The Stata Journal (2018)
18, Number 4, pp. 924–936

gencrm: A new command for generalized continuation-ratio models

Shawn Bauldry
Purdue University
West Lafayette, IN
sbauldry@purdue.edu

Jun Xu
Ball State University
Muncie, IN
jxu@bsu.edu

Andrew S. Fullerton
Oklahoma State University
Stillwater, OK
andrew.fullerton@okstate.edu

Abstract. A continuation-ratio model represents a variant of an ordered regression model that is suited to modeling processes that unfold in stages, particularly those in which a return to a previous stage is not possible (for example, educational attainment, job promotion, or disease progression). The parameters for covariates in continuation-ratio models may be constrained to be equal, vary by a set of common factors (that is, proportionality constraints), or freely vary across stages. Currently, there are three community-contributed commands that fit continuation-ratio models. Each of these commands fits some subset of continuation-ratio models involving parameter constraints, but none of them offer complete coverage of the range of possibilities. The new `gencrm` command expands the options for continuation-ratio models to include the possibility for some of or all the covariates to be constrained to be equal, to freely vary, or to vary by a set of common factors across stages. `gencrm` relies on Stata's maximum likelihood routines for estimation and avoids reshaping the data. `gencrm` includes options for three link functions (`logit`, `probit`, and `cloglog`) and supports Stata's multiple-imputation suites of commands.

Keywords: st0546, gencrm, generalized continuation-ratio models, stage models, sequential logit models, stopping-ratio models

1 Introduction

Outcomes measured by a set of ordered categories arise frequently in social science research. In some cases, such outcomes likely reflect an underlying continuous measure, such as Likert responses to attitude measures, the response categories to standard self-rated health measures, or intervals of income. In other cases, the ordered responses may reflect truly discrete social phenomena, such as educational degrees, steps in the process of voting, or stages of a disease. When one analyzes such outcomes, it can be important to account for the ordinal scale of the outcome using some form of an ordered regression model (Agresti 2010; Long 1997; McKelvey and Zavoina 1975; Winship and Mare 1984).

When selecting a specific ordered regression model, analysts have two primary choices to make. The first is how the probabilities of interest are defined. The standard or cumulative approach, which is implemented in Stata's `ologit` and `oprobit` commands, models the probability of being at or below a given value m $[\Pr(y \leq m)]$ (McCullagh and Nelder 1989). The standard cumulative approach is particularly suited to modeling measures such as Likert scale items that represent an underlying continuous

 st0546

distribution. The second approach models the probability of a given value m relative to the probability of the next higher value $[\Pr(y = m|y = m \text{ or } y = m+1)]$. These models are referred to as adjacent-category models (Agresti 2010; Fullerton 2009; Goodman 1983; Sobel 1997). The adjacent-categories approach has been applied in contexts in which, for instance, a middle category (for example, neither agree or disagree) or an incremental increase in a response level is of particular interest (Sobel 1997). A third approach models the probability of being at a given value m, given that a case has advanced to that stage of the sequential process $[\Pr(y = m|y \geq m)]$. These models are referred to as continuation-ratio models (or stage models, sequential logit models, or stopping-ratio models) (Fienberg 2007; Fullerton 2009; Fullerton and Xu 2016; McCullagh and Nelder 1989; Tutz 2012; Yee 2015). This approach is most appropriate for modeling social processes that unfold in stages, particularly those in which it is not possible to return to a previous stage, such as educational attainment.

The second decision involves the extent to which the parallel lines assumption (or proportional-odds assumption in the case of logit link) holds across the independent variables (IVs) (Long 1997). The standard approach, again as implemented in Stata's official commands for ordered logit and probit models, has the parallel lines assumption hold for all IVs—that is, regression coefficients are constrained to be equal across all cutpoint equations. If this assumption does not hold in the data, then the constraints can lead to substantially biased coefficients across cutpoint equations (Williams 2016). In addition, researchers may have substantive or theoretical reasons to expect that regression coefficients will vary across cutpoint equations. For instance, an intervention might shift people from one stage to the next (for example, precontemplation to contemplation) but no further (for example, contemplation to action) (Hedeker and Mermelstein 1998). One can relax the parallel lines assumption for all or a subset of IVs in two ways. Coefficients can either vary across cutpoint equations by a set of commons factors (that is, proportionality constraints) or freely vary across cutpoint equations.

