Understanding Society

Working Paper Series

No. 2015-02

May 2015

BICOP: a Stata command for fitting bivariate ordinal regressions with residual dependence characterised by a copula function and normal mixture marginals

Monica Hernández-Alava

School of Health and Related Research Health Economics and Decision Science, University of Sheffield

Stephen Pudney

Institute of Social and Economic Research, University of Essex

Non-technical summary

When analysing survey data, we often want to examine the way that two observed outcomes vary together across individuals. Often, we are not able to observe these outcomes directly, but only in the form of ordinal indicators – in other words, measures which tell us whether one outcome is larger than another, but not by how much. So, for example, a survey question that asks how well people are managing financially may permit only answers like "living comfortably" or "just getting by"; we know that the former is better than the latter, but not how much better.

When analysing this kind of data, it is important to avoid making unnecessarily strong assumptions – to allow the data to "speak for themselves" as far as possible, rather than imposing unnecessary constraints on the analysis.

This paper describes a new statistical modelling technique, implemented for the widely-used statistical software package Stata, which allows statistical researchers to analyse jointly two ordinal survey measures, in a less restrictive way than is usual.

The paper contains an example of the use of the new software to carry out a simple analysis of the relationship between expectations of households' future financial wellbeing and their current financial state. The example is based on data from wave 3 of *Understanding Society* and the new analysis method reveals much more clearly than existing methods the considerable degree of pessimism about the future among families who are currently experiencing financial difficulty.

This new software has been developed as part of the analysis methods strand of the *Understanding Society* programme, and is being made freely available to all Stata users – anyone who wishes to use the current beta version should contact one of the authors: Monica Hernández-Alava (monica.hernandez@sheffield.ac.uk) or Steve Pudney (spudney@essex.ac.uk).

BICOP: A Stata command for fitting bivariate ordinal regressions with residual dependence characterized by a copula function and normal mixture marginals

Mónica Hernández-Alava
School of Health and Related Research
Health Economics and Decision Science
University of Sheffield
Sheffield, UK
monica.hernandez@sheffield.ac.uk

Stephen Pudney
Institute for Social and Econmic Research
University of Esex
Colchester, UK
spudney@essex.ac.uk

Abstract. This article describes a new Stata command, bicop, for fitting a model consisting of a pair of ordinal regressions with a flexible residual distribution, with each marginal distribution specified as a two-part normal mixture, and stochastic dependence governed by a choice of copula functions. The bicop command generalizes the existing biprobit and bioprobit commands which assume a bivariate normal residual distribution. The command and post estimation options are presented and explained with an illustrative application to data on financial wellbeing from the UK *Understanding Society* Panel Survey.

Keywords: st0001, bicop, bivariate ordinal regression, copula, mixture model

1 Introduction

The article is organised as follows: section 2 gives an overview of the generalized bivariate ordinal regression model and the approach we use to allow for non-normality in the residual distribution. Section 3 explains the predictors which are provided post-estimation and section 4 describes the bicop syntax and options, including the syntax for predict. Section 5 concludes with an empirical example to illustrate the use of the bicop command.

2 The Generalized Bivariate Ordinal Regression Model

The model is as follows:

$$Y_{i1}^* = X_{i1}\beta_1 + U_{i1} \tag{1}$$

$$Y_{i2}^* = X_{i2}\beta_2 + U_{i2} \tag{2}$$

where: Y_{i1}^* and Y_{i2}^* are latent variables; X_{i1} and X_{i2} are row vectors of covariates and β_1 and β_2 are conformable column vectors of cefficients. U_{i1}, U_{i2} are unobserved residuals which may be stochastically dependent and non-normal.

The observable counterparts of Y_{i1}^*, Y_{i2}^* are generated by the following threshold-crossing conditions:

$$Y_{ij} = r \quad \text{iff} \quad \Gamma_{rj} \leq Y_{ij}^* < \Gamma_{r+1j} \;; \qquad r = 1...R_j \;;\; j = 1,2 \tag{3}$$

where R_j is the number of categories of Y_{ij} and the Γ_{rj} are threshold parameters, with $\Gamma_{1j} = -\infty$ and $\Gamma_{R_j j} = +\infty$. (Note that in practice the Y_{ij} do not have to be scored as $1, 2, 3...$; `bicop` will work, whatever numerical values are used to index outcomes – only their ordering matters.)

