MLE: Categorical and Limited Dependent Variables

Unit 2: Ordered, Multinomial, Count, and Limited Dependent Variables

1. Ordered Logit and Probit Models

PS2730-2020

Week 4

Professor Steven Finkel

Models for Non-Binary, Non-Continuous Variables

- Many other kinds of non-continuous variables aside from the dichotomous or binary variables that we have considered so far with logit and probit models
  - Ordinal Variables: more than two ranked categories without necessarily equal distance between the categories
  - Multinomial Variables: more than two unranked categories
  - Count Variables: more than two non-negative integer categories
  - Censored Variables: continuous up to (or down to) a threshold
  - Truncated or Sample-Selected Variables: continuous but observed only if past a threshold on itself or another variable
- All involve extensions of either (or both) the non-linear specification or the latent variable framework for modeling dichotomous dependent variables via logit and probit regression
- All estimated via ML methods, applying the framework we've discussed so far

Modeling Ordinal Variables

- Ordinal variables have multiple categories that can be ranked
  - Social class: Low, Medium, High
  - 4-5 category "Strongly Agree" to "Strongly Disagree" survey questions
  - Outcomes of civil conflict: peace, low-level conflict, civil war
- Can you treat these variables as interval and estimate via OLS?
  - NO! OLS assumes equal distances between the categories, as in "every unit change in X brings about a β unit change in Y". The units in Y must be equal, i.e., β at one point in the scale must be the same change as β at another point, and this is not necessarily the case with ordinal variables
  - Actually, the scale categories for ordinal variables are completely arbitrary anyway so the "unit change" idea is pretty meaningless. We could, e.g., assign a value of -400 to "low class", 6,225 to "middle class", and 4,500,823 to "high class", or we could assign "1' "2" "3".
- So we estimate instead with "ordered probit" or "ordered logit" and ML methods

Ordered Probit

- Ordered Probit is a straightforward extension of the latent variable framework to take ordered categories into account. Instead of only one τ threshold for Y* at 0 to distinguish observations of "0" or "1", we allow for multiple τ thresholds that distinguish observations of "category 1", "category 2", "category 3", etc.
- As in dichotomous probit model:

$$Y_i^* = \Sigma \beta X_i + \varepsilon_i$$

$$Y_i^* = X\mathrm{B} + \varepsilon_i$$

$$\mathrm{E}(Y_i^* \mid X) = X\mathrm{B}$$

- So Y* is continuous but unobserved. We map the observed variable Y_i to Y* via the "measurement equation" that says if Y* is above a certain threshold, observed Y will be 1; if Y* is above the next threshold, observed Y will be 2; above the next threshold, observed Y will be 3, and so on, depending on the number of categories

- Assume a three category ordinal variable
- Assign the zero threshold (τ_0) to be negative infinity ($-\infty$) and the threshold for the last category (τ_3) to be positive infinity (∞)
- Then the full model is:

$$Y_i^* = \Sigma\beta X_i + \varepsilon_i$$

$$Y_i^* = X\mathrm{B} + \varepsilon_i$$

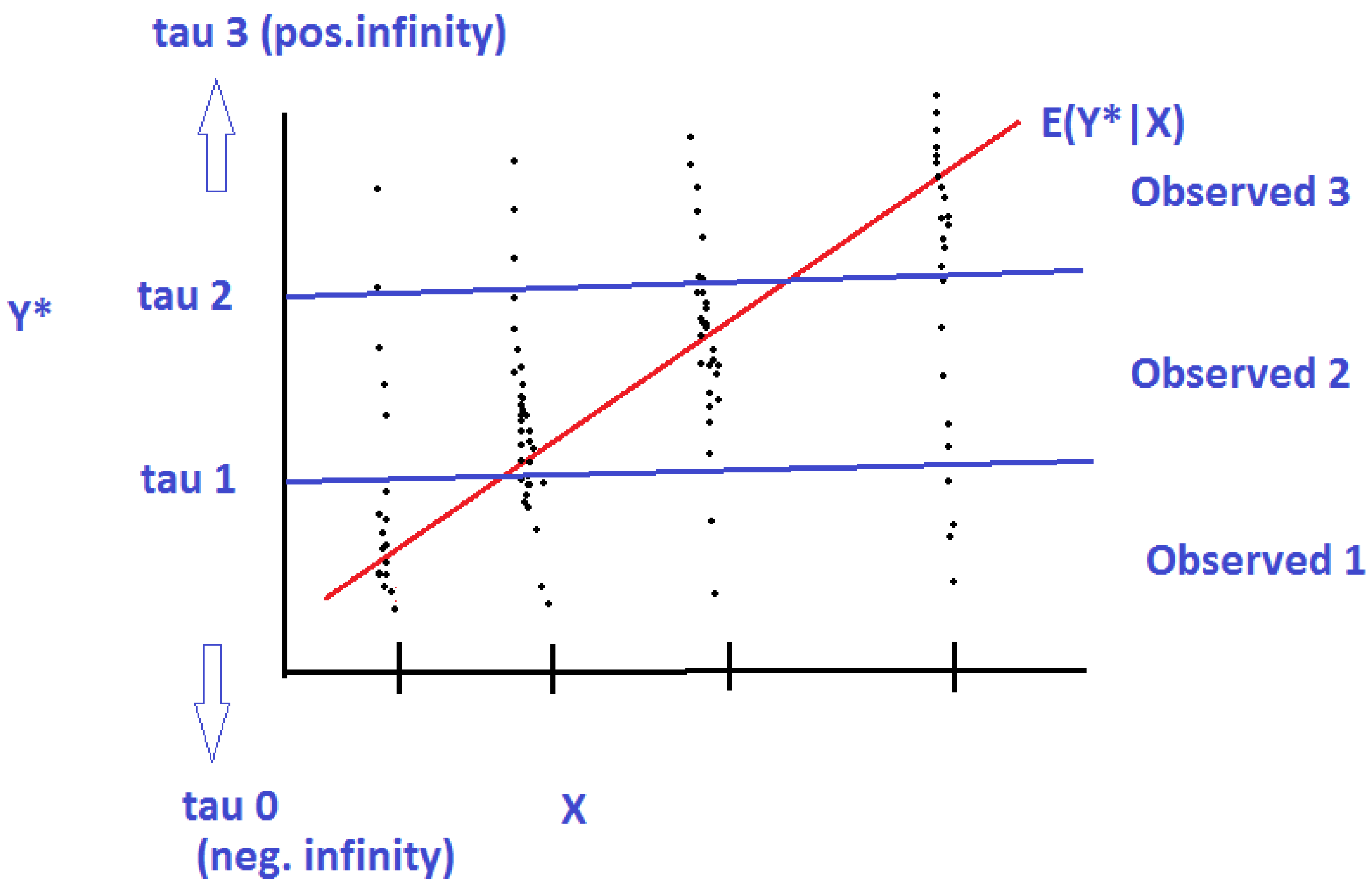
$$\mathrm{E}(Y_i^* \mid X) = X\mathrm{B}$$

