Measuring Voter Ideology: Descriptive Regression Measurement of the Left-Right Spectrum

June 9, 2015

J. Mark Ramseyer and Eric B. Rasmusen

Abstract

For scholars studying the political attitudes of the general public, someone's position on the ideological spectrum is a good place to start. Typically, scholars identify that position through factor analysis on survey questions, making the assumption that the most important artificially constructed factor indicates the person's position on the liberal-conservative spectrum. The leading attitudinal surveys, however— the GSS, the CCES, and the ANES— include a variable giving a respondent's self-identified ideology, a variable given no special prominence by factor analysis. We suggest a new ideology measure: the individual's fitted value from a regression of self-identified ideology on other variables. We describe various ways to choose those other variables. This approach gives proper priority to the usefulness of self-reported ideology. It lets us test whether voters identify their own ideology through identity-group variables; avoids the bias introduced in choosing which issue variables to include in factor analysis; and shows which positions the average American— as opposed to the analyst— thinks 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.

This paper: http://rasmusen.org/papers/spectrum-ramseyer-rasmusen.pdf.

Keywords: Liberalism, conservatism, ideology, political index, political spectrum, identity politics, factor analysis, LASSO, prediction, best-subsets regression.

We thank participants in a seminar at Oberlin College for helpful comments. The central idea in this paper is from a talk by James Lindgren in Urbana some years ago. This research was supported in part by Lilly Endowment, Inc., through its support for the Indiana University Pervasive Technology Institute, and in part by the Indiana METACyt Initiative. The Indiana METACyt Initiative at IU is also supported in part by Lilly Endowment, Inc.

Note: This is an earlier draft than we'd usually show to other people, but we'd like a chance to see what Harvard people think about it before Eric goes back to Bloomington at the end of June, especially since political science and political psychology journals are not our usual audiences. You'll see that it gets rather sketchy after page 18.

In particular, we'd welcome votes on which is (a) the most promising, and (b) the least promising of the Extensions at the end of the paper.

We also have long been searching for packaged software that will do Best Subsets regression (all possible regressions of 5 variables drawn from 30, to find the combination with the highest R2).

1 Introduction

Who is liberal? Who is conservative? Scholars, like the general public, use these terms to refer to the ends of the standard uni-dimensional political spectrum, but pinning down what they mean is difficult and contentious.

Because the terms refer to the ends of that spectrum the two questions are intimately related. To follow our data source, we will use "Who is conservative?", but we treat the question exactly the same as we would "Who is liberal?" onsider 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, http://reason.com/archives/1975/07/01/inside-ronald-reagan.

> "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 observers 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. It 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.

Other observers take a synthetic approach. They start by specifying a set of positions on issues as conservative, and then try to determine what the positions have in common. This, too, is an approach a political theorist might use, and 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 the online quizzes and the politician ratings.

What is needed for political science, however, is a way to operationalize political ideology, a way to rank views numerically from the most liberal to the most conservative.The public, too, would like to operationalize it, as evinced by the number of online quizzes to rate one's own degree of conservatism.[1] Journalists too pay close attention to liberal and conservative ratings of politicians.[2] The question comes up chiefly in two contexts: for politicians and other government officials (e.g. judges), and for voters. Politicians are relatively small in number and heavily documented by news reports and official votes. We will not consider them in this paper. Voters— and people generally— are analyzed using survey data. They will be our subject here.

Political scientists commonly take a different synthetic approach 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 their degree of conservatism. Given this assumption, they can estimate a person's conservatism by estimating the underlying latent variable that best explains his observed positions on political issues. Typically, they do this through factor analysis. Note that the measures by which these scholars identify conservatives come entirely from the data at hand – from the positions people take on various issues. They may (or may not) include the respondents' self-identified degree of conservatism, but if they do they do not give that variable special weight. The well-known Aldrich & McKelvey (1977) scaling takes this approach.

We propose a new synthetic approach that eliminates several of the problems introduced by factor analysis. Like those scholars who would identify conservative survey respondents through the technique, we do not define what it means to be conservative ex ante. Like them, we assume that the respondents take the positions they do because of their conservatism. But where other scholars either (a) omit self-identified ideology and rely exclusively on issue variables, or (b) include the self-identified ideology as one more variable in a mix with the issue variables in factor analysis, we treat it as a dependent variable in regression analysis. Of the many survey questions, self-identified ideology most clearly reflects the respondent's own sense of what it means to be conservative. Scholars who ignore this throw away valuable information.

In using regression analysis, we take a statistical technique that reflects the way respondent think. To a respondent, ideological identity reflects the positions he takes on various ideological issues. When we regress self-identified identity (as the dependent variable) on the various issue variables (as independent variables), we capture that dynamic. In essence, we estimate what the average person thinks is the conservative position on each issue. We thus start with a survey (the Cooperative Congressional Election Study, CCES) in which each respondent describes his own

[1] See, for example, http://gotoquiz.com/conservative_or_liberal.

[2] The American Conservative Union, for instance, has rated Congressmen since 1971. See http://conservative.org/legislative-ratings/.

degree of conservatism. We then use linear regression to examine 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 matches self-identified conservatism in the survey. We apply the resulting coefficients to an individual's issue positions to measure where he lies on the ideological scale from liberal to conservative.

Note several things. First, we describe the statistical technique by which to discover the issue variables that most clearly identify conservative respondents. By identifying the most closely predictive variables, we reduce the random error (or bias) introduced through the selection of variables to include in factor analysis. Second, by tethering the calculation of the coefficients on issue variables to self-identified conservatism, we again reduce the random error (or bias) introduced through variable selection. Third, we do not account only for a conservative's view of what conservatism might be. Instead, we average over the views of conservatives, moderates, and liberals. Inter alia, we explore whether conservatives define conservatism differently than liberals do. Last, we ask whether respondents call themselves conservative not because of their positions on issues, but because of their perceived (or desired) self-image – e.g., that perhaps a white Southern man calls himself a conservative because he is white, Southern, and male, even if he takes liberal positions on most issues.