Models of this type are not distribution-free. The likelihood function requires evaluation of the probability that (Y_{i1}^*, Y_{i2}^*) falls in a rectangle corresponding to the observed values of (Y_{i1}, Y_{i2}). For given parameter values, that probability can be computed from knowledge of the joint distribution function $F(U_{i1}, U_{i2})$, allowing the likelihood to be maximised numerically. However, if the assumed form for $F(U_{i1}, U_{i2})$ is incorrect, the probabilities appearing in the likelihood function will be misspecified, and the (pseudo-) ML estimator is inconsistent. This means that the standard approach based on a bivariate normal form for $F(.,.)$ is potentially vulnerable to bias. On the other hand, a full nonparametric specification for $F(.,.)$ would be complicated and unlikely to provide reliable estimates except in very large samples, so an intermediate degree of flexibility is desirable.

The model specification is based on a copula representation of the joint distribution of the residuals:

$$F(u_1, u_2) = c(F_1(u_1), F_2(u_2); \theta) \tag{4}$$

where: $F_1(U_{i1}) \equiv F(U_{i1}, \infty)$ and $F_2(U_{i2}) \equiv F(\infty, U_{i2})$ are the marginal distribution functions of U_1 and U_2; $c(.,.;\theta)$ is a copula function; and θ is a parameter governing the stochastic dependence of U_1 and U_2. The `bicop` command generalizes the standard bivariate normal model in two ways:

Marginals The marginal distributions $F_1(.)$ and $F_2(.)$ can each be specified as mixtures of two normal components. In the most general form:

$$F_j(u) = \pi_j \Phi\left(\frac{u - \mu_{j1}}{\sigma_{j1}}\right) + (1 - \pi_j)\Phi\left(\frac{u - \mu_{j2}}{\sigma_{j2}}\right) \tag{5}$$

where: π_j is the mixing parameter; (μ_{j1}, μ_{j2}) and $(\sigma_{j1}, \sigma_{j2})$ are location and dispersion parameters constrained to satisfy the mean and variance normalisations $\pi_j\mu_{j1} + (1 -$

$\pi_j)\mu_{j2} \equiv 0$ and $\pi_j\left(\sigma_{j1}^2+\mu_{j1}^2\right)+(1-\pi_j)\left(\sigma_{j2}^2+\mu_{j2}^2\right)=1$. These normal mixtures are able to capture a wide range of distributional shapes, especially those involving skewness or bimodality.

The `bicop` command carries out the optimization with respect to $\ln\left[\pi_j/(1-\pi_j)\right]$ rather than π_j, but both values are reported in the output. In the Stata output log, the mixing parameters $\pi_j, (1-\pi_j), \mu_{j1}, \mu_{j2}, \sigma_{j1}, \sigma_{j2}$ are labelled `pu1, pu2, mu1, mu2, su1, su2` for equation 1 and `pv1, pv2, mv1, mv2, sv1, sv2` for equation 2.

Dependence The copula function characterising the pattern of stochastic dependence between U_{i1} and U_{i2} can be any of the following 1-parameter functional forms.

Gaussian: $c(u_1,u_2)=\Phi\left(\Phi^{-1}(u_1),\Phi^{-1}(u_2);\theta\right)$, where $\Phi(.,.;\theta)$ is the distribution function of the bivariate normal with correlation coefficient $-1\le\theta\le 1$, and $\Phi^{-1}(.)$ is the inverse of the univariate $N(0,1)$ distribution function.