The combination of the two decision points creates a 4×3 table that defines a family of ordered regression models (Fullerton 2009; Fullerton and Xu 2016). Table 1 provides the official Stata commands and community-contributed commands that are currently available for each type of ordered regression model. The standard choice for an approach to comparisons, labeled cumulative, is well represented by Stata's official commands `ologit` and `oprobit` and William's (2006) `gologit2` command. The only missing option involves allowing a subset of IVs to have coefficients that vary across cutpoint equations by a set of common factors, but this option can be obtained using Stata's `constraint` command if the researcher can specify the common factors in advance.

Table 1. Typology of ordered regression models with existing commands and `gencrm`

	Approach to comparisons		
Parallel lines	Cumulative	Stage	Adjacent
for all IVs	`ologit`, `gologit2`	`ccrlogit`, `ocratio`, `gencrm`	`adjcatlogit`
for some IVs (freely vary)	`gologit2`	`gencrm`	–
for some IVs (common factors)	–	`gencrm`	–
no IVs	`gologit2`	`seqlogit`, `ucrlogit`, `gencrm`	`mlogit`

NOTE: Based on Fullerton (2009, table 1). Probit and complementary log-log links are available for some of the models. In some cases, one can also use Stata's `constraint` commands to fit a model in which the coefficients for some IVs vary by a set of common factors. In addition, in some cases, one can translate stage or continuation-ratio models into a discrete-time survival analysis framework. The stereotype logit model (Stata's `slogit`) is another ordered regression model related to the adjacent approach to comparisons. We did not include it in this table, because it does not precisely fit the parallel lines dimension as we have outlined. With the stereotype logit, all the IVs vary by the same common factors, but one can include multiple dimensions.

There are several official Stata commands and community-contributed commands that fit variants of models for the stage and adjacent approaches. In particular, Wolfe's (1998) older `ocratio` command and Fagerland's (2014) newer `ccrlogit` command fit continuation-ratio models in which all IVs have coefficients subject to the parallel lines constraint. In addition, Buis's (2007) `seqlogit` command and Fagerland's (2014) `ucrlogit` command fit continuation-ratio models in which all the IVs have coefficients that freely vary across all cutpoint equations. Fagerland's (2014) `adjcatlogit` fits adjacent-category models in which the coefficients for all the IVs are constrained to be equal across cutpoint equations. Finally, Stata's `slogit` command fits a stereotype logit model, which is a form of an adjacent-category model, with a proportionality constraint. Stata's `mlogit` command fits an adjacent-category model in which all the IVs have coefficients that freely vary.

In this article, we introduce a new command, `gencrm`, to fit generalized continuation-ratio models that include all forms of the parallel lines assumption (see table 1). In contrast to existing commands, the new command provides the option of maintaining the parallel lines assumption for all IVs or any subset of IVs. In addition, `gencrm` provides the option for allowing a subset of IVs to have coefficients that vary by a set of common factors. Section 2 provides a brief overview of the generalized continuation-ratio model. Section 3 introduces the new command and discusses various options, some of which are also not currently available for any of the existing commands. Section 4 provides an extended empirical example illustrating the use of the new command.

