether the respondent voted for President Obama in 2012. Since President Obama is on the left, presumably a conservatism measure should predict not voting for him. This, of course, introduces the sort of subjective definition of conservatism that we criticized at the start of the paper, but it gives us another test for our index.

Table 7 presents two kinds of prediction methods a given conservatism measure might use: least squares and logit. Least squares give the best linear predictor, but it would yield biased coefficients and standard errors, so ordinarily a nonlinear method like logit would be used. Here, our purpose is more restricted— s to see how our measures predict the presidential vote relative to each other— so least squares is acceptable, but we have included the McFadden pseudo-R^2's from logit as well.

The R^2's in Table 7 range from .46 for *Conservative-Self* to .77 using all the issue and party variables. Our favored *Conservative-5* has an R^2 of .60, compared with .62 for *Conservative-10* and .68 for the universal issues regression. *Conservative-Factor* has an R^2 of .63, comparable to *Conservative-10*'s .62

Party identification per se does relatively poorly, with an R^2 of .52. The variable *Republican-7*, self-identified position on the Democrat-to-Republican spectrum, does better, with an R^2 of .65 that approaches the .68 of the universal regression. Using party affiliation is perhaps unfair, however, for predicting vote for a presidential candidate, especially since the strength of one's affiliation with one's party will depend on one's enthusiasm for its nominee.

We tried two other variants besides the regressions in Table 7. The first variant uses the bottom 5 variables in the top 10 instead of the top 5. This generates an R^2 of .51 instead of .60, indicating that variable choice does matter even among top variables. The second variant uses the best-5 variables, but not the regression coefficients from their regression. Instead, the possible responses are ordered so bigger numbers indicate more conservative answers, and then are added together. Call this measure *Conservative-Average*. Despite arbitrary coefficients, it has an R^2 of .55, comparable to *Conservative-5*'s .60. The success of *Conservative-Average* shows that if an

ideology index uses a set of suitable variables, it does not matter much if they are weighted equally rather than with carefully estimated regression coefficients. It recalls the finding in Ansolabehere, Rodden & Snyder (2008) that an index composed of a respondent's answers to several variants of an issue question is much more stable across time than his answers to single questions. The result also mirrors a well-known result in the psychology of decisionmaking that quite good decisions can be made by giving numerical ratings to various factors and adding them up for each alternative even without optimal weights— better decisions than when the decisionmaker uses the factors to make a non-mechanical, subjective decision. See Robyn M. Dawes, *Rational Choice in an Uncertain World*, Harcourt Brace (1988). In our setting, the results of using the bottom-5 and of *Conservative-Average* suggest that picking the best variables for the ideology index is more important than weighting them optimally.

TABLE 7

PREDICTIONS OF VOTE FOR OBAMA USING VARIOUS MEASURES

Explanatory variables	R^2 (least squares)	Pseudo R^2 (logit)
Republican and Democrat	.52	.46
Republican-7	.65	.61
Issue variables (universal regression)	.68	.68
Issues, Republican-7, Rep., Dem.	.77	.79
Conservative-factor	.63	.64
Conservative-Self	.46	.42
Conservative-5	.60	.56
Conservative-10	.62	.61
Conservative-lasso	.52	.48
Conservative-5, Republic, Democrat	.69	.69
Conservative-5, Republican7	.73	.74
Conservative-10, Republic, Democrat	.71	.72
Conservative-10, Republican7	.74	.75

Our conclusion from the results of these various specification and measures is that 5 variables

23

are enough for a reasonably good prediction of one's vote for president in 2012. We prefer OLS to ordered logit because it is less parametric. It provides the best linear predictor, which does not depend on errors following the logistic distribution as logit does, and it is simpler. The reader can examine Table 7 for himself and decide what tradeoff between explanatory power, complexity, and parsimony suits his preferences.

IDENTITY VARIABLES

A question of interest is whether a person's self-identified conservatism is determined by his beliefs or his identity. It might be, for example, that a woman self-identifies as conservative because she is black, female, and a union member and she thinks that someone like her ought to be a liberal, even though her stands on issues are conservative.