Clayton: $c(u_1,u_2)=\left[\max\left\{u_1^{-\theta}+u_2^{-\theta}-1,0\right\}\right]^{-1/\theta}$ for $-1\le\theta<0$ and $0<\theta\le\infty$ and $c(u_1,u_2)=u_1u_2$ for $\theta=0$

Frank: $-\frac{1}{\theta}\ln\left(1+\frac{(e^{-\theta u_1}-1)(e^{-\theta u_2}-1)}{e^{-\theta}-1}\right)$ for $\theta\neq 0$ and $c(u_1,u_2)=u_1u_2$ for $\theta=0$

Gumbel: $exp\left(-\left[(-\ln u_1)^\theta+(-\ln u_2)^\theta\right]^{1/\theta}\right)$ for $\theta\ge 1$

Joe: $1-\left[(1-u_1)^\theta+(1-u_2)^\theta-(1-u_1)^\theta(1-u_2)^\theta\right]^{1/\theta}$ for $\theta\ge 1$

These copulas are capable of representing a range of dependence structures. The Gaussian and the Frank copulas are similar in the sense that both of them allow for positive and negative dependence and dependence is symmetric in both tails. However, compared to the Gaussian copula, the Frank copula exhibits weaker dependence in the tails and dependence is strongest in the middle of the distribution. In contrast, the Clayton, Gumbel and Joe copulas do not allow for negative dependence and dependence in the tails is asymmetric. The Clayton copula exhibits strong left tail dependence and relatively weak right tail dependence. Thus, if two variables are strongly correlated at low values but not so correlated at high values, then the Clayton copula is a good choice. The Gumbel and Joe copulas display the opposite pattern with weak left tail dependence and strong right tail dependence. The right tail dependence is stronger in the Joe copula than in the Gumbel and thus the Joe copula is closer to the opposite of the Clayton copula.

`bicop` maximizes the likelihood with respect to the following transformation $\delta(\theta)$ rather than θ itself:

$$\delta(\theta)=\begin{cases} tanh^{-1}(\theta) & \text{Gaussian}\\ e^{\theta}-1 & \text{Clayton}\\ \theta & \text{Frank}\\ e^{\theta}+1 & \text{Gumbel, Joe}\end{cases}$$

The output from `bicop` reports both δ (labelled as `/depend`) and θ.

Both mixture and copula models have been found to be difficult to estimate in some circumstances (see McLachlan and Peel (2000) on the former and Trivedi and Zimmer (2005) on the latter). There are two distinct problems awaiting the unwary. Non-convergence of the likelihood optimizer often occurs in copula models, typically for some choices of copula function but not others. The problem tends to arise when the chosen copula function does a poor job of representing the pattern of dependence between the two residuals, and can often be resolved by switching to a different copula function. But poor starting values can also cause nonconvergence, and restarting the optimizer from a different point in the parameter space will work in some cases.

Another possible reason for nonconvergence is local non-identification of the mixture parameters. For the normal mixture (5), the parameter π_j is not identified at interior points in the parameter space where $\mu_{j1} = \mu_{j2}$ and $\sigma_{j1} = \sigma_{j2}$. Boundary problems also arise, since μ_{j1}, σ_{j1} are not identified when $\pi_j = 0$, nor are μ_{j2}, σ_{j2} identified when $\pi_j = 1$. All three regions correspond to a pure $N(0,1)$ distribution; consequently, if either of the marginal distributions is approximately normal, identification will be weak and non-convergence a likely result. These cases usually become evident if the `mlo(log showstep)` option is used to display current parameter values during the course of the optimization. Once spotted in this way, the relevant marginal can be respecified as an unmixed normal in a subsequent run.

Related to this last type of nonconvergence problem is the problem of testing for the appropriate number of mixture components. Standard likelihood ratio tests of $H_0 : U_j \sim N(0,1)$ against a 2-component normal mixture do not work correctly in this non-regular context (Titterington et al. 1985, p. 154) and we are not aware of any alternative formal procedure that is entirely satisfactory.

The problem of multiple optima is less obvious than nonconvergence – and therefore more dangerous. The existence of multiple optima poses problems for likelihood maximization in many mixture models, and should be assumed to be a potential pitfall here. The `bicop` command offers the standard Stata optimization options for starting values (see [R] **maximize**), and the application in section 5 provides an example of a recommended starting-values strategy.