with measurement equations

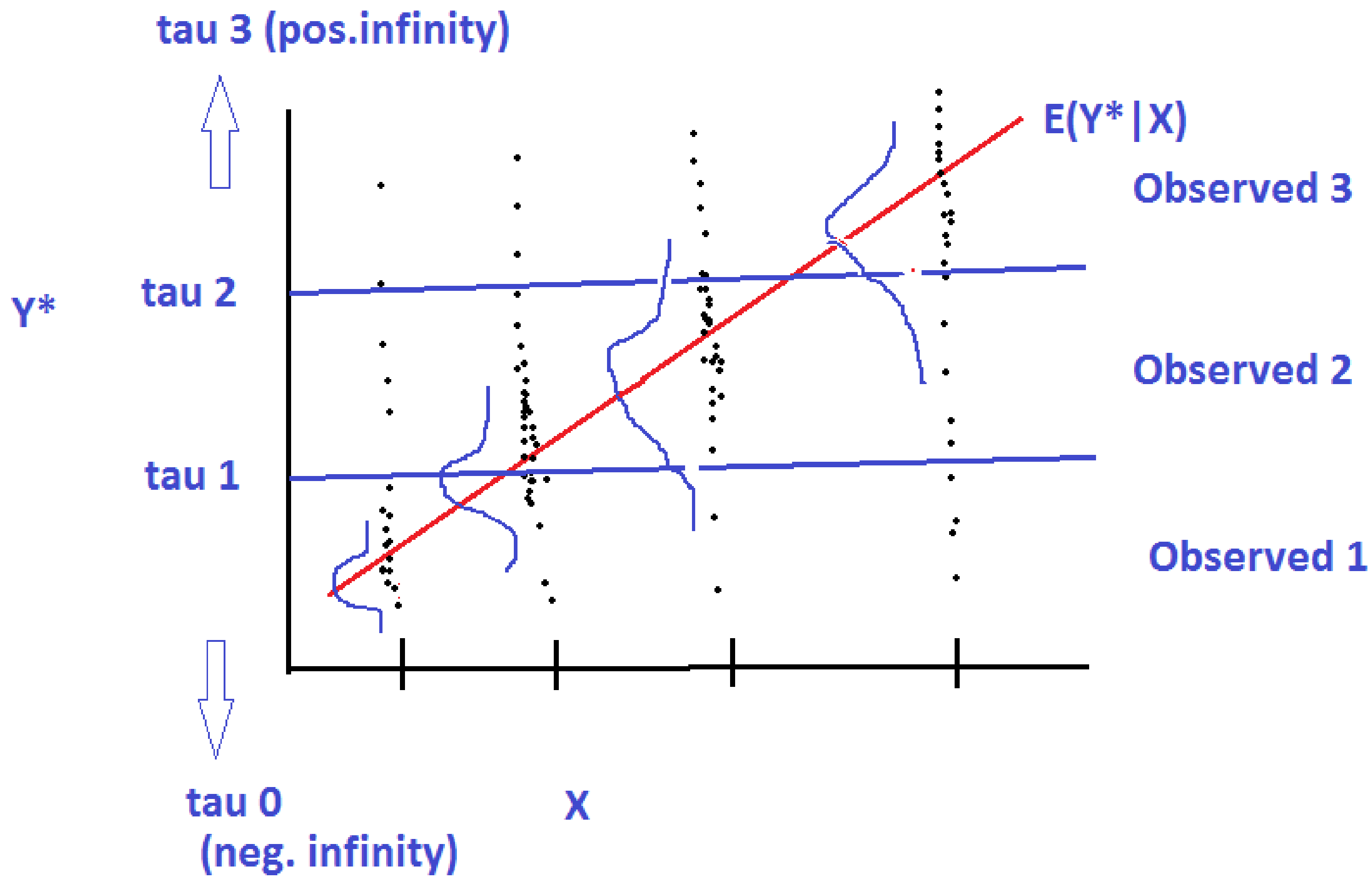
$$Y_i = 1 \text{ if } \tau_0(-\propto) <= \mathrm{Y}_i^* < \tau_1$$

$$Y_i = 2 \text{ if } \tau_1 <= \mathrm{Y}_i^* < \tau_2$$

$$Y_i = 3 \text{ if } \tau_2 <= \mathrm{Y}_i^* < \tau_3(\propto)$$


tau 3 (pos.infinity)
E(Y* | X)
Observed 3
Y*
tau 2
Observed 2
tau 1
Observed 1
tau 0
(neg. infinity)
X

