Models for Ordered Outcomes

Political Science 200C
Spring 2000
Simon Jackman

"A model is unordered if it is not ordered." (Amemyia 1985, 292).

1 Ordered Outcomes

Often dependent variables are ordinal, but are not continuous in the sense that the metric used to code the variables is substantively meaningful. For instance, it is customary to employ a 7-point scale when measuring party-identification in the U.S., assigning the numerals $\{0, \ldots, 6\}$ to the categories {"Strong Republican", "Weak Republican", ..., "Strong Democrat"}. But the metric underlying party identification is not necessarily the same as the linear metric relating the numerals zero through 6 (i.e., the real line). In substantive terms, the difference between 0 and 2 on the coded party identification scale (moving from "Strong Republican" to "Republican Leaner") may be quite different from the difference between 2 and 4 ("Republican Leaner" to "Democrat Leaner"), or 4 and 6 ("Democrat Leaner" to "Strong Democrat"). These variables are sometimes also called "polychotomous" (as opposed to "dichotomous").

When such a variable appears on the left-hand side of a statistical model it is obvious that LS regression will suffer from many of the short-comings we saw LS regression to face in the binary case: i.e., heteroskedasticity, predicted probabilities outside the unit interval, etc.

2 The Ordered Probit Model

A widely used approach to estimating models of this type is an ordered response model, which almost allows employs the probit link function. This model is thus often referred to as the "ordered probit" model. Like many models for qualitative dependent variables, this model has its origins in bio-statistics (Aitchison and Silvey 1957) but was brought into the social sciences by two political scientists (McKelvey and Zavoina 1975; both PhD candidates at the University of Rochester at the time, incidentally).

The central idea is that there is a latent continuous metric underlying the ordinal responses observed by the analyst. Thresholds partition the real line into a series of regions corresponding to the various ordinal categories. The latent continuous variable, y^* is a linear combination of some predictors, $\mathbf{x}$, plus a disturbance term that has a standard Normal distribution:

$$y_i^* = \mathbf{x}_i\boldsymbol{\beta} + e_i, \quad e_i \sim N(0,1), \ \forall\, i = 1, \ldots, N. \tag{1}$$

y_i, the observed ordinal variable, takes on values 0 through m according to the following scheme:

$$y_i = j \iff \mu_{j-1} < y_i^* \leq \mu_j,$$

where $j = 0, \ldots, m$, and by slight abuse of notation in the pursuit of completeness I define $\mu_{-1} = -\infty$, and $\mu_m = +\infty$.

Like the models for binary data, we are concerned with how changes in the predictors translate into the probability of observing a particular ordinal outcome. Consider the probabilities of each ordinal outcome:

$$\begin{aligned}
P[y_i = 0] &= P[\mu_{-1} < y_i^* \leq \mu_0], \\
&= P[-\infty < y_i^* \leq \mu_0], \\
&= P[y_i^* \leq \mu_0], \\
& \text{ substituting from (1),} \\
&= P[\mathbf{x}_i\boldsymbol{\beta} + e_i \leq \mu_0], \\
&= P[e_i \leq \mu_0 - \mathbf{x}_i\boldsymbol{\beta}], \\
&= \Phi(\mu_0 - \mathbf{x}_i\boldsymbol{\beta}); \\
P[y_i = 1] &= P[\mu_0 < y_i^* \leq \mu_1], \\
&= P[\mu_0 < \mathbf{x}_i\boldsymbol{\beta} + e_i \leq \mu_1], \\
&= P[\mu_0 - \mathbf{x}_i\boldsymbol{\beta} < e_i \leq \mu_1 - \mathbf{x}_i\boldsymbol{\beta}], \\
&= \Phi(\mu_1 - \mathbf{x}_i\boldsymbol{\beta}) - \Phi(\mu_0 - \mathbf{x}_i\boldsymbol{\beta}).
\end{aligned}$$

It is straightforward to see that

$$P[y_i = 2] = \Phi(\mu_2 - \mathbf{x}_i\boldsymbol{\beta}) - \Phi(\mu_1 - \mathbf{x}_i\boldsymbol{\beta}),$$

and that generically

$$P[y_i = j] = \Phi(\mu_j - \mathbf{x}_i\boldsymbol{\beta}) - \Phi(\mu_{j-1} - \mathbf{x}_i\boldsymbol{\beta}).$$

For $j = m$ (the "highest" category) the generic form reduces to

$$\begin{aligned} P[y_i = m] &= \Phi(\mu_m - \mathbf{x}_i\boldsymbol{\beta}) - \Phi(\mu_{m-1} - \mathbf{x}_i\boldsymbol{\beta}), \\ &= 1 - \Phi(\mu_{m-1} - \mathbf{x}_i\boldsymbol{\beta}). \end{aligned}$$

To estimate this model we use MLE, and so first we need a log-likelihood function. This is done by defining an indicator variable Z_{ij}, which equals 1 if $y_i = j$ and 0 otherwise. The log-likelihood is simply

$$\ln \mathcal{L} = \sum_{i=1}^{N} \sum_{j=0}^{m} Z_{ij} \ln[\Phi_{ij} - \Phi_{i,j-1}],$$

where $\Phi_{ij} = \Phi[\mu_j - \mathbf{x}_i\boldsymbol{\beta}]$ and $\Phi_{i,j-1} = \Phi[\mu_{j-1} - \mathbf{x}_i\boldsymbol{\beta}]$.