We can test for that by adding identity variables to our analysis and seeing if they enter into the top ten. We have 36 issue variables and 17 identity variables. A regression of *Conservative-Self* on all of them yields an R^2 of .53. The variables with t-statistics over 2 are:

Issue variables: Abortion, Gay Marriage, ACA Health Plan, Global Warming, Taxes v. Spending, Iraq Mistake, Gun Control, Immigpatrol, Immigpolice, Immigservices, Jobsenvironment, Affirmative Action, Balanced Budget, Ryan Budget, Tax Cut, Tax Hike Act, Birth Control, Repeal ACA, Gay Military, Keystone Pipeline, Troops–allies, Troops-UN, Income v. Sales Tax, Black Favors, Black Class

Identity variables: Birthyear, Gender, Education, Registered to Vote, Donated, Union, Born Again, Atheist-Agnostic, Religious

Of all of these, the five with the biggest t-statistics are all issue variables: *Abortion, Gay Marriage, ACA Health Plan, Global Warming*, and *Taxes v. Spending.*

Our universal regression including both issue and identity variables has an R^2 of .53, only slightly higher than the .52 with just the issue variables. In contrast, if we drop the issue variables and retain just the identity variables, the R^2 falls to .19. Apparently, the identity variables help explain a few observations, but do not explain *Conservative-Self* more generally.

The top identity variable in simple regressions is *Religious*, with an . With 53 variables, best subsets regression becomes considerably more difficult—- it took over an hour for our office computer to run the routine (there are some 19 billion possible sets of 10 variables in competition with each other for the highest R^2, though the algorithm does not need to check each set separately). The only identity variable that appears in the top ten is *Religious* (which is also the best identity variable for a simple regression, with an R^2 of .09). *Religious* is only in last place among the top ten, displacing *Immigpatrol.*

TABLE 8

PREDICTORS OF VOTING FOR OBAMA IN 2012: IDENTITY VARIABLES

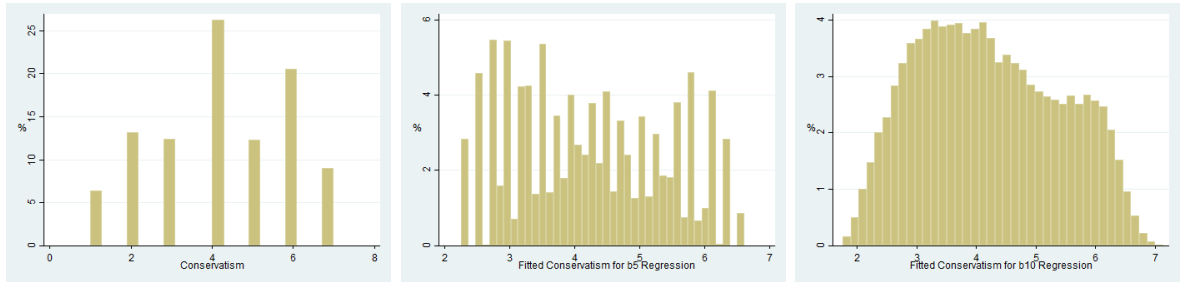| Explanatory variables	R^2 (least squares)	Pseudo R^2 (logit)
Identity variables	.23	.20
Issue variables	.68	.68
Issue, Identification variables	.69	.70
Issue, Identification, Republican-7	.77	.79
Conservative-5	.60	.56
Conservative-10	.62	.61
Conservative-9 , Religious	.62	.60

We conclude that a person's demographic variables are not good predictors of whether he is conservative. Issues make the conservative or liberal, not demographics. Note, however, that this is not the same as saying that positions on issues make someone conservative rather than being conservative makes someone adopt position that he thinks a conservative is supposed to take. Even if one's conservativivism is established by one's position on a few issues, general philosophy, or temperament, one might then adopt positions on other issues because they are labelled as conservative. Weber & Saris (2014) find this. Using data from the European Social Survey, they conclude that issues important to a person affect his left-right orientation but they then use that orientiation to choose positions on issues less important to them. We do not attempt to show causality of that kind in the present paper.

OTHER APPLICATIONS OF THE *Conservative-5* MEASURE

Figure 2 shows histograms of three measures of conservatism. The first is *Conservative-Self*; the second and third are *Conservative-5* and *Conservative-10.* The second and third figures lack the peaks in the center and at the right, and have a mode at the far left (for *Conservative-5*) and the moderate left (for *Conservative-10*). This confirms the well-known result that Americans do not like to label themselves as on the left. Thus, although the mean of *Conservative-Self* is 4.25, more conservative than the 4.00 halfway between 1 and 7, in fact the modal political belief is on the left. Americans do not like to label themselves as liberal even if they take the issue positions that they attribute to liberals. This suggests that self-identified ideology is not as good a measure of someone's ideology as asking them about a few issues and weighting their responses. We also see that the distribution of American beliefs about these issues is more evenly distributed than

25

one might think.

FIGURE 2
DISTRIBUTIONS OF THREE MEASURES OF CONSERVATISM

Note: These percentages are adjusted for sampling weights.

Regional Differences Table 9 shows the levels of *Conservative-5* by region. The Northeast is the most liberal region, with Massachusetts and Vermont the most liberal states (not counting DC). The South is the most conservative region, with Alabama and Oklahoma the most conservative states.