- As with dichotomous probit, Y* is unobserved so cannot use OLS
- We also need to make assumptions about the error term ε. If we assume standard normal, we arrive at the "ordinal probit" specification. If we assume logistic distribution, we arrive at "ordinal logit" (though we will also arrive at ordinal logit by extending the non-linear probability framework a little later)
- Idea: XB takes Y* to some expected value, and then, depending on the size of the normally distributed error and whether it takes Y* past given thresholds, the observed Y will be 1, 2, or 3. We can use normal curve properties to calculate the probability, given XB, of obtaining an error term sufficiently large to put Y* over the τ_1 threshold, and over the τ_2 threshold, which would result in an observed Y of 2 or 3 respectively. Otherwise observed Y will be 1.



- Given XB, if ε is large enough, it will put even a very low Y* above the τ_1 or τ_2 thresholds; thus Y=2 or 3
- Given XB, if ε is small enough, it will put even a very high Y* below the τ_2 or τ_1 thresholds; thus Y=2 or 1

- The probability of observing, e.g., a "1" is:

$$P(Y=1|X) = P(\tau_0 \leq Y^* < \tau_1 | X)$$
$$P(Y=1|X) = P(\tau_0 \leq X\text{B} + \varepsilon < \tau_1 | X)$$
$$P(Y=1|X) = P(\tau_0 - X\text{B} \leq \varepsilon < \tau_1 - X\text{B} | X)$$

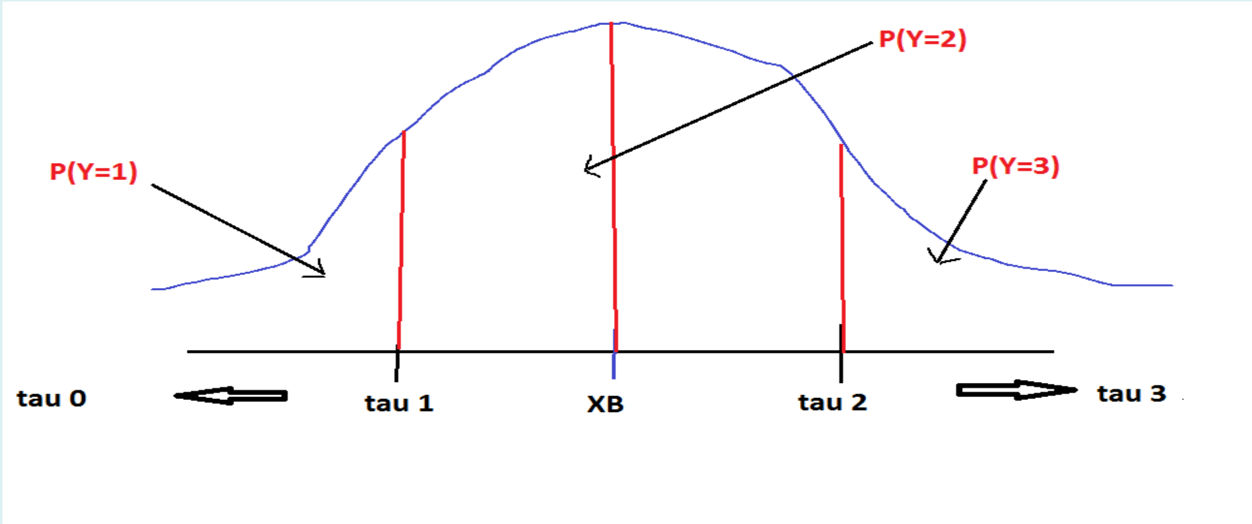
- What is the probability that a normally distributed error term lies between two points? It is the difference in the cumulative probability associated with each of those points – like the distance between two z-scores.
- Intuitively: XB puts Y*, e.g., at 2. If τ_1 is, say, .5 , then we know that anytime there is an error term less than (.5-2=)-1.5, the person will be under the first threshold. Can there be a error term that will put the case under the τ_o ? No, would have to be smaller than negative infinity. So we say that the probability of being in category 1 is the probability of the error term being less than (τ_1 – XB).

- Formally:

$$P(Y=1|X)=\Phi(\tau_1 - XB) - \Phi(\tau_0 - XB)$$

- The probability of observing a 1 is the proportion of the CDF (the cumulative normal distribution) associated with the first threshold minus XB, minus the proportion of the CDF associated with the zero threshold minus XB. We know that the latter term must be 0, since the zero threshold is negative infinity and therefore cuts off none of the CDF.
- So: $$P(Y=1|X)=\Phi(\tau_1 - XB)$$

- The probability of observing a 1 is the proportion of the CDF associated with the first threshold minus XB. This gives the probability of obtaining an ε large enough to push Y* over the negative infinity threshold but not so large as to push (or keep) Y* over τ_1.



