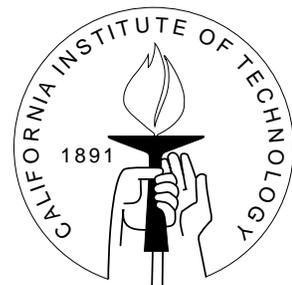# AN EMPIRICAL BAYES APPROACH TO ESTIMATING ORDINAL TREATMENT EFFECTS

R. Michael Alvarez
California Institute of Technology

Delia Bailey
Washington University in St. Louis

Jonathan N. Katz
California Institute of Technology

An Empirical Bayes Approach to Estimating Ordinal Treatment Effects

R. Michael Alvarez   Delia Bailey   Jonathan N. Katz

Abstract

Ordinal variables — categorical variables with a defined order to the categories, but without equal spacing between them — are frequently used in social science applications. Although a good deal of research exists on the proper modeling of ordinal response variables, there is not a clear directive as to how to model ordinal treatment variables. The usual approaches found in the literature for using ordinal treatment variables are either to use fully unconstrained, though additive, ordinal group indicators or to use a numeric predictor constrained to be continuous. Generalized additive models are a useful exception to these assumptions (Beck and Jackman 1998). In contrast to the generalized additive modeling approach, we propose the use of a Bayesian shrinkage estimator to model ordinal treatment variables. The estimator we discuss in this paper allows the model to contain both individual group level indicators and a continuous predictor. In contrast to traditionally used shrinkage models that pull the data toward a common mean, we use a linear model as the basis. Thus, each individual effect can be arbitrary, but the model "shrinks" the estimates toward a linear ordinal framework according to the data. We demonstrate the estimator on two political science examples: the impact of voter identification requirements on turnout (Alvarez, Bailey, and Katz 2007), and the impact of the frequency of religious service attendance on the liberality of abortion attitudes (e.g., Singh and Leahy 1978, Tedrow and Mahoney 1979, Combs and Welch 1982).

1. INTRODUCTION

Ordinal variables — categorical variables with a defined order to the categories, but without equal spacing between them — are commonly used in social science empirical applications. There has been published research that discusses how to appropriately model ordinal response variables, for example, by the use of ordered discrete choice models (e.g., Aitchison and Silvey 1957; McKelvey and Zavonia 1975; Aldrich and Cnudde 1975; Maddala 1983; Agresti 1990; Jones and Westerland 2006). However, researchers frequently use ordinal treatment variables: for example, studies have investigated the impact of partisan identification on vote choice (Miller 1991), or the role of education and the closing date of voter registration on voter turnout (Nagler 1991), or more recently how the degree of local flooding caused by Hurricane Katrina affected voter turnout in the 2006 mayoral elections in New Orleans (Alvarez et al. 2008). Despite the amount of statistical research on methods for modeling ordinal response variables, there is not a clear directive as to how to model ordinal treatment variables.

When faced with the use of an ordinal treatment variable in some statistical model, applied researchers have resorted to many different approaches. Sometimes researchers will simply use the ordinal variable itself as a response variable, thus implicitly using an unconstrained and additive treatment variable. In other situations, the researcher may avoid the linearity assumption by using subcategories of the original ordinal indicator as a vector of treatment variables, though typically in such applications the researcher will still assume additivity. These are but some of the more common ways applied researchers will deal with ordinal treatment variables in statistical models.

A good deal of the confusion regarding how to use an ordinal treatment variable in any particular statistical model arises from measurement problems. Many ordinal variables in social science have a latent structure that may be interval or even continuous that the measurement instrument fails to capture. This may be due to poor measurement design, such as in the case of survey questions, or perhaps due to sparse data. In other cases, a lack of strong theory may make it difficult for the researcher to specify how an ordinal treatment should be used in any particular application.

An important example of how applied researchers often proceed is found in the seminal paper on voter turnout in the United States, Rosenstone and Wolfinger's study (1978). In their paper, they study the effect of demographic and institutional covariates on probability that individuals in the 1972 Current Population Survey reported turning out to vote in that year's presidential election. One of the important demographic covariates in their analysis is education; their research pointed out the importance of education levels in predicting individual-level participation. Rosenstone and Wolfinger treat each respondent's reported level of education as an ordinal variable, coding their education covariate as 1 for those who report 0 to 4 years of education, 2 for those who report 5 to 7 years of education, and so on. Subsequent work that has examined directly the effect of education on voter turnout has continued using this ordinal classification scheme (e.g.,

Nagler 1991).

In this context, however, it is also important to note that because of the lack of theory or a solid measurement approach, some of the common ways applied researchers use ordinal treatment variables may be incorrect — and they thus may make incorrect inferences from the results of their models. For example, some survey questions will ask respondents which option, from a list, the respondent might prefer; while the designer of the survey and the eventual user of the survey data might assume that respondents necessarily perceive that the response options have some sort of natural ordering (thus allowing their use as an ordinal treatment variable), the respondents to the question might not see the response options in that assumed order in practice. If that is the case, then using the ordinal indicator itself as a linear and additive treatment variable might lead to incorrect inferences.

