Week 4: Ordered Logit Model

POLI803

Howard Liu

Week 4, 2024

University of South Carolina

Outline

Ordered Logistic Regression

- Random utility representation
- Ordered logit (probit) models
- Marginal effect of x

Review: logit model

We have a binary DV:

$$Y = \begin{cases} 1 & \text{(vote)} \\ 0 & \text{(not vote)} \end{cases}$$

Review: logit model

We have a binary DV:

$$Y = \begin{cases} 1 & \text{(vote)} \\ 0 & \text{(not vote)} \end{cases}$$

The values of Y are determined by P or $\Pr(Y = 1)$,

Review: logit model

We have a binary DV:

$$Y = \begin{cases} 1 & \text{(vote)} \\ 0 & \text{(not vote)} \end{cases}$$

The values of Y are determined by P or $\Pr(Y = 1)$,

$$Y^* = \alpha + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \cdots + \beta_k X_k$$
$$\hat{P} = \Lambda(Y^*)$$

Review: logit model

We have a binary DV:

$$Y = \begin{cases} 1 & \text{(vote)} \\ 0 & \text{(not vote)} \end{cases}$$

The values of Y are determined by P or $\Pr(Y = 1)$,

$$Y^* = \alpha + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \cdots + \beta_k X_k$$
$$\hat{P} = \Lambda(Y^*)$$

- Y^* = latent utility (propensity).
- Y^* can range between $-\infty$ and ∞, but $\hat{P}$ ranges between 0 and 1.
- We don't care about the actual $(Y^*) \rightarrow$ but care about the more interpretable latent probability [0,1] $\hat{P}$ (the s-curve)

Random utility representation (thresholds)

There is an alternative interpretation of the same model

$$Y = \begin{cases} 1 & \text{(vote)} \\ 0 & \text{(not vote)} \end{cases}$$

Random utility representation (thresholds)

There is an alternative interpretation of the same model

$$Y = \begin{cases} 1 & \text{(vote)} \\ 0 & \text{(not vote)} \end{cases}$$

We think of Y^* as an unobservable random utility of voting, whereas Y is the actual observation

$$Y^* = \boldsymbol{X}\boldsymbol{\beta} + \epsilon$$

Random utility representation (thresholds)

There is an alternative interpretation of the same model

$$Y = \begin{cases} 1 & \text{(vote)} \\ 0 & \text{(not vote)} \end{cases}$$

We think of Y^* as an **unobservable random utility** of voting, whereas Y is the actual observation

$$Y^* = \boldsymbol{X}\boldsymbol{\beta} + \epsilon$$

- Actor votes ($Y = 1$) when Y^* is greater than some threshold (usually 0); $\Pr(Y^* > \mathit{threshold}) \rightsquigarrow \Pr(Y^* = 1)$
- Conceptual steps: Xs $\Rightarrow Y^* \Rightarrow Y$

Random utility representation

$$Y^* = \boldsymbol{X\beta} + \epsilon$$

$$Y = \begin{cases} 1 & (\text{when } Y^* > 0) \\ 0 & (\text{when } Y^* \leq 0) \end{cases}$$

- Systematic component:
  $\boldsymbol{X\beta} = \alpha + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \cdots + \beta_k X_k$
- Stochastic component: ϵ follows a logistic distribution

Random utility representation (an example)

Consider a simple model as an example:

$$Y^* = \alpha + \beta_1 X_1 + \epsilon$$

where X_1 takes three values: 0, 1, 2, and $\hat{\alpha} = 0$ and $\hat{\beta} = 1$

Random utility representation (an example)

Consider a simple model as an example:

$$Y^* = \alpha + \beta_1 X_1 + \epsilon$$

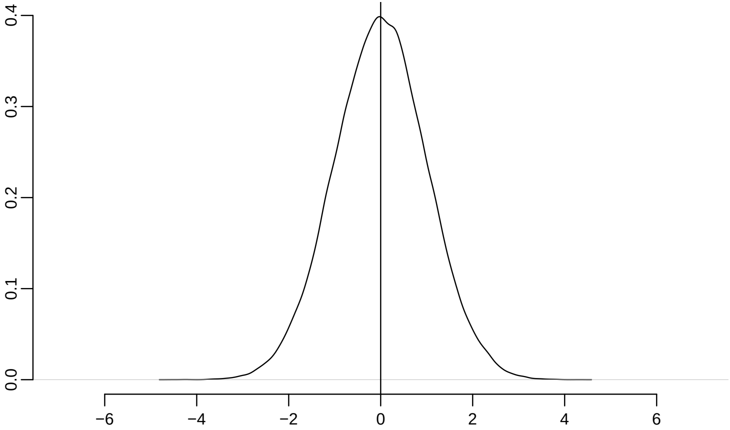
where X_1 takes three values: 0, 1, 2, and $\hat{\alpha} = 0$ and $\hat{\beta} = 1$

$$Y^* = \begin{cases} 0 + \epsilon & (\text{when } X_1 = 0) \\ 1 + \epsilon & (\text{when } X_1 = 1) \\ 2 + \epsilon & (\text{when } X_1 = 2) \end{cases}$$