Maps to: τ_1-XB 0 τ_2-XB

- Wherever XB takes Y*, an error term small enough will carry it below τ_1 and produce an observed Y of 1. This occurs with probability= $\Phi(\tau_1 - XB)$. If XB=4 and τ_1 equals 3, then error terms less than -1 will put Y* below the threshold, and Y=1. The probability of this occurring in any normal distribution is Φ (-1), or .16
- Our example: XB=2, τ_1 =.5, then P(Y=1)= Φ (-1.5)=.067. So there is a 6.7% chance of observing Y=1 for a person with XB at 2 and τ_1 at .5. **P(Y=1|X)=.067**
- We can similarly work out the Ps associated with observing Y of 2 and 3

$$P(Y=2 \mid X) = P(\tau_1 - XB \mid \leq \varepsilon < \tau_2 - XB) \mid X)$$

- which is the difference in the proportion of the CDF associated with each of the points

$$P(Y=2|X) = \Phi(\tau_2 - XB) - \Phi(\tau_1 - XB)$$

- Take case of XB=2 again. If $\tau_2 = 3$, then the chance of being in category 2 is equal to the probability the error term is less than (3-2= 1) **and** greater than (.5-2= –1.5), which is the cut-point for getting into category 1. So Y=2 is whenever the error term is between –1.5 and 1.
- Verify this on the diagram. We have the prop. CDF associated with τ_2 in a normal distribution with XB as the mean, so it is the prop. CDF associated with (τ_2 – XB) in the standard normal distribution. That gives P of being at or below the τ_2 threshold, or at or below 3. This is the *cumulative probability* of P(<=2). We then subtract from that the P of being in category 1, which is $\Phi(\tau_1 - XB)$, to get P(Y=2) exactly.
- $\Phi(1)=.841$ $\Phi(-1.5)=.067$, so **P(Y=2|X)=.774**

$$\mathrm{P}(Y=3 \mid X) = \mathrm{P}(\tau_2 - X\mathrm{B} | \leq \varepsilon < \tau_3 - X\mathrm{B}) \mid X)$$

- Which is the probability of getting an error term large enough to put Y* over the τ_2 threshold but not so large to put Y* over the τ_3 threshold. Since τ_3 is positive infinity, it is impossible to get an error term larger, so we only really need to know whether the error term is greater than (τ_2 -XB). In terms of cumulative probabilities, the P(Y=3|X) would be the entire CDF minus the proportion of the CDF associated with P(Y<=2).

$$\mathrm{P}(Y=3 \mid X) = \Phi(\tau_3 - X\mathrm{B}) - \Phi(\tau_2 - X\mathrm{B})$$

$$\mathrm{P}(Y=3 \mid X) = 1 - \Phi(\tau_2 - X\mathrm{B})$$

- Our example: **P(Y=3|X)=1-.841 = .159**
- We can generalize all the probabilities as:

$$\mathrm{P}(Y=M \mid X) = \Phi(\tau_m - X\mathrm{B}) - \Phi(\tau_{m-1} - X\mathrm{B})$$

ML Estimation of the Ordered Probit Model

- Given $\quad P(Y = M \mid X) = \Phi(\tau_m - X\mathrm{B}) - \Phi(\tau_{m-1} - X\mathrm{B})$
- we want to find the B, such that they maximize the joint probability of having observed the 1s, 2s, and 3s that we did observe in the sample

  $L = \Pi P_i$

  $L = \Pi(\Phi(\tau_m - X\mathrm{B}) - \Phi(\tau_{m-1} - X\mathrm{B})),$

  for whatever category m that case i is in
- So if case 1 is in category 1, we use the P for M=1 in the likelihood, if case 2 is in category 3, we use the P for M=3, etc.
- We find the τ and the β that maximize:

  $\ln L = \Sigma \ln(\Phi(\tau_m - X\mathrm{B}) - \Phi(\tau_{m-1} - X\mathrm{B}))$

  for the specific category j that individual i is in,

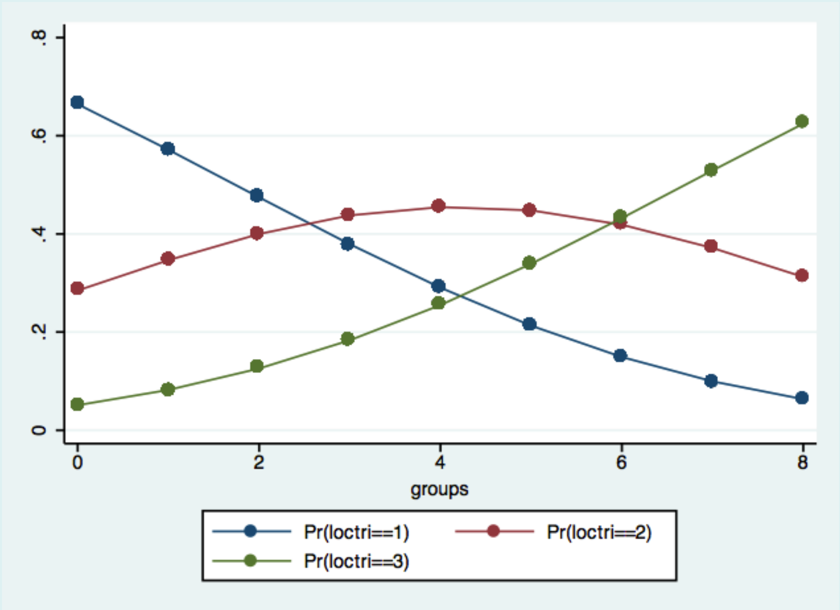
  summed over all the j categories

```
. oprobit loctri groups

Iteration 0:   log likelihood = -984.70332
Iteration 1:   log likelihood = -890.23197
Iteration 2:   log likelihood = -889.97202
Iteration 3:   log likelihood = -889.97199

Ordered probit regression                       Number of obs     =        940
                                                LR chi2(1)        =     189.46
                                                Prob > chi2       =     0.0000
Log likelihood = -889.97199                     Pseudo R2         =     0.0962

------------------------------------------------------------------------------
      loctri |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      groups |   .3386263   .0251088    13.49   0.000      .289414    .3878386
-------------+----------------------------------------------------------------
       /cut1 |   .6068731   .0742094                      .4614253    .7523208
       /cut2 |   1.844961   .0888391                      1.670839    2.019082
------------------------------------------------------------------------------
```

