Notes on Ordered Logit Regression

Basics – OLS Problems with Ordered Dependent Variables

OLS regression assumes interval level measurement for the dependent variable. This is not appropriate for modeling ordinal (rank-order) variables.

Nevertheless, it is common for researchers to treating ordinal dependent variables as interval-level variables and model them using OLS regression. This modeling approach assumes effects are linear and additive. This is inappropriate and in some situations can lead to obvious problems such as predictions that are not logically possible (i.e., predicted values below the minimum possible score and/or predicted values above the maximum possible score).

This can lead to several problems.

Estimates of effects can be biased because the outcomes should be modeled as nonlinear and nonadditive. Predictions can be inaccurate, most obviously at high and low levels.

Statistical hypothesis tests may be unsound because standard OLS assumptions about the error term are not met.

If the underlying variable is truly interval-level, the use of an ordinal-level measure and treating it as an interval-level measure in OLS regression introduces measurement error which can be random and/or nonrandom – especially in the tails of the distribution.

Best case scenario for using OLS regression:

The underlying conceptual variable is interval-level in nature.

The ordinal representation of the variable uses many categories (e.g., 8 or more) and cases are distributed somewhat evenly across the categories.

Top-coding and bottom-coding does not distort the tails of the distribution.

The ordinal categories generally reflect evenly-spaced, interval level distances.

Even so, using categories instead of interval scores will introduce random error by obscuring interval variation within categories.

This will attenuate estimates of effects (i.e., it will bias b's toward zero).

Similarly, it will attenuate estimates of strength of association and predictive success (i.e., it will bias standardized coefficients, r-square, and correlations toward zero).

Intrinsic nonlinearity and nonadditivity in the effects of X's will be relatively modest.

Statistical hypothesis tests may be considered "informative." But they are not strictly sound because standard OLS assumptions about the error term are not met. This is less of a concern when sample sizes are very large.

Worst case scenario of OLS regression

The underlying conceptual variable is not truly interval with "distances" between categories either being not meaningful or being uncertain or inconsistent.

This violates an important OLS assumption that numeric scores for the categories reflect equal differences.

Categories are limited to only a small number (e.g., 2-5)

Top-coding and bottom-coding significantly truncates the tails of the distribution.

Cases are concentrated in a few categories.

The above three items result in larger amounts of random measurement error in the dependent variable compared to true interval level measurement.

As noted earlier, this will attenuate estimates of effects (i.e., it will bias b's toward zero).

Similarly, it will attenuate estimates of strength of association and predictive success (i.e., it will bias standardized coefficients, r-square, and correlations toward zero).

The difference here is that these unwanted consequences will be more dramatic.

Intrinsic nonlinearity and nonadditivity in the effects of X's will be very strong.

Statistical hypothesis tests will be clearly unsound because standard OLS assumptions about the error term are not met.

Alternatives to OLS Regression – Logit Regression and Ordered Logit Regression

In the extreme version of the worst case scenario, the dependent variable is measured using only two categories. OLS regression will potentially have severe problems and limitations. Logit regression would be the better option.

In other worst case scenarios (e.g., with three or more categories that do not capture the interval variation in Y very well), OLS regression will have severe problems and limitations.

If the situation is so bad the rank ordering of categories of Y is questionable, multinomial logit regression is be the better option.

If rank ordering of categories is defensible, ordered logit regression is the preferred option. It is superior to two-category logit regression because it retains more information about the dependent variable.

In the best case scenario, the problems and limitations of OLS regression may be moderate. Nevertheless, ordered logit regression may still be preferred because it is superior for performing statistical tests.

Overview of Ordered Logit Regression (Assumes familiarity with Multinomial Logit Regression)

Ordered logit regression is a method developed to fulfill the goal of modeling the relative frequency distribution of cases across three or more ranked categories of the dependent variable (Y).

In contrast to multinomial logit regression, which assumes categories cannot be ranked, ordered logit regression incorporates the assumption that there is a definite ordering of the categories of the dependent variables.

When this assumption is appropriate, the ordered logit regression is simpler and easier to interpret than the multinomial logit regression. At the same time, it has more sound statistical properties – especially with regard to statistical hypothesis testing – than OLS regression.

The specific modeling task of ordinal logit regression is to predict the expected relative frequency distribution of cases across the ranked categories of the dependent variable under any combination of values on relevant independent variables (X's).

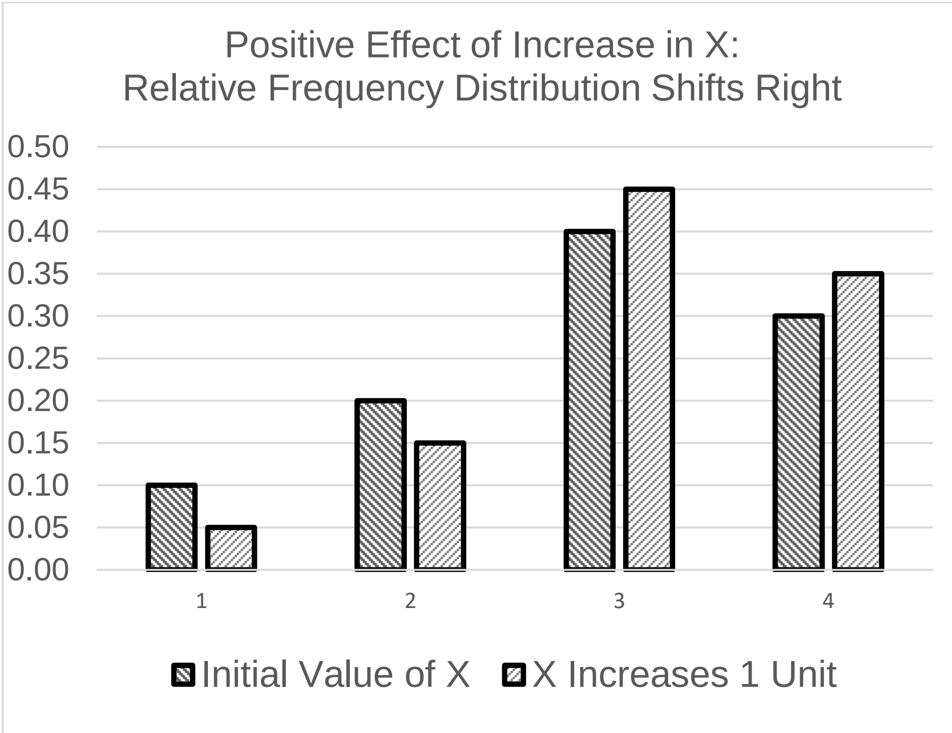
We saw in multinomial regression, the distribution of cases across K categories can be perfectly captured using K-1 relative risk ratio equations. Thus, the simplest multinomial equation (with no predictor (X) variables) has K-1 equations, each one of which estimates a "constant" which reflects the log of the relative risk ratio of a given category to the chosen baseline category.