We can see how our measure of conservatism matches regional ideas of what it means to be conservative. The Northeast is a liberal part of the country, so someone who would be called a conservative there might be called a liberal in the South, in which case we would underestimate the difference between the two regions. Or, it might be that the idea of what is conservative are the same in both regions, so self-identified conservatism does accurately measure the difference in ideology. Our data can help distinguish between these two possibilities. Table 9's last seven columns show respondents' overestimates of how conservative they would rate on the national scale.

To determine how overestimates vary across regions, however, we must correct for regression to the mean. Look back to the histograms in Figure 2. The *Conservative-5* and *Conservative-10* indices do not have as many extreme liberals—1's and 2's— as the direct survey responses of *Conservative-Self*. When least squares regression analysis constructs predicted values, it tends to avoid extreme predictions, because when they are wrong the squared error is large. This reflects the fact that someone's high self-evaluation of his conservatism has two components. First, it is likely that someone with a high value of *Conservative-Self* really is more conservative, even according to the views of the population at large rather than his own idea of conservatism. Second, that person's measurement error is likely to be more positive— that is, more in the conservative direction. He is more likely to be someone with an idiosyncratic view of how conservative he is, compared with how other Americans would rate him. Thus, the best estimate of his conservatism in the sense of what the general population would think is below 7, and that is

what the regression indices provide, on average. For $Conservative\text{-}Self = 7$ respondents, the average value of $Conservative\text{-}5$ is 5.60, not 7. For t $Conservative\text{-}Self=1$ respondents, their average value of $Conservative\text{-}5$ is 3.02, not 1.

Our interest is in comparing overestimates across regions, so Table 9 subtracts the effect of regression to the mean. This allows us to compare regional differences across levels of $Conservative\text{-}Self$ even though values such as $Conservative\text{-}Self = 7$ show more regression to the mean than $Conservative\text{-}Self = 4$. Our definition of "overestimate" will be how much lower a region's average conservatism is compared with the national average for respondents with a particular level of self-identified conservatism:

Overestimate = (National mean of $Conservative\text{-}5$ for respondents with $Conservative\text{-}Self = C_i$)

- (Region i's average value of $Conservative\text{-}5$ for respondents with $Conservative\text{-}Self = C_i$)

For $Conservative - Self = 7$ (the most conservative), Table 9 shows the Northeast with an overestimate of .16, because in the Northeast the average value of $Conservative\text{-}5$ across respondents who chose $Conservative\text{-}Self = 7$ is 5.44, whereas nationally the average for respondents who chose $Conservative\text{-}Self = 7$ is 5.60. People who consider themselves extreme conservatives in the Northeast are not as conservative as extreme conservatives nationally. Equivalently, Northeastern people who consider themselves extreme conservatives define "conservative" slightly differently than people elsewhere in America.

Even moderate conservatives in the Northeast are more liberal than elsewhere in the country, it seems. Those with $Conservative\text{-}Self= 4,5,6,7$ are all less conservative than their equivalents elsewhere. Interestingly enough, though, the overestimates are smaller for the values 1,2, and 3. These, the liberals, have a view of their degree of liberalism similar to the rest of the country's. A moderate liberal in the Northeast is not an extreme liberal according to the rest of the country.

Other patterns emerge. The South's pattern is the diametric opposite of the Northeast's. Southern liberals underestimate their conservatism, and just as in the Northeast, the misestimate extends to middle-of-the-road people— $Conservative\text{-}Self= 1,2,3,4$— but according to the rest of the country, Southern conservatives are no more or no less conservative. The Midwestern view of the right-left spectrum is close to the U.S. average– little underestimate or overestimate. The West is quite different. Western liberals are more liberal than they think, and Western conservatives more conservative. The West's average value of $Conservative\text{-}5$ is 4.18, the second most liberal. Western extreme liberals, though, are more extreme than they think, compared to extreme liberals elsewhere. A $Conservative\text{-}Self = 1$ liberal in the West has a value of $Conservative\text{-}5$ that is .19 points higher than the national average for the most extreme liberals, a large amount given that the difference between the means for the Northeast and the West is only .18. Indeed all

across the range *Conservative-Self* = 1,2,3,4, Western liberals are more to the left than equivalent degrees of liberalism elsewhere. Extreme conservatives in the West underestimate their conservatism, on the other hand, though this is a weaker effect and only for *Conservative-Self* = 5,6. It is not that the West has extremes in its levels of self-identified conservatism: 9% are 1's and 13% are 7's in the West, compared to 8% and 13% for the nation. Rather, within the extreme categories the views are more extreme, or some people who elsewhere would respond with 1 or 7 respond with 2 and 6 in the West, leaving the extreme categories more extreme.