2.1 Identification Constraints

As it stands, optimization of this log-likelihood will not result in a unique solution. Without some constraints on $\boldsymbol{\beta}$ or the threshold parameters μ an algorithm trying to maximize the log-likelihood would endlessly circle on a "plateau" of equally-likely combinations of $\hat{\boldsymbol{\beta}}$ and μ parameters. Formally, these parameters of the model are said to be "unidentified". Intuitively, this arises because both $\boldsymbol{\beta}$ and μ are "location" parameters that calibrate the mapping from the observed predictors to the latent y_i^*. There is no unique combination of $\hat{\mu}$ and $\boldsymbol{\beta}$ that maximizes the fit to the data. Put differently, for any given $\hat{\boldsymbol{\beta}}$ there exists a $\hat{\mu}$ that produces a likelihood equal to that obtained from at least one other $\hat{\boldsymbol{\beta}}$ and $\hat{\mu}$.

To get around this problem, a number of identifying restrictions are possible (see Table 1). The most common usual identification constraint is to set $\mu_0 = 0$ (LIMDEP and SST do this by default, and this is often in the very definition of the model in some texts) or else to suppress the intercept in the model. In any event either one of the thresholds must be "anchored" *a priori* or the intercept-term dropped; we have to assume something so as to get a toe-hold in calibrating $\mathbf{x}_i\boldsymbol{\beta}$ with the latent variable y_i^*.

The other identification constraint is to do with the "dispersion" or

	$\boldsymbol{\beta}$	σ	$\boldsymbol{\mu}$
1	unconstrained	fixed e.g., $\sigma = 1$	one μ_j fixed e.g., $\mu_1 = 0$
2	drop intercept	fixed e.g., $\sigma = 1$	unconstrained
3	unconstrained e.g., see Krehbiel and Rivers (1988) or Bartels (1991)	unconstrained	two μ_j fixed

Table 1: **Ordered Probit Model, Identification Constraints.**

variance parameter, σ^2, or more technically, the standard deviation, σ. If the variance of y_i^* were also something to be estimated then the model's parameters are unidentified; even with μ_0 "anchoring" the mapping of $\mathbf{x}_i\boldsymbol{\beta}$ to y_i^*, allowing σ^2 and $\boldsymbol{\beta}$ to both be "free parameters" would also result in an infinite collection of estimates that fit the data equally well. For any candidate $\hat{\boldsymbol{\beta}}$ there is no unique scaling of y_i^* via a $\hat{\sigma^2}$ maximizing fit to the data. Setting the variance to a known constant *a priori* circumvents this problem. Standard practice is to set σ^2 to 1 rather than an arbitrary known constant, since this simplifies the Φ_{ij} terms in evaluating the log-likelihood function.

Likewise, we could identify the μ and $\boldsymbol{\beta}$ parameters a variety of ways, and as I make clear above, different implementations of this model use different approaches. Setting $\mu_0 = 0$ is some respects highly arbitrary, and done largely for programming convenience only, since it is the first threshold encountered in an ordered probit model no matter how many ordinal categories the user may pass to a computer program designed to estimate these models.

2.2 Exploiting Identification Constraints

It is important to remember that these identification constraints are nonetheless arbitrary, and in the hands of a skillful analyst this can be a useful way with which to extract substantive mileage from the results of an ordered probit model. It is sometimes possible to re-define the latent variable y_i^* as substantively meaningful quantity, such as money, votes, numbers of soldiers, hours worked, etc, and set the thresholds to cut points in terms of this metric, rather than in terms of the probit metric. In re-calibrating the thresholds one also re-calibrates the $\boldsymbol{\beta}$ terms, so that now they are interpreted as the effects of the predictors have in the units of the substantively interesting quantity.

3 Example: using the ordered probit model to estimate ideal points

To see how the ordered probit model can be exploited in this fashion, I consider how one might use the model to estimate legislator's unobserved ideal points on a policy dimension. Two recent examples exploiting this property of the data are Krehbiel and Rivers's (1988) and Bartels' (1991) analyses of sequences of votes in the US Congress. In both articles the parameter estimates of the ordered probit model are rescaled to fit a substantively meaningful metric---dollars: in the Krehbiel and Rivers case this metric corresponds with senators' ideal points on the minimum wage (dollars per hour); in Bartels' article the latent variable of interest is a representatives ideal points for defense spending (billions of dollars in defense appropriations).

These ideal points are unobserved by the analyst. For various reasons few Senators or representatives directly report their preferred level of the minimum wage, or the defense budget. Nonetheless these ideal points are of obvious political significance. Assessing the effects of public opinion on ideal points over the defense budget is one of the chief concerns of the Bartels article. In the Krehbiel and Rivers piece the focus is on less directly on the effects of constituency-specific characteristics on ideal points over the minimum wage, and more to do with a comparison of committee ideal points with ideal points in the Senate as a whole. In both cases the unobserved ideal points are posited as the latent continuous variables underlying an observed sequence of votes.

The observed votes over a sequence of related motions allows the analyst to code the legislators ordinally. This follows from some assumptions about the legislators' utility functions over the policy dimension:

1. each legislators' utility function is symmetric and single-peaked, i.e.,

$$u_i(\theta) = \phi_i(|\theta - x_i|)$$

where ϕ_i is any monotone decreasing function (a weaker assumption than the standard assumption about quadratic utility functions), θ is a policy option ($\theta \in \Omega \subset \mathbb{R}$), and x_i is the unobserved ideal point of legislator i, $i = 1, \ldots, N$;

2. voting is sincere (both Krehbiel and Rivers and Bartels are quick to discount the possibility of sophisticated voting in their respective contexts, with recourse to both theoretical arguments and close inspection of the sequence and nature of the votes under consideration).