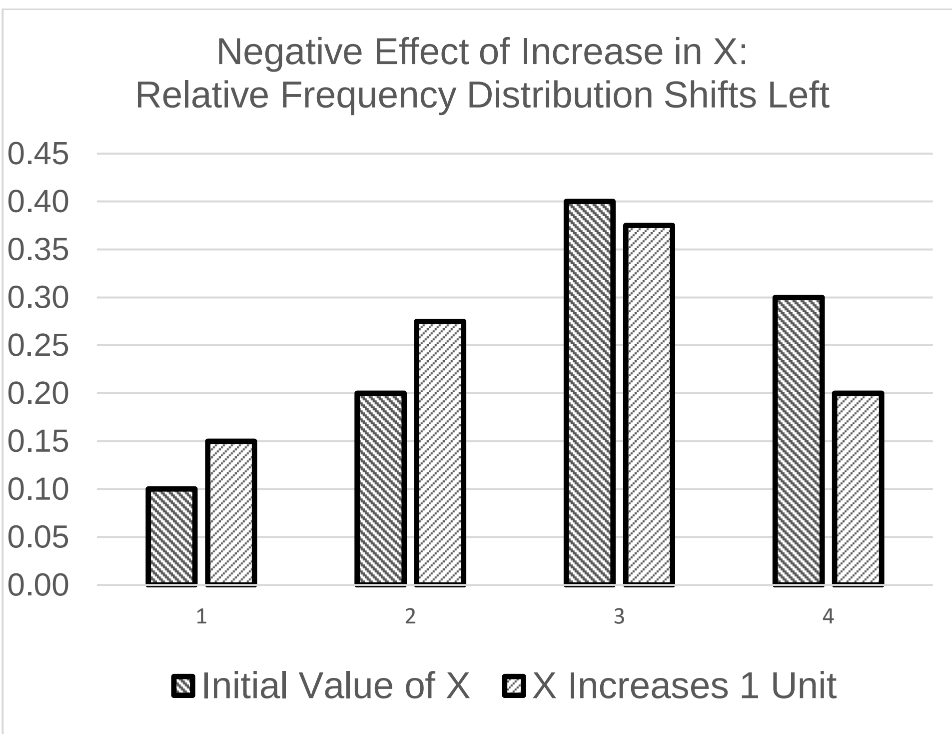
When one or more predictors (X's) is added to the model, a separate coefficient is estimated for each the predictor in each one of the K-1 risk ratio equations. The problem with the resulting equation is that the ordinal quality of the dependent variable is not recognized. As a result, the effects of predictors (X's) can change the shape of the relative frequency distribution across values of the dependent variable (Y) in complex ways that are not consistent with the ordinal nature of the variable.

Ordered logit regression incorporates the assumption that the dependent variable has ordinal properties. Thus the effect of a predictor (X) will shift the distribution of cases across the categories of the dependent variable (Y) in a systematic direction toward higher or lower categories. Specifically, the relative frequency distribution of cases will systematically shift toward higher categories if X has a positive effect and it will systematically shift toward lower categories if X has a negative effect. These outcomes are depicted with in the example figures below.

Ordered logit regression is simpler than multinomial logit regression because effects of independent variables (X's) are estimated by a single coefficient. The coefficient changes the relative frequency distribution of Y by increasing (or decreasing) the values of a set of "cut-point" coefficients. The "cut-point" coefficients reflect the expected ratios of cases across the "cut-points" in the distribution of Y when all X's are zero. This can be understood as the "baseline" or "reference" shape of the relative frequency distribution of cases across categories of Y.

The coefficient estimated for a predictors (X) impacts the relative frequency distribution of cases across categories of Y as follows. A one-unit change in X increments *all* of the cut-point coefficients by an amount equal to the value of the coefficient for X. This has the consequence of shifting cases up toward higher categories of Y when the effect of X is positive and shifting cases toward lower categories of Y when the effect of X is negative.


Positive Effect of Increase in X:
Relative Frequency Distribution Shifts Right
0.50
0.45
0.40
0.35
0.30
0.25
0.20
0.15
0.10
0.05
0.00
1
2
3
4
Initial Value of X
X Increases 1 Unit


Negative Effect of Increase in X:
Relative Frequency Distribution Shifts Left
0.45
0.40
0.35
0.30
0.25
0.20
0.15
0.10
0.05
0.00
1
2
3
4
Initial Value of X
X Increases 1 Unit

The initial or baseline relative frequency distribution for Y is captured by K-1 logit "cut point" coefficients. These are analogous to the K-1 "constants" (logit coefficients) that are estimated in the multinomial logit model. But there is an important difference. The logit coefficient constants in the multinomial logit model are logarithms of relative risk ratios for the expected cases in a given category to the expected cases in a single, arbitrarily chosen "base" or "reference" category.

In contrast, the logit cut-point coefficients in the ordinal logit model reflect the natural logarithm of the ratio of the predicted fraction of cases above the cut-point to the fraction of cases below the cut-point. A separate coefficient is estimated for each logically possible "cut-point" in the distribution of Y.

In the multinomial logit regression model the constants predict the expected relative frequency distribution of cases across categories of the dependent variable (Y) when all predictors (X's) are zero (0).

The same is true for the constants in the ordinal logit regression model. Specifically, for a dependent variable with K ranked categories, the (K-1) cut point coefficients predict the relative frequency distribution of the dependent variable (Y) when all predictors (X's) are zero (0).

An important difference between the multinomial logit model and the ordered logit model is seen when the model incorporates a predictor (X). In the multinomial model, K-1 equations are estimated – one equation for each relative risk ratio – and a separate coefficient for X is estimated for each equation. This estimates a large number of effect coefficients which can capture complex patterns in ratios of cases across different categories of the dependent variable (Y).

In the ordered logit model, a single coefficient for X is estimated. This single coefficient is applied to all of the cut-point values. This assures that the ratios of cases in higher categories to cases in lower categories is systematically shifting up (right) or down (left).

If the dependent variable (Y) has true ordinal properties, the single, ordered logit coefficient will capture the pattern of change in the relative frequency distribution of the dependent variable (Y) across values of the predictor (X). The key assumptions are that: (a) the effect of X is systematic with respect to direction and (b) the effect of X is proportional across all cut points.

If the effect of X is not systematic with respect to direction, the multinomial logit model will be superior because it can capture a wider range of patterns while ordered logit model can only capture effects that are systematic with respect to direction.

Similarly, if the effect of X has a uniform direction, but is not proportional across all cut points, the multinomial logit model will again be superior because it can capture systematic directional effects while allowing them to vary in more complex ways than can be captured using the basic ordered logit model.

The ordered logit model is "nested" under the multinomial logit model. So tests comparing the results of the two models will indicate whether the assumptions of simpler ordered logit regression are reasonable. An application of this kind of test is presented in a later section of this document.

Importantly, note that the ordered logit model can capture more complex patterns if one incorporates nonlinear effects of predictors (X's) and interactions among predictors (X's).

The next sections discuss the empirical examples of how ordinal logit regression coefficients reflect "baseline" relative frequency distributions of Y and effects of X's.

Stata Programs with Examples and Demonstration Analyses

The Stata program "w11_ologit_demo1.do" provides examples of how to run ordered logit models, perform significance tests for the effects of predictors (X's), and generate predicted relative frequency distributions that reveal the substantive implications of the model.

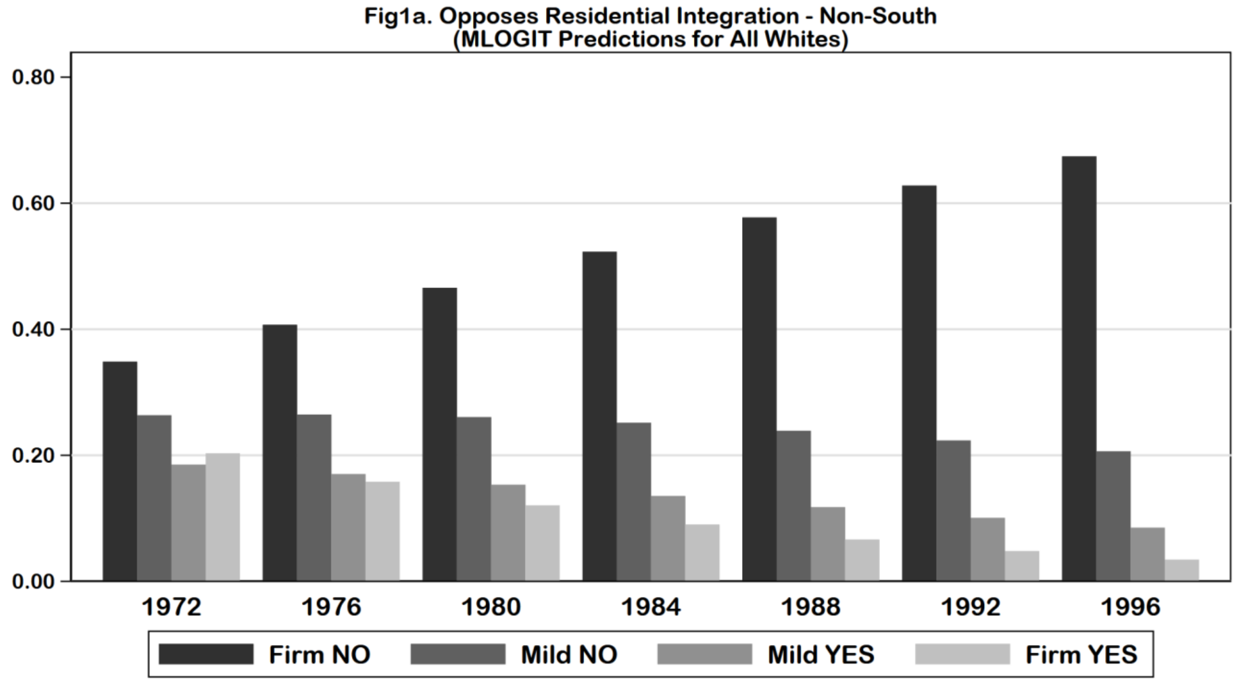
The program "w11_ologit_demo2.do" provides additional examples of logit regression analyses.