- Can generate predicted P(Y=1), P(Y=2), P(Y=3) for all cases, depending on the level of X. These will yield non-linear P relationships with X for each outcome
- Display normal(.607-.339*4) is P(Y=1) is .227 for a person in 4 groups
- Display normal(1.845-.339*4)-normal(.607-.339*4) is P(Y=2) is .461 for a person in 4 groups
- Display 1- normal (1.845-.339*4)is P(Y=3) is .312 for a person in 4 groups

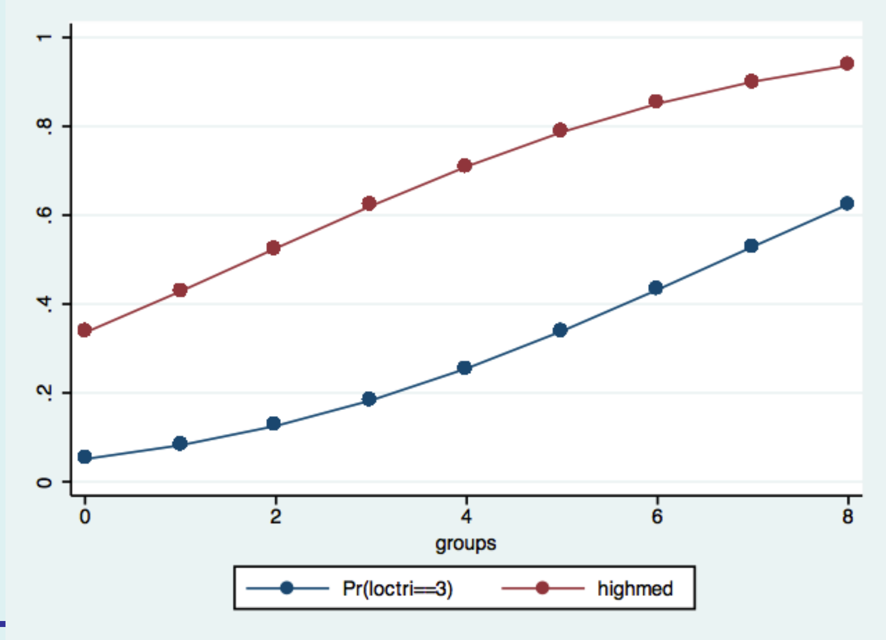
Additional Interpretations and Model Fit

- Individual significance tests based on LR or Wald
- Summary statistics for model and model comparisons
  - LR tests of nested models
  - McFadden's R-squared, McKelvey-Zavoina
  - Count and Adjusted Count R-squared
  - AIC and BIC entropy measures
- Use "MCHANGE" for changes in P(Y=M) as X changes by 1 unit, 1 standard unit, or marginal change
- Plot effects for better visualization
  - Effects of changes in variables on probability of being in categories 1/2/3 etc
  - Marginal effects of variables on all categories (mchangeplot in SPOST/STATA)
- Can calculate effects on Y* from either a unit change or a standard unit change in X using "listcoef"
  - Use standardized Y* given that variance of Y* is affected by inclusion or exclusion of sets of variables (as we discussed in context of binary outcomes)



Predicted P for all categories at all levels of X

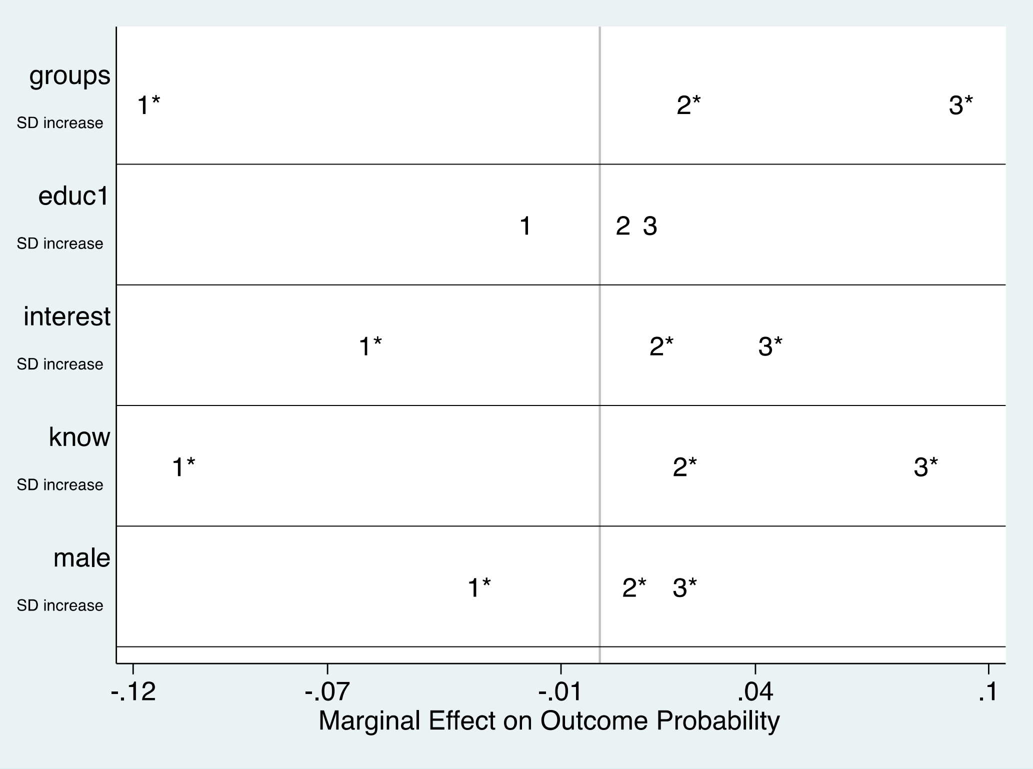
Predicted P of being in highest category, and highest or medium category, at all levels of X