2. Survey Responses and Self-Identification.

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 questions about their political identity honestly. They do not always share a common sense of what the various political 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 a conservative Republican). 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 (as Ansolabehere, Rodden & Snyder (2008) point out in the context of political surveys – showing that the average of a person's answers to various questions is much more stable over time than his answers to individual questions).

2.1. Factor Analysis.

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

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

The underlying assumption of these analyses is that the respondent's position on issues is determined by one or more factors, underlying ideological variables that are uncorrelated with each other and that we can interpret as corresponding to such ideas such as "conservatism", "economic conservatism", "populism", and so forth. Thus, to use the analysis, it is necessary to match the artificially constructed factor to the political idea, a process that requires both interpretation and the assumption that each factor does correspond to some idea we can understand. Without that process, we are left with an artificial index of dubious utility, a linear combination of every variable in the survey. Moreover, the usual assumption that the factors are uncorrelated with each other means that they cannot possibly correspond to ideas such as "economic conservatism" and "social conservatism" which are commonly held by the same group of people— conservatives. Heckman & Snyder (1997) make this point in arguing for their structural approach.

"Rotating" the factors is one response to the problem of interpretation. When using two or more factors it is possible to construct the factors in many different ways that yield the same fit to the data. It is routine to start by assigning the most possible fit to factor 1, then to find factor 2 as the linear combination of the issue variables that is (a) uncorrelated with factor 1, and (b) explains the most possible variance of the issue variables. "Rotation" is based on the idea that there are many other ways to construct two artificial variables with the same total amount of variance explanation by not insisting on giving factor 1 the best possible fit and letting factor 2 explain more of the variance instead. The analyst can eyeball the factors and the factor loadings and try to come up with two factors such that factor 1 is more correlated with one set of issues and factor 2 with another set such that the factors can be assigned meaningful labels. Of course, this brings in the analyst's prior notion of what issues should be associated with each other. For present purposes, we are only interested in one factor, a measure to correspond to the idea of liberal vs. conservative generally, so the rotation question does not arise. What does raise concern is the implicit assumption that the single variable which best fits the data is liberal vs. conservative ideology. That is plausible, but it is also plausible that the leading factor is distrust of elites, dissatisfaction with the status quo, or some other form of ideology. Indeed, it is not obvious that there exists one dominant ideological variable. The correct model of Americans might be that their political positions are determined by, say, five different latent ideological variables, just as in the standard "Big Five" view of personality psychologists that the answers to survey questions are best explained by five latent variables of approximately equal importance (extraversion, conscientiousness, etc.)[3]

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.

[3]The Big Five factors are extraversion, agreeableness, conscientiousness, neuroticism,and openness. See Sanjay Srivastava (undated) "Measuring the Big Five Personality Factors," http://psdlab.uoregon.edu/bigfive.html for an excellent short history, description, and links to current tests that measure the factors. Psychologists are keenly interested in balancing accuracy against practicality. For an extreme, see Rammstedt & John (2007), which is titled, "'Measuring Personality in One Minute or Less: A 10-Item Short Version of the Big Five Inventory in English and German."

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 the unobserved preference dimension via an ordinal item-response model." 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."

With 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. Treier, Hillygus and Jackman's 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." Moreover, simple tools often yield results close to those from theoretically more rigorous techniques. 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 below 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 (Mark hasn't seen this section yet)

We would like to emphasize that simplicity is a major goal for any summary measurement such as "conservatism". If accuracy were the only goal, the measure that is clearly best would be to present the reader with 100% of the data— a given individual's answers to every one of the survey questions, for example. Accuracy is easy to maximize. But in fact the entire project is to find a way to increase simplicity while sacrificing only a little accuracy. Having 100% of the data may be optimal for a computer, but not for a human, however smart. Our brains are limited and our time has to be allocated among many tasks. Thus, between the alternatives of "The height of every American", "The number of Americans for each inch-long interval of height", and "The average height of Americans", for most occasions we would find the average the most useful, even though it is the least accurate and the least informative.

Much of the theory of business accounting is about how to deal with this. If you want to know the financial health of 1,000 firms, you do not want to have someone send you an email with 1,000 annual reports as attachments. Most likely, all you want is 1,000 numbers— the return on assets of each firm. Or, perhaps the return on equity is better, or the stock return; it depends on your purpose.

The IQ test is another example. Suppose 1,000 Chinese immigrants apply for unskilled work at your company. One thing you could do is to ask them to submit their resumes and school transcripts and then verify these with every employer and school, but that is very expensive, especially since the schools and employers have little incentive to tell the truth. Or, you could hire someone who knows Chinese and who is corruption-proof and have them interview every candidate for 6 hours. An alternative would be to give each applicant a 2-hour IQ test (or whatever sort of test is appropriate for the job— a strength test, for example). Or, one might decide to offer a 10-minute test instead. A 10-minute test would likely be less accurate than a 2-hour test (unless–perhaps— one wants to measure snap judgments) but it would be cheaper.

As noted in a footnote above, psychologists who design tests for actual use in decisionmaking think hard about the tradeoff between length and informativeness; one paper, Rammstedt & John (2007), is titled "Measuring Personality in One Minute or Less: A 10-Item Short Version of the Big Five Inventory in English and German." The General Social Survey, the preeminent survey in sociology, includes an IQ test with just 10 questions, all verbal, which in fact is just a small subset of questions from one of the well-known IQ tests. Nonetheless, this 10-question test has a correlation of .71 with IQ, compared with .30 for the father's education, .29 for the father's occupational prestige, and .51 for the subject's education (see Wolfle [1980]).[4] People could game a simple test like that, of course, but since people do not take GSS-prep courses, it serves well for scholarly purposes. For personality, or, even better, position on the political spectrum, one need not even worry about strategic behavior by the subjects.