The data for the analysis executed in these programs are taken from the General Social Survey. The analysis data set is for the subset of white respondents. The dependent variable is a categorical variable (OPPINT4) that measures a respondent's opposition to racial residential integration based on four ranked categories. Specifically, OPPINT4 measures a respondent's level of agreement with the view that "It is OK for Whites to oppose neighborhood integration". The four categories of OPPINT4 are coded 1 for the lowest level of agreement and 4 for the highest level of agreement.

> The exact GSS question wording is "White people have a right to keep Blacks out of their neighborhoods if they want to, and Blacks should respect that right."
>
> Responses are coded as follows: Strongly Disagree (firm no) = 1, Disagree (mild no) = 2, Agree (mild yes) = 3, and Strongly Agree (firm yes) = 4.

The analyses use the Stata ologit procedure to estimate how opposition to integration (or, alternatively, support for maintaining segregation) varies over time and region. Positive effects shift the relative frequency distribution toward the higher categories; negative effects shift the relative frequency distribution toward the lower categories. If positive effects are very powerful, cases will be concentrated in the highest category (category 4). If negative effects are very powerful, cases will be concentrated in the lowest category (category 1).

Figures generated by the program follow next. The figures represent the effects of the independent variables by showing how a bar chart representing the predicted distribution of cases across four ranked categories changes shape across time and region. (See the associated log file for other related results generated by the program (e.g., lists of predicted values).



Source: National Opinion Research Center General Social Survey. Notes: Sample consists of White adults.

Figure 1a (above) depicts predictions from multinomial logit (mlogit) for the subset of respondents in the Non-South. The figure provides insight into whether ordered logit is appropriate. If ordered logit is appropriate, changes will follow a specific form; the height of the leftmost bars will change in the opposite direction from the height of the rightmost bars. This pattern is clearly evident. As survey year increases, the leftmost bar increases in height and the rightmost bar decreases in height. Accordingly, ordered logit appears appropriate.

Figure 1b (below) depicts predictions from multinomial logit (mlogit) for the subset of respondents in the South. As before, the figure provides insight into whether ordered logit is appropriate. If ordered logit is appropriate, changes will follow a specific form; the height of the leftmost bars will change in the opposite direction from the height of the rightmost bars. Again, this pattern is clearly evident. As survey year increases, the leftmost bar increases in height and the rightmost bar decreases in height. Ordered logit appears to be appropriate.

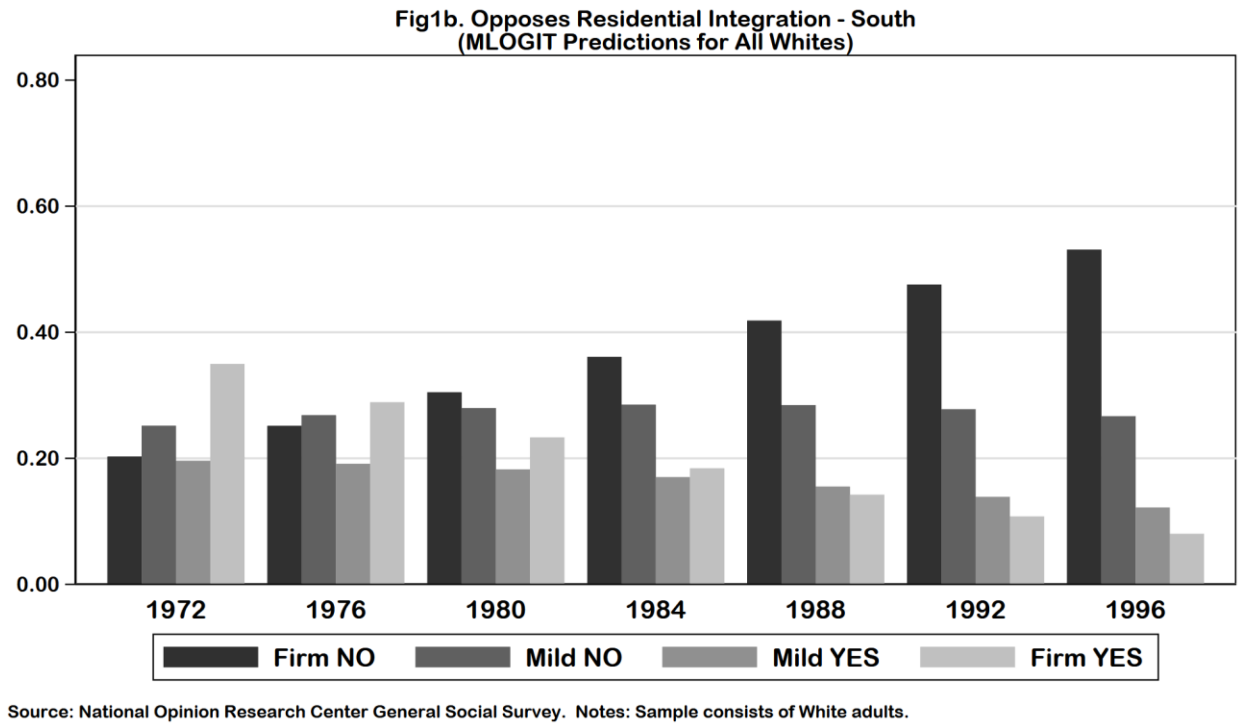
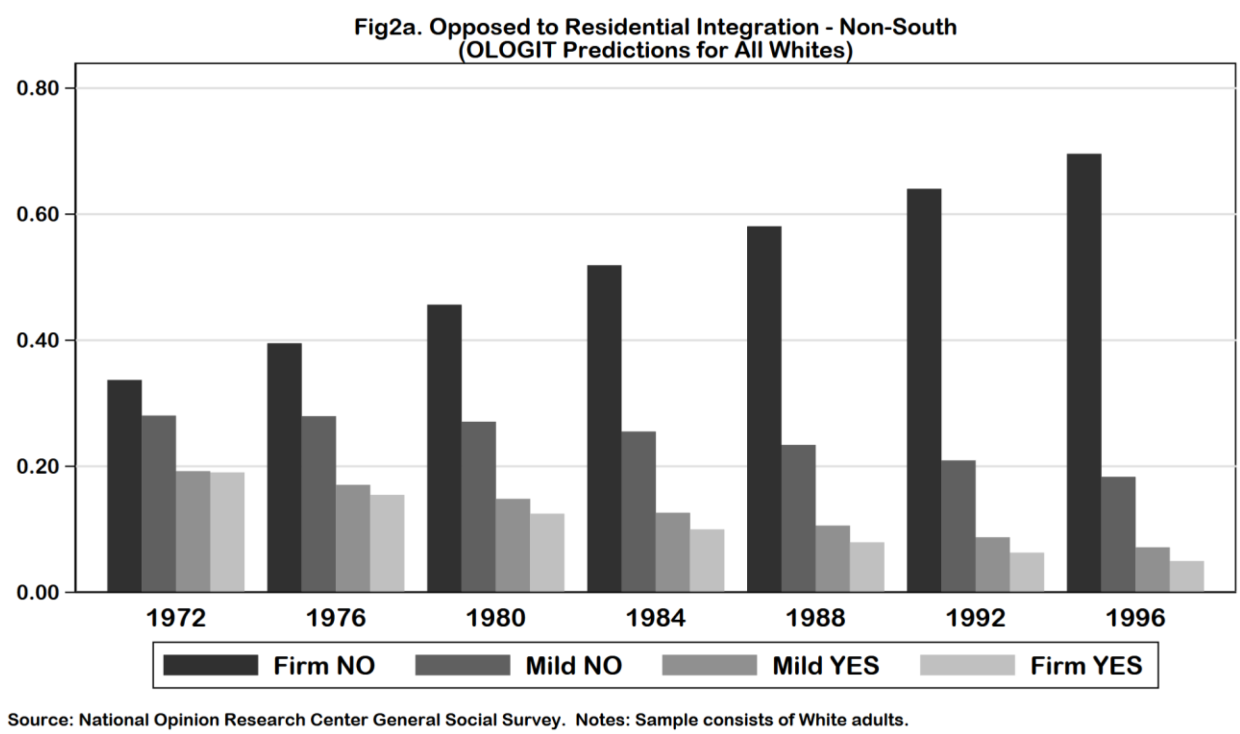
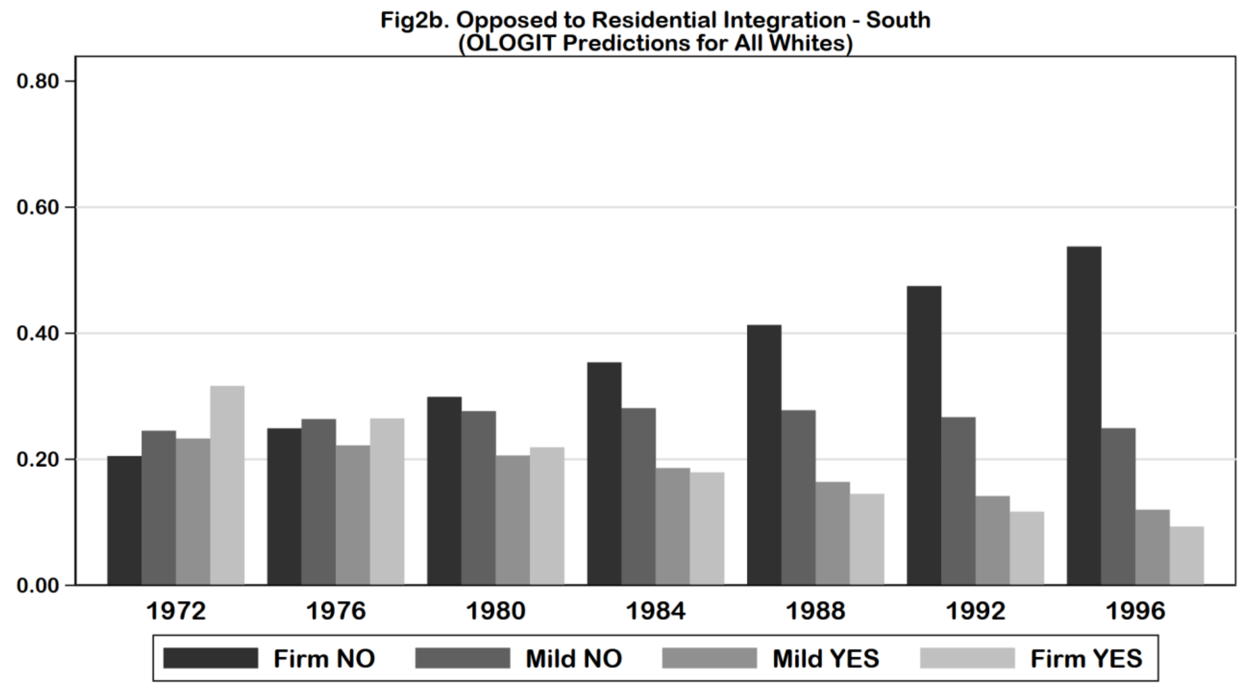