Generalized additive models (GAM) are the exception to these practices, however, and adding nonlinearities to the classical constrained model can be useful.[1] Our approach differs from both conventional ad hoc assumptions, and the GAM appoach, as we use a Bayesian shrinkage estimator as a modeling option for sparse, ordinal independent variables. A major advantage of our empirical Bayes estimator is that it allows the model to contain both individual group level indicators and a continuous predictor. Thus, each individual effect can be arbitrary, but the model "shrinks" the estimates toward a constrained ordinal framework according to the data.[2] We present the estimator in detail in the following section, and then below we present two examples from contemporary research to show the practical application of our approach for dealing with ordinal treatment variables.

2. THE MODEL

Consider a typical regression setting (using the standard generalized linear model framework), where we wish to uncover the effect of an ordinal treatment variable on a dependent variable controlling for other observable covariates. Then, for a dependent variable Y, a link function $g(\cdot)$, and observables X, we have:

$$\mathrm{E}[Y_i|X] = g^{-1}(\alpha_{j[i]} + X_i\beta)$$

where $i = 1, \ldots, N$ indexes observations and $\alpha_{j[i]}$ represents the effect on Y_i of an ordinal variable of interest, T_i with $j = 1, \ldots, J$ categories.[3] One approach to estimating $\alpha_{j[i]}$ is

---

[1]See Beck and Jackman (1998) for an exposition of this approach for political science examples.

[2]In contrast to traditionally used shrinkage models that pull the data toward a common mean, in each of our examples, we use a linear model as the basis.The methodology is flexible, however, and the group-level indicators can be pooled toward any monotonic function of the data. For an interesting non-political science example of multilevel regression with an ordinal explanatory variable, see Gelman and Hill 2006, (Section 21.3).

[3]For a standard linear regression, the link function is the identity function and we can simplify the expression to be

$$\mathrm{E}[Y_i|X] = \alpha_{j[i]} + X_i\beta$$

to completely pool the levels and to assume a constant additive effect for each level of the treatment variable:

$$\alpha^P_{j[i]} = T_i\alpha.$$

A second approach is completely unpooled: the regression includes separate indicator variables for each level of the ordinal variable:

$$\alpha^U_{j[i]} = \mathbb{I}\{T_i = j\}.$$

In the classical regression framework, it is generally not possible to combine these two approaches. In a Bayesian setting, however, the effect of T_i on Y_i can be constrained to have a common mean that is a monotonic function of the levels j, but with individual deviations — or random effects — at each level. The size of the deviations is determined by the data. Intuitively, what the estimator is doing is performing a weighted average of the pooled and unpooled models above, with the weights being proportional to the data. The weighted average, then, is:

$$\alpha_{j[i]} = \omega\alpha^P_{j[i]} + (1 - \omega)\alpha^U_{j[i]},$$

$$\omega = \left(\frac{\sigma^2_{\alpha^P}}{\sigma^2_{\alpha^P} + \sigma^2_{\alpha^U}}\right)^{-1}$$

where the weights, ω and $(1 - \omega)$ are the relative precisions of the pooled and unpooled $\alpha_{j[i]}$.

In practice, however, we would estimate the model using Markov Chain Monte Carlo (MCMC) methods. We would model the treatment effect, $\alpha_{j[i]}$, as a random parameter drawn from a probability distribution, F, with a constant mean:

$$\alpha_{j[i]} \overset{\text{iid}}{\sim} F(\mu_j, \sigma^2_\alpha),$$

$$\mu_j = h(T_i).$$

Examples of possible forms of μ_j are linear:

$$\mu_j = \gamma_0 + \gamma_1 T_i,$$

quadratic:

$$\mu_j = \gamma_0 + \gamma_1 T_i + \gamma_2 T_i^2,$$

or logarithmic:

$$\mu_j = \ln(T_i).$$

A final consideration is the interpretation of the parameters associated with μ_j. If a trend is present in the random effects that is also present in the monotonic function, μ_j, there will be an identification problem. The random effects are generally pooled toward zero, solving part of the identification problem. But, any trend in the random

effects must be constrained to be zero as well. "Post-processing" the estimates provides a computationally convenient way to constrain the effects. For a concrete example, consider a model such that:

$$\alpha_{j[i]} = \gamma^0 + \gamma^1 T_i + \nu_j;$$
$$\nu_j \overset{\text{iid}}{\sim} N(0, \sigma_\alpha).$$

That is, the estimated treatment effects are pulled towards a linear effect with deviation being drawn from a Normal distribution. The parameters γ^0 and γ^1 are only partially identified between any linear trend that exists in the random effects, ν_j. The identification is partial, as the random effects are pooled toward zero, but with a small number of levels, convergence of the MCMC chains while constraining the random effects to have both zero mean and slope is time consuming. To solve this problem, after estimation, the data is "post-processed" to obtain finite population parameters based on the regression of α_j on T. This is equivalent to constraining the random effects to have mean zero and slope zero (Gelman and Hill, 2006).

3. EMPIRICAL EXAMPLES AND PRACTICAL APPLICATIONS

We present two examples below applying this estimator: first, the effect of religious service attendance on the liberality of attitudes toward abortion, and second, the impact of voter identification requirements on voter turnout at the polls.