2 Continuation-ratio model

Following Fullerton and Xu (2016, 65), the general form of the continuation-ratio model is given by

$$\Pr(y = m | y \geq m, \mathbf{x}) = F(\tau_m - \mathbf{x}_1\boldsymbol{\beta} - \mathbf{x}_2\boldsymbol{\gamma}_m - \phi_m\mathbf{x}_3\boldsymbol{\lambda}) \qquad (1 \leq m < M)$$

where y is an ordered outcome with $m = 1, \ldots, M$ categories, $F(\cdot)$ is the logistic, probit, or complementary log-log cumulative distribution function, and $\mathbf{x} = [\mathbf{x}_1 \quad \mathbf{x}_2 \quad \mathbf{x}_3]$ is a vector of IVs partitioned into three sets corresponding to whether the parallel lines assumption is maintained, relaxed, or partially relaxed (that is, the coefficients are allowed to vary by a set of common factors). The parameters in the model include a vector of cutpoints τ_m, a vector $\boldsymbol{\beta}$ of coefficients that do not vary across cutpoint equations, a vector $\boldsymbol{\gamma}_m$ of coefficients that vary across cutpoint equations, a vector $\boldsymbol{\lambda}$ of coefficients that vary across cutpoint equations by a set of common factors, and a vector ϕ_m of common factors. To identify the model, we assume that $\phi_1 = 1$ and $\phi_M = 0$; thus, there are $M - 2$ common factors to estimate.

The probability for any given value (m) of the outcome (y) conditional on the covariates is the product of the probability that $y = m$ for the current stage and the probabilities that $y > m$ for all earlier stages. This is given by

$$\Pr(y = m|\mathbf{x}) = \begin{cases} F\{\tau_1 - g(\mathbf{x})\} & m = 1 \\ \left(\prod_{j=1}^{m-1} [1 - F\{\tau_j - g(\mathbf{x})\}]\right) F\{\tau_m - g(\mathbf{x})\} & 1 < m \leq M \\ \prod_{j=1}^{M-1} [1 - F\{\tau_j - g(\mathbf{x})\}] & m = M \end{cases} \qquad (1)$$

where $F(\cdot)$ is the logistic, probit, or complementary log-log cumulative distribution function and $g(\mathbf{x}) = \mathbf{x}_1\boldsymbol{\beta} + \mathbf{x}_2\boldsymbol{\gamma}_m + \phi_m\mathbf{x}_3\boldsymbol{\lambda}$. In the `gencrm` command, (1) is the basis for writing the likelihood function that is used with Stata's maximum likelihood (ML) estimators to fit generalized continuation-ratio models.[1]

1. Some of the other commands that fit continuation-ratio models in Stata use different forms of the likelihood function. `seqlogit` uses an inverse form of the likelihood function, and `ccrlogit` and `ucrlogit` begin with the opposite direction for defining the continuation ratios. We verified that `gencrm` provides numerically consistent results with all of these programs accounting for the differences in the likelihood functions.

3 The gencrm estimation command

3.1 Syntax

`gencrm` *depvar* [*indepvars*] [*if*] [*in*] [*weight*] [, `factor(`*varlist*`)` `free(`*varlist*`)` `link(`*string*`)` `vce(`*vcetype*`)` `or` `eform` *display_options* *maximize_options*]

3.2 Description

`gencrm` fits a continuation-ratio regression model of an ordered outcome, *depvar*, on a set of IVs, *indepvars*, that includes variables with coefficients constrained to be equal across all cutpoint equations (that is, subject to the parallel lines assumption), coefficients allowed to vary by a set of common factors across cutpoint equations, and coefficients allowed to freely vary across cutpoint equations.

3.3 Options

`factor(`*varlist*`)` specifies the IVs, if any, that have coefficients that vary by a set of common factors across cutpoint equations (that is, proportionality constraint).

`free(`*varlist*`)` specifies the IVs, if any, that have coefficients that freely vary across cutpoint equations. Any IVs not appearing in either `factor(`*varlist*`)` or `free(`*varlist*`)` have coefficients that are constrained to be equal across cutpoint equations.

`link(`*string*`)` specifies the link function. The default is `link(logit)`. Users may also specify a `probit` or `cloglog` link.