Mchange and Mchangeplot for a Multivariate Model

Expression: Pr(loctri), predict(outcome())

	1	2	3
groups			
+1	-0.075	0.018	0.056
p-value	0.000	0.000	0.000
+SD	-0.116	0.023	0.093
p-value	0.000	0.000	0.000
Marginal	-0.076	0.024	0.052
p-value	0.000	0.000	0.000
educ1			
+1	-0.014	0.004	0.010
p-value	0.189	0.174	0.198
+SD	-0.019	0.006	0.013
p-value	0.188	0.166	0.200
Marginal	-0.014	0.004	0.009
p-value	0.191	0.193	0.191
interest			
+1	-0.084	0.020	0.064
p-value	0.000	0.000	0.000
+SD	-0.059	0.016	0.044
p-value	0.000	0.000	0.000
Marginal	-0.086	0.028	0.058
p-value	0.000	0.000	0.000
know			
+1	-0.056	0.015	0.041
p-value	0.000	0.000	0.000
+SD	-0.107	0.022	0.084
p-value	0.000	0.000	0.000
Marginal	-0.057	0.018	0.039
p-value	0.000	0.000	0.000
male			
1 vs 0	-0.066	0.020	0.046
p-value	0.010	0.008	0.013



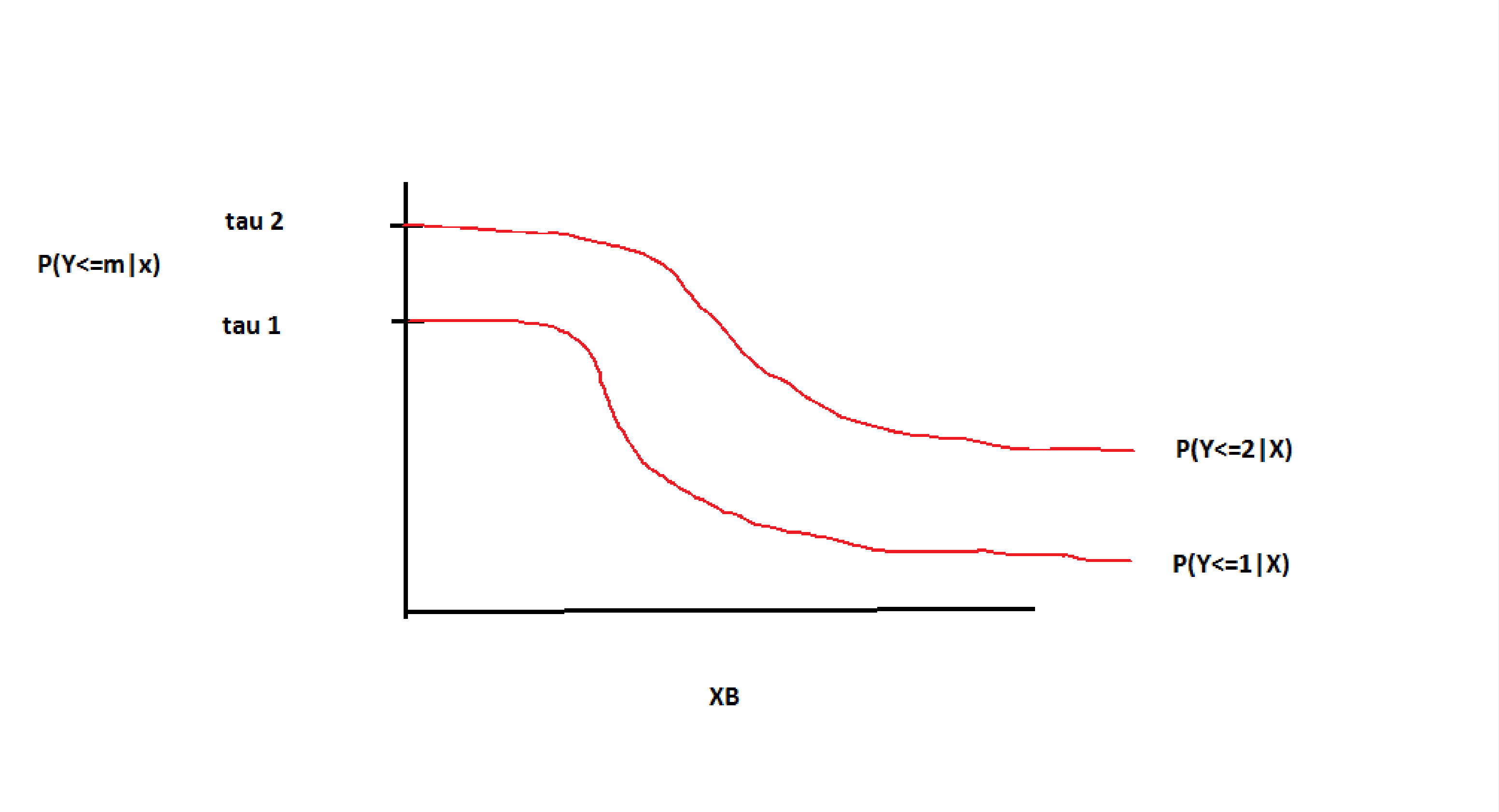
Note: Default change in mchangeplot is 1 SD change in the IV – can change this