TABLE 9

CONSERVATISM AND OVERESTIMATION OF CONSERVATISM BY REGION

Region	*Conservative-5*	Value of *Conservative-Self*						
		1	2	3	4	5	6	7
Midwest	4.28	.00	-.02	-.03	-.04	-.03	-.02	-.04
Northeast	4.01	.03	.02	**.05**	**.08**	**.10**	**.17**	**.16**
South	4.42	**-.19**	**-.10**	**-.08**	**-.08**	-.04	-.03	.01
West	4.18	**.16**	**.12**	**.08**	**.09**	.02	**-.06**	**-.07**

Note: For the definition of "overestimate", see the text. Magnitudes of .05 or greater in either direction are boldfaced.

TABLE 9
CONSERVATISM AND OVERESTIMATION OF CONSERVATISM BY REGION AND STATE

Region	*Conservative-5*	Overestimate	Sample size
midwest	4.28	.007	12,269
northeast	4.01	.041	10,902
south	4.42	-.066	18,607
west	4.18	-.016	12,757
Alabama	4.64	-0.08	749
Oklahoma	4.64	-0.16	617
Wyoming	4.63	-0.37	141
Idaho	4.63	-0.13	378
Utah	4.61	-0.01	539
New York	3.94	0.07	2,834
Rhode Island	3.90	0.10	262
Vermont	3.84	0.06	159
Massachusetts	3.78	0.04	1,149
D.C.	3.34	0.14	109

Note: For the definition of "overestimate" see the text.

Restricting Best Subsets to the Extremes. One might think that people who rate themselves as more extreme liberals and conservatives would be better informed about what issues correspond with the left-right spectrum. Dropping the middle-of-the-road respondents, whether those with *Conservative-Self* = 4 or with the three values 3, 4 and 5, the top five issues are *Climate Change, Gay Marriage Abortion, Obamacare,* and *BlackFavors*, just as in the entire sample. The R^2 dropping *Conservative-Self* values of 3, 4, and 5 is .55, higher than the .47 from the entire sample. Dropping just *Conservative-Self* = 4, the R^2 rises to .60.

An alternative split includes just *Conservative-Self* = 1, 2 (extreme liberals) or just *Conservative-Self* = 6, 7 (extreme conservatives). Using either, the R^2 from using five issues to compute an ideology measure falls below .05. Although the issues are useful for distinguishing between liberals and conservatives in the general population, they are unable distinguishing between small gradations of the ideologically committed.

5. CONCLUDING REMARKS

For a wide variety of empirical projects, scholars need a simple way to summarize a respondent's ideological commitment. Some scholars will hope to use the resulting measure as a dependent variable: they will try to explain why various people hold the political preferences that they do. Other scholars will try to use the measure as an independent variable: to measure the impact that ideological commitment can have on other facets of observed behavior.

The techniques used to create this measure of ideological commitment present a variety of problems. One way is to rely on self-identification: ask how respondents characterize themselves. Unfortunately for the scholar, people do not share a common definition of conservative and liberal; many people are averse to applying the labels to themselves; and any scholar who relies on a single response necessarily invites measurement error. Many scholars use factor analysis instead. This method, however, does not give self-identification variables their due. Scholars using the method either ignore self-identification entirely, or include it with the issue variables. The first approach throws away information. The second fails to treat it with the distinction it deserves as the respondent's own summary of the commitments he holds. Other scholars have developed additional techniques. While many of these methods offer conceptual advances, they leave the user uncertain quite what he has actually computed.

We offer a computationally tractable, conceptually simple technique that gives self-identified ideological commitment its due. Select the 5 or 10 issue variables that best predict a respondent's ideological commitment. Regress that commitment on those variables. Use the resulting coefficients to calculate fitted values for each respondent.

Linear regression with specification selection is a more transparent way to measure conservatism than factor analysis, and requires fewer survey questions. A set of five well-chosen issue questions measures conservatism almost as well as a much larger set in our five-fold cross-validation. Narrowing down issues in this way, it seems Americans tend to define liberal and conservative by social issues more than questions of economic or foreign policy. Moreover, the ideology measure thus defined— the fitted values from a regression using the five top variables— nicely predicts whether a respondent voted for Barack Obama in 2012. Using the fitted values from the top 5 issue variable regression, the resulting R^2 for prediction of vote for Mr. Obama is .60, considerably better than the .46 from self-identified conservatism and almost equal to factor analysis's .63. We also found evidence suggesting that Americans tend not to pick their ideology according to their identity group: demographics variables do not predict self-identified conservatism as well as issue variables.

