intreg hypothetical example

Here is a hypothetical example using `intreg`. y is a continuous var that ranges from about -70 to 88. It is normally distributed. ycat is a collapsed, ordinal version of y. y1 and y2 are the upper and lower bounds of the y intervals.

```
. use "http://www.nd.edu/~rwilliam/xsoc73994/statafiles/intreg.dta", clear
(Hypothetical data for intreg example)

. des

Contains data from D:\Soc73994\Statafiles\intreg.dta
  obs:         1,000                          Hypothetical data for intreg
                                                example
 vars:             7                          6 Nov 2006 07:57
 size:        32,000 (99.9% of memory free)
-------------------------------------------------------------------------------
              storage  display     value
variable name   type   format      label      variable label
-------------------------------------------------------------------------------
y               float  %9.0g                  Continuous Y, ranges from -70.4
                                                to 88.06
ycat            float  %10.0g      ycat       Y collapsed into 5 intervals
y1              float  %9.0g                  Lower bound of Y interval
y2              float  %9.0g                  Upper bound of Y interval
x1              float  %9.0g
x2              float  %9.0g
x3              float  %9.0g
-------------------------------------------------------------------------------
Sorted by:

. sum y

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
           y |      1000    14.01144    25.05774  -70.36776    88.0509

. tab1 ycat

-> tabulation of ycat

Y collapsed |
     into 5 |
  intervals |      Freq.     Percent        Cum.
------------+-----------------------------------
       LE 0 |        287       28.70       28.70
    0 to 15 |        224       22.40       51.10
   15 to 30 |        203       20.30       71.40
   30 to 45 |        183       18.30       89.70
 45 or more |        103       10.30      100.00
------------+-----------------------------------
      Total |      1,000      100.00
```

```
. * intreg with collapsed Y
. intreg  y1 y2 x1 x2 x3

Fitting constant-only model:

Iteration 0:   log likelihood = -1688.3436
Iteration 1:   log likelihood = -1574.6026
Iteration 2:   log likelihood = -1565.5637
Iteration 3:   log likelihood = -1565.5603
Iteration 4:   log likelihood = -1565.5603

Fitting full model:

Iteration 0:   log likelihood = -1508.2373
Iteration 1:   log likelihood = -1379.0543
Iteration 2:   log likelihood = -1372.4038
Iteration 3:   log likelihood = -1372.3949
Iteration 4:   log likelihood = -1372.3949

Interval regression                               Number of obs   =       1000
                                                  LR chi2(3)      =     386.33
Log likelihood = -1372.3949                       Prob > chi2     =     0.0000

------------------------------------------------------------------------------
             |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          x1 |   1.221547   .2544077     4.80   0.000     .7229169    1.720177
          x2 |   .8989353   .0799428    11.24   0.000     .7422503     1.05562
          x3 |   .9384835   .2191945     4.28   0.000     .5088702    1.368097
       _cons |   .0771196   1.451354     0.05   0.958    -2.767483    2.921722
-------------+----------------------------------------------------------------
    /lnsigma |   3.003777   .0320312    93.78   0.000     2.940997    3.066557
-------------+----------------------------------------------------------------
       sigma |   20.16155   .6457982                      18.93472    21.46787
------------------------------------------------------------------------------

  Observation summary:       287  left-censored observations
                               0     uncensored observations
                             103 right-censored observations
                             610       interval observations

. * OLS regression with original Y
. reg  y x1 x2 x3

      Source |       SS       df       MS              Number of obs =    1000
-------------+------------------------------           F(  3,   996) =  188.94
       Model |  227500.386     3  75833.4619           Prob > F      =  0.0000
    Residual |  399761.928   996  401.367397           R-squared     =  0.3627
-------------+------------------------------           Adj R-squared =  0.3608
       Total |  627262.313   999  627.890204           Root MSE      =  20.034

------------------------------------------------------------------------------
           y |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          x1 |   1.120216   .2308738     4.85   0.000     .6671616    1.573271
          x2 |   .9312722   .0706904    13.17   0.000      .792553    1.069991
          x3 |   .8474134   .1983744     4.27   0.000     .4581337    1.236693
       _cons |    .196622   1.245274     0.16   0.875    -2.247039    2.640284
------------------------------------------------------------------------------
```

```
. * oprobit with collapsed Y
. oprobit  ycat x1 x2 x3

Iteration 0:   log likelihood = -1561.9813
Iteration 1:   log likelihood = -1370.1889
Iteration 2:   log likelihood = -1368.7383
Iteration 3:   log likelihood = -1368.7378

Ordered probit regression                         Number of obs   =       1000
                                                  LR chi2(3)      =     386.49
                                                  Prob > chi2     =     0.0000
Log likelihood = -1368.7378                       Pseudo R2       =     0.1237

------------------------------------------------------------------------------
        ycat |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          x1 |   .0604916   .0126526     4.78   0.000      .035693    .0852902
          x2 |   .0445961    .004006    11.13   0.000     .0367445    .0524476
          x3 |   .0466968   .0108907     4.29   0.000     .0253514    .0680421
-------------+----------------------------------------------------------------
       /cut1 |   .0091044   .0732018                     -.1343684    .1525773
       /cut2 |   .7462179   .0751763                      .5988751    .8935608
       /cut3 |   1.415098   .0809962                      1.256348    1.573848
       /cut4 |   2.285878   .0952678                      2.099156    2.472599
------------------------------------------------------------------------------

.
```

Several things to note about the above:

- The nice thing about `intreg`, as opposed to other ordinal methods, is that you interpret its parameters the same way you do the parameters from an OLS regression. The sigma that `intreg` reports is equivalent to the root mean square error (i.e. the standard error of the residuals) from an OLS regression
- In this particular example, `intreg` does remarkably well. Its coefficients, standard errors, etc. are very similar to those produced by OLS regression on the un-collapsed y variable.
- Also, `intreg` produces almost the exact same log-likelihood as does `oprobit`, and also the same z values. (NOTE: You should compare the log-likelihoods rather than the model chi-squares when comparing `intreg` and `oprobit`.) But, the coefficients from `intreg` are much easier to interpret.
    - As the Stata manual points out, if `oprobit` fit much better, you might want to modify the `intreg` model (e.g. take logs of the interval points) or use `oprobit` or `ologit` or some other ordinal method instead.
- I caution, however, that the example is "rigged" in `intreg`'s favor, in that the assumptions it makes about normality are true in the constructed data set. You can't always count on it working this well. As the Stata manual notes, `intreg` assumes normality.