Figure 2a (below) depicts predictions from ordered logit (ologit) for respondents in the Non-South. The ordered model inherently *requires* that effects impact the relative frequency distribution of opposed to integration in a particular way; specifically, the height of leftmost bars *must* change in the opposite direction from the height of rightmost bars. Accordingly, as survey year increases, the leftmost bar increases in height and the rightmost bar decreases in height.



Here time (year) has a clear negative effect on support for opposing integration. As survey year progresses from lower to higher, the bar charts representing level of opposition to integration systematically shift left toward lower values of the dependent variable.

The comparison with the earlier figure for multinomial logit predictions (Figure 1a) suggests the simpler ordered logit model captures the pattern of change about as well as the more complex multinomial logit model. In this situation, the ordinal logit model is preferred because it is more parsimonious (simpler) and is easier to interpret.



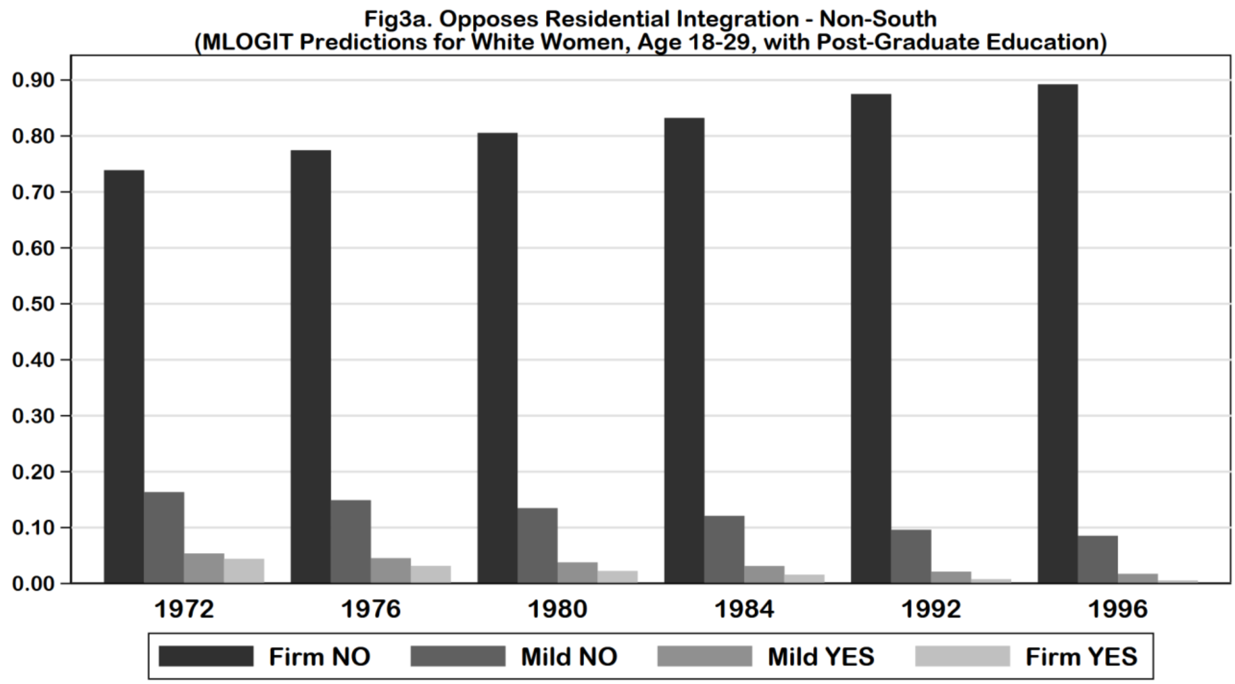
Source: National Opinion Research Center General Social Survey. Notes: Sample consists of White adults.

Figure 2b depicts predictions from ordered logit (ologit) for respondents in the South.

Again, the ordinal logit model inherently *requires* that effects impact the relative frequency distribution of opposition to integration in a specific way; the height of leftmost bars *must* change in the opposite direction from the height of rightmost bars. This pattern is clear. As survey year increases, the leftmost bar increases in height and the rightmost bar decreases in height.

Again time (year) has a negative effect on support for opposing integration. As survey year progresses from lower to higher, the bar charts representing opposition to integration systematically shift left toward lower values of the dependent variable.

Also as before, the comparison with the figure for multinomial logit predictions (Figure 1b) suggests the simpler ordered logit model captures the pattern of change as well as the more complex multinomial logit model. In this situation, the ordinal logit model is preferred because it is more parsimonious (simpler) and is easier to interpret.