3.1 Krehbiel and Rivers: 2 votes, known alternatives

Under these assumptions, inspecting a pattern of votes tells us something about the regions of the policy space in which legislators' ideal points lie. For instance, Krehbiel and Rivers consider a sequence of two votes, each vote a binary choice between well-specified points in a given unidimensional policy space. First, legislators vote between θ_1 and θ_3, and say θ_3 wins. Then θ_2 is voted against θ_3. With θ_1, θ_2, and θ_3 known, it is then possible to rank order the legislators in terms of where their unobserved ideal points lie in $\Omega \subset \mathbb{R}$.

To see this, note that a legislator votes for θ_1 over θ_3 in the first vote if $u_i(\theta_1) > u_i(\theta_3)$ and otherwise votes against θ_1 (in the case of the utilities being equal we assume that the legislator prefers the status quo policy, here labelled θ_1, without loss of generality.) In the second vote, a legislator votes for θ_2 over θ_3 if $u_i(\theta_2) > u_i(\theta_3)$ and otherwise votes against θ_2. Theoretically there are four possible patterns of voting here, 2^2 combinations of the "yea" and "nay" possibilities.

In general, legislator i votes for option θ_j over θ_k if and only if

$$\phi_i(|\theta_j - x_i|) > \phi_i(|\theta_k - x_i|).$$

Since ϕ_i is strictly decreasing in θ about x_i, this condition can be rewritten as

$$|\theta_k - x_i| > |\theta_j - x_i|$$

i.e., proposal θ_k is further away from legislator i's ideal point than proposal θ_j. These conditions imply the following:

$$x_i < \frac{\theta_j + \theta_k}{2} \quad \Longleftrightarrow \quad \theta_j \leq \theta_k,$$

$$x_i > \frac{\theta_j + \theta_k}{2} \iff \theta_j > \theta_k.$$

That is, if legislator i votes for θ_j over θ_k then we can infer something about where that legislator's ideal point lies conditional on knowing something about the relative positions of θ_j and θ_k in the policy space. If we assume (again, without loss of generality) that $\theta_1 < \theta_2 < \theta_3$ then we can summarize the possibilities as follows:

$$\begin{aligned}
\text{Vote 1}: \theta_1 \equiv \text{“Yea”} &\iff x_i < \frac{\theta_1 + \theta_3}{2} \\
\theta_3 \equiv \text{“Nay”} &\iff x_i \geq \frac{\theta_1 + \theta_3}{2} \\
\text{Vote 2}: \theta_2 \equiv \text{“Yea”} &\iff x_i < \frac{\theta_2 + \theta_3}{2} \\
\theta_3 \equiv \text{“Nay”} &\iff x_i \geq \frac{\theta_2 + \theta_3}{2}
\end{aligned}$$

Given the assumptions that $\theta_1 > \theta_2 > \theta_3$ and that legislators vote sincerely, it is fairly clear that the sequence {Vote 1, Vote 2} = {“Yea”, “Nay”} is theoretically impossible, since this implies

$$\frac{\theta_2 + \theta_3}{2} \leq x_i < \frac{\theta_1 + \theta_3}{2}$$

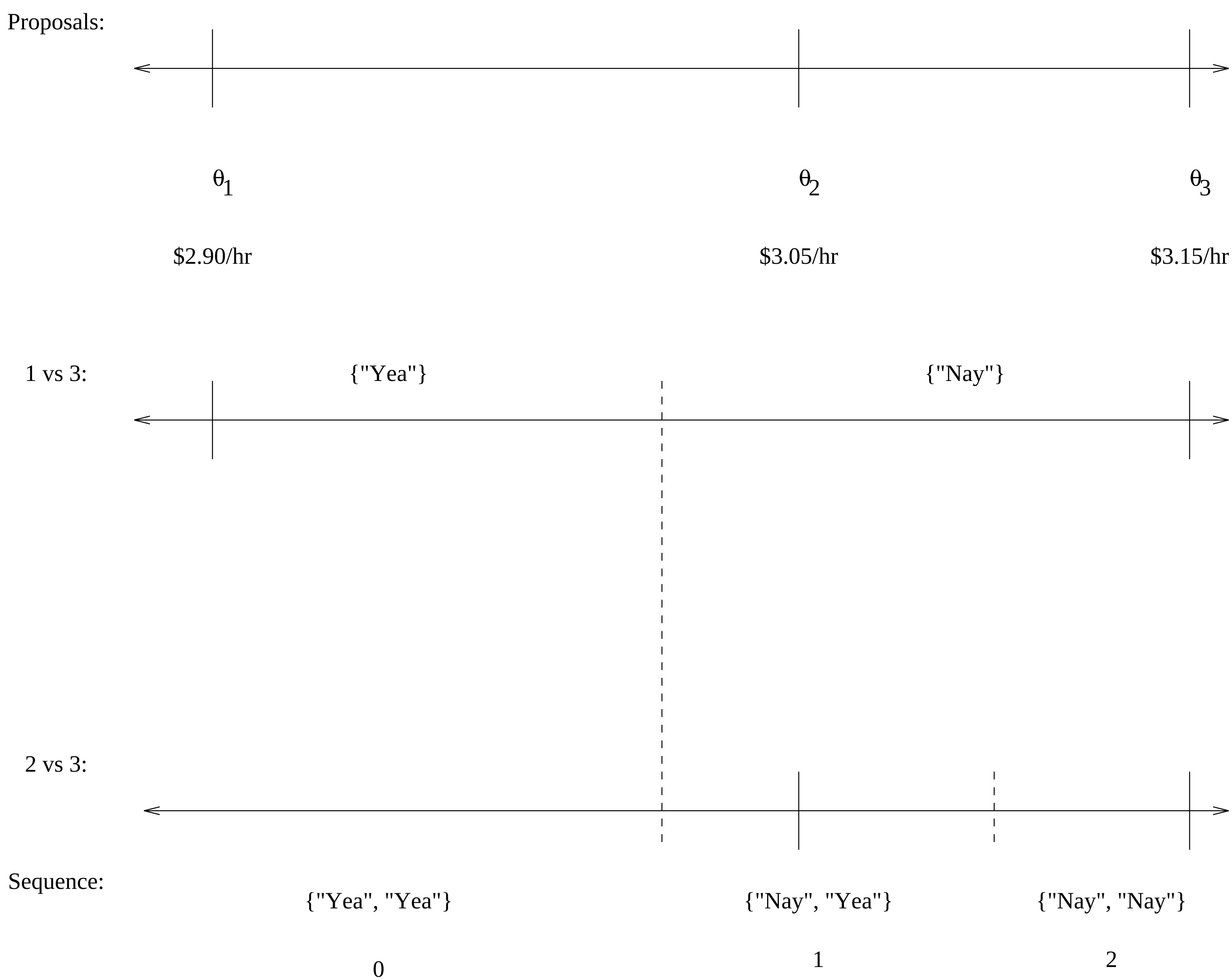
which in turn implies the contradiction $\theta_2 \leq \theta_1$.