We can see that:

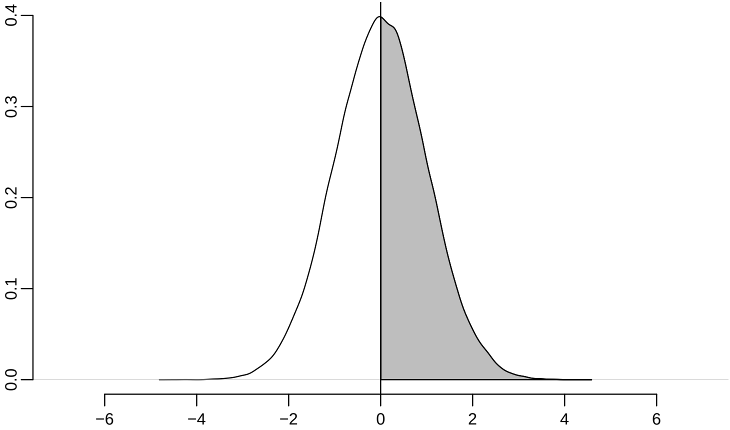
- As X_1 gets bigger, Y^* gets bigger ($Y^* = \alpha + \beta_1 X_1 + \epsilon$)
- As Y^* gets bigger, it is more likely to satisfy the condition: $Y^* > 0$, hence more likely that $Y = 1$

When $X_1 = 0$ and thus $Y^* = 0 + \epsilon$



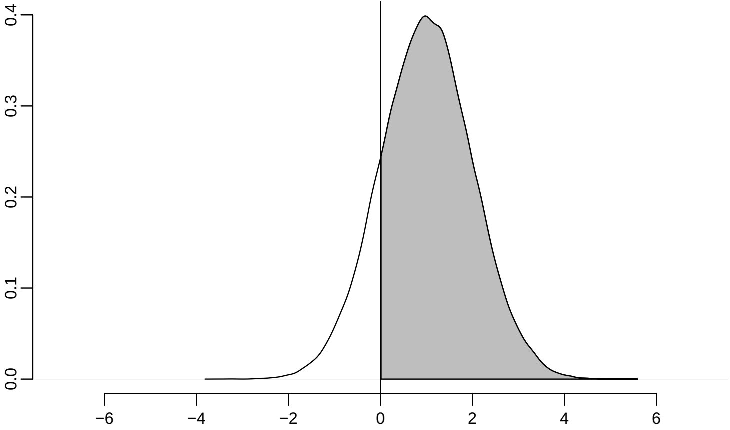
When $X_1 = 0$, about half of the cases satisfy $Y^* > 0$
Look at the area under the logistic curve (Λ)–also called cumulative distribution function (cdf)

When $X_1 = 0$ and thus $Y^* = 0 + \epsilon$



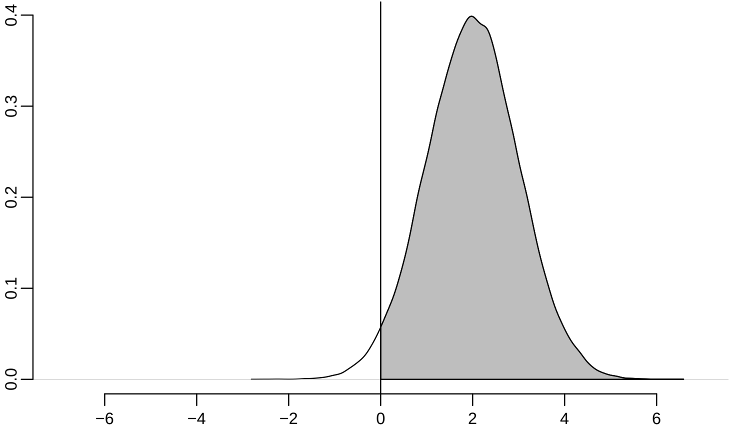
When $X_1 = 0$, about half of the cases satisfy $Y^* > 0$
Look at the area under the logistic curve (Λ)–also called cumulative distribution function (cdf)