Similarly, in measuring conservatism our goal is to find a measure that is well correlated with numerous features of a subject's political ideology, is also correlated with what the average American thinks is conservatism, is transparent, and is quick to measure.

3. The Data and Method

We take our data from the 2012 Cooperative Congressional Election Study (CCES). The data

[4]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. Based on percentile, the only levels possible are 0, 63, 69, 75, 80, 94, 103, 111, 117, or 127 (from data at De La Jara, [2006], Half Sigma [2011]).

are in many ways similar to 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 and the care that goes into its questions, but we could make the same points with the GSS or ANES.[5] Note that large sample size is useful not only because it allows such things as splitting up the sample into regional or racial subsamples but also because it permits use of recently developed "machine learning" techniques of analysis that replace conventional confidence intervals with the division of the sample between "training" subsamples used for estimation and "testing" subsamples used for verification.

1. The self-identification variable. The CCES asks respondents to locate themselves along an ideological spectrum from 1 (very liberal) to 7 (very conservative). It asks this question both before and after a given election. We take this question as our basic respondent self-identification question, and call it *Conservative-self.*

> "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

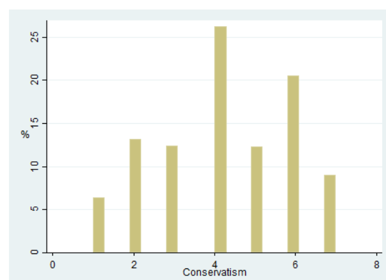

TABLE 1 AND FIGURE 1
THE ANSWERS TO THE SELF-IDENTIFIED CONSERVATISM QUESTION, *Conservative-self*, AND THE RESPONSE PERCENTAGES (N = 51,598) [6]

2. Issue variables. We use the 37 issue variables in Table A1 of the Appendix. These are the CCES questions that were more narrowly ideological (Was the Iraq invasion a mistake?) than

[5]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.

[6] These percentages are adjusted for sampling weights. The survey oversamples certain groups in order to get a representative sample overall.

specifically partisan (Is President Obama to blame for the economy?). The questions cover such issues as the Iraq war, gun control, immigration, abortion, environmentalism, gay marriage, affirmative action, tax policy, free trade, the Affordable Care Act (ACA Health Plan), and the Keystone pipeline.

3. Identity variables. We use the 17 identity variables in Table A2 of the Data Appendix. They include such questions 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 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 predictor of *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 factor analysis. In most 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 the issue questions. Marriage 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. In the discussion below, we observe that including the identity variables in the factor analysis reduced the proportion of variance explained by the first factor. It did so 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. The project. 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 *Conservative-self*, which two variables best predict it, which 3 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.

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?"

We will do it both ways. Unconditional correlation is easy. Conditional correlation (regression) is more challenging but also potentially more useful to the scholar. The unconditional correlation approach depends heavily on the original set of variables. It might be, for example, that abortion best predicts conservatism. A scholar taking the unconditional approach might then select five questions about abortion – since all have the highest correlation. By contrast, a scholar following the regression approach would pick only one of the five.

Because we compare regressions using different explanatory variables, missing values present a special problem in the regression approach. 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. Because this increases the number of observations used to calculate standard errors, it biases the value of those errors; because it adds no new information about the missing variables, it leaves the point estimates unchanged (see Little [1992]). Crucially, the mean value that we impute will not help explain the variance. Hence, any increase in the R^2 results from the actual values for the variables added.

Given that we care only about goodness of fit rather than hypothesis testing, we do not care about any bias in standard errors. We thus use mean imputation as the simplest and easiest to understand imputation technique.[7] Alternatively, we could use "regression imputation." By this technique, we would regress the variable with missing observations, X_1, on the other X variables and replace the missing values with the fitted values. This does make more use of the available information. Unfortunately, because it uses all the explanatory variables to imput the missing values for any single variable, it is inconsistent with our goal of finding the best estimate of ideology from a limited number of explanatory variables. It also, of course, adds complexity and

[7] No technique for imputing missing values is entirely accepted. It is most common simply to drop observations with missing values. We could probably do that here, and maybe should, for this large dataset. Multiple imputation has wide support. It creates estimates of the missing value using correlations of that variable with the other explanatory variables, adds noise so that the standard errors reflect the imprecision of the estimates, and bootstraps the standard errors.

reduces transparency.

4. Picking Variables by Factor Analysis and Other Methods

We start with the standard technique, factor analysis. Because we have 37 variables, we could – hypothetically – generate as many as 37 factors. In fact, STATA does that for us automatically. Scholars always stop with fewer, however, since most use the technique primarily to reduce the number of explanatory variables. Factor analysis on the 54,535 observations yields factor 1 explaining .69 of the variance, factor 2 explaining .16, and factor 3 explaining .11. We are interested only in factor 1. It is a linear combination of all 37 variables, but we show only the top 10 and bottom 5 in Table 2. Following scholarly custom, we exclude the identity variables.

Variable	Factor Loading (Correlation with Factor 1)	Regression Score
Global Warming	-.72	-.108
ACA Health Plan	-.71	-.098
Repeal ACA	.67	.084
Affirmative Action	-.67	-.082
Black Favors	.66	.090
Mandatory Birth Control Insurance	.65	.081
Taxes vs. Spending	-.64	-.078
Black Class	-.63	-.076
Gay Marriage	-.63	-.083
Immig–arrest	.62	.075
...	...	...
Troops–Genocide	-.10	-.007
Korea Trade	.05	.005
Tax Cut	-.02	-.0004
Fiscal	-.01	-.0004
Simpson Budget	.01	.002

Table 2
Factor Analysis of the 37 Issue Variables: The Top Ten and Bottom Five Loadings on the First Factor, and Regression Scores for Constructing Fitted Values

The Table 2 "factor loading" shows the Pearson correlation coefficient between a variable and the estimated factor. Necessarily, it takes values between -1 and +1. Note the "regression score" in the Table. To use an estimated factor, the analyst must first construct coefficients with the factor loadings in order to multiply each of the variables. Scholars have used many techniques: e.g., weighted sum scores, Bartlett scores, and Anderson-Rubin scores. The simplest is the "sum score" method: use the factor loadings (which, remember, are simply the correlation coefficients) as coefficients, multiply a respondent's issue values by the factor loadings, and sum them to generate a score. The "regression score" method constructs the coefficients "by multiplying the observed variable correlation matrix by the matrix of factor loadings" (DiStefano, Zhu & Mindrila, 2009, p. 4). Obviously, this adds yet another layer of opacity to factor analysis.