Ordered Logit

- We can arrive at the same kind of probability model for ordered categorical variables by extending our earlier logit framework
- In the binary case we modeled the odds or the probability that a person/case is in category 1 versus category 0
- In the ordered case, we model the odds for the *cumulative probability* that a case is in category 1 or below, category 2 or below, etc.
- For a three-category variable (low (1), Medium (2), High (3)):
- We say that the:
  - *cumulative probability* of being in category 1 is P(Y=1)
  - *cumulative probability* of being in category 2 is P(Y=1)+P(Y=2)
  - *cumulative probability* of being in category 3 is P(Y=1)+P(Y=2)+P(Y=3)=1
- If J is the number of categories, *j* is the depiction of individual categories j=1…J, and m is category m, then:

$$P(Y \le m \mid X) = \Sigma P(Y = j \mid X) \text{ for j=1 to m}$$

- So if 3 categories, we get 2 cumulative probabilities P(Y<=2), P(Y<=1)
- If these cumulative probabilities are non-linear in relationship to X, bounded by 0 and 1 with Xs being unbounded, we can arrive at the same non-linear specification as in binary logit



- Why is curve sloped downward? We assume a positive relationship with X, that means that as X increases, the cumulative probability of being in category 1 **decreases**, and the cumulative probability of being in category 2 **decreases** as well (because Y will be higher)
- The ordered logit model:

$$P(Y \leq m \mid X) = \frac{\exp^{\tau_m - X\text{B}}}{1 + \exp^{\tau_m - X\text{B}}}$$

- **Having a negative XB here means that increases in X make the cumulative p smaller, i.e., a *positive* substantive relationship**
- NOTE: This would have looked the same had we done **regular dichotomous DV logit** by predicted P(Y=0) versus P(Y=1), instead of P(Y=1) versus P(Y=0), with τ=0

$$P(Y = 0 \mid X) = \frac{\exp^{-X\text{B}}}{1 + \exp^{-X\text{B}}} = \frac{1}{1 + \exp^{X\text{B}}}$$

- In terms of odds:

Cumulative Odds of being in category less than or equal to m, compared to being in category greater than m:

$$\frac{P(Y \leq m \mid x)}{P(Y > m \mid x)} = \exp^{\tau_m - X\mathrm{B}}$$

- Taking logs of both sides gives the log-cumulative odds as:

$$\ln \frac{P(Y \leq m \mid x)}{P(Y > m \mid x)} = \tau_m - X\mathrm{B}$$

- So ordered logit is linear in the cumulative logits, or log-cumulative odds that Y is in category m or lower
- Increases in X lead the cumulative logit to **decrease** by B amount, which means that the odds are smaller that the case is in lower categories as X gets larger (if B is positive, that is)

- We can derive probabilities of being in each category

$$P(Y=1\,|\,X)=\frac{\exp^{\tau_1-X\mathrm{B}}}{1+\exp^{\tau_1-X\mathrm{B}}}$$ which is the same as the cumulative P(Y<=1)

$$P(Y=2\,|\,X)=\frac{\exp^{\tau_2-X\mathrm{B}}}{1+\exp^{\tau_2-X\mathrm{B}}}-\frac{\exp^{\tau_1-X\mathrm{B}}}{1+\exp^{\tau_1-X\mathrm{B}}}$$

or the cumulative probability of P(Y<=2) minus the cumulative P(Y<=1)

$$P(Y=3\,|\,X)=1-\frac{\exp^{\tau_2-X\mathrm{B}}}{1+\exp^{\tau_2-X\mathrm{B}}}$$

which is 1 minus the cumulative probability of P(Y<=2)

Example of Ordered Logit

```
Iteration 0:   log likelihood = -984.70332
Iteration 1:   log likelihood = -890.69135
Iteration 2:   log likelihood = -889.30178
Iteration 3:   log likelihood = -889.29858
Iteration 4:   log likelihood = -889.29858

Ordered logistic regression                     Number of obs     =        940
                                                LR chi2(1)        =     190.81
                                                Prob > chi2       =     0.0000
Log likelihood = -889.29858                     Pseudo R2         =     0.0969
```

loctri	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
groups	.5789225	.0442994	13.07	0.000	.4920973	.6657477
/cut1	1.033894	.12517			.788565	1.279223
/cut2	3.129991	.1637335			2.80908	3.450903

- So cumulative logit for category 1: 1.034-.579*groups
- Cumulative logit for category 2: 3.13-.579*groups
- Predicted Probability for category 1: exp(1.034-.579*groups)/(1+ exp(1.034-.579*groups))
- Predicted Probability for category 2:

exp(3.13-.579*groups)/(1+ exp(3.13-.579*groups)) - exp(1.034-.579*groups)/(1+ exp(1.034-.579*groups))

- Predicted Probability for category 3: 1- (exp(3.13-.579*groups)/(1+ exp(3.13-.579*groups))

- Odds Interpretation:
- As groups increase by one unit, the cumulative odds of being in category *m or below* versus being *above category m* changes by a constant factor of exp(-β).
- Here it is exp(-.579)=.561
- So: probabilities for groups=4 person: category 1: .217 2: .476 3: .307

  cumulative odds for groups=4 1: .274 2: 2.25 3: 1
- So: probabilities for groups=3 person: category 1: .331 2: .470 3: .199

  cumulative odds for groups=3: 1: .495 2: 4.03 3: 1
- Cumulative odds of being at or below category 1 or below versus above for a groups=4 person compared to a groups=3 person is .274/.495**=.55**
- Cumulative odds of being at or below category 2 or below versus above for a groups=4 compared to groups=3 person is 2.25/4.05**=.56**
- Easier interpretation is to express in terms of the **greater** odds of being above category *m* versus in a category *at or smaller than category m* as exp(β), or 1.78 in this case
- This is what is given by "listcoef" in STATA, with other standardized effects