3 Prediction

The `bicop` command allows the usual Stata prediction options post-estimation, through the evaluation of the linear indexes $X_{i1}\beta_1$ and $X_{i2}\beta_2$, the associated prediction standard errors and probabilities of specific outcomes for (Y_{i1}, Y_{i2}) conditional on the covariates (X_{i1}, X_{i2}). However `bicop` goes further than this and has options for conditional prediction. This can be used, for instance, as a way of converting (or "mapping" or "crosswalking") a measurement scale represented by the dependent variable Y_{i1} into another scale represented by Y_{i2}. Following use of `bicop`, the `predict` command does this by constructing estimates of the distribution of one dependent variable conditional on the observed outcome for the other. For example:

$$Pr(Y_{i2} = s|Y_{i1} = r, X_{i1}, X_{i2}) = \frac{Pr(Y_{i1} = r, Y_{i2} = s|X_{i1}, X_{i2})}{\sum_{s=1}^{R_2} Pr(Y_{i1} = r, Y_{i2} = s|X_{i1}, X_{i2})} \tag{6}$$

where $r \in [1, R_1]$ and $s \in [1, R_2]$ are specified levels for the two outcomes.

4 Command syntax

4.1 bicop

Syntax

`bicop` *depvar* [*indepvars*] [*if*] [*in*] [*weight*] , [`mixture(`*mixturetype*`)` `copula(`*copulatype*`)` `constraints(`*numlist*`)` `vce(`*vcetype*`)` `level(`#`)` `mlopts(`*maximize_options*`)` `from(`*init_specs*`)`]

Description

`bicop` is a user-written program which fits a generalized bivariate ordinal regression model using maximum likelihood estimation. It is implemented as a `d1 ml` evaluator. The model involves a pair of latent regression equations, each with a standard threshold-crossing condition to generate ordinal observed dependent variables. The bivariate residual distribution is specified to have marginals each with the form a two-part normal mixture, and a choice of copula functions to represent the pattern of dependencee between the two residuals.

Options

`mixture(`*mixturetype*`)` specifies the marginal distribution of each residual. There are five choices for *mixturetype*: `none` specifies each marginal distribution as a $N(0, 1)$ form; `mix1` specifies the residual from equation 1 to have a 2-part normal mixture distribution, but the residual from equation 2 to be $N(0, 1)$; `mix2` specifies $N(0, 1)$ for equation 1 and a normal mixture for equation 2; `both` allows each residual to have a different normal mixture distribution; and `equal` specifies that both residuals have the same normal-mixture distribution. If omitted, the `none` option is the default.

`copula(`*copulatype*`)` specifies the copula function to be used to control the pattern of stochastic dependence of the two residuals. There are five choices for *copulatype*: `gaussian` specifies the Gaussian copula. The four non-Gaussian options are `clayton`, `frank`, `gumbel` and `joe`. If omitted, the Gaussian copula is used by default. Note that if both the `mixture` and `copula` options are omitted, the `bicop` command produces the same results as the existing `bioprobit` and (if both dependent variables are binary) `biprobit` commands.

`vce(`*vcetype*`)` specifies how to estimate the variance-covariance matrix corresponding to the parameter estimates. The supported options are `oim`, `opg`, `robust` or `cluster`.

The current version of the command does not allow `bootstrap` or `jacknife` estimators. See [R] **vce_option**.

`level(`*#*`)` sets the significance level to be used for confidence intervals; see [R] **estimation options**.

`from(`*init_specs*`)`, where *init_specs* is either *matname* the name of a matrix containing the starting values, or *matname*`,[copy/skip]`. The `copy` sub-option specifies that the initialization vector be copied into the initial-value vector by position rather than by name, and the `skip` sub-option specifies that any irrelevant parameters found in the specified initialization vector be ignored.

`mlopts(`*maximize_options*`)` specifies the maximization options; *maximize_options* can include: technique(*algorithm_spec*), iterate(#), [no]log, trace, gradient, showstep, hessian, showtolerance, tolerance(#), ltolerance(#), gtolerance(#), nrtolerance(#), nonrtolerance, difficult; see [R] **maximize**. We recommend routine use of the `difficult` option.

4.2 predict

Syntax

`predict` *varname* [*if*] [*in*] [`,` *predicttype* `outcome(`*outcome*`)`]

Description

Following `bicop`, Stata's `predict` command can be used to construct a number of alternative predictions. They include the linear indexes $X_{i1}\beta_1$ and $X_{i2}\beta_2$ and corresponding standard errors; probabilites of the form $Pr(Y_{ij} = r|X_{ij})$ or $Pr(Y_{i1} = r, Y_{i2} = s|X_{i1}, X_{i2})$; and conditional probabilities of the form $Pr(Y_{ij} = r|Y_{ik} = s, X_{i1}, X_{i2})$.