Source: National Opinion Research Center General Social Survey. Notes: Sample consists of White adults.

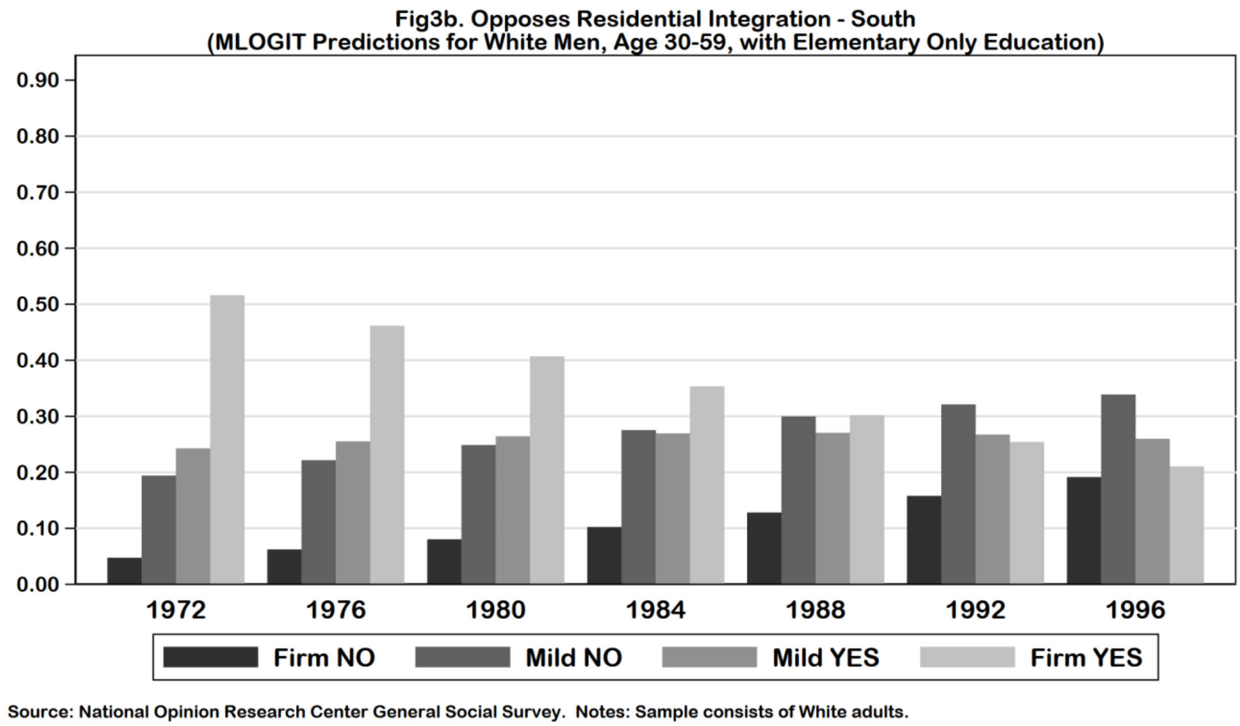
Figure 3a (above) depicts predictions from ordered logit (ologit) for a subset of respondents that does not support opposing integration. The predictions are for White women who are young (ages 18-29), highly educated (have post graduate education), and live in the Non-South.

The combination of values on the X variables has a strong negative impact on opposition to integration. As a consequence, the height of leftmost bars is very high in comparison to the height of rightmost bars. By 1996 respondents who are non-southern, women, young, and highly educated are almost completely concentrated in the lowest category for opposition to integration.

Figure 3b (below) depicts predictions from ordered logit (ologit) for a set of respondent that have fairly high levels of support for opposition to integration. The predictions are for White men who are older (ages 30-59), are lesser educated (elementary only – did not attend high school), and live in the South.

This combination of values on the X variables creates a strong positive effect on opposition to integration in some cases, especially at earlier survey years. The consequence is that height of leftmost bars is relatively low in comparison to the height of rightmost bars. In 1972 respondents who are southern, men, older, and lesser educated are heavily concentrated in the highest category for opposition to integration.

The comparison of this figure with the previous figure (Figure 3a) establishes that the effects of the independent variables in the analysis (time, region of residence, age, sex, and education) have important effects on opposition to integration.



Accordingly, the predicted relative frequency distribution of respondents across categories of opposition to integration is very different for some cases compared to other cases.

Because ologit assumes the categories are ordered, the differences are manifest as cases being concentrated in leftmost bars versus being concentrated in rightmost bars.

Interpreting Results of Ordered Logit Regression

In ordered logistic regression, the dependent variable (Y) is conceived as having ranked categories. The goal is to predict how cases are distributed across the ranked categories of Y based on knowledge of each case's values on independent variables X_1, X_2, ... X_k.

As an example, the distribution of cases across OPPINT4, an ordered dependent variable measuring strength of agreement with the sentiment that it is "OK for Whites to oppose neighborhood integration" is shown in Table 1.

> Please note: To simplify analysis and calculations, the analyses from this point forward are performed using only the cases for the survey year 1972.

Table 1. Distribution of Dependent Variable OPPINT4 (in 1972)

OPPINT4	N	p	cum p
1 Firm NO	440	0.3534	0.3534
2 Mild NO	305	0.2450	0.5984
3 Mild YES	222	0.1783	0.7767
4 Firm YES	278	0.2233	1.0000
	1,245	1.0000	

The values of OPPINT4 are ordinal. They register higher and lower levels of opposition to integration. However, in contrast to interval measurement, the scale "distance" from one level to the next is not assumed to be precisely reflected by the quantitative values of the scores associated with the categories.

The distribution of cases across these ordered categories can be summarized in terms of "cut points" in the ascending cumulative distribution of cases across the rank order values of Y. Each cut point divides the distribution of four ordinal scores into a simple lower-higher dichotomy. With four ranks, three low-high divisions can be created as follows.

Cut point 1 compares the cases in the category with the lowest value on OPPINT4 (1="Firm No") with the cases in the three categories with higher values (2="Mild No", 3="Mild Yes", & 4="Firm Yes");

Cut point 2 compares the cases in two categories with the lowest values on OPPINT4 (1="Firm Yes" & 2="Mild Yes") with the cases in the two categories with the highest values (3="Mild No" & 4="Firm No"); and

Cut point 3 compares the cases in the three categories with the lowest values on OPPINT4 (1="Firm Yes", 2="Mild Yes", & 3="Mild No") with the cases in the category with the highest value on OPPINT4 (4="Firm No).

The ratios of cases "below the cut" to cases "above the cut" at each of these "cut points" are summarized in Table 2.

Table 2. Adding Cut Point Odds Ratios and Logits

OPPINT4	N	p	cum p	cut points	p < cut	p > cut	odds ratio	logit = ln(odds)
1 Firm NO	440	0.3534	0.3534	-----	-----	-----	-----	-----
2 Mild NO	305	0.2450	0.5984	cut1	0.3534	0.6466	0.5466	-0.6041
3 Mild YES	222	0.1783	0.7767	cut2	0.5984	0.4016	1.4900	0.3988
4 Firm YES	278	0.2233	1.0000	cut3	0.7767	0.2233	3.4784	1.2466
	1,245	1.0000						

Latent Variable Interpretation of Ranked Categories

Ordinal logit can be understood as modeling a latent continuous measure (y^*) that has been grouped into ranked categories based on ordered but not necessarily "equally spaced" thresholds or "cut points" in the distribution of the latent variable. In the example under consideration here, three cut points divide the distribution on the continuous latent variable y^* into four categories.

We might view the underlying continuous variable y^* as being measured on an interval scale where distances between scale points are meaningful. Or we might view the underlying continuous variable y^* as being measured on a highly granular rank order scale with many more steps than the four levels in our example. But here we have only the ability to say that a given case in one of four ranges of the categories in the more detailed distribution.

The rank order numbers assigned to the ordered categories allows us to say cases in category 1 have lower values on y^* than cases in category 2. But we cannot necessarily assume that the values of y^* that delimit the thresholds or cut points that determine how cases are assigned to the four ranked levels of y are "equally spaced". We assume only that $y_1^* < y_2^* < y_3^*$. Thus, when we establish that respondents move from a lower category to a higher category, we can only say they are higher than before. We cannot say with quantitative precision how much higher they are.