The possibilities are summarized graphically in Figure 1. At the bottom of the figure are listed the resulting pairs of vote outcomes, corresponding to the ordinal categories forming the dependent variable for the ordered probit analysis:

Vote 1	Vote 2	Numeral
“Yea”	“Yea”	0
“Nay”	“Yea”	1
“Nay”	“Nay”	2
“Yea”	“Nay”	-

When the vote choices θ_1, θ_2, and θ_3 denote points in a policy-space that has a substantively-interpretable metric, running ordered probit with a dependent variable defined as above will result in estimates that are readily

Figure 1: ***Krehbiel and Rivers analysis of 1977 minimum wage amendments.***



interpretable in terms of that metric. For instance, the votes analyzed by Krehbiel and Rivers refer to proposed levels of the minimum wage, considered by the Senate in debating the Fair Labor Standard Amendments Act (1977, S. 1871). The θ quantities in this case are actual dollar amounts corresponding to the 1980 minimum wage: θ_1 = \$2.90 per hour (an amendment proposed by Dewey Bartlett), θ_2 = \$3.05 per hour (an amendment proposed by John Tower), and θ_3 = \$3.15 per hour (the level proposed by the Labor and Human Resources Committee, then under the chairmanship of Harrison A. Williams, and ultimately approved by the Senate, 76-14).

Given that we can associate these dollar figures with the options with the votes, it is also possible to associate dollar figures with the cut points between the categories defining the dependent variable. In this case, $\mu_0 = (\theta_1 + \theta_3)/2$ =\$3.025 per hour, and $\mu_1 = (\theta_2 + \theta_3)/2$ = \$3.10 per hour. With these exact restrictions on the μ parameters Krehbiel and Rivers are able to modify the standard identification restrictions for the ordered probit model. In particular, this exact knowledge on μ means that we can let σ^2 be a free parameter.

In practice, one could either write out and program a log-likelihood function with these identifying constraints on the μ "hard-wired" (and σ^2 a free parameter), or use a canned ordered probit routine and re-standardize the estimates to suit the constraints on the thresholds. This involves a linear transformation of the parameter estimates, noting that the difference between $\mu_0 = 0$ and $\hat{\mu}_1$ is now just the difference between \$3.025 per hour and \$3.10 per hour, or .075 of a dollar per hour. One would re-scale (i.e., multiply) the probit estimates by a constant *m* such that the transformed $\hat{\mu}_1$ equals .075, and then add 3.025: i.e.,

$$z^* = mz + c, \tag{2}$$

where *z* is a location estimate from the probit model (a threshold, a slope estimate, or an estimated ideal point), *m* is the re-scaling constant, and *c* is the location shift. Note that when re-scaling dispersion parameters like σ^2 or the standard errors of the parameter estimates one needs only the scale shift given by the multiplier *m*, and not the location shift given here by *m* = 3.025. Table 1 of the Krehbiel and Rivers article includes estimates of σ produced by this scale shift, and since $\sigma = 1$ is the typical parameterization, the scaling constant *m* is simply the estimate of σ in Table 1. That is, Krehbiel and Rivers could have obtained the estimates in Table 1 by taking the usual probit output, multiplying the estimates by .105, and adding 3.025 to the result for location parameters (slope parameters and the threshold parameters).