Options

predicttype specifies the type of prediction required. If *predicttype* is `xb1` or `xb2`, the variable *varname* is constructed as $X_{i1}\beta_1$ or $X_{i2}\beta_2$ respectively. Set *predicttype* to `std1` or `std2` to construct *varname* as the corresponding prediction standard error. If *predicttype* is entered as `pr`, the prediction is calculated as a probability $Pr(Y_{i1} = r|X_{ij}), Pr(Y_{i2} = r|X_{ij})$ or $Pr(Y_{i1} = r, Y_{i2} = s|X_{i1}, X_{i2})$ with r and s specified by the `outcome` option. The options `pcond1` and `pcond2` specify the conditional probabilities $Pr(Y_{i1} = r|Y_{i2} = s, X_{i1}, X_{i2})$ or $Pr(Y_{i2} = s|Y_{i1} = r, X_{i1}, X_{i2})$ respectively, with r and s supplied by `outcome`.

`outcome(`*r*`,`*s*`)` specifies the outcome levels to be used in predicting probabilities for Y_{i1} and Y_{i2}. The possibilities for *predicttype* and `outcome(`*r*`,`*s*`)` are as follows.

Option	Predicted probability
, pr outcome(r, .)	$Pr(Y_{i1} = r \| X_{i1})$
, pr outcome(. ,s)	$Pr(Y_{i2} = s \| X_{i2})$
, pr outcome(r,s)	$Pr(Y_{i1} = r, Y_{i2} = s \| X_{i1}, X_{i2})$
, pcond1 outcome(r,s)	$Pr(Y_{i1} = r \| Y_{i2} = s, X_{i1}, X_{i2})$
, pcond2 outcome(r,s)	$Pr(Y_{i2} = s \| Y_{i1} = r, X_{i1}, X_{i2})$

5 An illustrative application

We now show how to use the `bicop` command to model bivariate ordinal data. Our example is based on data from *Understanding Society*: the UK Household Longitudinal Survey (UKHLS). See Knies (2014) for a detailed description of the survey. The main UKHLS sample began in 2009 with approximately 30,000 households. Interviewing proceeds continuously through the year with households interviewed annually, but each wave takes two years to complete and thus overlaps with the preceding and succeding waves. We use data on 40,294 individuals in 26,594 households, observed at wave 3, covering calendar years 2011-12. We analyze the responses to the following two questions about financial wellbeing (FWB), and construct the variables Y_1 and Y_2 as the corresponding 5-level and 3-level ordinal indicators, both recoded to give scales increasing in current or expected FWB (see Pudney (2011) for discussion and analysis of this FWB measure).

"*How well would you say you yourself are managing financially these days? Would you say you are...*" [1 Living comfortably; 2 Doing alright; 3 Just about getting by; 4 Finding it quite difficult; 5 or finding it very difficult?]

"*Looking ahead, how do you think you will be financially a year from now, will you be...*" 1 [Better off; 2 Worse off than you are now; 3 or about the same?]

There are ten explanatory covariates, comprising four continuous variables (age/10 and age squared/100 and log and log squared of household gross income equivalized by the modified OECD scale), and six binary variables (distinguishing people who are: female, homeowners, employed/self-employed, unemployed, retired, and long-term sick/disabled). Standard errors and test statistics are adjusted for clustering of individuals within households which are identified by the variable `hidp`.