If a researcher believes the categories are delimited by threshold values of y^* that are evenly spaced on a true interval scale, linear regression could be considered as an alternative option to ordered logit.

If a researcher believes the categories are delimited by threshold values of y^* that are evenly spaced on a true rank-order or quantile scale, fractional regression could be considered as an alternative option to ordered logit.

If these strong assumptions are not reasonable, ordered logit stands as the best available option to model the underlying latent variable y^* which is measured imprecisely by the rank order values assigned to the categories of y.

Stata OLOGIT Results

The following results from Stata's ologit procedure estimated with no X variables in the model demonstrates that the "coefficients" correspond to the logit values in the "cut point table" (Table 2).

In this situation (i.e., when no independent variable is specified), the coefficients estimated by the ordered logit model simply serve to express the relative frequency distribution of the dependent variable (OPPINT) in terms of the three "logit" cut points in its observed cumulative distribution. (This would be analogous to running regression without an independent variable in which case the intercept would be estimated to be equal to the sample mean.)

```
. ologit oppint4

Iteration 0:   log likelihood =   -1686.23
Iteration 1:   log likelihood =   -1686.23

Ordered logistic regression                     Number of obs     =      1,245
Log likelihood =   -1686.23                     Pseudo R2         =    -0.0000

------------------------------------------------------------------------------
     oppint4 |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       /cut1 |  -.6040676   .0592871                     -.7202682   -.4878669
       /cut2 |   .3987761   .0578125                      .2854658    .5120865
       /cut3 |   1.246577   .0680533                      1.113195    1.379959
------------------------------------------------------------------------------
```

Note that the "cut-point" coefficients reported in the ordered logit regression results are identical to the logit values reported in Table 2 above.

The relative frequency distribution on Y can be generated from the logit values for the cut points (-0.6041, 0.3988, and 1.2466) from this equation as follows.

1. Exponentiate (take the antilog of) the logits(e^L) to obtain the corresponding cut point odds ratios (e.g., $e^{-0.6041} = 0.5466 = o$)
2. For all cut points, calculate the "proportion above the cut" from the odds ratio (o) based on $p = o/(1+o)$ (e.g., for cut1 0.5466/(1.5466) = 0.3534)
3. For all cut points calculate the "proportion below the cut)" from 1 minus "proportion above cut" (e.g., for cut1, 1.0-0.3534 = 0.6466)
4. Assign the proportion for the first category (category 1) based on the "proportion above the cut" calculated for the first cut point (e.g., 0.3534 for cut1)
5. Assign the proportion for the last category (category 4) based the "proportion below the cut" calculated for the last cut point (e.g., 0.2233 for cut3).
6. Assign the proportion for the intermediate categories (categories 2 and 3) based on the difference of p2 and p1 where p1 is the "proportion above the cut" for the category in question and p2 is the "proportion above the cut" for the next category (e.g., for category 2 it is 0.2450 = 0.5984-0.3534)

Implementing these calculations yields the results presented in Table 3 (below). Note that the results reproduce the values obtained by "manual" calculations previously reported in Table 2.

Table 3. Obtain Relative Frequency Distribution from Logits

OPPINT4	cut points	logit = ln(odds)	odds ratio	p < cut	p > cut	p
1 Firm NO	-----	-----	-----	-----	-----	0.3534
2 Mild NO	cut1	-0.6041	0.5466	0.3534	0.6466	0.2450
3 Mild YES	cut2	0.3988	1.4900	0.5984	0.4016	0.1783
4 Firm YES	cut3	1.2466	3.4784	0.7767	0.2233	0.2233

An ordinal variable "increases" when the predicted probability distribution across categories of Y shifts from lower categories to higher categories level and it "decreases" when the predicted probability distribution across categories of Y shifts higher categories to lower categories.

Accordingly, an independent variable has a positive effect on an ordinal dependent variable (Y) when an increase in the independent variable (X) leads the relative frequency distribution of cases to shift toward higher categories of Y. An independent variable has a negative effect on an ordinal dependent variable when an increase in the independent variable leads the relative frequency distribution of cases to shift toward lower levels of Y.

Systematic (directional) shifts in the predicted relative frequency distribution of cases across categories of Y are produced by increasing or decreasing the values of the logit coefficients for the cut points.

In Stata, the ologit model "cut point" values are defined in terms of the ratio of cases below the cut point to cases above the cut point. Shifting the values of the cut points "down" by a given amount shifts the implied probability distribution of cases toward higher levels of Y. This is illustrated in Table 4 which documents how reducing the "initial" values of the logit coefficients for the cut points (previously given in Table 3) down by 0.5 produces changes in the cut-point odds ratios, the proportion of cases above and below he cut points, and the proportion of cases in different categories.

Table 4. Decrease Logits by 0.5 and Recalculate Relative Frequency Distribution

OPPINT4	cut points	initial logit = ln(odds)	initial logit−0.5 =ln(odds)	odds ratio	p < cut	p > cut	p
1 Firm NO	-----	-----	-----	-----	-----	-----	0.2490
2 Mild NO	cut1	-0.6041	-1.1041	0.3315	0.2490	0.7510	0.2257
3 Mild YES	cut2	0.3988	-0.1012	0.9037	0.4747	0.5253	0.2037
4 Firm YES	cut3	1.2466	0.7466	2.1098	0.6784	0.3216	0.3216

Similarly, shifting the values of the cut points in the above logit model "up" by a fixed amount shifts the implied probability distribution of cases toward lower levels of Y. This is documented in Table 5 which increases the values of the initial logit coefficients by 0.5.

Table 5. Increase Logits by 0.5 and Recalculate Relative Frequency Distribution

OPPINT4	cut points	initial logit = ln(odds)	initial logit+0.5 =ln(odds)	odds ratio	p < cut	p > cut	p
1 Firm NO	-----	-----	-----	-----	-----	-----	0.4740
2 Mild NO	cut1	-0.6041	-0.1041	0.9012	0.4740	0.5260	0.2367
3 Mild YES	cut2	0.3988	0.8988	2.4566	0.7107	0.2893	0.1408
4 Firm YES	cut3	1.2466	1.7466	5.7349	0.8515	0.1485	0.1485

The ordered logit model represents effects of independent variables on Y in terms of logit coefficients indicating the amount by which the log-odds for the cut points in the ordered distribution of Y will shift (up or down) with a unit change in the independent variable.

As demonstrated above, these effects in turn produce shifts in the predicted relative frequency distribution of cases across the ordered categories of Y.

Next we examine results from an ordered logit regression that predicts Y by the independent variable SOUTH (coded 1 if residing in the South and 0 otherwise). First, however, we use simple cross tabulation analysis to investigate how opposition to integration varies across region.

```
. tab oppint4 south , chi2 lrchi2

OK to Oppose
Integration    Resides in South
(No 1-4 Yes)      0=No      1=Yes       Total
-----------+----------------------+----------
   Firm-No |       365         75 |       440
   Mild-No |       231         74 |       305
  Mild-Yes |       166         56 |       222
  Firm-Yes |       165        113 |       278
-----------+----------------------+----------
     Total |       927        318 |     1,245

          Pearson chi2(3) =  50.3302   Pr = 0.000
 likelihood-ratio chi2(3) =  48.6714   Pr = 0.000


. tab oppint4 south , col nofreq

OK to Oppose
Integration    Resides in South
(No 1-4 Yes)      0=No      1=Yes       Total
-----------+----------------------+----------
   Firm-No |     39.37      23.58 |     35.34
   Mild-No |     24.92      23.27 |     24.50
  Mild-Yes |     17.91      17.61 |     17.83
  Firm-Yes |     17.80      35.53 |     22.33
-----------+----------------------+----------
     Total |    100.00     100.00 |    100.00
```

Now we examine the results of the relevant ordered logit regression.

```
. ologit oppint4 south

Iteration 0:   log likelihood =   -1686.23
Iteration 1:   log likelihood = -1663.7359
Iteration 2:   log likelihood = -1663.7036
Iteration 3:   log likelihood = -1663.7036

Ordered logistic regression                     Number of obs     =      1,245
                                                LR chi2(1)        =      45.05
                                                Prob > chi2       =     0.0000
Log likelihood = -1663.7036                     Pseudo R2         =     0.0134

------------------------------------------------------------------------------
     oppint4 |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       south |   .7953073   .1191109     6.68   0.000     .5618542     1.02876
-------------+----------------------------------------------------------------
       /cut1 |  -.4230379   .0653703                     -.5511613   -.2949144
       /cut2 |   .6040889   .0663011                      .4741412    .7340366
       /cut3 |   1.475783   .0776177                      1.323655    1.627911
------------------------------------------------------------------------------
```

Here the coefficients for the "cut points" (i.e., cut1, cut2, and cut3) summarize the expected relative frequency distribution of OPPINT4 when SOUTH is 0 (that is, for respondents not residing in the South). This is demonstrated in the Panel for South=0 in Table 6 below.