3.1. *Frequency of Religious Service Attendance and Abortion Attitudes*

Researchers have investigated the determinants of attitudes toward abortion, including gender, socioeconomic status, religious affiliation, and religious service attendance (e.g., Singh and Leahy 1978, Tedrow and Mahoney 1979, Combs and Welch 1982). For this example we take up the question of the impact of the frequency of religious service attendance on liberality of abortion attitudes. The data come from the 2004 General Social Survey (GSS, NORC 2004). Among the many questions that comprise the GSS, respondents were asked their opinions on abortion issues, the frequency of their attendance at religious services, and a battery of socioeconomic and demographic profile questions. The dependent variable on which we focus is the question, "Please tell me whether or not you think it should be possible for a pregnant woman to obtain a legal abortion if the woman wants it for any reason?" which takes the values "Yes" and "No." The ordinality of the religious attendance variable is given by the response options provided during the survey. To the prompt, "How often do you attend religious services?" respondents are given the choices: "Never," "Less than once a year," "Once a year," "Several times a year," "Once a month," "Two to three times a month," "Nearly every week," "Every week," and "More than once a week." The relative frequency of responses in each category in the sample are given in Table 1.

Response	Raw Counts	Percentage
Yes	346	40.6
No	507	59.4
Total	853	100.0

Table 1: Responses to the question "Please tell me whether or not you think it should be possible for a pregnant woman to obtain a legal abortion if the woman wants it for any reason?" Source: GSS 2004.

We estimate the effect of frequency of religious service attendance on the probability of supporting access to legal abortion for any reason with a logistic regression controlling for reported years of schooling and age.[4]

$$\Pr(Y_i = 1) = \text{logit}^{-1}(\alpha_{j[i]} + \beta^1 X_i),$$
$$\text{for } j = 1, \ldots, 9; \quad i = 1, \ldots, N$$

where j indexes frequency of religious service attendance and i indexes the respondents. The variable Y_i is binary and equal to one if the respondent responded "Yes" to the abortion opinion question. The vector of covariates, X_i, includes the following:

Education: the highest year of school completed;
Age: the respondent's age in years.

As noted above, we could model the impact of the variable of interest, *Church*, as an unpooled additive effect (e.g., indicator variables for each level of frequency), or alternatively, constrain the effect to be linear. Rather than commit to either extreme, we effectively combine the first two approaches into a sort of weighted average, where the weighting variable is determined by the data:

$$\alpha_{j[i]} = \alpha^0 + \alpha^1 Church_i + \nu_j,$$
$$\nu_j \overset{\text{iid}}{\sim} N(0, \sigma_\alpha).$$

That is, for each religious attendance level, j, the estimated impact on the probability of responding "Yes" to the abortion attitude question is a random intercept term, ν_j, and is pooled toward a group linear impact, $\alpha^0 + \alpha^1 Church_i$.

The MCMC estimation is implemented with a Gibbs sampling algorithm via the statistical software *JAGS* (Plummer 2007). Independent conjugate priors are assumed for each element of β and γ and for the variances. Specifically, each β and γ is assumed to be distributed normally, with mean zero and precision parameter 0.0001. The parameters σ_y and σ_α are assumed to be uniformly distributed between 0 and 100. We let the

---

[4]We estimated an extended model also controlling for gender, marital status, and having children. Those characteristics were not significant determinants of abortion opinions. We dropped them from the model in the sake of parsimony. It does not affect the estimates of church attendance.

algorithm run for 25,000 iterations as a burn-in, then 50,000 iterations with a thinning interval of 10. We use the resulting 5,000 draws from the posterior distribution as the basis of our estimates.[5]

A final consideration is in interpretation of the γ^0 and γ^1 parameters. As mentioned above, these parameters are partially not identified because of the linear trend in the ν_j parameters. The identification is partial,however, as the ν_j parameters are pooled toward zero, but with only $J = 9$ convergence would be slow to impose the constraint. To correct for this problem, after estimation, the data is "post-processed" to obtain finite population parameters based on the regression of α_j on $Church$. This is equivalent to constraining the ν_j parameters to have mean zero and slope zero (Gelman and Hill, 2006).

Table 2 in the Appendix presents the the coefficient estimates and standard errors from a completely unpooled logit, a logistic regression constraining the effect of religious service attendance to be linear, and the estimates from applying the shrinkage estimator. As would be expected, the effects of education and age are constant across the three models. Adding an additional year of school increases the probability of a more liberal attitude toward abortion, as does increasing the respondent's age, but it is a much smaller effect. The constrained estimate of the effect of religious service attendance is -0.32 on the logit scale, which translates to approximately a 5 percentage point decrease in the probability of supporting legal access to abortion for any reason, regardless of the change in attendance level: switching from "Never" to "Less than once a year," has the same effect as changing from "Every week" to "More than once a week." The coefficient estimates presented for the unpooled logit model are the individual intercepts estimated for each level. The estimated effect from the shrinkage model is the linear effect *and* the deviations from that linear effect for each level.