Note how *ACA Health Plan* and *Repeal ACA* are both in the top ten, with almost identical factor loadings. This is because they are closely related, with a correlation of -.51. Similarly, the correlation between *Affirmative Action* and *Black Favors* is -.49. Indeed, even *ACA Health Plan* and *Global Warming* have a correlation of .48.

A regression of *Conservative-self* on the factor scores (which we will henceforth call *Conservative-factor* yields an R^2 of .48 (or, equivalently, a correlation of 0.70). This will be useful for comparisons later.

One could also include the 17 identity variables in the factor analysis. This would make sense if one thought that the identity variables were correlated with the underlying ideology, whether because they proxied for omitted issues or because of identity politics. If we add those 17 identity variables to the 37 issue variables, the resulting factor 1 explains .53 of the variance instead of .69. Factor 2 now explains .14 instead of .16, and factor 3 explains the same .11 as it did before. When we regress *Conservative-Self* on this new *Conservative-Factor*, we obtain an R^2 of .50 instead of .48. Given the small size of this increase, we return to dropping the identity variables from the factor analysis.

Regression Methods Using Self-Identified Ideology

Turn, then, 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 value for each survey respondent. Note that we could use the technique to estimate the conservatism for respondents outside the sample as well. We shall explore several ways to select the appropriate issue variables.

Although we use ordinary least squares, we realize that even a non-structural model of conservatism should employ ordered probit instead. After all, 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 more intensive, and would be less transparent than least squares. Ordered probit would generate better estimates of the standard errors— but we are not using those estimates anyway. After all, we aim not to test hypotheses but to describe the data, predict, and create an index variable. We aim to replace 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.[8] We do show the results of various ordered probit regressions later, however, for comparison of different measures of conservatism.

Method 1: One Big Regression: Use All the Variables

One way to choose variables is not to choose— just use them all. We have 37 issue variables and 17 identity variables. We could regress *Conservative-self* on all of them, or we could restrict ourselves to the issue variables. If we use all the variables, those with t statistics of 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, Athest-Agnostic, Religious*

Of these, the 5 variables with the biggest t-statistics are *Abortion, Gay Marriage, ACA Health Plan, Global Warming, and Taxes v. Spending.*[9]

The regression with 54 independent variables has 51,598 observations and an $R^2 = .53$. If we include only the issue variables, the R^2 falls to .52. By contrast, if we drop the issue variables and retain 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.

Although this method is more transparent than factor analysis and tests whether identity

[8]A possible second stage would be to take our "best" estimation equations and see whether changing the intervals between answers would increase the R^2— e.g., recode "1,2,3" as "1, 2, 9" and see whether R^2 increases.

[9]We will see later that these are all in the best-10 regression that we find with stepwise regression, and that the first four of these make up the best-4 regression.

affects *Conservative-self*, it is cumbersome. Moreover, in a regression with this many variables, interpretation of t-statistics is problematic. After all, 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 54 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.

Method 2: Picking Variables by Unconditional Correlation

Instead of using the single big regression, suppose we examined how each variable performed individually. Suppose we examined, that is, the unconditional correlation between each variable and *Conservative-self*. Table 3 shows a correlation matrix of the top five, arranged by the size of the unconditional correlation. Note that the top five are all issue variables. The top identity variable is *Religious*, with a correlation of .30.

	Correlation with Cons-self	Warming	ACA	Gay	Repeal	Mand
Global Warming	.52	1.00				
ACA Health Plan	.52	.50	1.00			
Gay Marriage	.51	.41	.39			
Repeal ACA	-.46	-.44	-.53	-.36	1.00	
Mand Birth Ctrl Ins	-.46	-.44	-.45	-.42	.46	1.00

Table 3
A Correlation Matrix of the Top Five Variables by Unconditional Correlation

A regression of *Conservative-self* on the top 5 variables yields an R^2 of .46, which is less than the .48 of *Conservative-factor*, the latent variable from factor analysis.

Method 3: Picking Variables by Conditional Correlation

To find the variables that together best predict *Conservative-self* as a set, we take still another approach. We mix forward and backwards stepwise regression. [10] We begin by identifying the variable with the highest correlation coefficient with *Conservative-self.* This is, of course, the variable that would produce the highest R^2 in a univariate regression. We then take the residuals from that regression, and identify the variable most highly correlated with the residuals. This is again the variable that produces the highest R^2 when regressed against the residuals. We take the residuals from the second regression and identify the variable most highly highly correlated with them. We do the same with a third regression, and so forth.

With each regression, we ask whether the t-statistics on any of the already-included variables are less than the t-statistic on the new variable. If it is, we try omitting that already-included variable to see if the R^2 rises with the new variable instead. If it does, we replace the best-$(n-1)$ regression with the higher R^2 one, and continue as before, checking each time to see if old variables become inferior.

We suggest two ways to estimate the accuracy of the resulting R^2. The R^2 statistic from even a single regression has a complicated distribution. Cramer (1987) showed that the density function for R^2 is an infinite weighted sum of Beta densities with Poisson weights.[11] To explore the accuracy of the obtained R^2, we first employ a bootstrapping technique.

Table 4 shows the sets of variables that stepwise selected, of size one to ten. The last column shows the unconditional correlations of each variable with *Conservative-self.* Note that the R^2 in b1 is the square of the unconditional correlation.