The logit coefficient for SOUTH indicates the amount by which the logit values for the cut points should be adjusted to generate the expected relative frequency distribution of OPPINT4 when SOUTH is 1. The adjusted logit coefficients for South are shown in the Panel for South=1 in Table 6 along with the implied odds and probabilities of being above and below the cut points.

Table 6. Predicted Cut Point Logits When South = 0 and 1 Applied to Obtain Relative Frequency Distributions

			Predictions				
OPPINT4	cut points	logit = ln(odds)	logit–b(X) = ln(odds)	odds ratio	p < cut	p > cut	p
South = 0							
1 Firm NO	-----	-----	-----	-----	-----	-----	0.3958
2 Mild NO	cut1	-0.4230	-0.4230	0.6551	0.3958	0.6042	0.2508
3 Mild YES	cut2	0.6041	0.6041	1.8296	0.6466	0.3534	0.1673
4 Firm YES	cut3	1.4758	1.4758	4.3744	0.8139	0.1861	0.1861
South = 1							
1 Firm NO	-----	-----	-----	-----	-----	-----	0.2282
2 Mild NO	cut1	-0.4230	-1.2184	0.2957	0.2282	0.7718	0.2241
3 Mild YES	cut2	0.6041	-0.1912	0.8259	0.4523	0.5477	0.2115
4 Firm YES	cut3	1.4758	0.6805	1.9748	0.6638	0.3362	0.3362

This illustrates how the logit coefficients in the results for the ologit model imply predictions regarding the relative frequency distribution of Y by region (as shown in Table 6).

The predicted proportions for the distribution of cases across categories of Y can be generated more easily by using Stata's "predict" command. This is illustrated as follows.

```
PREDICTIONS FROM OLOGIT OPPINT4 SOUTH

ologit oppint4 south
  ... regression results omitted (previously presented above)

predict pvo2p1 pvo2p2 pvo2p3 pvo2p4 , p

. table south , c(mean pvo2p1 mean pvo2p2 mean pvo2p3 mean pvo2p4) format(%8.4f)

Resides          FIRM NO       MILD NO      MILD YES      FIRM YES
in South     mean(pvo2p1)  mean(pvo2p2)  mean(pvo2p3)  mean(pvo2p4)
-------------------------------------------------------------------
     0=No         0.3958        0.2508        0.1673        0.1861
    1=Yes         0.2282        0.2241        0.2115        0.3362
-------------------------------------------------------------------
```

A Technical Aside on OLOGIT Coefficients and Predictions

Having Stata generate predictions from ologit is a great convenience. However, before one relies on software-generated results, one should fully understand how the predictions are generated to assure one will be able to use them correctly and confidently. With this in mind, I call attention to a detail noted in Table 6.

The detail is that the calculation to obtain predicted values for logits based on the ologit coefficients do not follow the usual conventional form

$Y = b_0 + b_1X_1$.

Instead the calculations are implemented in the unconventional form of

$Y = b_0 - b_1X_1$.

(This unusual calculation is also noted in Long and Freeze 2006: 187).

Accordingly, the logit prediction for cut point 1 in the South is obtained from

$Y = b_0 - b_1X_1. = -0.4230 - 0.7953 = -1.2184$

First off, as noted below, this choice regarding how to implement the calculation is unconventional but it is technically fine. It deviates from convention but it yields the correct result. The fact that the approach does not follow convention can be confusing to researchers when they first begin using Stata's ologit procedure. So, What accounts for the use of this unconventional equation format? The answer is not especially satisfying. The answer is that, while this approach is unconventional for how most regression predictions are formulated. It is a practice that happens to be common in the sub-literature that develops the method of ordered logit regression.

The practice in the literature can be characterized in the following way. All regression models can be "parameterized" in a variety of mathematically equivalent ways that (by definition of being mathematically equivalent) will yield identical results for predictions and other fundamental model estimates. This is true for models of ordered logit regression as well. Thus, there are many mathematically equivalent parameterizations of the ordered logit regression model and all will yield correct predictions based on the full set of coefficients for the independent variables and the coefficients for the cut points.

One available parameterization of the model would lead to prediction equations to follow the convention of $Y = b_0 + b_1X_1$ used in OLS regression. But, for obscure reasons, in the technical literature for ordered logit regression, the parameterization of the model that has gained widest use, and which happens to be implemented in the Stata ologit procedure, adopts the logit prediction equations that take the form

$Y = b_0 - b_1X_1$.

There is nothing incorrect or even technically controversial about this choice for implementing calculations. But it is not the familiar, conventional form. This can potentially lead to confusion and error if researchers fail to notice the choice and try to calculate predicted values from estimated model parameters using the standard form of the prediction equation. However, all is well, when one takes note of the different approach implemented in the Stat ologit procedure and uses the appropriate prediction equation for this implementation.

(Note: I checked with Stata technical consultants to ask about the choice to adopt this particular parameterization of ordered logit regression when the ologit procedure was implemented. I learned that the original code for this procedure was developed by a statistician who was a specialist in ordered logit regression. This person was comfortable with the non-standard equation form because (a) it was accepted and used among other specialists working on these methods and (b) it is mathematically correct and equivalent to the more conventional parameterization of the model. So it goes.)

Substantive Importance of Effects of X

These results from the ologit regression suggest that southerners and non-southerners are appreciably different on level of opposition to integration. This is evident across the full distribution of cases across categories of Y. But it especially clear in the fact that the proportion of expected “Firm YES’s” for the statement it is "OK to oppose integration" is about 0.15 higher in the South.

The Z value for SOUTH (and its associated probability value) indicates that the “effect” of region is statistically significant. This speaks only to the question of whether the differences between the relative frequency distribution across categories of Y for southerners and non-southerners could have occurred by chance if we had drawn the sample from a population where in fact the relative frequency distribution was the same for both groups.

An equally if not more important question is how to evaluate the “magnitude of the effect” of region. Ultimately, one should compare the full relative frequency distributions across categories of Y for the two groups (Non-South and South). The figures reviewed at the beginning of this discussion demonstrate how graphs can be used to highlight differences across comparisons.

Another approach is to compare distributions quantitatively as shown in Table 7 below. One simple basis for comparison is to review the absolute value of the differences in the predicted relative frequency distributions across categories of Y for the South and the Non-South.

This is show separately for the observed regional distributions from simple cross-tabulation analysis and also for the predicted regional distributions based on the ordered logit regression. One clear point of difference is that the regions differ at the low and high ends of the distribution; by over 0.15 at the lowest category and by over 0.15 at the highest category in both the observed distribution and the predicted distribution.

The Dissimilarity Index (D) quantifies differences over the full relative frequency distribution on Y. The value of D is equal to one-half of the sum of the absolute differences between Non-south and South cumulated over the categories of the dependent variable. The value of D can be interpreted as follows;

> D indicates the minimum proportion of one group that would have to be transferred to another category of Y to bring about identical regional distributions on Y.

The value of D of 0.1943 thus indicates that almost 20% of southerners would have to shift from higher categories to lower categories on opposition to integration to match the relative frequency distribution predicted in the North.

The minimum possible value of D is 0.0. This value would result if region had no effect on opposition to integration; in which case, the relative frequency distributions would be predicted to be identical for both regions.