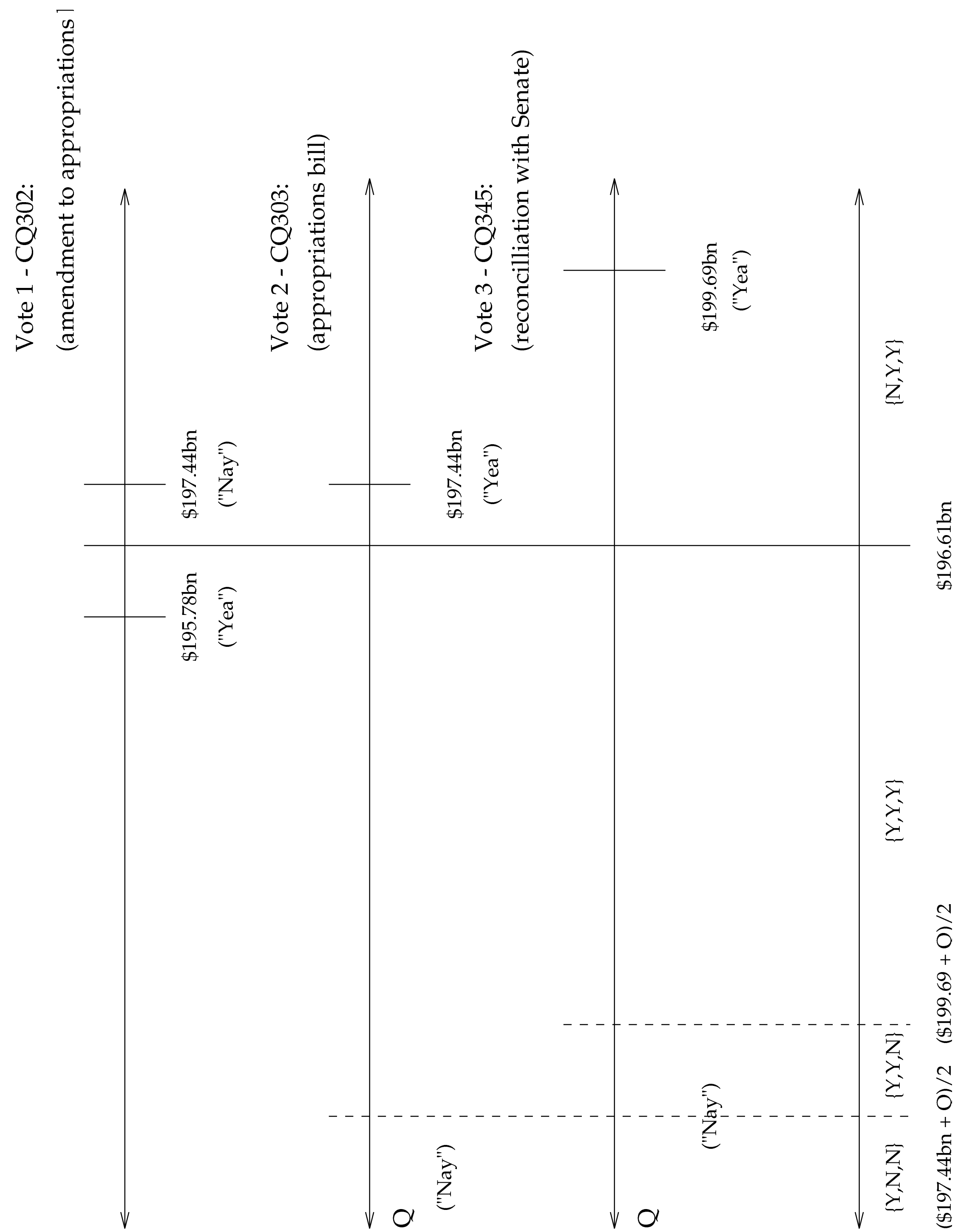
3.2 Bartels: three votes, same unknown alternative on two

The situation studied by Krehbiel and Rivers seems the exception to the rule. Only under fairly specialized circumstances are legislative options or responses to a survey precisely defined in terms of a substantively interesting metric. It is often the case that legislators choose between some specific quantity and an unknown or yet-to-be decided reversion point. Or in the survey research context, responses of the form "Strong Democrat" or "Weakly Opposed" etc can be ordered but it is unclear what the thresholds between these categories correspond to in terms of an underlying, substantively-meaningful metric.

The Bartels application has some of these features. Bartels analyses a sequence of three votes in the House of Representatives in November and December 1981 on the increased defense expenditures President Reagan requested in his first budget. However, unlike the situation analyzed by Krehbiel and Rivers the cut-points between the votes are not clearly defined. Instead, legislators choose between an alternative stated in dollar figures, and an unknown reversion point, Q. In two of the three votes, the cut-point for two of the three votes is known only as the midpoint between \$197.44 billion and Q, which is just \$98.72+Q/2 billion (for CQ 303, a vote on the appropriations bill) and as the midpoint between \$199.69 billion (the figure reported after conference with the Senate) and Q, which is just \$99.85+Q/2 billion (for CQ 345). The first vote in the sequence, CQ 302, was a vote on an amendment seeking to cut the funds appropriated for weapons procurement and research and development by 2%, and so the cut-point for this vote is known *a priori* to be \$196.61 billion, (195.78 + 197.44)/2 billion dollars, again subject to the fairly uncontroversial assumptions about the legislators' utility functions and sincere voting. It is straightforward to order these cut-points as well, which allows us to partition the real line into categories suitable for analyzing legislator's votes through the sequence via ordered probit.