The remaining options are all standard Stata options for choosing the variance–covariance estimator (the `gencrm` command supports robust and cluster–robust standard errors), exponential forms for coefficients, and options for displaying results and other ML options. In addition, the command supports the usual Stata options for selecting subsets of cases and incorporating various types of weights and factor-variable prefixes.

The `gencrm` command is integrated with Stata's multiple-imputation suite of commands with the `mi` prefix. However, `gencrm` is not currently integrated with Stata's `predict` and `margins` functions. Thus, it is also not integrated with Stata's survey suite of commands (though it permits the use of weights).

3.4 Stored results

In addition to standard results returned from Stata's ML commands, `gencrm` stores the following in `e()`:

Scalars
- `e(k_cat)` number of values of *depvar*

Macros
- `e(cmd)` `gencrm`
- `e(factor)` variables allowed to vary by common factor or factors (if specified)
- `e(free)` variables allowed to freely vary (if specified)
- `e(link)` link function (if specified)

Matrices
- `e(b)` coefficient vector
- `e(V)` variance–covariance matrix of the estimators

Functions
- `e(sample)` marks estimation sample

4 Example

In this section, we provide an extended example of fitting a continuation-ratio model with the `gencrm` command. Our example involves fitting a model for educational attainment, which was one of the early uses of a continuation-ratio model (Mare 1980). The specific example is adapted from Fullerton and Xu (2016) and uses 2016 data from the General Social Survey (Smith, Hout, and Marsden 2016). The outcome is educational attainment (`degree`) measured as a series of education degrees: 1) less than high school, 2) high school, 3) junior college degree, 4) bachelor's degree, and 5) graduate degree. Predictors include age, female (`fem`), white (`wht`; versus other races), mother's education (`maeduc`), and father's education (`paeduc`). Readers interested in replicating the example can download `gencrm-gss-data.dta`.

We begin with a model in which each of the covariates predicts successively higher degrees and all the coefficients are constrained to be equal across degrees. The following code illustrates fitting a baseline model with just `degree`, then the model with covariates, and then conducting a likelihood-ratio test for an improvement in model fit from adding predictors to the model.

```
. use gencrm-gss-data
. quietly gencrm degree, link(logit)
. estimates store m1
. gencrm degree age fem wht paeduc maeduc, link(logit) or
initial:       log likelihood = -3668.1349
alternative:   log likelihood = -3645.9171
rescale:       log likelihood = -3633.1208
rescale eq:    log likelihood = -2539.6218
Iteration 0:   log likelihood = -2539.6218
Iteration 1:   log likelihood = -2442.7855
Iteration 2:   log likelihood = -2441.9627
Iteration 3:   log likelihood = -2441.9619
Iteration 4:   log likelihood = -2441.9619
Ordered Logit Estimates                          Number of obs     =      1,942
                                                 Wald chi2(5)      =     319.94
Log likelihood = -2441.9619                      Prob > chi2       =     0.0000

------------------------------------------------------------------------------
      degree | Odds Ratio   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         age |   1.015591   .0022791     6.89   0.000     1.011134    1.020068
         fem |   1.149795   .0818703     1.96   0.050     1.000026    1.321994
         wht |   1.050093   .0945003     0.54   0.587     .8802912    1.252649
      paeduc |   1.136152   .0144333    10.05   0.000     1.108212    1.164796
      maeduc |   1.086111   .0150384     5.97   0.000     1.057032    1.115989
-------------+----------------------------------------------------------------
       /tau1 |   .6107924   .2129124     2.87   0.004     .1934918    1.028093
       /tau2 |   3.585688   .2270397    15.79   0.000     3.140699    4.030678
       /tau3 |   2.000827     .23791     8.41   0.000     1.534532    2.467122
       /tau4 |   4.199473   .2498289    16.81   0.000     3.709817    4.689129
------------------------------------------------------------------------------
Note: Estimates are transformed only in the first equation.
. estimates store m2
. lrtest m2 m1, force
Likelihood-ratio test                                 LR chi2(5)  =    355.68
(Assumption: m1 nested in m2)                         Prob > chi2 =    0.0000
```

We see from both the likelihood-ratio test and the Wald test that the addition of covariates significantly improves model fit. Note that the `force` option is required for the `lrtest` command because of the way we have written the likelihood functions for the baseline model and the model with covariates. Turning to the parameter estimates (odds ratios in this case), we find statistically significant positive relationships for all the covariates with the exception of the indicator for whites. For instance, women have about 15% greater odds of achieving each higher degree level than men.