[10] We do this manually, as we found tht the Stata stepwise command does not operate as theoretically required. The STATA command we used was "stepwise forward pe(.001) pr(.001)". See http://www.stata.com/manuals13/rstepwise.pdf. It started by including a variable that did not have the highest t-statistic in a one-variable regression, however, contrary to the way it is supposed to work–whether due to some error of ours, or to a bug in the command.

[11] See David Giles, "Good Old R-Squared!" *Econometrics Beat: Dave Giles's Blog*, http://davegiles.blogspot.com/2013/05/good-old-r-squared.html.

Best-k Predictors	b1	b2	b3	b4	b5	b6	b7	b8	b9	b10	Corr.
1. Global warming is not a problem *(Global Warming)*	.2702		.4328	.4547	.4708	.4821	.4881	.4939	.4982	.5019	.52
2. Against gay marriage *(Gay Marriage)*		.3844	.4328	.4547	.4708	.4821	.4881	.4939	.4982	.5019	.51
3. For ACA Health Care *(ACA Health Care)*		.3844	.4328	.4547	.4708	.4821	.4881	.4939	.4982	.5019	.52
4. Blacks should get special help *(Black Favors)*				.4547	.4708	.4821	.4881	.4939	.4982	.5019	-.43
5. Abortion should be legal *(Abortion)*					.4708	.4821	.4881	.4939	.4982	.5019	-.46
6. Spending cuts better than tax increases (*Tax v. Spending)*						.4821	.4881	.4939	.4982	.5019	.45
7. Mandatory Birth Control Insurance *(Birth Control)*							.4881	.4939	.4982	.5019	.33
8. Invading Iraq was not a mistake (*Iraq Mistake)*								.4939	.4982	.5019	.31
9. Year of birth *(Birthyear)*									.4982	.5019	-.16
10. Immigration— Increase border patrol*(Immig–patrol)*										.5019	-.32

Table 4

An R^2 Matrix of the Best-K Regressions for *Conservative-self* (n=42,855)[12]

In other words, Table 4 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 picked *Global Warming*, but the best-2 regression dropped the variable for *ACA Health Plan* and *Gay Marriage.* Apparently, *Global Warming* correlates highly with both *ACA Health*

[12]The R^2 in the binomial regression would be the square of the unconditional correlation of the last column. The descriptions in this table are short summaries; for the exact wordings, see Appendix 3.

Plan and *Gay Marriage*. It is the best single variable to use if one explanatory variable is allowed. *ACA Health Plan* and *Gay Marriage* each explain different aspects of *Conservative-self*, however, they perform better in combination than either one does with *Global Warming*. Observe also that R^2 generally increases at a decreasing rate.

Of all the identity variables, only *Birthyear,* appears in the best-10, and only as the 9th most useful variable. Apparently, group identity does not drive self-identification as a conservative. None of the top 5 variables directly involves taxes or regulation. Of course, both *ACA Health Plan* and *Global Warming* do implicate taxation and government regulation. In the best-10 regression, the t-statistics range from 10.4 to 22.7. They do not have their usual meaning, because we have selected for the variables with the largest coefficients and smallest standard errors. In effect, we have deliberately created a biased sample.

Unfortunately, stepwise does not necessarily find the best-fit regression. To see the problem, suppose we rule out identity variables a priori. *Immigration–patrol* will now replace *Birthyear* in the best-9 regression, though R^2 rounds to the same value of .4982. When the program moves from this new best-9 regression to the best-10, however, it adds *Affirmative Action* rather than *Birthyear*. In turn, this choice generates an R^2 of .5022— higher than than the best-10 selected through the original process. In effect, once stepwise starts down a given path, it misses better fitting combinations that lie off that path.

Table 5 shows the best-5 regression, redone to drop missing values rather than imputing them. This would be the appropriate equation for a sample with no missing data. It has 41,597 observations and an R^2 of .51. Table 5 shows the coefficients.

Regressor	Coefficient	Possible values
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 get special favors (*Black Favors*)	-.23	1,2,3,4,5
Abortion should be legal (*Abortion*)	-.24	1,2,3,4
Constant	2.52	1

Table 5

The Best-5 Regression for *Conservative-self*[13]

[13]The descriptions in this table are summaries; for the precise questions see Appendix 3.

For reference, here is the best-5 correlation matrix.

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

Table 6
The Best-5 Predictor Correlation Matrix

A variant is to use your favorite method to pick the best-5, but then simply average the responses rather than use the coefficients. This is crude, but in psychology it has worked, and it's another way to simplify. We have not done that here. We did, however, try using the bottom 5 of the top ten issue variables instead of the best-5. These bottom 5 were *Tax v. Spending, Mand. Birth Control Insur., Iraq Mistake, Immig–patrol,* and *Affirmative Action.* That regression has an R^2 of .39 as opposed to .47 for the best-5, a substantially worse performance.

Method 4: Best Subsets

Stepwise regression and the best-5 predictor is our preferred technique. We would prefer to use a more transparent method and one with less discretion than stepwise (both less discretion by the scholar, and less randomness in the program). The ideal would be to find, by almost brute force, what k regressors give the biggest R^2 when regressed on *Conservative-self.* ("Almost" because Hocking & Leslie (1967) noted that if (v1,v2,v3) has higher R^2 than (v2, v3, v4, v5, v6, v7) then we can throw out all the 3-variable combinations of that second set.) The idea is simple, but we have never seen it used in economics or political science, though under the name of "best subsets" regression it is well known to statisticians. In the 60's and 70's, scholars noted that maximizing the Akaike Information Criterion is with minimal assumptions consistent and efficient as a way of finding the true set of explanatory variables. The Akaike is log(estimate of variance of the error term) + penalty-function-for-adding-RHS-variables. Maximizing adjusted R^2 is consistent but not efficient, See Castle, Qin & Reed. Because we fix k (the number of explanatory variables) in the best-k regression, maximizing the Akaike is equivalent to maximizing R^2.