When $X_1 = 1$ and thus $Y^* = 1 + \epsilon$



When $X_1 = 1 \quad \Rightarrow \quad Y^* \uparrow \; \Rightarrow \quad$ more cases satisfy $Y^* > 0$

When $X_1 = 2$ and thus $Y^* = 2 + \epsilon$



When $X_1 = 2 \;\Rightarrow\; Y^* \uparrow \;\Rightarrow\;$ even more cases satisfy $Y^* > 0$

Random utility representation

- We can think of $\Pr(Y^* > 0)$ as $\Pr(Y = 1)$

Random utility representation

- We can think of $\Pr(Y^* > 0)$ as $\Pr(Y = 1)$
  - As X_1 gets bigger Y^* gets bigger $\rightsquigarrow$ P gets bigger

Random utility representation

- We can think of $\Pr(Y^* > 0)$ as $\Pr(Y = 1)$
  - As X_1 gets bigger Y^* gets bigger $\rightsquigarrow$ P gets bigger
- When we have one threshold, it becomes a logit regression
- When we have multiple thresholds, it becomes an ordered logit regression

Ordered Logistic Regression

Let's say we are interested in roll call voting in the US congress

$$Y = \begin{cases} 0 & \text{(vote Nay)} \\ 1 & \text{(abstain)} \\ 2 & \text{(vote Yay)} \end{cases}$$

Ordered Logistic Regression

Let's say we are interested in roll call voting in the US congress

$$Y = \begin{cases} 0 & \text{(vote Nay)} \\ 1 & \text{(abstain)} \\ 2 & \text{(vote Yay)} \end{cases}$$

We can think of unobservable random utility of supporting the bill:

$$Y^* = \boldsymbol{X}\beta + \epsilon$$

Ordered Logistic Regression

Let's say we are interested in roll call voting in the US congress

$$Y = \begin{cases} 0 & \text{(vote Nay)} \\ 1 & \text{(abstain)} \\ 2 & \text{(vote Yay)} \end{cases}$$

We can think of unobservable random utility of supporting the bill:

$$Y^* = \boldsymbol{X}\beta + \epsilon$$

- As Y^* gets bigger, $\Pr(Y = 2)$ increases

Ordered Logistic Regression

Let's say we are interested in roll call voting in the US congress

$$Y = \begin{cases} 0 & \text{(vote Nay)} \\ 1 & \text{(abstain)} \\ 2 & \text{(vote Yay)} \end{cases}$$

We can think of unobservable random utility of supporting the bill:

$$Y^* = \boldsymbol{X}\beta + \epsilon$$

- As Y^* gets bigger, $\Pr(Y = 2)$ increases
- As Y^* gets bigger, $\Pr(Y = 0)$ decreases

Ordered Logistic Regression

Let's say we are interested in roll call voting in the US congress

$$Y = \begin{cases} 0 & \text{(vote Nay)} \\ 1 & \text{(abstain)} \\ 2 & \text{(vote Yay)} \end{cases}$$

We can think of unobservable random utility of supporting the bill:

$$Y^* = \boldsymbol{X}\beta + \epsilon$$

- As Y^* gets bigger, $\Pr(Y = 2)$ increases
- As Y^* gets bigger, $\Pr(Y = 0)$ decreases
- As Y^* gets bigger, $\Pr(Y = 1)$ increases relative to $\Pr(Y = 0)$ but decreases relative to $\Pr(Y = 2)$

Ordered logistic regression

- Recall, with logit models we needed one threshold (0) to classify two values

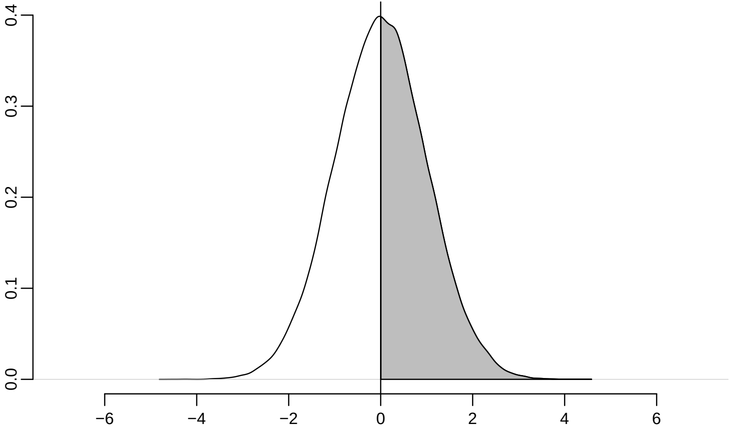
Ordered logistic regression

- Recall, with logit models we needed one threshold (0) to classify two values
- When we have 3 categories, we need 2 thresholds

$$Y = \begin{cases} 0 & \text{(vote Nay) when } Y^* \leq c_1 \\ 1 & \text{(abstain) when } c_1 < Y^* \leq c_2 \\ 2 & \text{(vote Yay) when } Y^* > c_2 \end{cases}$$