With our next model, we consider the possibility that the coefficients for age vary across different degrees. To allow for this, we indicate that `age` is free to vary in the `gencrm` command as follows. In addition, we also conduct a likelihood-ratio test comparing this model with the previous model and illustrate a Wald test for the joint significance of the coefficients across each cutpoint equation.

```
. gencrm degree age fem wht paeduc maeduc, link(logit) free(age)
initial:       log likelihood = -3668.1349
alternative:   log likelihood =   -3339.37
rescale:       log likelihood = -3271.2404
rescale eq:    log likelihood = -2595.8049
Iteration 0:   log likelihood = -2595.8049
Iteration 1:   log likelihood = -2446.8028
Iteration 2:   log likelihood = -2431.5829
Iteration 3:   log likelihood = -2431.5317
Iteration 4:   log likelihood = -2431.5317
Ordered Logit Estimates                             Number of obs     =      1,942
                                                    Wald chi2(8)      =     337.15
Log likelihood = -2431.5317                         Prob > chi2       =     0.0000

------------------------------------------------------------------------------
      degree |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
parallel     |
         fem |   .1444237   .0713972     2.02   0.043     .0044877    .2843597
         wht |   .0425363   .0905248     0.47   0.638    -.134889     .2199616
      paeduc |   .1289287   .0127438    10.12   0.000     .1039513     .153906
      maeduc |   .0822023   .0138752     5.92   0.000     .0550074    .1093973
-------------+----------------------------------------------------------------
eq1          |
         age |   .0040033   .0051552     0.78   0.437    -.0061007    .0141074
-------------+----------------------------------------------------------------
eq2          |
         age |   .0119869    .002952     4.06   0.000      .006201    .0177729
-------------+----------------------------------------------------------------
eq3          |
         age |    .020851   .0059033     3.53   0.000     .0092809    .0324212
-------------+----------------------------------------------------------------
eq4          |
         age |   .0339279   .0051812     6.55   0.000      .023773    .0440828
-------------+----------------------------------------------------------------
       /tau1 |   .0405321     .32246     0.13   0.900    -.5914778    .6725421
       /tau2 |   3.422866   .2437559    14.04   0.000     2.945113    3.900618
       /tau3 |   2.269478   .3547533     6.40   0.000     1.574174    2.964782
       /tau4 |   5.132067   .3471922    14.78   0.000     4.451583    5.812552
------------------------------------------------------------------------------

. estimates store m3
. lrtest m3 m2
Likelihood-ratio test                                 LR chi2(3)  =     20.86
(Assumption: m2 nested in m3)                         Prob > chi2 =    0.0001
. test _b[eq1:age] = _b[eq2:age] = _b[eq3:age] = _b[eq4:age]
 ( 1)  [eq1]age - [eq2]age = 0
 ( 2)  [eq1]age - [eq3]age = 0
 ( 3)  [eq1]age - [eq4]age = 0
           chi2(  3) =   20.59
         Prob > chi2 =    0.0001
```

The output for this version of the continuation-ratio model includes separate panels for the coefficients that are constrained to be equal across cutpoint equations (labeled `parallel`) and the coefficients that vary across cutpoint equations (labeled by equation number). In this case, there are four cutpoint equations, and we have coefficients for age for each of the cutpoint equations labeled as `eq1` through `eq4`.