Figure 1 compares the estimated probability of supporting legal access to abortion for any reason averaged over the entire sample for each level of religious service attendance. The black line denotes the linear trend, and the dashed lines are the 95% confidence interval around it. The blue circles present the estimated probability from the shrinkage model, and the blue bars are the 95% confidence intervals around those estimates. The black triangles are the estimated probabilities from the unpooled logit model, and the black bars are the 95% confidence intervals from that model. The point estimates for the unpooled logit are jittered slightly to the right for visibility. Both the unpooled logit and the shrinkage estimator suggest a stronger S-shaped probability curve than the linear constraint allows, with the suggestion that attending services "2-3 times a week" has a smaller effect on the conservatism in abortion attitudes, than does attending services "Once a month." The uncertainty about the empirical Bayes estimates is larger than under the constrained linear model, but smaller than under the completely unpooled model — as we would be expected, as the shrinkage estimator contains more information than the unpooled model, but with less constraint than just the linear form. In this case, the estimates from the shrinkage model are closer to the estimates from the model

---

[5]After examining trace plots, Geweke diagnostics, and Gelman diagnostics of parallel chains, we determined that the algorithm had indeed converged.
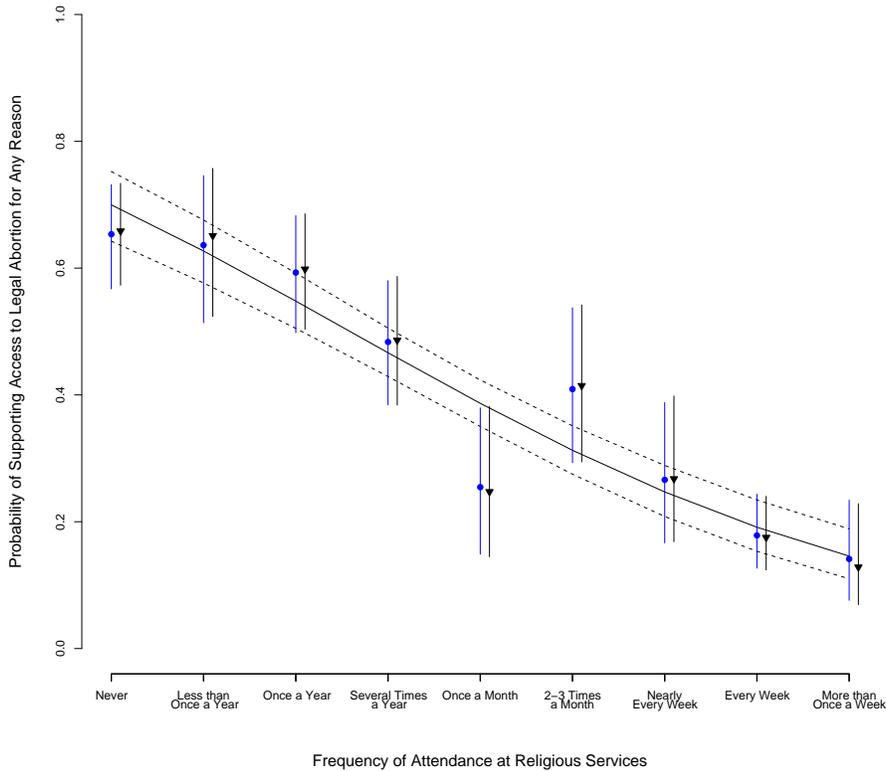
Figure 1: *Estimated probability of supporting the statement "Please tell me whether or not you think it should be possible for a pregnant woman to obtain a legal abortion if the woman wants it for any reason?" relative to the frequency of attending religious services for averaged over all respondents in the sample from the 2004 General Social Survey. The solid line is the linear trend that the identification effects are shrunk towards. The dashed lines are the 95% confidence region for the linear trend. The blue dots are the point estimates from the shrinkage estimator and the blue bars represent the 95% confidence intervals for the effect. The black triangles are the point estimates from the unpooled logistic regression and the black bars represent the 95% confidence intervals for the effect. The unpooled point estimates are jittered to the right for visibility.*

including indicator variables for each level of treatment in the model. In our next example, the estimates shrink much more closely back to the linear trend, and the uncertainty estimates shrink accordingly as well.

3.2. *Voter Identification Laws in the States*

Our second example comes from our own research. In order to document the effect of voter identification requirements on registered voters as they were imposed in states in

the 2000 and 2004 presidential elections, and in the 2002 and 2006 midterm elections, we use four election cycles and individual responses to the Current Population Surveys allows us to isolate the effect of voter identification requirements on voter turnout (Alvarez et al. 2007). Despite the recent decision by the U.S. Supreme Court in the *Crawford v. Marion County Election Board* case, there have only been a few unpublished studies of how different voter identification regulations might influence voter participation.[6] The state-level panel data allows us to control for changes in the electoral environment both across states and across time — which we could not do with only one year of data — and the individual-level data allows us to answer questions about whether certain sub-populations are disproportionately effected by these regulations — which is not possible using aggregate data.

As a starting point for our analysis, we developed a classification scheme for the different voter identification regimes that exist in the United States. Since the enactment of HAVA, there are eight basic types of requirements to vote *at the polls*. They are in listed in order of increasing stringency:

1. Voter must state his/her name.
2. Voter must sign his/her name in a poll-book.
3. Voter must sign his/her name in a poll-book and it must match a signature on file.
4. Voter is requested to present proof of identification or voter registration card.[7]
5. Voter must present proof of identification or voter registration card.[8]
6. Voter must present proof of identification and his/her signature must match the signature on the identification provided.
7. Voter is requested to present photo identification.[9]
8. Voter is required to present photo identification.

Combinations of the above requirements are often in place, such as requiring a voter to both state *and* sign his/her name. In our analysis, cases are coded at the level of requirement that is more stringent. In this example, the case would be coded as a signature requirement. Most states in 2004 required that first-time voters who registered by mail to present identification (per HAVA requirements), but here we are interested in the effect of requirements on all registered voters.