Issue	CCES code	Description
Iraq Mistake	cc305	Invading Iraq was a mistake.
Afghanistan mistake	cc306	Afghanistan— mistake
Gun Control	cc320	Gun laws should be stricter.
Global Warming	cc321	Global warming is not a problem.
Immig–legal	cc322_1	Immigration — Grant legal status
Immig-patrol	cc322_2	Immigration— Increase border patrol
Immig–police	cc322_3	Immigration — Allow police to question
Immig–business	cc322_4	Immigration— Fine US businesses
Immig–services	cc322_5	Immigration— Prohibit services
Immig–citizenship	cc322_6	Immigration — Deny automatic citizenship
Abortion	cc324	Abortion should be entirely legal.
Jobs v. Environment	cc325	Jobs-Environment
Gay Marriage	cc326	Gay marriage should be legal.
Affirmative Action	cc327	Affirmative action
Balanced budget	cc328	Balanced Budget Pref 1
Fiscal	cc329	Fiscal Preference— #2
Ryan budget	cc332a	Roll Call Votes - Ryan Budget Bill
Simpson budget	cc332b	Simpson-Bowles Budget Plan
Tax Cut	cc332c	Middle Class Tax Cut Act
Tax hike act	cc332d	Tax Hike Prevention Act
Mand. Brth Ctrl Ins.	cc332e	Birth Control Exemption
US Korea trade	cc332f	U.S.-Korea Free Trade Agreement
Repeal ACA	cc332g	Repeal Affordable Care Act
Keystone Pipeline	cc332h	Keystone Pipeline
ACA Health Plan	cc332i	Affordable Care Act of 2010
Gays in military	cc332j	End "Don't Ask, Don't Tell"
Troops–oil	cc414_1	Approve troops to — Ensure the supply of oil
Troops–terrorist	cc414_2	Approve troops to — Destroy a terrorist camp
Troops–genocide	cc414_3	Approve troops to— Genocide or a civil war
Troops–democracy	cc414_4	Approve troops to — Assist democracy
Troops–allies	cc414_5	Approve troops to— Protect allies
Troops–UN	cc414_6	Approve troops to — Help UN
Troops–none	cc414_7	Approve troops to —None
Tax or Spend	cc415r	Spending cuts preferred to tax increases.
Income or Sales tax	cc416r	Income tax preferred to sales tax
Black Favors	cc422a	Blacks should not get special favors
Black Class	cc422b	Conditions hard for blacks to leave lower class

Issue	CCES name	Description
Hispanic	hispanic	Of hispanic descent
Registered to vote	votereg	Registered to vote
Birthyear	birthyr	Year of birth.
Female	gender	Female=2
Education	educ	6 choices for education level
Donated	cc417a_4	Made political donations
Union	unionhh=3	Household member a union member
Black	race=2	black
Govworker	employercat=3	Employed by a government
Married	marstat=1	Married
Divorcedsep	marstat=2,3	Divorced or separated
Religious	pew_religimp=1	Religion not important in your life.
Born Again	pew_bornagain	Born again Christian
Atheist or Agnostic	religpew=9, 11	Atheist or agnostic
Family Income	faminc	Family income
Not Military	milstat_5	No member of family in military
Has Child	child18	Has a child under 18

The Phrasing of the Questions in *Conservative-10, Religious, Democrat, Republican*, and *Republican7*

Issue Variables

Global Warming (cc321). From what you know about global climate change or global warming, which one of the following statements comes closest to your opinion?

14764 1 Global climate change has been established as a serious problem, and immediate action is necessary.

16378 2 There is enough evidence that climate change is taking place and some action should be taken.

11461 3 We don't know enough about global climate change, and more research is necessary before we take any actions.

8693 4 Concern about global climate change is exaggerated. No action is necessary.

3075 5 Global climate change is not occurring; this is not a real issue.

Gay Marriage (cc326) . Do you favor or oppose allowing gays and lesbians to marry legally.

NO=0

YES=1

ACA Health Plan (cc332i). Congress has considered many specific bills this year. We'd like to know how you would have voted on [this bill].

Affordable Health Care for all Americans Act: Requires all Americans to obtain health insurance. Allows people to keep current provider. Sets up national health insurance option for those without coverage. Paid for with tax increases on those making more than $500,000 a year."

NO=0

YES=1

Black Favors (cc422a). Do you agree or disagree with the following statement[s]?

"The Irish, Italians, Jews, and many other minorities overcame prejudice and worked their way up. Blacks should do the same without any special favors."

19422 1 Strongly agree

8829 2 Somewhat agree

8339 3 Neither agree nor disagree

4790 4 Somewhat disagree

3552 5 Strongly disagree

Abortion (cc324). Which one of the opinions on this page best agrees with your view on abortion?