We find that the likelihood-ratio test and Wald test are significant, and both suggest that the coefficient for age does vary across degrees. In addition, we note that the coefficients for age across the different degree equations are statistically significant except for the first cutpoint equation. Thus, in the absence of substantive or theoretical reasons for constraining the coefficients to be equal, we would consider relaxing the parallel lines assumption for `age` (see Fullerton and Xu [2016] for further discussion).

Finally, past studies of the effects of socioeconomic background on educational transitions have argued that the effects exhibit a proportional decline across transitions (Hauser and Andrew 2006). We allow for this possibility for both father's and mother's education by including them in the `factor()` option of the `gencrm` command as follows:

```
. gencrm degree age fem wht paeduc maeduc, link(logit) factor(paeduc maeduc)
initial:       log likelihood = -3668.1349
alternative:   log likelihood = -3308.1636
rescale:       log likelihood = -3242.1982
rescale eq:    log likelihood = -2526.2919
Iteration 0:   log likelihood = -2526.2919  (not concave)
Iteration 1:   log likelihood = -2490.3498  (not concave)
Iteration 2:   log likelihood = -2471.9621  (not concave)
Iteration 3:   log likelihood = -2457.4778  (not concave)
Iteration 4:   log likelihood = -2452.1061
Iteration 5:   log likelihood = -2449.2655  (not concave)
Iteration 6:   log likelihood = -2427.5344  (not concave)
Iteration 7:   log likelihood = -2425.7204
Iteration 8:   log likelihood = -2421.2151
Iteration 9:   log likelihood = -2412.9341
Iteration 10:  log likelihood = -2410.5602
Iteration 11:  log likelihood = -2409.3258
Iteration 12:  log likelihood = -2409.2145
Iteration 13:  log likelihood = -2409.2125
Iteration 14:  log likelihood = -2409.2125
Ordered Logit Estimates                           Number of obs     =      1,942
                                                  Wald chi2(5)      =     213.27
Log likelihood = -2409.2125                       Prob > chi2       =     0.0000
```

degree	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
parallel						
age	.0149736	.0022572	6.63	0.000	.0105495	.0193977
fem	.1435732	.0715348	2.01	0.045	.0033676	.2837789
wht	.020399	.0904489	0.23	0.822	-.1568776	.1976756
factor						
paeduc	.1959092	.0205751	9.52	0.000	.1555827	.2362357
maeduc	.1061246	.019652	5.40	0.000	.0676073	.1446418
/tau1	1.408306	.2648227	5.32	0.000	.8892628	1.927349
/tau2	3.864981	.2926254	13.21	0.000	3.291446	4.438516
/tau3	1.272505	.388327	3.28	0.001	.5113985	2.033612
/tau4	1.849416	.3888958	4.76	0.000	1.087194	2.611637
/phi2	.7822944	.0826679	9.46	0.000	.6202683	.9443205
/phi3	.5131833	.0981178	5.23	0.000	.320876	.7054905
/phi4	.1413287	.0837833	1.69	0.092	-.0228835	.3055409

The output for this version of a continuation-ratio model is presented in three panels. The first panel, labeled `parallel`, includes the variables with coefficients constrained to be equal across cutpoint equations. The second panel, labeled `factor`, includes the variables with coefficients that are allowed to vary by a set of common factors across cutpoint equations. The third panel has been expanded to include both the estimates for the thresholds (labeled as `taus`) and the estimates for the common factors (labeled as `phis`).

```
. test /phi2 == 1
 ( 1)  [/]phi2 = 1
           chi2(  1) =    6.94
         Prob > chi2 =    0.0085
. test /phi3 == 1
 ( 1)  [/]phi3 = 1
           chi2(  1) =   24.62
         Prob > chi2 =    0.0000
. test /phi4 == 1
 ( 1)  [/]phi4 = 1
           chi2(  1) =  105.04
         Prob > chi2 =    0.0000
. test /phi2 = /phi3 = /phi4 == 1
 ( 1)  [/]phi2 - [/]phi3 = 0
 ( 2)  [/]phi2 - [/]phi4 = 0
 ( 3)  [/]phi2 = 1
           chi2(  3) =  124.57
         Prob > chi2 =    0.0000
```