Thus, we want to measure the extent to which voter identification requirements affected voter participation at the polls, but there are many methodological problems

---

[6]For some of the unpublished literature, see Atkeson et al. (2007); Barreto, Nuno and Sanchez (2007); Lott (2006); Mycoff, Wagner and Wilson (2007); and Vercellotti and Anderson (2006).

[7]An affidavit may be signed in lieu of presenting identification and a regular (non-provisional) ballot may still be cast.

[8]The range of acceptable proof of identification ranges across the states, but in addition to a form of government-issued photo identification, other acceptable pieces of identification include utility bills, social security cards, student identification cards, paychecks, and bank statements, as well as hunting and fishing licenses and gun permits.

[9]An affidavit may be signed in lieu of presenting photo identification and a regular (non-provisional) ballot may still be cast.

unique to this data, one of which is the ordinality of voter identification requirements. As is apparent from the listing of the types of regimes, it is not the case that a state either requires identification to vote, or does not. States require many different levels of identification from simply stating one's name to showing a picture identification. This further complicates the question, as we must determine not just one effect but several potentially incremental effects. Second, states may differ in their implementation of similar requirements. While one state may consider a student identification card or discount club membership card to be valid photo identification, another state may only recognize government-issued photo identification cards. Third, the data we have to answer this question is relatively sparse. That is, since the changes in voter identification requirements have really only started since the passage of HAVA in 2002 and the law we are most interested in — photo identification requirements — was only implemented in 2006, we have only a small amount of information in the available data about how each type of voter identification requirement might affect participation. Finally, identification requirements are not randomly assigned across states. This is a problem if states with historically lower turnout also tend to adopt stricter identification requirements, then we will have trouble isolating whether the low level of turnout is due to the identification requirement or to other factors that lead a given state to have lower turnout rates.

The estimation strategy used exploits the temporal and geographic variability in voter identification requirements to sidestep the problem on non-random assignment. This is referred to as a difference-in-differences estimator and our analysis is built on a generalization of this procedure. In particular, we use a multilevel model — also referred to as a random effects model — to assess how voter identification requirements affect participation by registered voters, using data from four years of recent CPS Voter Supplement data. While multilevel models have seen many applications in fields outside of political science, only in relatively recent years have we seen the use of multilevel models in political science applications and journals (e.g., Steenbergen and Jones 2002; Raudenbush and Bryk 2002; Western 1998).[10] The multilevel model allows us to control for the constant factors that cause turnout rates to vary within states and for the cyclical changes in turnout over time.

In addition to using a much richer dataset than previous studies with a generalization of a difference-in-differences estimator to minimize the problem of non-random assignment, we also attempt handle the sparse and ordinal nature of the data. The data is sparse because with eight different types of identification requirements and only fifty states, we do not observe that many elections under a given type of procedure. The standard approach around this problem is to assume some sort of linear (or other parametric) effect. That is, if we consider our list presented at the beginning of the section, we would assume that the effect of a signature match was three times that of merely stating one's name on an individual's probability of voting, since it is third on the list. While the ordering of the list seems plausible, the linear growth (or dose-response curve) is a very

---

[10]More recently, a special issue of *Political Analysis* was devoted to the topic of multilevel modeling in political methodology, with applications to a wide variety of important substantive problems (Kedar and Shively 2005).

strong assumption that seems implausible. We, instead, leverage the ordinal nature of the data to allow for deviations for this linear effect insofar as the data suggest using our Bayesian shrinkage estimator. A benefit of our Bayesian estimator is its flexibility. Nesting the estimation of the ordinal effect within the hierarchical model of state and year effects on turnout is straightforward.

The ordinal variable analysis is nested within a multilevel logistic regression of turnout.[11] Because we are interested in the effect of identification requirements *at the polls* and not the various unobserved barriers to voting associated with the registration process, the estimation is conditioned on the subset of respondents who are registered to vote. Our logistic model takes the form:

$$\Pr(Y_{it} = 1) = \text{logit}^{-1}(\alpha_{j[i]} + \beta^0 + \beta^1 X_{it}),$$
$$\text{for } j = 1, \ldots, 8; \quad i = 1, \ldots, N; \text{ and } t = 1, \ldots, 4;$$

where j indexes identification regime, i indexes the respondents, and t indexes years. The variable Y_{it} is binary and equal to one if the respondent reported voting in that year's election. The variable β^0 is an intercept term. The vector of covariates, X_{it}, includes the following:

South: an indicator equal to unity if the respondent resides in a southern state;
Female: an indicator equal to unity if the respondent is female;
Education: a ordinal variable indicating the reported level of education — 'some high school,' 'high school graduate,' 'some college,' or 'college graduate';
Education^2: the squared value of *Education*;
Age: the respondent's age in years;
Age^2: the squared value of *Age*;
Income: an ordinal variable indicating the reported level of household family income that takes on 13 values — ranging from 'Less than \$5,000' to 'More than \$75,000';
Non-White: an indicator equal to unity if the respondent reported a race other than White.

This covariate vector replicates what we consider to the be canonical model of voter turnout in the literature that uses CPS Voter Supplement data (e.g., Nagler 1991).