5684 1 By law, abortion should never be permitted

14146 2 The law should permit abortion only in case of rape, incest or when the woman's life is in danger

7174 3 The law should permit abortion for reasons other than rape, incest, or danger to the woman's life, but only after the need for the abortion has been clearly established

27111 4 By law, a woman should always be able to obtain an abortion as a matter of personal choice

Tax or Spend (cc415r). If your state were to have a budget deficit this year it would have to raise taxes on income and sales or cut spending, such as on education, health care, welfare, and road construction. What would you prefer more, raising taxes or cutting spending? Choose a point along the scale from 100% tax increases (and no spending cuts) to 100% spending cuts (and no tax increases). The point in the middle means that the budget should be balanced with equal amounts of spending cuts and tax increases. If you are not sure, or don't know, please check the 'not sure' box.

Write a number from 0=All from tax increases to 100=All from spending cuts.

Mandatory Birth Control Insurance (cc332e). Congress has considered many specific bills over the past two years. For each of the following tell us whether you support or oppose the legislation in principle.

"Birth Control Exemption. A bill to let employers and insurers refuse to cover birth control and other health services that violate their religion beliefs."

20915 1 Support
32425 2 Oppose

Iraq Mistake (cc305). "All things considered do you think it was a mistake to invade Iraq?"

NO=0
YES=1

Affirmative Actionl (cc327) "Armative action programs give preference to racial minorities in employment and college admissions in order to correct for past discrimination. Do you support or oppose armative action?"

Strongly support, Somewhat support, Somewhat oppose, Strongly oppose.

Border Patrol (cc322_2). What do you think the U.S. government should do about immigration? Select all that apply. [only one is given here] "Increase the number of border patrols on the US-Mexican border."

NO=0
YES=1

IDENTITY VARIABLES:

Religious (pew_religimp) " How important is religion in your life?"

1=Very important, Somewhat important, Not too important, 4=Not at all important

Party variables:

Republican7 (pid7) Would you call yourself a strong Democrat or a not very strong Democrat?Would you call yourself a strong Republican or a not very strong Republican?Do you think of yourself as closer to the Democratic or the Republican Party?"

1= Strong Democrat (13,723), 2= Not very strong Democrat, 3=Leans Democrat, 4= Independent (6,205), 5 = Leans Republican, 6 = Not a very strong Republican, 7 = Strong Republican (9,640).

Republican, Democrat (pid3) Generally speaking, do you think of yourself as a

1= Democrat, 2=Republican, 3= Independent, 4= Other (open textbox) (2,313), 5= Not sure.

References

Aldrich, John H. & Richard D. McKelvey (1977) "A Method of Scaling with Applications to the 1968 and 1972 Presidential Elections," *The American Political Science Review,* 71(1): 111–130.

Alford, J. R., C. L. Funk, & J. R. Hibbing (2005) "Are Political Orientations Genetically Transmitted?" *The American Political Science Review*, 99: 153–168.

Amodio, D. M., J. T. Jost, S. L. Masters & C. M. Lee (2007) "Neurocognitive Correlates of Liberalism and Conservatism," *Nature Neuroscience*, 10: 1246–1247.

Alwin, Duane F. & Jon A. Krosnick (1985) "The Measurement of Values in Surveys: A Comparison of Ratings and Rankings," *Public Opinion Quarterly,* 49: 535-552.

Ansolabehere, Stephen (2013) "Guide to the 2012 Cooperative Congressional Election Survey March 11, 2013."

Ansolabehere, Stephen, Jonathan Rodden & James M. Snyder, Jr. (2008) "The Strength of Issues: Using Multiple Measures to Gauge Preference Stability, Ideological Constraint, and Issue Voting," *The American Political Science Review*, 102: 215–232.

Belloni, Alexandre & Victor Chernozhukov (2013) "Least Squares after Model Selection in High-Dimensional Sparse Models," *Bernoulli*, 19(2): 521–547.

Bouchard, T. J., N. L. Segal, A. Tellegen, M. McGue, M. Keyes & R. Krueger (2003) "Evidence for the Construct Validity and Heritability of the Wilson-Patterson Conservatism Scale: A Reared-Apart Twins Study of Social Attitudes," *Personality and Individual Differences*, 34: 959–969.

Carmines, Edward G., Michael J. Ensley & Michael W. Wagner (2012a) "Political Ideology in American Politics: One, Two, or None?" *The Forum,* 10: art. 4.

Carmines, Edward G., Michael J. Ensley & Michael W. Wagner (2012b) "Who Fits the Left-Right Divide? Partisan Polarization in the American Electorate," *American Behavioral Scientist,* 20: 1–23.

Carmines, Edward G. & Nicholas D'Amico (2015) "The New Look in Political Ideology Research," *American Review of Political Science*, 18: 205–16.

Castle, Jennifer L., Xiaochuan Qin & W. Robert Reed (2013) "Using Model Selection Algorithms to Obtain Reliable Coefficient Estimates," *The Journal of Economic Surveys*, 27(2): 269-296 (April 2013).