A likelihood-ratio test is not appropriate for testing this model against a model in which all the coefficients are constrained to be equal across cutpoint equations because under the null hypothesis that the coefficients equal zero, the common factors (the ϕ's) are not identified (Fullerton and Xu 2018, 185). However, one can use Wald tests to determine whether any given common factor significantly differs from 1 or a joint test for whether any of the common factors differ from 1. Note that the z statistics and significance tests reported in the table of estimates are for the null hypothesis that the factors differ from 0, which is not a particularly useful null hypothesis in this context. For this example, we find that all the common factors significantly differ from 1. In addition, they exhibit the theoretically expected pattern of a declining effect of father's and mother's education across higher levels of education degrees.

Note that `gencrm` does not force the estimates for common factors to range between 0 and 1 (as is true of Stata's `slogit` command). Researchers should be aware that such estimates being returned is evidence of model misspecification, and we would recommend removing the proportionality constraint and instead recommend either constraining the coefficients to be equal or allowing them to vary freely across cutpoint equations.

5 Conclusion

Our new command, `gencrm`, fits generalized continuation-ratio models and expands the capabilities available for researchers using Stata to study processes that unfold in a series of discrete stages. In particular, `gencrm` allows for a specification involving three different subsets of IVs: those whose coefficients are constrained to be equal across cutpoint equations (that is, subject to the parallel lines assumption), those whose coefficients are allowed to vary by a set of common factors across cutpoint equations, and those whose coefficients are allowed to freely vary across cutpoint equations.

In our view, the choice to fit a model that explicitly accounts for a stage-based process is best justified on a substantive or theoretical basis. At a minimum, the measure should index an underlying process that proceeds through a series of steps in which it is not possible to return to an earlier step. Beyond that, in some cases continuation-ratio models can provide a nice substantive match with certain types of research questions or theoretical consideration. For instance, sociological and economic theories of educational progression often emphasize decision points at each degree or credential level.

In addition to the choice of fitting a continuation-ratio model, it is best to use substantive knowledge or theory to make decisions about which coefficients, if any, to allow to freely vary across cutpoint equations and which coefficients, if any, to constrain to have proportional effects across cutpoint equations. In some cases, research questions dictate specific coefficients to vary across cutpoint equations. For instance, some studies of educational attainment have posited that a woman's advantage emerges more clearly at higher levels of education. This possibility can be explored directly by allowing the coefficient for women to vary across education thresholds. In other cases, theory might suggest a proportionality constraint is appropriate as in our example in which the effects of mother's and father's education are thought to diminish across education thresholds. In absence of such substantive or theoretical concerns, one can also treat the `free()` option as simply a means of testing the parallel lines assumption for a subset of or for all the predictors in a continuation-ratio model.

6 References

Agresti, A. 2010. *Analysis of Ordinal Categorical Data*. 2nd ed. Hoboken, NJ: Wiley.

Buis, M. L. 2007. seqlogit: Stata module to fit a sequential logit model. Statistical Software Components S456843, Department of Economics, Boston College. http://ideas.repec.org/c/boc/bocode/s456843.html.

Fagerland, M. W. 2014. adjcatlogit, ccrlogit, and ucrlogit: Fitting ordinal logistic regression models. *Stata Journal* 14: 947–964.

Fienberg, S. E. 2007. *The Analysis of Cross-Classified Categorical Data*. 2nd ed. New York: Springer.

Fullerton, A. S. 2009. A Conceptual Framework for Ordered Logistic Regression Models. *Sociological Methods and Research* 38: 306–347.

Fullerton, A. S., and J. Xu. 2016. *Ordered Regression Models: Parallel, Partial, and Non-Parallel Alternatives*. Boca Raton, FL: Chapman & Hall/CRC.