Even with modern computers, this concept is surprisingly hard to implement. First, we need all 5-combinations of regressors. Mathematica can compute the number, but actually listing the combinations with many regressors is a herculean task. With 30 variables the number of possible 5-combinations $30/(5((30-5))) = 142,506$ and with 40 it is $40/(5((40-5))) = 658,008$ and with 72 it is 14 million. The number of 10-combinations for 40 variables is $40/(10((40-10))) = 847,660,528$ and for 72 variables it is 536 billion. Also, we were unable to find a satisfactory computer package to do the computation, and our own programming skills are limited. Programming this would perhaps nt exceed our limited abilities, but programming it to run efficiently would, and this level of computational intensity needs a professional to squeeze out every unnecessary function call.[14]

Method 5: Lasso

"How about using LASSO?" has been the common response of econometricians when we describe this project to them. LASSO is a relatively young technique that is well known in statistics but 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 absolute values of the coefficients do not exceed a threshold level. This drives down the coefficients of some variables to zero. It also reduces the coefficients of the variables that remain in the regression; it 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.

"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 ($bias^2 + sample\ variance$). 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 are a tradeoff, accepting some bias in return for reducing sampling error. The fact that they do not maximize R^2 is a feature, not a bug; it means they depend less heavily on the particular sample you have on hand.[15]

[14] The user-created package "eleaps" in R does best subsets regression, but we found that our stepwise procedure generated a regression with higher R^2, which implies that eleaps did not in fact check all subsets. Our inquiry to the package's creator met with no response.

[15] The best-known shrinkage estimator is the use of the biased but smaller estimator $\hat{\sigma^2} = \frac{\Sigma_{i=1}^n (y-\overline{y})^2}{n+1}$ for population variance instead of the conventional unbiased $\frac{\Sigma_{i=1}^n (y-\overline{y})^2}{n-1}$ or the obvious estimator, the sample variance $\frac{\Sigma_{i=1}^n (y-\overline{y})^2}{n}$. See https://en.wikipedia.org/wiki/Shrinkage_estimator. 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

The idea of shrinkage estimators, in fact, is similar 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 "one big 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 one big regression would add random noise to the prediction— random, because the irrelevant variable's mistaken effect might be either negative or positive.

What LASSO does is to combine the ideas of variable selection and shrinkage. Or, one could use LASSO just for the variable selection, to select 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).[16]

Variable	Bound							
	.02	.20	.21	.215	.225	.23	.30	.40
Global Warming	.0186	.1333	.1389	.1409	.1438	.1451	.1594	.1765
ACA Health Plan	.0335	.3206	.3349	.3401	.3475	.3506	.3803	.4120
Gay Marriage	.0206	.3433	.3570	.3617	.3694	.3730	.4159	.4728
Abortion			-.0036	-.0063	-.0107	-.0128	-.0398	-.0771
Tax v. Spend				.0001	.0003	.0004	.0017	.0032
Mand. B. C. Ins.					-.0007	-.0010	-.0510	-.1067
Repeal ACA						-.0018	-.0415	-.0929
Black Favors							-.0175	-.0501
Affirmative Action								.0131

Table 7

LASSO Coefficients as the Size Penalty Is Relaxed

The two best-3 estimates show how a less tight bound allows the coefficients to rise in magnitude even though no new variables pass the threshold to enter the regression with non-zero

intution 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 3 or more variables with normal distributions and identical variances has lower mean squared error than the sample mean. See the original Stein & James (1961), the *Scientific American* article by Efron & Morris (1977), or the pedagogic Rasmusen (2015).

[16]See chapter 3 of Hastie, Tibshirani & Friedman for a good explanation and comparison with stepwise and best subsets regression. We used the software package R with the user-written package "lasso2" and the command "l1ce". Note that this package cannot be used when there are missing values for some variables.

coefficients. Thus, if coefficient size penalty is relaxed from .02 to .20, the coefficient on *Global Warming* rises from .0186 to .1333, but the coefficient on *Abortion* stays constant at 0. When the penalty is relaxed to .21, *Abortion* enters with a coefficient of -.0036, and the coefficient on *Global Warming* rises from .1333 to .1389.

The top five variables using LASSO are not quite the same as the stepwise best-5. We will use the coefficient estimates to generate fitted values that we will use later in our comparisons of how the various measures of conservatism correlate with *Voted for Obama*.

Discussion of the Results

Figure 2 shows histograms of three measures of conservatism. The first is *Conservative-self*. The second and third are the fitted values from the best-5 and best-10 regressions— that is, conservatism as measured using the equations from the regressions based on 5 variables and on 10. We denote these fitted-value variables as *Conservative-5* and *Conservative-10*. The fitted values lack the peaks in the center and at the right, and have a mode at the far left (for best-5) and the moderate left (for best-10). We speculate that this is because people do not like to label themselves as "liberal" even if they take the issue positions that they attribute to liberals. If this is true, then self-identified ideology is not as good a measure of a person's ideology as asking them about a few issues and weighting their responses. Also, it seems that the distribution of American beliefs about these issues is more evenly distributed than one might think.

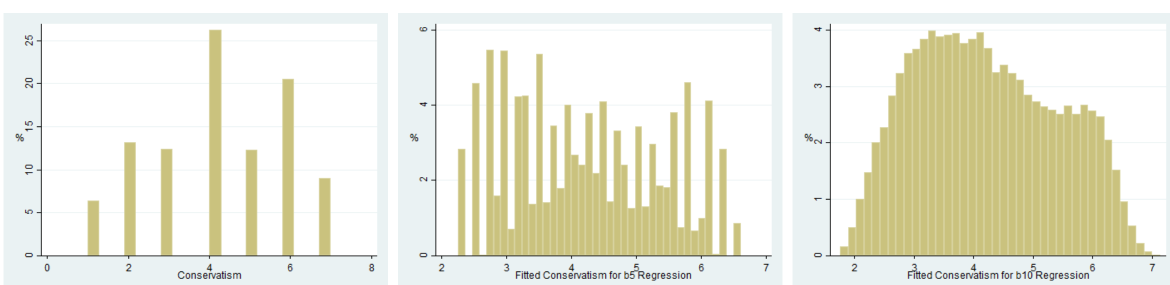

FIGURE 2
DISTRIBUTIONS OF THREE MEASURES OF CONSERVATISM[17]