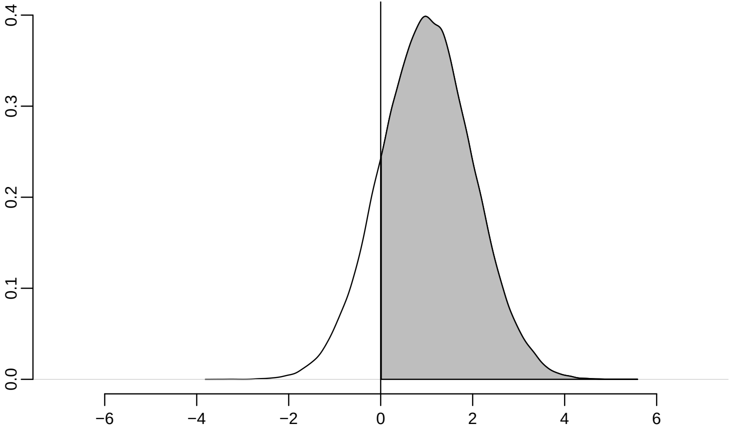
Ordered logistic regression

When we have a binary DV:



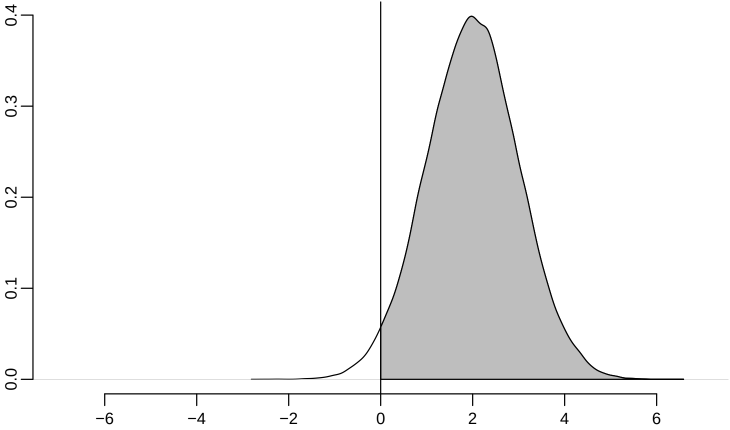
Ordered logistic regression

When we have a binary DV:



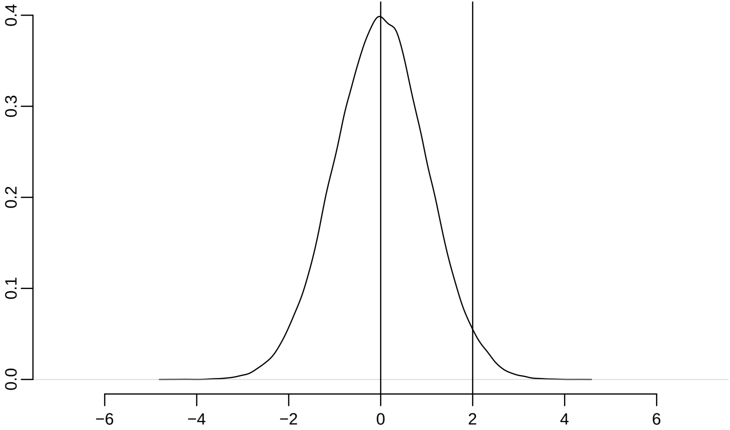
Ordered logistic regression

When we have a binary DV:



Ordered logistic regression

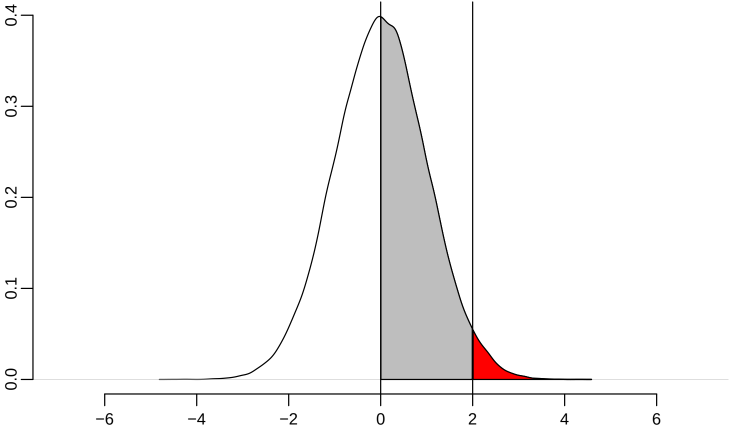
When we have a three-category ordered DV:



Red = $\Pr(Y = 2)$, Gray = $\Pr(Y = 1)$, White = $\Pr(Y = 0)$

Ordered logistic regression

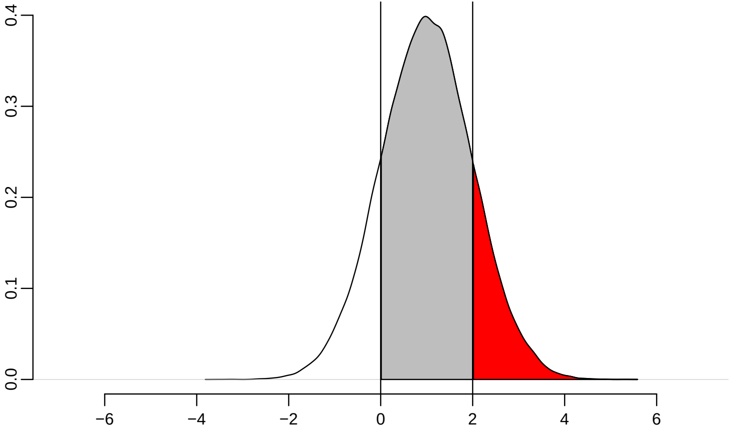
When we have a three-category ordered DV:



Red = $\Pr(Y = 2)$, Gray = $\Pr(Y = 1)$, White = $\Pr(Y = 0)$

Ordered logistic regression

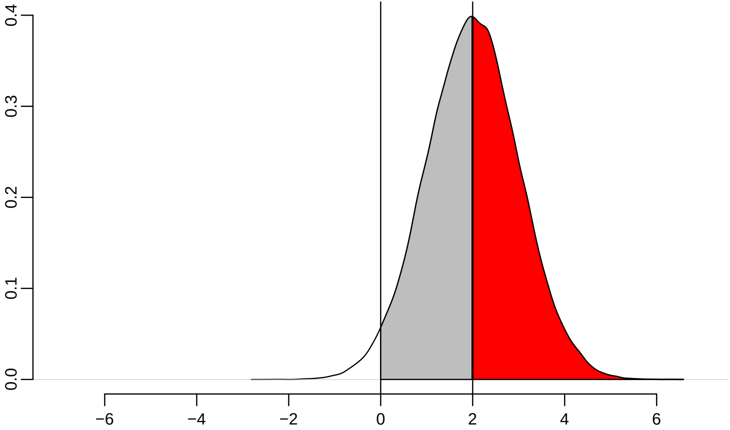
When we have a three-category ordered DV:



Red = $\Pr(Y = 2)$, Gray = $\Pr(Y = 1)$, White = $\Pr(Y = 0)$

Ordered logistic regression

When we have a three-category ordered DV:



Red = $\Pr(Y = 2)$, Gray = $\Pr(Y = 1)$, White = $\Pr(Y = 0)$

Ordered logit / probit

DV = ordered categorical variable

- A lot of applications in public opinion research
  - Eur. Social Survey: people's attitude toward immigration, support for welfare spending
  - E.g. Strongly approve, somewhat approve, neutral, somewhat disapprove, strongly disapprove
- Applications in conflict research
  - No violence, repression, civil war
  - Lose, draw, win in war
- Roll call voting (nay, abstain, yay)

Table 1. Ordered Logit Estimates of Approval of Vice-Presidential Candidate Selections, Post-VP Debate, by Viewership of *SNL* Debate Spoof (standard errors in parentheses)