Figure 2 depicts the possibilities in this instance. The two cut-points that depend on Q are depicted with dotted lines, while the known cut-point from the first vote is shown with a solid line. The reversion point Q is somewhere to the left of the graphed area, constant over the three votes, but unknown *a*

Figure 2: ***Bartels' analysis of 1981 defense appropriation votes.***



priori. The four sequences of votes are shown at the bottom of the figure, with the leftmost sequence {"Yea", "Nay", "Nay"} corresponding to the lowest category in the ordered probit analysis. Bartels reports that of the 108 legislators in his sample (those legislators whose districts were included in the sampling frame for the 1980 National Election Study), none exhibited a pattern of votes inconsistent with an assumption of fixed preferences, symmetric, single-peaked utility functions, and sincere voting. In other words, the four patterns of voting listed at the bottom of Figure 2 are the only logical possibilities under these assumptions and were the only patterns observed in the data.

3.3 Re-calibrating estimates from the ordered probit model

The dependent variable in the ordered probit takes on the values 0 through 3, corresponding to the four categories at the bottom of Figure 2. Following standard practice, the first probit threshold is set to 0 for identification purposes, and σ^2 to 1 (see Table A2, at p470 of the Bartels article). Bartels obtains the following estimates of the other thresholds, against which I list the corresponding values in terms of known dollar amounts, and/or the reversion level Q:

Threshold	Probit	\$bn
μ_0	0	$98.72 + Q/2$
$\hat{\mu}_1$	·183	$99.85 + Q/2$
$\hat{\mu}_2$	1·148	196.61

Converting the thresholds (and hence, all the probit parameters estimates as well) to dollar amounts is accomplished using the method we employed for the Krehbiel and Rivers case; i.e., a linear transformation will re-calibrate the probit estimates to the dollar scale. We need to know a slope and intercept parameters for this linear transformation, plus also solve for the unknown reversion amount Q. We can solve for all these parameters exactly, since the above information can be represented as a system of three equations in three unknowns,

$$\begin{aligned} 98.72 &= m\mu_0 + -.5Q + c, \\ 99.85 &= m\hat{\mu}_1 + -.5Q + c, \\ 196.61 &= m\hat{\mu}_2 + 0Q + c, \end{aligned}$$

where m is the slope and c is the intercept of the linear transformation linking the probit scale with the dollar (billions) scale. Solving for the unknowns Q, m and c is made easier by re-writing the system in matrix terms,

$$\begin{bmatrix} 1 & -\cdot 5 & \mu_0 \\ 1 & -\cdot 5 & \hat{\mu}_1 \\ 1 & 0 & \hat{\mu}_2 \end{bmatrix} \begin{bmatrix} c \\ Q \\ m \end{bmatrix} = \begin{bmatrix} 98\cdot 72 \\ 99\cdot 85 \\ 196\cdot 61 \end{bmatrix},$$

or substituting for μ_0, $\hat{\mu}_1$ and $\hat{\mu}_2$,

$$\begin{bmatrix} 1 & -\cdot 5 & 0 \\ 1 & -\cdot 5 & \cdot 183 \\ 1 & 0 & 1\cdot 148 \end{bmatrix} \begin{bmatrix} c \\ Q \\ m \end{bmatrix} = \begin{bmatrix} 98\cdot 72 \\ 99\cdot 85 \\ 196\cdot 61 \end{bmatrix},$$

and applying Cramer's Rule to solve for the vector of unknown parameters. (Cramer's Rule states that the solution for a system of equations of the form $Ax = b$, with x unknown, is simply $\hat{x} = A^{-1}b$, provided the matrix of A exists, and indeed, the inverse of a matrix is often defined this way, i.e., as the solution to a system of linear equations; see Chiang 1984, 103--12).

In this case this yields the solution vector $(\hat{c}, \hat{Q}, \hat{m})$ = (189.52, 181.60, 6.17). Accordingly, all the predicted values from the probit model can be interpreted in billions of dollars by multiplying by 6.17 and adding 189.52. Bartels in fact only reports the results of his analysis in terms of dollars; the results in the "raw" metric of the probit are relegated to the Appendix of the article. Below I reproduce Table 1 of his article, showing the effects of the independent variables in terms of billions of dollars of defense appropriations.

Note that when transforming the standard errors from the probit scale to the $bn scale we only multiply; this is because the standard errors measure spread about a fixed point and in this sense are "dispersion" parameters, not "location" parameters like a mean or the intercept or the threshold parameters μ. When re-standardizing the location parameters we first re-scale them by multiplying by $\hat{m}$, and then re-locate them on the new metric with the intercept term $\hat{c}$. Since standard errors measure spread (and are in this sense insensitive to location) it is only necessary to re-scale them via multiplication by $\hat{m}$. Likewise for slope parameters, which need only be "scaled up" or "scaled down" by $\hat{m}$ after the location shifts are applied to the thresholds and intercept parameters. Note also that it is the standard errors of the parameter estimates that are re-scaled, not the variances; this

Table 2: ***Bartels (1991) Table 1: Sources of Support for Pentagon Appropriations.***

Variable	Units	Estimated Effects ($billions)
Intercept	-	183·96
		(13·40)
Constituency opinion	NES 7-point scale	12·87
		(5·82)
Constituency × competitiveness	NES scale × loser's % vote	·0112
		(·0497)
Tax burden	$1,000s per capita	-4·14
		(3·72)
Pentagon outlays	$1,000s per capita	7·70
		(3·68)
Partisanship	Republican = 1	3·87
	Democrat = 0	(1·79)
Presidential influence	vote difference in 100,000s	4·69
		(3·45)

is because the ordered probit model is defined in terms of a latent "*z*" score --- a location measure divided its standard deviation --- the usual assumption being that the errors on the probit scale are iid standard Normal deviates (mean 0, standard deviation 1).