How closely do the variables predict whether a respondent voted for Obama in 2012? That is shown by Table 8. The correlations range from .59 (R^2=.35 in a univariate regression) for *Democrat* to -.80 (R^2=.64) for the fitted value from the monster regression using all the issue and all the identity variables. Notably, our best-5 fitted value (*Conservative-5*) has a correlation of .78,

[17] These percentages are adjusted for sampling weights. The survey oversamples certain groups in order to get a representative sample overall.

compared with .79 for the best-10 predictor. The correlation is higher than for party affiliation, because *Democrat* has a correlation of just .59 and *Republican* of -.66 . Probably the correlation coefficients on the *Democrat* and *Republican* variables are low because they cannot predict the vote of an Independent, unlike *Conservative-5*.[18] The correlation coefficient on the best-5 fitted value is also higher than the coefficient on self-reported conservatism, *Conservative-self*, which is -.68 (R^2=.46). Note that *Conservative-factor* (the latent value calculated from factor analysis) has a correlation of -.8004, better than the -.7988 of the fitted value from the conservatism monster regression, but worse than the fitted value when only the issue variables are used (.8036).

Conservative-lasso, the predictor found using lasso and using the lasso shrinkage coefficients, has a correlation of .79 with *Voted for Obama*. *Conservative-average*, which takes the best-5 variables but simply adds them up rather than using the regression coefficients, has a correlation of .76, almost as good as *Conservative-5*. We conclude that while it is important to choose the right variables for a conservatism index, it does not matter much if all of them are weighted equally instead of using carefully estimated regression coefficients. This is significant, since for back-of-the-envelope calculations the unweighted average is much easier to compute.[19]

	Obama	Con-self	Con-5	Con-10	Con-factor	Con-lasso	Rep	Dem
Voted for Obama	1.00							
Conservative-self	-.70	1.00						
Conservative-5	-.79	.76	1.00					
Conservative-10	-.81	.78	.97	1.00				
Conservative-factor	-.81	.77	.93	.95	1.00			
Conservative-lasso	-.79	.74	.97	.94	.89	1.00		
Republican	-.66	.58	.59	.59	.58	.58	1.00	
Democrat	.59	-.53	-.51	-.52	-.54	-.51	-.48	1.00

TABLE 8
CORRELATIONS BETWEEN VOTE FOR OBAMA, CONSERVATISM, AND PARTY AFFILIATION
(N=32,287)

We also tried regressing *Voted for Obama* on all the identity variables, which yields an R^2 of

[18] [xxx We could also try a three-value variable 0-1-2 with 1 for independent.]

[19] The result mirrors a well-known result in the psychology of decisionmaking that quite good decisions can be made simply by giving numerical ratings to various factors and adding them up for each alternative rather than figuring out optimal weights— much better decisions than by looking at the factors and then making a non-mechanical, subjective decision. See Robyn M. Dawes, *Rational Choice in an Uncertain World*, Harcourt Brace (1988).

just .23. Regressing it on all the issue variables directly yields R^2= .68. Regressing it on issue and identity variables both yields R^2=.69. Adding *Republican* and *Democrat* brings the R^2 up to .74, the highest of all. Table 9 shows the R^2 resulting from various regressions of *Voted for Obama* (all including an intercept), the conservatism measures of Table 8, a few other measures, and larger sets of variables.

Explanatory variables	R^2
Conservative-self	.49
Conservative-5	.64
Conservative-10	.62
Conservative-factor	.64
Conservative-lasso	.62
Conservative-issues+identity	.64
Conservative-issues	.64
Conservative-identity	.50
Conservative-average	.58
Republican	.44
Democrat	.35
Issue, Identification, Party variables	.74
Issue, Identification variables	.69
Issue variables	.68
Identity variables	.23

TABLE 9
R^2 FROM REGRESSIONS OF OBAMA VOTE ON OTHER VARIABLES

For estimation, linear regression is not appropriate when the dependent variable is Yes/No as it is with *Obama Vote*. But we have more intuition for R^2 than for the likelihood value that is generated by a logit regression. Table 7.5 shows goodness of fit measures from logit regressions of*Obama Vote* using different single explanatory variables.

Explanatory variables	Pseudo-R^2	Log Pseudolikelihood
Conservative-self	.42	-15,039
Conservative-5	.56	-11,755
Conservative-10	.59	-10,865
Conservative-factor	.64	-9,647
Conservative-issues+identity	.63	-9,918
Conservative-issues	.63	-9,763
Conservative-identity	.14	-22,856
Conservative-average	.52	-12,478
Conservative-lasso	.56	-10,174
Republican	.36	-17,141
Democrat	.28	-19,119
Issue, Identification, Party variables	.76	-6,284
Issue, Identification variables	.70	-7,932
Issue variables	.68	-8,561
Identity variables	.20	-21,518

Table 10
Logit Regressions of Obama Vote on Other Variables

Our conclusion from the results of these various specification and measures is that issues are a good predictor of whether someone voted for Obama and that 5 variables are enough for a reasonably good fit. 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— and it is simpler. The reader can examine Tables 9 and 10 for himself and decide what tradeoff between explanatory power, complexity, and parsimony suits his own preferences.

Further Analysis of the Best-5 Regression

Tuesday lunch people: If you voted, which one of these would you vote for as the most promising, and which as the least promising extension?

This section would have lots of fun subsections delving into the relation between the Best-5 and the other variables. Some things we might do:

(a) Test the Top 5 on 20%-size samples of the data and see how well it fits, as a robustness check. Use the entire stepwise procedure on 20% samples and see what comes out.