As the level of turnout in a state may vary due to yearly shocks or regional trends,

---

[11]Estimating the empirical Bayes model nested inside a hierarchical framework on a dataset of over 260,000 observations was computationally difficult. Practical issues with memory aside, convergence on a subset of the data was obtained only after the inclusion of an additive redundant parameterization. The entire model was estimated, then, in a maximum likelihood framework, relying on the software package *lme4* for the statistical software *R* (Bates 2007; R Core Development Team 2007).

random effects are included for state and year.

$$\beta^0 = \gamma^0_{s[i]} + \gamma^1_{t[i]};$$

$$\gamma^0_{s[i]} \overset{\text{iid}}{\sim} N(0, \sigma_{\gamma_s});$$

$$\gamma^0_{t[i]} \overset{\text{iid}}{\sim} N(0, \sigma_{\gamma_t});$$

$$\text{for } s = 1, \ldots, S \text{ and } t = 1, \ldots, T.$$

That is, each individual i in state s and year t share a common intercept term, with each level of intercepts pooled toward zero and with common variance.

As noted above, we could model the impact of the variable of interest, ID, as an unpooled additive effect (e.g., indicator variables for each regime), or alternatively, constrain the effect to be linear. Rather than commit to either extreme, we effectively combine the first two approaches into a sort of weighted average, where the weighting variable is determined by the data:

$$\alpha_{j[i]} = \alpha^0 + \alpha^1 ID_{it} + \nu_j,$$

$$\nu_j \overset{\text{iid}}{\sim} N(0, \sigma_\alpha).$$

That is, for each identification requirement level, j, the estimated impact on turnout is a random intercept term, ν_j, and is pooled toward a group linear impact, $\alpha^0 + \alpha^1 ID_{it}$.

Figure 2 plots the average marginal effect of voter identification regimes on the probability that a respondent turns out to vote. The horizontal axis represents the voter identification requirements. The vertical axis plots the probability of turning out to vote. We note that the estimated probabilities are high, but recall we are looking at registered voter only and not eligible citizens, as is often done. Turnout rates about among eligible citizens is well below a half in recent elections, but in our sample of registered voters nearly 80% report turning out to vote. The line represents the probability of voting for a mean respondent in our sample, for each identification requirement being in place. The points on the graph denote the deviation from the linear trend estimated for each requirement and the vertical bars denote the 95% intervals of uncertainty around each. Interestingly, we see that the requirements for signature matching, requiring an identification card and requiring a photo identification card have a more negative effect on participation than suggested by the simple linear model. Requesting identification cards and requesting photo identification cards is less strict than suggested by the linear trend. These estimates first indicate that indeed, voter identification requirements do not have a simple linear effect on the likelihood that a voter participates. In addition, we see that the stricter requirements — requirements more than merely presenting a non-photo identification card — are significant negative burdens on voters, relative to a weaker requirement, such as merely signing a poll-book.

Figure 3 compares the estimated average marginal effect of the voter identification variable from the empirical Bayes model with estimates from a hierarchical model of turnout with no constraints on the voter identification variable.
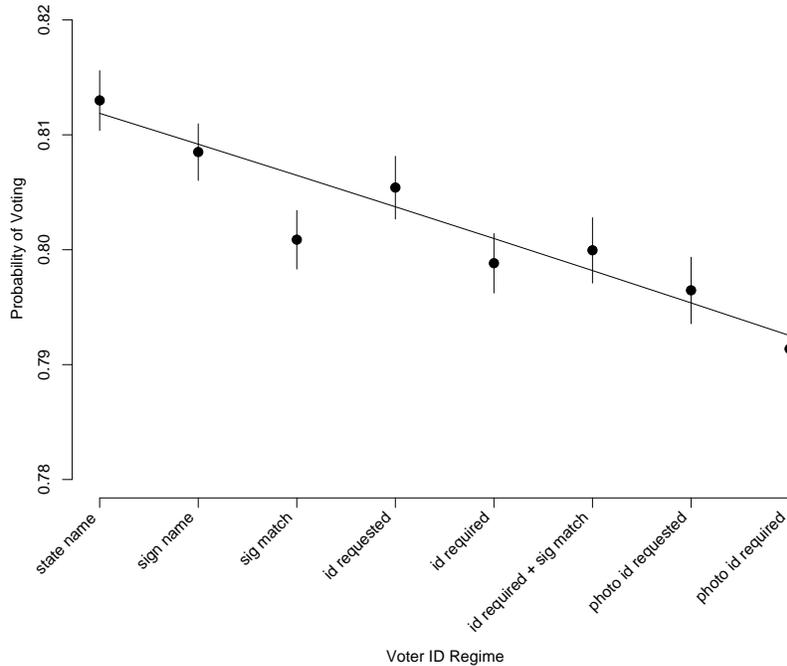
Figure 2: *Average estimated probability of voting by identification requirement. The graph plots the average impact from our sample of registered voters from the Current Population Survey (2000-2006). The estimates come from a logistic regression of the probability of voting controlling for demographic characteristics. The solid line is the linear trend that the identification effects are shrunk towards. The dots are the point estimates and the bars represent the 95% confidence intervals for the effect.*

4. CONCLUSIONS

The use of ordinal treatment variables is common in social science research. Because of measurement problems, or the lack of strong theory, researchers often have ordinal treatment indicators they wish to use in a statistical model, but they do not have any good advice about how to incorporate these ordinal treatment indicators into their analysis. In this paper we have presented an empirical Bayes estimator that gives researchers a flexible tool with which they can estimate the appropriate functional form for their ordinal treatment variables.