	"Do you approve or disapprove of John McCain's pick of Sarah Palin as his vice–presidential running mate?" (n = 1,731)	"Do you approve or disapprove of Barack Obama's pick of Joe Biden as his vice–presidential running mate?" (n = 1,731)
	(1 = Disapprove; 2 = Not Sure/Don't Know; 3 = Approve)	
Pre-Debate Approval	1.79**	1.86**
	(.10)	(.10)
Viewed *SNL* Debate Spoof	−.74**	.46*
	(.21)	(.19)
Party Identification (1 = strong Democrat to 5 = strong Republican)	.70**	−.30**
	(.10)	(.08)
Ideology (1 = very liberal to 5 = very conservative)	.38**	−.23
	(.09)	(.08)
Overall Media Exposure	−.06*	.11**
	(.03)	(.03)
Political Knowledge	−.13**	.07*
	(.04)	(.03)
White	.19	.16
	(.20)	(.17)
Male	−.33*	−.18
	(.14)	(.12)
Age	−.02	.01
	(.04)	(.05)
Constant 1	5.50	1.56
Constant 2	6.74	3.92
Chi-Squared	1291.23**	757.93**

$*p \leq .05$; $**p \leq .01$ (two-tailed).

Religion and Attitudes toward Redistributive Policies among Americans

Table 1. Ordered Logit and Regression Estimates for Models of Support for Redistributive Policies, 2013 Economic Values Survey (PRRI).

	Favor tax increases on the rich		Support repeal of ACA		Support increase in minimum wage		Government equality policy scale	
Variable	*b*	*z*	*b*	*z*	*b*	*z*	*b*	*z*
Religion variables								
Black Protestant	0.695	1.95*	0.214	0.58	0.085	0.23	0.078	0.52
Evangelical	0.133	0.73	0.083	0.44	0.401	2.20*	0.082	1.09
Catholic	0.063	0.37	0.149	0.83	0.260	1.50	0.034	0.47
Other faith	0.082	0.39	−0.179	−0.80	−0.139	−0.66	0.001	0.11
Secular	−0.105	−0.51	−0.061	−0.28	0.102	0.49	−0.008	−0.10
Religiosity scale	−0.104	−1.21	0.042	0.48	0.004	0.05	−0.011	−0.30
Religious left	0.052	0.23	−0.311	−1.37	0.014	0.06	0.010	0.11
Religious right	−0.319	−2.00*	0.286	1.68*	0.053	0.33	−0.125	−1.87*
Both religious left and right	0.521	1.61	0.249	0.76	0.401	1.24	0.189	1.44
Preserve traditional beliefs	−0.174	−2.05*	0.121	1.40	−0.289	−3.37***	−0.110	−3.14***
Jesus promotes just society	0.058	1.85*	−0.086	−2.63**	0.037	1.16	0.041	3.13***

Example: Conflict Mediation

Terris & Maoz (2005) "Rational Mediation: A Theory and a Test." *JPR*.

- RQ: What explains the occurrence / intensity of third-party mediation in international conflict?

Example: Conflict Mediation

Terris & Maoz (2005) "Rational Mediation: A Theory and a Test." *JPR*.

- RQ: What explains the occurrence / intensity of third-party mediation in international conflict?
- Some international conflicts experience more intrusive mediation, some experience less intrusive mediation, and others experience none

Example: Conflict Mediation

Terris & Maoz (2005) "Rational Mediation: A Theory and a Test." *JPR*.

- RQ: What explains the occurrence / intensity of third-party mediation in international conflict?
- Some international conflicts experience more intrusive mediation, some experience less intrusive mediation, and others experience none

$$Y = \begin{cases} 0 & \text{No mediation} \\ 1 & \text{Less intrusive mediation} \\ 2 & \text{More intrusive mediation} \end{cases}$$

Replication: Conflict Mediation

Theory: when the conflict is more versatile (susceptible to change), more intrusive forms of mediation become more likely

- Conflict versatility: likelihood that the underlying conflict can be converted into a cooperative game

Replication: Conflict Mediation

Theory: when the conflict is more versatile (susceptible to change), more intrusive forms of mediation become more likely

- Conflict versatility: likelihood that the underlying conflict can be converted into a cooperative game

Underlying (unobservable) random utility of mediation, Y^*

$$Y^* = \beta * \texttt{Conflict Versatility} + \boldsymbol{X\beta} + \epsilon$$

$$Y = \begin{cases} 0 & \text{when } Y^* \leq cut_1 \\ 1 & \text{when } cut_1 < Y^* \leq cut_2 \\ 2 & \text{when } Y^* > cut_2 \end{cases}$$

Example: Conflict Mediation

The data set is available online:

`http://vanity.dss.ucdavis.edu/~maoz/datasets.htm`

- DV (`medintrus`): "None" (0), "Information/Procedural" (1), "Directive" (2)
- Conflict versatility (`cumversatil`): 85.2 $\sim$ 948.0 (higher values = more versatile)
- Minimum Regime Score (`minreg302`): $-90 \sim 60$ (higher values = disputants are more democratic)
- Capability ratio (`caprat`): 1.002 $\sim$ 13439.462 (higher values = one disputant is stronger than the other)
- Alliance (`ally1`): dummy (1 if disputants are allied, 0 otherwise)
- Past mediation (`lagprmed`): 0 $\sim$ 12 (Number of past mediated conflicts)

Fitting an ordered logit model in R

```
library(MASS)
fit <- polr (y ~ x1 + x2 + x3,
     data = dataset.name)
```

The `polr` function is included in the MASS package

Example: Conflict Mediation

Fifth model in Table 2 (p. 579)

	Dependent variable:
Minimum Regime Score	0.007**
	(0.003)
Capability Ratio	−0.008*
	(0.005)
Alliance	1.066***
	(0.211)
Prior Mediation	0.263***
	(0.060)
Conflict Versatility	0.004***
	(0.0004)
Observations	1,382
Note:	*p<0.1; **p<0.05; ***p<0.01

Example: Conflict Mediation

A few things to note:

- In ordered logit model, the intercept will not be estimated (assumed to be zero)
- Instead, we estimate two cut-points. By default, the `stargazer` table doesn't show them
- We can use the `summary` function to find the values of the cut-points

Example: Conflict Mediation

```
Call:
polr(formula = medintrus ~ minreg302 + caprat + ally1 + lagprmed +
    cumversatil, data = tm)

Coefficients:
                  Value Std. Error t value
minreg302      0.007349  0.0033115   2.219
caprat        -0.008497  0.0046378  -1.832
ally1Alliance  1.065726  0.2105948   5.061
lagprmed       0.263148  0.0596802   4.409
cumversatil    0.003942  0.0003839  10.267

Intercepts:
                                    Value   Std. Error t value
None|Information/Procedural          4.3434  0.3200    13.5723
Information/Procedural|Directive     5.0789  0.3353    15.1468

Residual Deviance: 898.8218
AIC: 912.8218
(385 observations deleted due to missingness)
```

Example: Conflict Mediation

How the table should look

	Dependent variable:
Minimum Regime Score	0.007^{**}
	(0.003)
Capability Ratio	-0.008^{*}
	(0.005)
Alliance	1.066^{***}
	(0.211)
Prior Mediation	0.263^{***}
	(0.060)
Conflict Versatility	0.004^{***}
	(0.0004)
Cut point 1	4.343^{***}
	(0.320)
Cut point 2	5.079^{***}
	(0.335)
Observations	1,382
Note:	$^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$

Example: Conflict Mediation

So we find that `Conflict Versatility` has a slope of 0.004 and it's highly statistically significant

Example: Conflict Mediation

So we find that `Conflict Versatility` has a slope of 0.004 and it's highly statistically significant

- But what does that mean in terms of different ordered categories? We can't tell from the table $\rightarrow$ need effect plots
- Recall this is the effect of `Conflict Versatility` on Y^*, which is NOT the quantity of interest in itself

Example: Conflict Mediation

So we find that `Conflict Versatility` has a slope of 0.004 and it's highly statistically significant

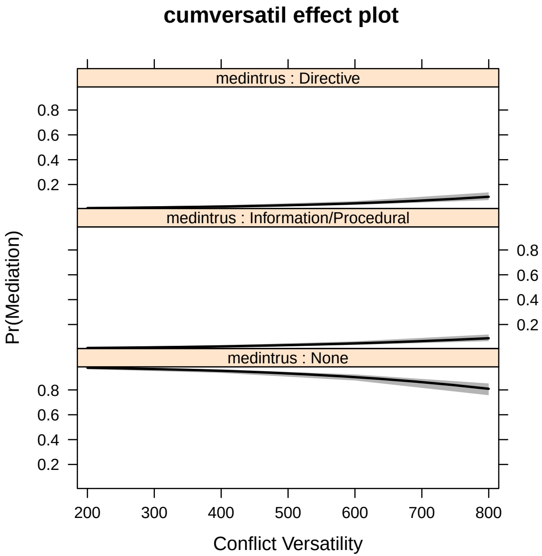
- But what does that mean in terms of different ordered categories? We can't tell from the table → need effect plots
- Recall this is the effect of `Conflict Versatility` on Y^*, which is NOT the quantity of interest in itself
- We need to know whether this induces a meaningful change in $\Pr(Y = 2)$ and/or $\Pr(Y = 1)$ relative to $\Pr(Y = 0)$