Cavanaugh, Joseph & Andrew Neath (2011) "Akaikes Information Criterion: Background, Derivation, Properties, and Refinements," *International Encyclopedia of Statistical Science*, 26-29.

Conover, Pamela Johnston & Stanley Feldman (1981) "The Origins and Meaning of Liberal/Conservative Self-Identifications," *The American Journal of Political Science*, 25: 617-645.

Conover, Pamela Johnston & Stanley Feldman (1984) "How People Organize the Political World: A Schematic Model," *The American Journal of Political Science*, 28: 95–126.

Conover, P. J., & S. Feldman (1981) "The Origin and Meaning of Liberal/Conservative Self Identification," *The American Journal of Political Science*, 25: 617–645.

Cramer, J. S. (1987) "Mean and Variance of R^2 in Small and Moderate Samples," *The Journal of Econometrics,* 35: 253–266.

Crowson, H. Michael (2009) "Are All Conservatives Alike? A Study of the Psychological Correlates of Cultural and Economic Conservatism," *Journal of Psychology,* 143: 449–463.

Dawes, Robyn M. (1980) *Rational Choice in an Uncertain World*, Harcourt Brace (1988)

Efron, Bradley & Carl N. Morris (1977) "Stein's Paradox in Statistics," *Scientific American*, 236(5): 119–127.

Evans G, A. Heath & M. Lalljee (1996) "Measuring Left-Right and Libertarian-Conservative Attitudes in the British Electorate," *British Journal of Sociology,* 47: 93–112.

Feldman, S. (1988) "Structure And Consistency in Public Opinion: The Role of Core Beliefs and Values," *The American Journal of Political Science*, 31: 416–440.

Feldman, Stanley & Christopher Johnston (2014) "Understanding the Determinants of Political Ideology: Implications of Structural Complexity," *Political Psychology,* 35: 337-358.

Feldman, Stanley & John Zaller (1992) "The Political Culture of Ambivalence: Ideological Responses to the Welfare State," *The American Journal of Political Science*, 36: 268-307.

Funk, Carolyn L., Kevin B. Smith, John R. Alford, Matthew V. Hibbing, Nicholas R. Eaton, Robert F. Krueger, Lindon J. Eaves & John R. Hibbing (2012) "Genetic and Environmental Transmission of Political Orientations," *Political Psychology*, 1–15 (2012).

Gentzkow, Matthew & Shapiro, Jesse M. (2008) "Competition and Truth in the Market for News," *The Journal of Economic Perspectives*, 22(2): 133-154 (Spring, 2008).

Gentzkow, Matthew & Jesse M. Shapiro, (2011) "Ideological Segregation Online and Offline,"

The Quarterly Journal of Economics Surveys, 126(4): 1799–1839 (Nov 2011).

Gentzkow, Matthew, Jesse M. Shapiro & Michael Sinkinson (2012) "Competition and Ideological Diversity: Historical Evidence from US Newspapers," National Bureau of Economic Research (2012/7/19).

Graham, J., J. Haidt & B. A. Nosek (2009) "Liberals and Conservatives Rely on Different Sets of Moral Foundations," *The Journal of Personality and Social Psychology*, 96: 1029–1046.

Haidt, J., & J. Graham (2007) "When Morality Opposes Justice: Conservatives Have Moral Intuitions that Liberals May Not Recognize," *Social Justice Research*, 20: 98–116.

Hastie, Trevor, Robert Tibshirani & Jerome Friedman (2003) *The Elements of Statistical Learning Data Mining, Inference, and Prediction*, Second edition.

Hatemi, P. K., Hibbing, J. R., Medland, S. E., M. C. Keller, Alford, J. R., Smith, K. B., Martin, N. G., & L. J. Eaves (2010) "Not by Twins Alone: Using the Extended Family Design to Investigate Genetic Influence on Political Beliefs," *The American Journal of Political Science*, 54: 798–814.

Heath, Anthony, Geoffrey Evans & Jean Martin (1994) "The Measurement of Core Beliefs and Values: The Development of Balanced Socialist/Laissez Faire and Libertarian/Authoritarian Scales," *The British Journal of Political Science,* 24: 115-132.

Heckman, James J. & James M. Snyder, Jr. (1997) "Linear Probability Models of the Demand for Attributes with an Empirical Application to Estimating the Preferences of Legislators," *The RAND Journal of Economics,* 28: S142-S189.

Hocking, R. R. & R. N. Leslie (1967) "Selection of the Best Subset in Regression Analysis," *Technometrics*, 9(4): 531-540 (Nov. 1967).

Inbar, Y., D.A. Pizarro & P. Bloom (2008) "Conservatives Are More Easily Disgusted than Liberals," *Cogn. Emot.*.