In sum, the empirical Bayes estimator that we presented above allows the model to contain both individual group level indicators and a continuous predictor. In contrast to traditionally used shrinkage models that pull the data toward a common mean, we use a linear model as the basis. Thus, each individual effect can be arbitrary, but the model "shrink" the estimates toward a linear ordinal framework according to the data. We see this as a simple and flexible way for researchers to specify how an ordinal treatment
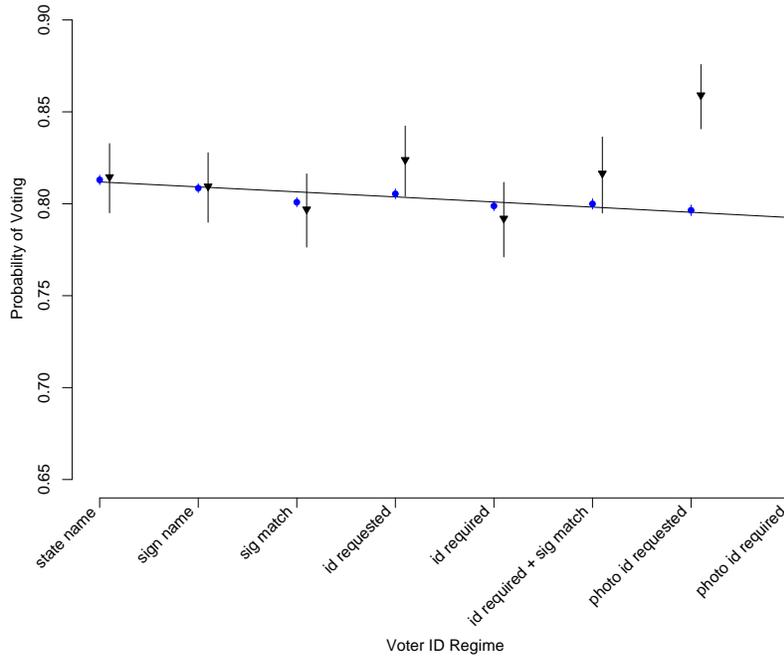
13

Figure 3: *Average estimated probability of voting by identification requirement. The graph plots the average impact from our sample of registered voters from the Current Population Survey (2000-2006). The estimates come from a logistic regression of the probability of voting controlling for demographic characteristics. The solid line is the linear trend that the identification effects are shrunk towards. The blue dots are the point estimates from the shrinkage estimator and the blue bars represent the 95% confidence intervals for the effect. The black triangles are the point estimates from the unpooled logistic regression and the black bars represent the 95% confidence intervals for the effect. The unpooled point estimates are jittered to the right for visibility.*

variable should be used in their particular response model.

In addition to developing this model in this paper, we presented two empirical applications of our empirical Bayes estimator, using examples drawn from contemporary research problems. One of these problems involved an ordinal treatment variable drawn from a widely used survey (abortion preferences from the GSS), while the other focused on a variable that measured a dimension of state election administration policies (voter identification policies). We showed with each application the utility of our approach.

We believe that our methodology should have widespread utility in many other areas of social science research. Ordinal treatment variables are widely used in comparative political studies, and of course in other social science disciplines. Using a flexible strategy for dealing with ordinal treatment variables, like we develop in this paper, should be adopted instead of some of the more rigid and assumption-driven practices seen in the

applied literature today.

5. REFERENCES

Agresti, Alan. 1990. *Categorical Data Analysis.* New York:Wiley-Interscience.

Aldrich, J. and C. F. Cnudde. "Probing the Bounds of Conventional Wisdom: A Comparison of Regression, Probit, and Discriminant Analysis." *American Journal of Political Science*, 19, 3; 571-608.

Alvarez, R. Michael, Delia Bailey and Jonathan N. Katz. 2007. "The Effect of Voter Identification Laws on Turnout." Manuscript, California Institute of Technology and Washington University in St. Louis.

Alvarez, R. Michael, Thad E. Hall, and Betsy Sinclair. 2008. "Katrina's Voters: Floods, Representation, and Social Context." Paper presented at the 2008 Annual Meetings of the Midwest Political Science Association.

Aitchison, J. and S. D. Silvey. 1957. "The Generalization of Probit Analysis to the Case of Multiple Responses." *Biometrika*, 131-140.

Atkeson, Lonna Rae, Lisa A. Bryant, Thad E. Hall, Kyle L. Saunders, and R. Michael Alvarez. 2007. "New Barriers to Participation: Application of New Mexico's Voter Identification Law." Paper presented at the 2007 Annual Meeting of the American Political Science Association, Chicago, Illinois.

Barreto, Matt A., Stephen A. Nuno, and Gabriel R. Sanchez. 2007. "Voter ID Requirements and the Disenfranchisement of Latino, Black and Asian Voters." Paper presented at the 2007 Annual Meeting of the American Political Science Association, Chicago, Illinois.

Bates, Douglas. 2007. "lme4: Linear mixed-effects models using S4 classes." R package version 0.99875-9.

Beck, Nathaniel and Simon Jackman. 1998. "Beyond Linearity by Default: Generalized Additive Models." *American Journal of Political Science* 42(2): 596-627.

Burden, Barry C. 2001. "Voter Turnout and the National Election Studies." *Political Analysis* 8, 4, 389-398.

Combs, Michael W. and Susan Welch. 1982. "Blacks, White, and Attitudes Toward Abortion." *Public Opinion Quarterly* 46, 4, 510-520.