Example: Conflict Mediation

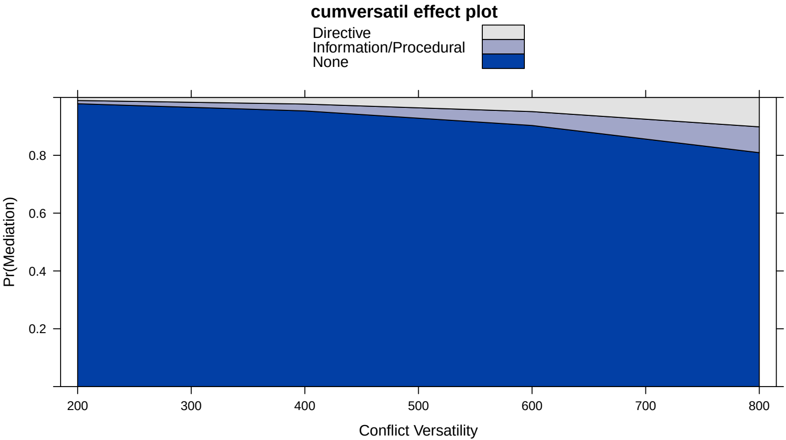
So we find that `Conflict Versatility` has a slope of 0.004 and it's highly statistically significant

- But what does that mean in terms of different ordered categories? We can't tell from the table $\rightarrow$ need effect plots
- Recall this is the effect of `Conflict Versatility` on Y^*, which is NOT the quantity of interest in itself
- We need to know whether this induces a meaningful change in $\Pr(Y = 2)$ and/or $\Pr(Y = 1)$ relative to $\Pr(Y = 0)$
- Recall also that the effects of `Conflict Versatility` on probabilities depend on the values of other independent variables
  - We usually set the values at their mean or median value
  - We should try setting them at other "interesting" values

Example: Conflict Mediation

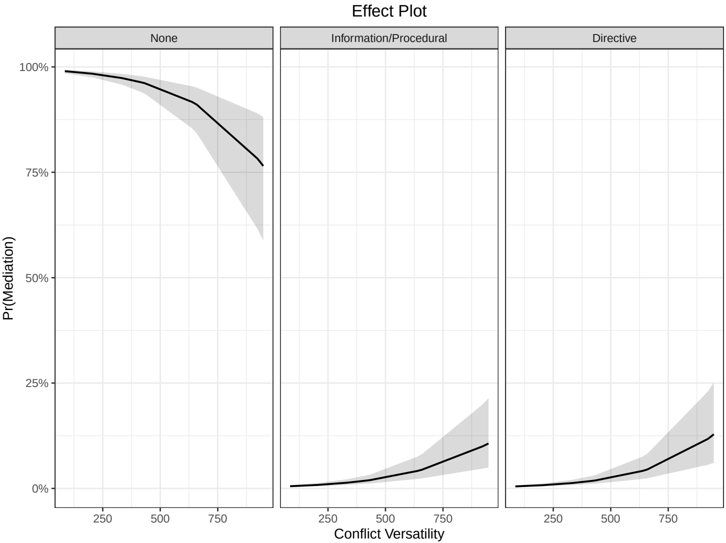


Example: Conflict Mediation



- Interpret the areas under the curves

Example: Conflict Mediation



Example: Conflict Mediation

```
> eff.cv

cumversatil effect (probability) for None
cumversatil
      200       400       600       800 
0.9783138 0.9535049 0.9031250 0.8090859 

cumversatil effect (probability) for Information/Procedural
cumversatil
       200        400        600        800 
0.01117349 0.02365755 0.04797716 0.08931031 

cumversatil effect (probability) for Directive
cumversatil
       200        400        600        800 
0.01051269 0.02283757 0.04889785 0.10160382 
> 
```

Summary

- When DV is an ordered categorical $\Rightarrow$ ordered logit model
  - Roll call voting (nay, abstain, yay)
  - Levels of support for a certain policy (survey research)
  - Military victory (lose, draw, win)

- After estimating the model, we need to investigate the substantive effects of our main independent variable using the `effect` function

Summary

- When DV is an ordered categorical $\Rightarrow$ ordered logit model
  - Roll call voting (nay, abstain, yay)
  - Levels of support for a certain policy (survey research)
  - Military victory (lose, draw, win)
- After estimating the model, we need to investigate the substantive effects of our main independent variable using the `effect` function
- In doing so, try setting the values of the other independent variables at interesting values, and see how the effects of the main IV change