Jacoby, William G. (2006) "Value Choices and American Public Opinion," *The American Journal of Political Science*, 50: 706-723.

Jost, J. T. (2009) "Elective Affinities?: On the Psychological Bases of Left-Right Differences," *Psychological Inquiry*, 20: 129–141.

Jost, John T., Jack Glaser, Arie W. Kruglanski, & Frank J. Sulloway (2003) "Political Conservatism as Motivated Social Cognition," *Psychological Bulletin*, 129: 339–375 (May 2003).

Layman, Geoffrey C. & Thomas M. Carsey (2002) "Party Polarization and "Conflict Extension" in the American Electorate," *The American Journal of Political Science*, 46: 786-802.

Lindgren, James (2012) "Who Fears Science?" working paper, Northwestern Law School (2012) http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2018806.

Little, Roderick J. A. (1992) TITLE *The Journal of the American Statistical Association,* 87(420): 1227-1237 (Dec. 1992).

McCann, James A. (1997) "Electoral Choices and Core Value Changes: The 1992 Presidential Campaign," *The American Journal of Political Science*, 41: 564-583.

McCarty, John A. & L.J. Shrum (2000) "The Measurement of Personal Values in Survey Research: A Test of Alternative Rating Procedures," *Public Opinion Quarterly,* 64: 271-298.

Miller, Alan S. (1992) "Are Self-Proclaimed Conservatives Really Conservative? Trends in Attitudes and Self-Identification among the Young," *Social Forces,* 71: 195-210 (September 1992).

Piurko, Y., S. H. Schwartz, & E. Davidov (2011) "Basic Personal Values and the Meaning of Left-Right Political Orientations in 20 Countries," *Political Psychology Surveys*, 32(4): 537–561 (2011).

Rammstedt, Beatrice & Oliver P. John (2007) "Measuring Personality in One Minute or Less: A 10-Item Short Version of the Big Five Inventory in English and German," *Journal of Research in Personality*, 41: 203–212.

Schiffer, Adam J. (2000) "I'm Not That Liberal: Explaining Conservative Democratic Identification," *Political Behavior,* 22: 293-310.

Smith, K. B., & P. K. Hatemi (2011) "OLS Is AOK for ACE: A Regression-Based Approach to Synthesizing Political Science and Behavioral Genetics Models," *Political Behavior*, 35: 383-408.

Srivastava, Sanjay (undated) "Measuring the Big Five Personality Factors," http://psdlab.uoregon.edu/bigfive.html.

Stein, Charles M. & James, W. (1961) "Estimation with Quadratic Loss," *Proc. Fourth Berkeley Symp. Math. Statist. Prob.*, 1: 361-379.

Swedlow, Brendon & Mikel L. Wyckoff (2009) "Value Preferences and Ideological Structuring of Attitudes in American Public Opinion," *American Politics Research,* 37: 1048-1087 (November 2009).

Treier, Shawn & D. Sunshine Hillygus (2009) "The Nature of Political Ideology in the

Contemporary Electorate," *Public Opinion Quarterly,* 73: 679-703.

Treier, Shawn & Simon Jackman. 2008) "Democracy as a Latent Variable," *The American Journal of Political Science*, 52: 201-217.

Verhulst, Brad, Lindon J. Eaves, & Peter K. Hatemi (2012) "Correlation not Causation: The Relationship between Personality Traits and Political Ideologies," *The American Journal of Political Science*, 56: 34-51.

Weber, Wiebke & Willem E. Saris (2014) "The Relationship between Issues and an Individual's Left-Right Orientation," *Acta Politica*, 1-21.

Wilson, G. D. (1973) "Development and Evaluation of the C-Scale," In G. D. Wilson (Ed.), *The Psychology of Conservatism*, 49–69, London: Academic Press

Wilson, G. D., & J. R. Patterson (1968) "A New Measure of Social Conservatism," *The British Journal of Social and Clinical Psychology*, 7: 264–269.

Wolfle, Lee M. (1980) "The Enduring Effects of Education on Verbal Skills," *Sociology of Education*, 53(2): 104–114 (Apr. 1980).

Zakrisson, I. (2005) "Construction of a Short Version of the Right-Wing Authoritarianism (RWA) Scale," *Personality and Individual Differences*, 39: 863–872.

Zechmeister, Elizabeth (2006) "What's Left and Who's Right? A Q-method Study of Individual and Contextual Influences on the Meaning of Ideological Labels," *Political Behavior*, 28(2): 151-173 (June 2006).

Zumbrunnen, John & Amy Gangl (2008) "Conflict, Fusion, or Coexistence? The Complexity of Contemporary American Conservatism," *Political Behavior*, 30(2): 199-221 (June 2008).