The "Parallel Regression" or "Proportional Odds" Assumption

- Important assumption of Ordered Logit or Probit: There is only one β for each X – that is, the lines are *parallel* for the cumulative Ps for all categories. This means that the effect of X on getting into category 1 is the threshold for category 1 – XB (or the change in the cumulative log-odds for category 1), the effect of getting into category 2 (or below) is the threshold for category 2-XB (or the change in the cumulative log-odds for category 2), etc., **and the β are the same for all of these calculations**. It is *not* the case that in predicting P(Y=2) that you use a different value of β than you use to predict P(Y=3), etc.
- So changing from groups =3 to groups=4 leads to
  - A factor change in the cumulative odds of being *at/below* versus being above category 1 of .56 (.274/.495) or a 1.78 factor change in the odds of being *greater* than category 1 versus at or below (.495/.274)
  - A factor change in the cumulative odds of being *at/below* versus being above category 2 of .56, or 2.25/4.03 or a 1.78 factor change in the odds of being *greater* than category 2 versus at or below (4.03/2.25)
- The change in cumulative odds at or below versus above is a constant .56 (or 1.78 change in being above versus at or below) **NO MATTER WHICH CATEGORY OR CATEGORIES YOU ARE TALKING ABOUT**
- **This is the "parallel regression" or "proportional odds" assumption**

- Could imagine situations, though, where cumulative P (or odds) of being at or below category 1, e.g., would be affected by X to one degree, and the cumulative P of being at or below category 2, e.g., would be affected by X to a different degree
- Could run a bunch of different bivariate logits based on the different cumulative P values, and compare the coefficients for X to see if they are (nearly) identical

- "Brant" Test available in Stata: brant,detail

```
Estimated coefficients from binary logits

------------------------------------
    Variable |   y_gt_1     y_gt_2
-------------+----------------------
      groups |    0.544      0.613
             |    10.84      10.09
       _cons |   -0.963     -3.257
             |    -7.22     -14.63
------------------------------------
                        legend: b/t

Brant test of parallel regression assumption

             |     chi2     p>chi2      df
-------------+-----------------------------
         All |     1.09      0.296       1
-------------+-----------------------------
      groups |     1.09      0.296       1

A significant test statistic provides evidence that the parallel
regression assumption has been violated.
```

- If significant, proportional regression assumption is violated, and you need to move to alternatives. There are many proposed alternatives.
- Possibilities depend on two factors – see Fullerton (2009), "A Conceptual Framework for Ordered Logistic Regression Models"
  - Whether the parallel regression/proportional odds assumption is thought to be violated by some, or by all independent variables
  - Whether the comparisons desired are between "cumulative odds" for different outcomes (as we've been discussing), or between odds for different stages or sequences on the dependent variable (e.g., educational attainment with BA as one "stage", "MA" another", "PhD" another), or between outcomes from adjacent categories only (e.g., a three category AGREE/NEITHER/DISAGREE attitudinal measure where the interest may be in "agree vs. neither" versus "disagree versus neither" (along with "agree versus disagree")

Generalized Ordered Logit

- Most general model relaxing proportional regression/odds assumption while maintaining some "ordinality" in the outcome categories is the "generalized ordered logit" model, implemented in Stata as "gologit2".
- Instead of $P(Y \leq m \mid X) = \frac{\exp^{\tau_m - XB}}{1 + \exp^{\tau_m - XB}}$ and $\ln \frac{P(Y \leq m \mid x)}{P(Y > m \mid x)} = \tau_m - XB$
- We allow the beta coefficients to vary for each *m* outcome (up to category J-1)

$$P(Y \leq m \mid X) = \frac{\exp^{\tau_m - XB_m}}{1 + \exp^{\tau_m - XB_m}} \qquad \ln \frac{P(Y \leq m \mid x)}{P(Y > m \mid x)} = \tau_m - XB_m$$

- So there are J-1*X additional parameters in the GOL model than traditional ordered logit. Can test the overall significance of the "full" (GOL) versus "reduced" (OL) model with LR or Wald tests
- Gologit2 can inductively test parallel regression for each IV and relax as needed – if only some IVs have different coefficients per category, we arrive at the "partial proportional odds" model, which is also nested within GOL

- We can derive probabilities of being in each category in the generalized ordered model

$$P(Y=1|X)=\frac{\exp^{\tau_1 - X\mathrm{B}_1}}{1+\exp^{\tau_1 - X\mathrm{B}_1}}$$ which is the same as the cumulative P(Y<=1)

$$P(Y=2|X)=\frac{\exp^{\tau_2 - X\mathrm{B}_2}}{1+\exp^{\tau_2 - X\mathrm{B}_2}}-\frac{\exp^{\tau_1 - X\mathrm{B}_1}}{1+\exp^{\tau_1 - X\mathrm{B}_1}}$$

or the cumulative probability of P(Y<=2) minus the cumulative P(Y<=1)

$$P(Y=3|X)=1-\frac{\exp^{\tau_2 - X\mathrm{B}_2}}{1+\exp^{\tau_2 - X\mathrm{B}_2}}$$

which is 1 minus the cumulative probability of P(Y<=2)

- Issues:
  - No formal constraint that the Ps be bounded by 0/1
  - No formal constraint that the predictions maintain ordinality of the outcome, so, e.g., increasing X by one unit can theoretically lead to very large increases in the probability of observing both categories 1 and 3 and decreases in the probability of observing category 2, depending on the specific β coefficients for that category. So the model becomes similar to multinomial logit in terms of the number of estimated parameters, and differs only in treating the cumulative nature of the Ps and odds as a conceptual assumption at the outset, though not constrained empirically to be in accordance with the assumptions
  - But the flexibility in imposing and relaxing constraints on the β is attractive, straightforward comparisons to ordinal logit and partial proportional odds models (which are nested within GOL), etc.