crreg: A New command for Generalized Continuation Ratio Models

Shawn Bauldry
Purdue University

Jun Xu
Ball State University

Andrew Fullerton
Oklahoma State University

Stata Conference
July 28, 2017

Introduction

- Ordered outcomes common in social science research
  - Reflect underlying continuous measure
  - Reflect discrete social phenomena/processes
- Beneficial to address ordered nature of outcome with model
- Two primary choices when selecting model
  - How probabilities of interest are defined
  - Extent of parallel lines (proportional odds) assumption

Introduction

Choices when selecting model

- How probabilities of interest are defined
  - cumulative: being at or below a given value

$$\Pr[y \leq m]$$

  - adjacent: being at a given value conditional on being at that or the next higher value

$$\Pr[y = m | y = m \text{ or } y = m + 1]$$

  - stage: being at a given value conditional on being at or above that value

$$\Pr[y = m | y \geq m]$$

Introduction

Choices when selecting model

- Extent of parallel lines (proportional odds) assumption
  - All coefficients constrained equal across cutpoint equations
  - Some coefficients freely vary across cutpoint equations
  - Some coefficients vary by a common factor across cutpoint equations
  - All coefficients allowed to freely vary across cutpoint equations

Introduction

Table: Stata commands for ordered regression models.

	approach to comparisons		
parallel lines	cumulative	stage	adjacent
for all IVs	`ologit`	`ccrlogit`	`adjcatlogit`
	`gologit2`	`ocratio`	
for some IVs (free)	`gologit2`	-	-
for some IVs (factor)	-	-	`slogit`
no IVs	`gologit2`	`seqlogit`,	`mlogit`
		`ucrlogit`	

Notes: Based on Fullerton (2009, Table 1). `gologit2` Williams (2006); `ocratio` Wolfe (1998); `seqlogit` Buis (2007); `ccrlogit`, `ucrlogit`, `adjcatlogit` Fagerland (2014).

Introduction

`crreg` relative to existing commands

- command for generalized continuation ratio (stage) models
- allows constrained, free, and common factor coefficients for all or subset of covariates
- allows choice of logit, probit, or complementary log-log link functions
- integrated with survey and multiple imputation commands

Continuation Ratio Model

Continuation Ratio Model

$$\Pr(y = m | y \geq m, \mathbf{x}) = F(\tau_m - \mathbf{x}_1\boldsymbol{\beta} - \mathbf{x}_2\boldsymbol{\gamma}_m - \phi_m\mathbf{x}_3\boldsymbol{\eta})$$

where

- y is an ordered outcome with $m = 1, \ldots, M$ categories
- $\mathbf{x} = [\mathbf{x}_1\ \mathbf{x}_2\ \mathbf{x}_3]$ is a partitioned vector of independent variables
- τ_m is the cutpoint for equation m
- $\boldsymbol{\beta}$ coefficients that do not vary across cutpoint equations
- $\boldsymbol{\gamma}_m$ coefficients that vary across cutpoint equations
- $\boldsymbol{\eta}$ coefficients that vary across cutpoint equations by a common factor
- ϕ_m is the common factor for equation m

crreg Command

`crreg` *depvar* [*indepvars*] [if] [in] [weight], [`prop`(*varlist*) `free`(*varlist*) `link`(*string*) `vce`(*vcetype*) `or rrr irrr hr eform`(*string*) *]

- *indepvars*: specify all IVs
- `prop`(*varlist*): specify IVs that have coefficients that vary by a common factor
- `free`(*varlist*): specify IVs that freely vary across cutpoint equations
- `link`(*string*): specify logit (default), probit, or cloglog link functions

Example

Example

Continuation ratio model for educational attainment

- General Social Survey data from 2014
- Educational attainment: (1) less than high school, (2) high school or junior college, (3) bachelor's degree, (4) graduate degree
- Predictors:
  - constrained: sex and race
  - vary by common factor: mother's and father's education
  - vary freely: age
- Logit link