Electionline, 2002. "Election Reform Briefing: Voter Identification." http://electionline. org/Portals/1/Publications/Voter%20Identification.pdf

Electionline, 2004. "Election Preview 2004: What's Changed, What Hasn't and Why." http://electionline.org/Portals/1/Publications/Election.preview.2004.report.final.update.pdf

Electionline, 2006. "Election Reform Briefing 15: The 2006 Election." http://www.electionline.org/Portals/1/Publications/EB15.briefing.pdf

Gelman, Andrew and Jennifer Hill. 2006. *Data Analysis Using Regression and Multilevel/Hierarchical Models.* New York: Cambridge University Press.

Jones, Bradford S. and Chad Westerland. 2006. "Order Matters(?): Alternatives to Conventional Practices for Ordinal Categorical Variables." Paper presented at the Annual Meeting of the Midwest Political Science Association. Chicago, IL.

Kedar, Orit and W. Phillips Shively. 2005. "Introduction to the Special Issue." *Political Analysis* 13, 4, 297-300.

Lott, John R. 2006. "Evidence of Voter Fraud and the Impact that Regulations to Reduce Fraud Have on Voter Participation Rates." Available at SSRN: http://ssrn.com/abstract=925611.

Maddala, G.S. 1983. *Limited-Dependent and Qualitative Variables in Econometrics.* Cambridge:Cambridge University Press.

McKelvey, R.D. and W. Zavonia. 1975. "A Statistical Model for the Analysis of Ordinal Level Dependent Variables." *Journal of Mathematical Sociology*, 4, 103-20.

Miller, Warren E. 1991. "Party Identification, Realignment, and Party Voting: Back to the Basics." *American Political Science Review* 85, 2, 557-568.

Moffit, Robert. 1991. "Program Evaluation with Nonexperimental data."*Evaluation Review* 15, 291–314.

Mycoff, Jason D., Michael Wagner, and David Wilson. 2007. "The Effect of Voter Identification Laws on Turnout." Paper presented at the 2007 Annual Meeting of the American Political Science Association.

Nagler, Jonathan. 1991. "The Effect of Voter Registration Laws and Education on U.S. Voter Turnout." *American Political Science Review* 85, 1393-1405.

National Association of Secretaries of State. 2003. "Voter Identification Requirements by State." Survey. http://www.nass.org/index.php?option=com_docman&task=doc_download&gid=80.

National Commission on Election Reform. 2002. *To Ensure Pride and Confidence in the Electoral Process: Report of the National Commission on Election Reform.* Washington, D.C.: Brookings Institution Press.

NORC. 2004 *General Social Survey.* Survey. http://www.norc.org/projects/General+Social+Survey.htm.

R Development Core Team. 2005. *R: A language and environment for statistical computing.* Version 2.6.1. Vienna, Austria: R Foundation for Statisical Computing. http://www.R-project.org

Raudenbush, Stephen W. and Anthony S. Bryk. 2002. *Hierarchical Linear Models: Applications and Data Analysis Methods*, 2nd. edition. Newbury Park: Sage.

Rosenstone, S.J. and R.E. Wolfinger, 1978. "The Effect of Registration Laws on Voter Turnout." *American Political Science Review*, 72, 1; 22-45.

Singh, B. Kirshna and Peter J. Leahy. 1978. "Contextual and Ideological Dimensions of Attitudes toward Discretionary Abortion." *Demography* 15, 381-88.

Steenbergen, Marco R. and Bradford S. Jones. 2002. "Modeling Multilevel Data Structures." *American Journal of Political Science* 46, 1, 218-237.

Tedrow, Lucky M. and E.R. Mahoney. 1979. "Trends in Attitudes toward Abortion: 1972-1976." *Public Opinion Quarterly* 43, 181-189.

Vercellotti, Timothy and David Anderson, 2006. "Protecting the franchise, or restricting it? The effects of voter identification requirements on turnout." Manuscript, Rutgers University.

Western, Bruce. 1998. "Causal Heterogeneity in Comparative Research: A Bayesian Hierarchical Modeling Approach." *American Journal of Political Science* 42, 1233-1259.

6. APPENDIX

	Unpooled Logit		Pooled Logit		Shrinkage Logit	
	b	se	b	se	b	sd
Education	0.19	0.03	0.19	0.03	0.19	0.03
Age	0.02	0.00	0.02	0.00	0.02	0.00
Church, Slope			−0.33	0.03	−0.32	0.03
Church - Never	−2.81	0.49			−2.50	0.48
Church - Less than once a year	−2.85	0.54			−2.25	0.53
Church - Once a year	−3.06	0.49			−2.12	0.48
Church - Several times a year	−3.53	0.51			−2.24	0.50
Church - Once a month	−4.58	0.57			−2.93	0.55
Church - 2-3 times a month	−3.81	0.53			−1.91	0.52
Church - Nearly every week	−4.48	0.58			−2.23	0.56
Church - Every Week	−5.02	0.55			−2.43	0.54
Church - More than once a week	−5.38	0.62			−2.39	0.59
Intercept			−2.23	0.46		

Table 2: Estimated logit coefficients and standard errors from the completely unpooled, completely pooled, and Bayesian shrinkage models. The slope parameter estimate for the shrinkage model is the finite population slope; the deviations from the linear model are based on the mean and standard deviation of 5000 draws from the posterior distribution.