———. 2018. Constrained and unconstrained partial adjacent category logit models for ordinal response variables. *Sociological Methods and Research* 47: 169–206.

Goodman, L. A. 1983. The analysis of dependence in cross-classifications having ordered categories, using log-linear models for frequencies and log-linear models for odds. *Biometrics* 39: 149–160.

Hauser, R. M., and M. Andrew. 2006. Another look at the stratification of educational transitions: The logistic response model with partial proportionality constraints. *Sociological Methodology* 36: 1–26.

Hedeker, D., and R. J. Mermelstein. 1998. A multilevel thresholds of change model for analysis of stages of change data. *Multivariate Behavioral Research* 33: 427–455.

Long, J. S. 1997. *Regression Models for Categorical and Limited Dependent Variables*. Thousand Oaks, CA: Sage.

Mare, R. D. 1980. Social background and school continuation decisions. *Journal of the American Statistical Association* 75: 295–305.

McCullagh, P., and J. A. Nelder. 1989. *Generalized Linear Models*. 2nd ed. London: Chapman & Hall/CRC.

McKelvey, R. D., and W. Zavoina. 1975. A statistical model for the analysis of ordinal level dependent variables. *Journal of Mathematical Sociology* 4: 103–120.

Smith, T. W., M. Hout, and P. V. Marsden. 2016. General social surveys, 1972–2014 [cumulative file] (ICPSR 36319). Ann Arbor, MI. Inter-university Consortium for Political and Social Research.

Sobel, M. E. 1997. Modeling symmetry, asymmetry, and change in ordered scales with midpoints using adjacent category logit models for discrete data. *Sociological Methods and Research* 26: 213–232.

Tutz, G. 2012. *Regression for Categorical Data*. Cambridge: Cambridge University Press.

Williams, R. 2006. Generalized ordered logit/partial proportional odds models for ordinal dependent variables. *Stata Journal* 6: 58–82.

———. 2016. Understanding and interpreting generalized ordered logit models. *Journal of Mathematical Sociology* 40: 7–20.

Winship, C., and R. D. Mare. 1984. Regression models with ordinal variables. *American Sociological Review* 49: 512–525.

Wolfe, R. 1998. sg86: Continuation-ratio models for ordinal response data, *Stata Technical Bulletin* 44: 18–21. Reprinted in *Stata Technical Bulletin Reprints*, vol. 8, pp. 149–153. College Station, TX: Stata Press.

Yee, T. W. 2015. *Vector Generalized Linear and Additive Models: With an Implementation in R*. New York: Springer.

About the authors

Shawn Bauldry is an associate professor of sociology at Purdue University. His statistical research primarily focuses on the development of structural equation models. His work in this area has been published in *Sociological Methods and Research*, *Sociological Methodology*, *Psychological Methods*, and *Psychometrika*.

Jun Xu is a professor of sociology at Ball State University. His primary research interests include Asian America, China, comparative sociology, health, and statistical modeling and programming. His work has appeared in journals such as *Social Forces*, *Social Science and Medicine*, *Sociological Methods and Research*, *Social Science Research*, and *The Stata Journal*. He is also a coauthor (with Andrew S. Fullerton) of a statistical monograph titled *Ordered Regression Models: Parallel, Partial, and Non-Parallel Alternatives* published by Chapman and Hall of Taylor & Francis Group.

Andrew S. Fullerton is a professor of sociology at Oklahoma State University. His research interests include work and occupations, political sociology, social stratification, and quantitative methods. His work has been published in journals such as *Social Forces*, *Social Problems*, *Sociological Methods and Research*, *European Sociological Review*, *Public Opinion Quarterly*, and *Social Science Research*. He also published a book on ordered regression models with Chapman & Hall/CRC Press. He is currently working on several substantive projects related to cross-national differences in job insecurity and a methodological project focused on a Bayesian approach to the cumulative ordered regression model and tests of the parallel regression assumption.