The maximum possible value for D is 1.00. This occurs when the two distributions do not overlap at all (e.g., if the predicted distribution for Non-South was 0.75, 0.25, 0.0, 0.0, and the predicted distribution for South was 0.0, 0.0, 0.25, 0.75). Here the value of D observed is nearly one fifth of the maximum possible value. This suggests the impact of region is substantial.

Table 7. Dissimilarity Calculations to Assess Importance of Regional Differences

	Observed Distribution			Predicted Distributions		
	(1)	(2)	(3)	(1)	(2)	(3)
			|Difference|			|Difference|
OPPINT4	Non-South	South	= |(2)-(1)|	Non-South	South	= |(2)-(1)|
1 Firm NO	0.3937	0.2358	0.1579	0.3958	0.2282	0.1676
2 Mild NO	0.2492	0.2327	0.0165	0.2508	0.2241	0.0267
3 Mild YES	0.1791	0.1761	0.0030	0.1673	0.2115	0.0442
4 Firm YES	0.1780	0.3553	0.1774	0.1861	0.3362	0.1501
	1.0000	1.0000	0.3547	1.0000	1.0000	0.3886
Dissimilarity Index			***0.1774***			***0.1943***

The Dissimilarity Index (D) also can be used to assess classification errors for Y that occur based on model predictions in comparison to what would occur under the expectations of the null hypothesis that X variables do not have effects.

Table 8 shows the calculations for comparing errors in classification observed under the expectations of the null hypothesis of no effect of region and errors of classification obtained under the predictions of a model that includes region. The null hypothesis expectation produces a D of 0.1774 indicating a minimum of 17% of cases would have to change categories to reproduce the observed distribution. In contrast, the model that incorporates region as a predictor produces a much lower value of D of 0.0472 which indicates the model has a misclassification rate that is far lower than under the null hypothesis of no effect of region.

Table 8. Dissimilarity Calculations to Assess Reduction in Classification Error Based on Model Predictions

	Model w/South Included			Null Model (No X's)		
	(1)	(2)	(3)	(1)	(2)	(3)
OPPINT4	Observed p	Predicted p	|Difference| = |(2)-(1)|	Observed p	Predicted p	|Difference| = |(2)-(1)|
Non-South						
1 Firm NO	0.3937	0.3958	0.0021	0.3937	0.3534	0.0403
2 Mild NO	0.2492	0.2508	0.0016	0.2492	0.2450	0.0042
3 Mild YES	0.1791	0.1673	0.0118	0.1791	0.1783	0.0008
4 Firm YES	0.1780	0.1861	0.0081	0.1780	0.2233	0.0453
	1.0000	1.0000		1.0000	1.0000	
South						
1 Firm NO	0.2358	0.2282	0.0076	0.2358	0.3534	0.1176
2 Mild NO	0.2327	0.2241	0.0086	0.2327	0.2450	0.0123
3 Mild YES	0.1761	0.2115	0.0354	0.1761	0.1783	0.0022
4 Firm YES	0.3553	0.3362	0.0191	0.3553	0.2233	0.1320
	1.0000	1.0000	0.0943	1.0000	1.0000	0.3547
Dissimilarity Index			***0.0472***			***0.1774***

Evaluating OLOGIT Relative to MLOGIT

When the dependent variable has three or more categories, ordered logit can approach but not surpass the goodness of fit achieved by multinomial logit. The reason for this is that multinomial logit imposes fewer constraints on the model and uses many more effect parameters to capture the relationships involved. In comparison, the ordered logit model assumes the dependent variable has an ordinal structure and fits a model with fewer parameters.

The ordered logit model is more parsimonious and easier to interpret than the multinomial logit model. So, if it is captures the effects of X well, it is an attractive option in comparison to multinomial logit.

With this in mind, it is useful to compare results of ologit with mlogit to see if the assumption of ordinal structure in the effects of X's on Y is justified. One approach to this is to compare likelihood values for the two models. Specifically, the goodness of fit for ologit compared to mlogit can be evaluated by comparing the likelihood value obtained by fitting the model with ologit to that obtained by fitting the model with mlogit.

L1 = the log-likelihood value reported by ologit

L0 = the log-likelihood value reported by mlogit

The mlogit estimates p(k−1) additional parameters (where p = the number of independent variables [excluding the constant] and k = the number of categories in the dependent variables). We can then perform a "likelihood-ratio test" based on $-2(\mathrm{L1} - \mathrm{L0})$ and evaluate it with reference to χ^2 at df[p(k − 2)].

This test is can be seen as a comparison of nested models. The ordered logit model imposes a simpler set of structured constraints when predicting the distribution of cases across categories of Y. The multinomial logit model relaxes these constraints and allows for more complex effects of X on the distribution of cases across categories of Y. Given this, a large value of $-2(\mathrm{L1} - \mathrm{L0})$ can be, be taken as evidence of poorness of fit when adopting a model that assumes relatively simple effects of X on an ordinal variable.

In the example under consideration, the comparison of ologit and mlogit can be conducted as follows. First, run the two separate models.

```
. mlogit oppint4 south

Iteration 0:   log likelihood =   -1686.23
Iteration 1:   log likelihood = -1662.4117
Iteration 2:   log likelihood = -1661.8945
Iteration 3:   log likelihood = -1661.8944

Multinomial logistic regression                 Number of obs     =      1,245
                                                LR chi2(3)        =      48.67
                                                Prob > chi2       =     0.0000
Log likelihood = -1661.8944                     Pseudo R2         =     0.0144

------------------------------------------------------------------------------
     oppint4 |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
Firm_No      |  (base outcome)
-------------+----------------------------------------------------------------
Mild_No      |
       south |   .4440556   .1841618     2.41   0.016     .0831051    .8050061
       _cons |  -.4574796   .0840757    -5.44   0.000    -.6222651   -.2926942
-------------+----------------------------------------------------------------
Mild_Yes     |
       south |   .4957722   .1998857     2.48   0.013     .1040034    .8875409
       _cons |  -.7879096   .0936153    -8.42   0.000    -.9713922    -.604427
-------------+----------------------------------------------------------------
Firm_Yes     |
       south |   1.203851   .1760205     6.84   0.000     .8588567    1.548844
       _cons |  -.7939519   .0938101    -8.46   0.000    -.9778163   -.6100875
------------------------------------------------------------------------------
```

```
. ologit oppint4 south if (year==7)

Iteration 0:   log likelihood =   -1686.23
Iteration 1:   log likelihood = -1663.7359
Iteration 2:   log likelihood = -1663.7036
Iteration 3:   log likelihood = -1663.7036

Ordered logistic regression                     Number of obs     =      1,245
                                                LR chi2(1)        =      45.05
                                                Prob > chi2       =     0.0000
Log likelihood = -1663.7036                     Pseudo R2         =     0.0134

------------------------------------------------------------------------------
     oppint4 |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       south |   .7953073   .1191109     6.68   0.000     .5618542     1.02876
-------------+----------------------------------------------------------------
       /cut1 |  -.4230379   .0653703                     -.5511613   -.2949144
       /cut2 |   .6040889   .0663011                      .4741412    .7340366
       /cut3 |   1.475783   .0776177                      1.323655    1.627911
------------------------------------------------------------------------------
```

Next calculate the relevant test statistics

L1 = −1663.7036 = the log-likelihood value reported by ologit

L0 = −1661.8944 = the log-likelihood value reported by mlogit

$$
\begin{aligned}
\chi^2 &= -2(L1 - L0) \\
&= -2[(-1663.7036) - (-1661.8944)] \\
&= (-1)2[(-1663.7036) - (-1661.8944)] \\
&= 2[(-1)(-1663.7036) - (-1)(-1661.8944)] \\
&= 2(1663.7036 - 1661.8944) \\
&= 3.6184
\end{aligned}
$$

This result is evaluated at df = p(k−2) = 1(4−2) = 2

The probability value of χ^2 of 3.6184 at df(2) is 0.164 and thus is not statistically significant at conventional levels. So the additional complexity of the multinomial model does not lead to a significantly better goodness of fit than the more parsimonious ordinal logit model.

END OF NOTES