(b) Look at the predicted values of conservatism. What kind of observations have big residuals in each direction? How do those residuals correlate with the other variables like *Male*? This would tell us about how some groups of people classify themselves as conservative even though everyone else would call them liberal.

(c) In the same vein as (b), look at subsets of the data based on the control variables such as *Male* and use our regression method to find the top 5 for each group of people. Do men view conservatism differently than women?

(d) Take issues that a priori one would think should be in the Top 5 but aren't. See how they correlate with *Conservative-5* and discuss why they aren't in the Top 5.

(e) Use the Best-5 method to figure out if Liberals mean different things by conservative than conservatives do. Maybe do this by first dropping all the 1,2 liberals from the sample and finding the top 5, then dropping all the 6,7 conservatives and finding a different (mabye) top 5. Also see what the wishy washy middle think— drop all the 3,4,5 and find the TOP 5. (related to (e). We played with that a little and got much higher R^2 as a result. Then run it ONLY using the middle-of-the-roaders.

(h) Find out the average ideology in different subpopulations by region, using the best-5 equation for the whole country.

(i) See how conservative Americans are relative to Europeans by this measure, by trying to find similar issue questions on European surveys.

(j) Do a survey of Harvard Law, Harvard Econ, or some other faculty or student group to see how they compare with Americans as a whole.

Conclusion

We hope we have demonstrated a number of methodological and substantive points:

1. Linear regression with specification selection is a better way to measure conservatism than factor analysis. It is more transparent and does not require as many survey questions.

2. Using just five issues to measure conservatism is almost as good as using 37 issue variables plus 17 identity variables. The R^2 using 5 variables is .47, compared to .50 with 10 (Table 4) and .53 with 54.

3. On average, Americans think social issues define liberal vs. conservative better than economic or foreign policy issues. The top five variables we found were Global Warming, Abortion, Black Favors, Gay Marriage, and ACA Health Plan, of which only ACA Health Plan might be considered an economic issue (Table 4). Of the top ten variables, only Tax v. Spending was an economic issue.

4. Americans do not pick their ideology based directly by identity group, although they may do so indirectly by letting their identity group determine their stance on issues. The only identity variable in the top 10 was Birthyear, at number 9 (Table 4).

5. The spread from liberal to conservative is not skewed towards conservatives, as consideration of self-identification on the political spectrum would suggest. In fact, the peak in Figure 2 is on the left, with a gradual decline in density towards the right. Moreover, there is a continuum, with no clear division in the population between liberals and conservatives.

6. Conservatism as measured by the top 5 variables is a good predictor of whether a person voted for Barack Obama in 2012, with an R^2 of .64 in a bivariate regression. That compares with .49 for self-identified conservatism, .35 for affiliation with the Democratic Party, .64 from conservatism measured using all 54 variables or from factor analysis and .74 from a multivariate regression with all issue, identity variables, and party variables (Table 10).

Appendix I–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

Appendix II–Identity Variables [20]

Issue	CCES origin	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

[20] *[Steve A. suggests adding a suburban/reigion interaction.]*

Appendix A3
The Phrasing of the Best-10 Survey Questions

Global Warming (cc321). "From what you know about global Global Warming or global warming, which one of the following statements comes closest to your opinion?" 6 responses, each a descriptive sentence. 1= most concern about global warming.

Gay Marriage (cc326) . "Do you favor or oppose allowing gays and lesbians to marry legally." Yes or no. 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 7 bills.

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."[21] Yes or no. Yes = 1.

Black Favors (cc422a). "The Irish, Italians, Jews, and many other minorities overcame prejudice and worked their way up. Blacks should do the same without any special favors." 5 possible answers. 1= strongly agree.

Abortion (cc324). "There has been some discussion about abortion during recent years. Which one of the opinions on this page best agrees with your view on this issue?" 4 responses, 1=ban it completely

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." Responses are 0 (all from tax increases) to 100, integers.

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." 2 possible responses.

Iraq Mistake (cc305). "All things considered do you think it was a mistake to invade Iraq?" Yes or no. Yes = 1.

Birthyear (birthyr). "In what year were you born?" Any integer.

Immig-patrol (cc322_2). "Increase the number of border patrols on the US-Mexican border." Yes or no. Yes = 1.

---

[21] This question is useful even though the ACA doesn't allow people to keep their current provider, doesn't have a national health insurance option, and is mostly paid for by mandating that individuals and employers purchase health insurance (though individual subsidies and Medicaid expansion will presumably be paid for by tax increases, if not by borrowing).

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): 521547.

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," *Am. Behav. Scientist,* 56: 1631–1653.

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.

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).

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)

De La Jara, Rodrigo (2006) "IQ Percentile and Rarity Chart," *IQ Comparison Site*, http://www.iqcomparisonsite.com/iqtable.aspx (2006).

DiStefano, Christine, Min Zhu & Diana Mindrila (2009) "Understanding and Using Factor Scores: Considerations for the Applied Researcher," *Practical Assessment, Research & Evaluation*, 14(20): 1-11 (October 2009).

Efron, Bradley & Carl N. Morris (1977) "Stein's Paradox in Statistics," *Scientific American*, 236(5): 119127.

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.

Half Sigma (2011) "A Word about Wordsum," *Half Sigma* blog, http://halfsigma.typepad.com/half_sigma/2011/07/a-word-about-wordsum.html (July 21, 2011).

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) *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,” *Soc. Forces,* 71: 195-210.

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: 203212.

Rasmusen, Eric “Shrinkage Estimators: The Zero, The Oracle, and The James-Stein,” working paper, http://www.rasmusen.org/papers/shrinkage-rasmusen.pdf, (June 9, 2015).

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.*

Sanjay Srivastava (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," *Am. Pol. Res.,* 37: 1048-1087.

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.

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 Gangi (2008) "Conflict, Fusion, or Coexistence? The Complexity of Contemporary American Conservatism," *Political Behavior*, 30:2 (June 2008).