Note also that the Bartels case is complicated by the presence of the unknown parameter Q. In general, re-scaling ordered probit parameters is somewhat simpler, involving two parameters in two unknowns; the intercept and slope parameters of the linear equation mapping from the probit scale to the substantively-interesting scale. Two pieces of information are required to recover these two parameters; i.e., we need to be able to set two (and only two) thresholds to values on the substantively-interesting match and then recover the slope and intercept parameters.

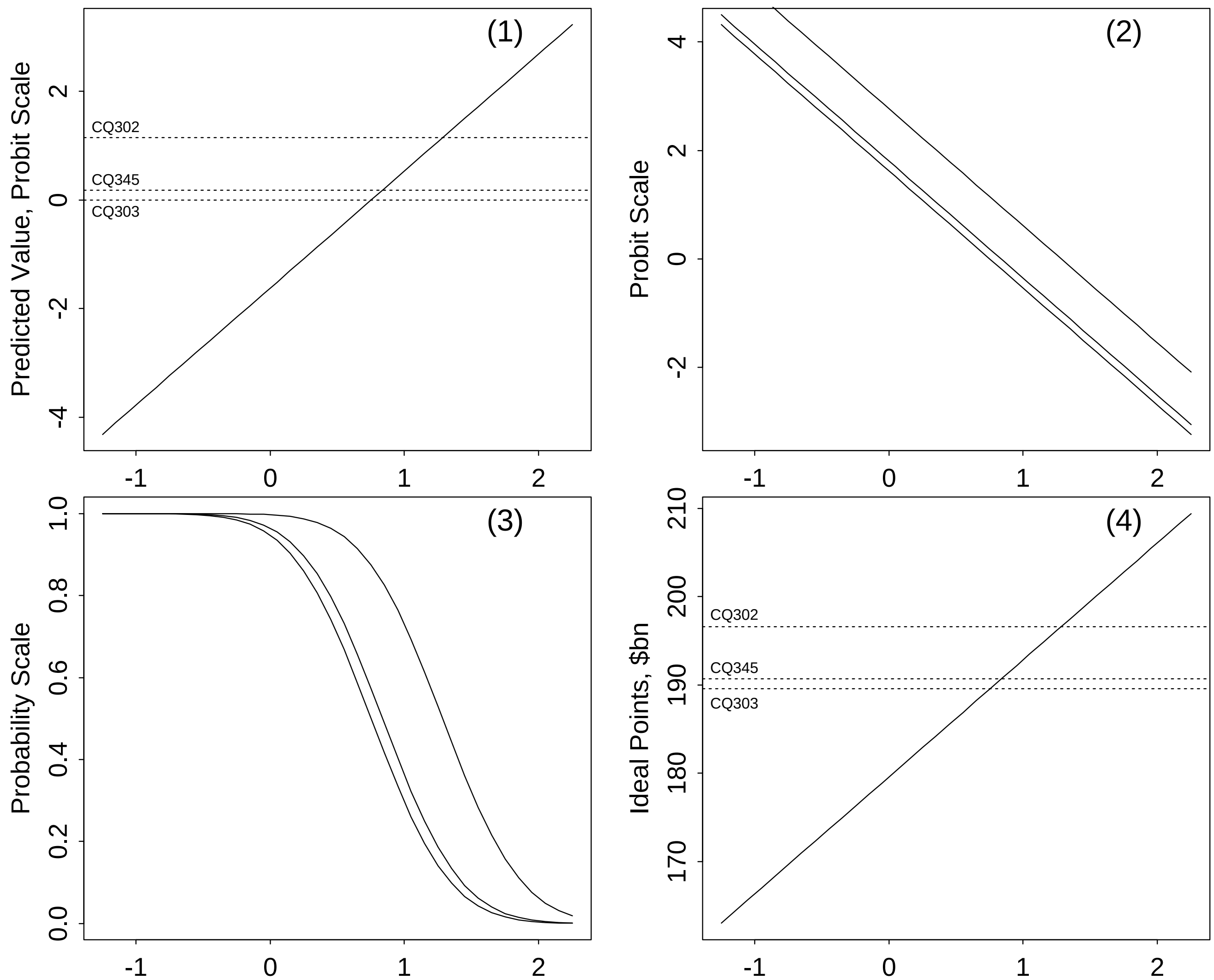
4 Interpretation

Ordered probit is just a generalization of the binary response model, slightly complicated by the threshold parameters and the identification restrictions necessary to recover estimates of them and the structural parameters simultaneously. Accordingly there is little to add beyond what we saw for the binary model.

Estimation is by maximum likelihood and the estimates have all the usual properties that MLEs have. Summary statistics are also available as for the binary model; likelihood ratio tests are also a convenient way of testing combinations of parameters and alternative specifications etc. The two-by-two table of "hits and misses" in the binary case generalizes to a m by m table in the ordered probit case.

Plotting out marginal effects of the independent variables results in m - 1 lines showing the cumulative estimated probability of being in the m (ordered) categories. An example using the Bartels case is shown in Figure 3. Four panels are shown in this figure; each showing the effects of the moving over the observed range in constituency opinion (from the lowest constituency preference, -1.25, to the highest constituency preference, 2.25) on the latent probit variable, a probability metric, and in terms of the legislators ideal points ($bn). All other independent variables are constant held at their sample means (this may or may not be an appropriate assumption for the other independent variables, but is about the best I can do without the original data; all these simulations are based summary statistics and the parameter

Figure 3: *S*imulated effects of change in constituency opinion.



estimates in the article).