Example

Command and iterations

```
. crreg deg age fem wht paeduc maeduc, free(age) prop(paeduc maeduc)

initial:       log likelihood = -2760.1121
alternative:   log likelihood = -2779.5687
rescale:       log likelihood = -2559.4785
rescale eq:    log likelihood = -1959.0203
Iteration 0:   log likelihood = -1959.0203  (not concave)
Iteration 1:   log likelihood = -1917.4637  (not concave)
Iteration 2:   log likelihood = -1895.3905
Iteration 3:   log likelihood = -1858.1663  (not concave)
Iteration 4:   log likelihood =   -1849.96
Iteration 5:   log likelihood = -1815.3368  (not concave)
Iteration 6:   log likelihood =      -1807
Iteration 7:   log likelihood = -1806.6132
Iteration 8:   log likelihood = -1805.2291
Iteration 9:   log likelihood = -1805.1555
Iteration 10:  log likelihood = -1805.1544
Iteration 11:  log likelihood = -1805.1544
```

Example

Output

```
Ordered Logit Estimates                           Number of obs     =      1,765
                                                  Wald chi2(2)      =       1.90
Log likelihood = -1805.1544                       Prob > chi2       =     0.3867

------------------------------------------------------------------------------
         deg |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
constrained  |
         fem |  -.0274121    .082374    -0.33   0.739    -.1888621    .1340379
         wht |  -.1438116   .1065851    -1.35   0.177    -.3527144    .0650913
-------------+----------------------------------------------------------------
factor       |
      paeduc |   .1692259   .0200097     8.46   0.000     .1300076    .2084441
      maeduc |   .1457927   .0198782     7.33   0.000     .1068321    .1847532
-------------+----------------------------------------------------------------
eq1          |
         age |    .017095   .0054125     3.16   0.002     .0064866    .0277034
       _cons |  -1.644997   .3791252    -4.34   0.000    -2.388069   -.9019255
-------------+----------------------------------------------------------------
eq2          |
         age |   .0211126    .003509     6.02   0.000     .0142351    .0279901
       _cons |  -4.749109   .3833059   -12.39   0.000    -5.500375   -3.997844
-------------+----------------------------------------------------------------
eq3          |
         age |   .0217539   .0054937     3.96   0.000     .0109864    .0325214
       _cons |  -2.296517   .5549287    -4.14   0.000    -3.384157   -1.208877
-------------+----------------------------------------------------------------
-continued-
```

Example

Output

```
-continued-
-------------+----------------------------------------------------------------
phi2         |
       _cons |   .8547806   .0913566     9.36   0.000     .6757251    1.033836
-------------+----------------------------------------------------------------
phi3         |
       _cons |   .2012101   .0925738     2.17   0.030     .0197687    .3826514
------------------------------------------------------------------------------
```

Interpretation

- Coefficients are standard log odds [depends on link function]
- Consider testing equality of freely varying coefficients
- Consider testing whether common factors equal 1

Example

Additional tests

- test equality of coefficients that freely vary

```
. test [eq1]:age = [eq2]:age = [eq3]:age

 ( 1)  [eq1]age - [eq2]age = 0
 ( 2)  [eq1]age - [eq3]age = 0

           chi2(  2) =    0.48
         Prob > chi2 =    0.7864
```

- test common factors equal 1

```
. test [phi2]:_cons = 1

 ( 1)  [phi2]_cons = 1

           chi2(  1) =    2.53
         Prob > chi2 =    0.1119

. test [phi3]:_cons = 1

 ( 1)  [phi3]_cons = 1

           chi2(  1) =   74.45
         Prob > chi2 =    0.0000
```

Conclusion

New `crreg` command for generalized ordered regression models

- allows constrained, free, and common factor coefficients for all or subset of covariates
- allows choice of logit, probit, or complementary log-log link functions
- integrated with survey and multiple imputation commands
- based on Stata's ML commands

Thank You

Beta version available

- https://github.com/sbauldry/crreg
- `net install crreg, from(https://github.com/sbauldry/crreg/raw/master) replace`
- email: sbauldry@purdue.edu