The first panel shows the familiar linear relationship between a predictor and the latent probit variable. The second panel's parallel lines are the effects of increasing relative to the three thresholds; each line is just $\mu_j - \mathbf{x}_i'\hat{\boldsymbol{\beta}}, j = 0, 1, 2$. Recall that the probability of being in the jth category is estimated as $\Phi(\cdot)$ of this quantity, less the probability of being in the immediately lower category. Stacking these quantities, as in the third panel of Figure 3 results in the three curved lines. The area below the lowest curved line is just the probability of being in category "0" in Bartels' analysis, the category of legislators whose voting patterns revealed relatively low levels of preferred defense appropriations. Notice how this probability falls away as constituency opinion increases, to be replaced by a greater probability that a legislator belongs to one of the higher categories.

The last panel shows the effects of increasing constituency opinion on legislators' ideal points, in terms of dollars. Since the dollar metric is a linear transformation of the probit metric, and the effects of the independent variables are linear on the probit metric, it is no surprise that the effects of the independent variables expressed on the ideal-point metric are linear also. The dotted horizontal lines show the thresholds between the categories used in the ordered probit.

Typically one is interested in plots as in panels (3) and (4). These are the types of comparative statics presented in practice; I provide the other plots for completeness. The `Splus` code I used appears below.

```
# simulate some predicted vals from Larry's 1991 piece

# means of data
x0_c(1,1.205,33.499,2.012,.454,.444,-.269)
names(x)_c("Intercept","Cnstncy Opn","ConstxComp","Tax Burden",
"Pentagon outlays","Partisanship","Pres Influence")
# partisanship = 0 Dem, 1 Repub in case want to play with that
# other vars defnd in apsr piece

# parameter estimates (probit)
b_c(-.91,2.095,.00183,-.673,1.253,.63,.764)

# threshold estimates
mu_c(0,.183,1.148)
```

```
# function for transforming to dollar metric
bart_function(x){6.145*x + 189.555}

# simulate change in const opinion
x_x0
xsim_seq(-1.25,2.25,.1)
yhat.out_NULL
muyhat.out_NULL
pyhat.out_NULL
money.out_NULL
for (i in 1:length(xsim))
        {x[2]_xsim[i]
         x[3]_x0[3]*xsim[i]
         yhat_x%*%b
         muyhat_mu-yhat
         pyhat_pnorm(muyhat)
         money_bart(yhat)
         yhat.out_c(yhat.out,yhat)
         muyhat.out_rbind(muyhat.out,muyhat)
         pyhat.out_rbind(pyhat.out,pyhat)
         money.out_c(money.out,money)
         }
# graphing options
par(mfcol=c(2,2),mar=c(2,4,0.5,2))

plot(xsim,yhat.out,type="l",
ylab="Predicted Value, Probit Scale",cex=.85)
abline(h=mu,lty=2)
yrange_par()$usr
yrange_.95*yrange[4]+.05*yrange[3]
text(-1.15,mu[1]-.2,"CQ303",cex=.5)
text(-1.15,mu[2]+.2,"CQ345",cex=.5)
text(-1.15,mu[3]+.2,"CQ302",cex=.5)
text(1.75,yrange,"(1)")

plot(xsim,pyhat.out[,1],type="l",
        ylab="Probability Scale",cex=.85)
```

```
yrange_par()$usr
yrange_.95*yrange[4]+.05*yrange[3]
lines(xsim,pyhat.out[,2])
lines(xsim,pyhat.out[,3])
text(1.75,yrange,"(3)")

plot(xsim,muyhat.out[,1],type="l",
        ylab="Probit Scale",cex=.85)
yrange_par()$usr
yrange_.95*yrange[4]+.05*yrange[3]
lines(xsim,muyhat.out[,2])
lines(xsim,muyhat.out[,3])
text(1.75,yrange,"(2)")

plot(xsim,money.out,type="l",
ylab="Ideal Points, $bn",cex=.85)
abline(h=bart(mu),lty=2)
yrange_par()$usr
yrange_.95*yrange[4]+.05*yrange[3]
text(-1.15,188,"CQ303",cex=.5)
text(-1.15,192,"CQ345",cex=.5)
text(-1.15,198,"CQ302",cex=.5)
text(1.75,yrange,"(4)")
```

5 References

Aitchison, J., and S. Silvey. 1957. The Generalization of Probit Analysis to the Case of Multiple Responses. *Biometrika*. 44:131--40.

Aldrich, John H. and Forrest D. Nelson. 1984. *Linear Probability, Logit, and Probit Models*. Sage University Paper series on Quantitative Applications in the Social Sciences, series np. 07-045. Sage. Newbury Park, California.

Amemiya, Takeshi. 1985. *Advanced Econometrics*. Harvard. Cambridge.

Bartels, Larry M. 1991. Constituency Opinion and Congressional Polic Making: The Reagan Defense Buildup. *American Political Science Review*. 85:457--74.

Krehbeil, Keith and R. Douglas Rivers. 1988. The Analysis of Committee Power: An Application of Senate Voting on the Minimum Wage. *American Journal of Political Science*. 32:1151--74.

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