The following (slightly edited) code demonstrates that `bicop` with the options `, mixture(none) copula(gaussian)` produces identical results to `bioprobit`.

```
. bicop finnow finfut age10 agesq100 female homeowner lnequinc lninc2 ///
>         empl unemp retired sick, copula(gaussian) mixture(none)  vce(cluster hidp)
LogL for independent ordered probit model -89654.704
initial:       log pseudolikelihood = -113388.49
rescale:       log pseudolikelihood = -103461.88
rescale eq:    log pseudolikelihood = -93210.201
Iteration 0:   log pseudolikelihood = -93210.201
...
Iteration 4:   log pseudolikelihood = -89564.842
Bivariate copula model for ordered variables (copula: gaussian, mixture: none)
                                                  Number of obs   =      40294
                                                  Wald chi2(20)   =    9807.62
Log pseudolikelihood = -89564.842                 Prob > chi2     =     0.0000
                               (Std. Err. adjusted for 26594 clusters in hidp)
------------------------------------------------------------------------------
             |               Robust
             |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
finnow       |
       age10 |  -.2981192   .0212167   -14.05   0.000     -.339703   -.2565353
    agesq100 |   .0337349   .0021674    15.56   0.000     .0294869    .0379828
      female |   .0446246   .0097811     4.56   0.000     .0254539    .0637953
   homeowner |   .5353464   .0144859    36.96   0.000     .5069546    .5637382
    lnequinc |   .3284021   .0142074    23.11   0.000      .300556    .3562482
      lninc2 |   .0316093   .0024536    12.88   0.000     .0268004    .0364183
        empl |   .1930021   .0227458     8.49   0.000     .1484212    .2375831
       unemp |  -.4110227    .033462   -12.28   0.000     -.476607   -.3454384
     retired |    .367516   .0302527    12.15   0.000     .3082219    .4268102
        sick |  -.4145216    .034477   -12.02   0.000    -.4820953   -.3469479
-------------+----------------------------------------------------------------
finfut       |
       age10 |  -.5780115   .0216186   -26.74   0.000    -.6203831   -.5356399
    agesq100 |   .0387898   .0020306    19.10   0.000     .0348098    .0427698
      female |  -.0496022   .0109948    -4.51   0.000    -.0711515   -.0280528
   homeowner |  -.1140781   .0154134    -7.40   0.000    -.1442879   -.0838684
    lnequinc |  -.0176062   .0111469    -1.58   0.114    -.0394537    .0042413
      lninc2 |   .0029657   .0013554     2.19   0.029     .0003093    .0056222
        empl |   .1149161   .0247606     4.64   0.000     .0663862    .1634459
       unemp |   .3558604   .0389355     9.14   0.000     .2795483    .4321726
     retired |   -.047609   .0301214    -1.58   0.114    -.1066459    .0114279
        sick |  -.2262982    .036674    -6.17   0.000    -.2981779   -.1544185
-------------+----------------------------------------------------------------
     /cut1_1 |  -2.323442   .0574358   -40.45   0.000    -2.436014    -2.21087
     /cut1_2 |  -1.637693   .0564935   -28.99   0.000    -1.748419   -1.526968
     /cut1_3 |  -.6068782    .056062   -10.83   0.000    -.7167577   -.4969987
     /cut1_4 |   .3814269   .0558544     6.83   0.000     .2719543    .4908995
     /cut2_1 |  -2.697588   .0615936   -43.80   0.000    -2.818309   -2.576867
     /cut2_2 |  -.9829583   .0596949   -16.47   0.000    -1.099958   -.8659584
     /depend |   .0814084   .0068055    11.96   0.000     .0680698     .094747
-------------+----------------------------------------------------------------
       theta |    .081229   .0067606                      .0679648    .0944645
------------------------------------------------------------------------------
Wald test of equality of coefficients chi2(df = 10 )=  7390.402 [p-value= 0.000 ]
. estimates store gaussian
```

```
. bioprobit  finnow finfut  age10 agesq100 female homeowner lnequinc lninc2 ///
>         empl unemp retired sick, vce(cluster hidp)
group(finfu |
         t) |      Freq.     Percent        Cum.
------------+-----------------------------------
          1 |      8,512       21.12       21.12
          2 |     23,193       57.56       78.68
          3 |      8,589       21.32      100.00
------------+-----------------------------------
      Total |     40,294      100.00
...
Iteration 2:   log pseudolikelihood = -89564.842
Bivariate ordered probit regression               Number of obs   =      40294
                                                  Wald chi2(10)   =    5553.80
Log pseudolikelihood = -89564.842                 Prob > chi2     =     0.0000
                               (Std. Err. adjusted for 26594 clusters in hidp)
------------------------------------------------------------------------------
             |               Robust
             |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
finnow       |
       age10 |  -.2981192   .0212166   -14.05   0.000     -.339703   -.2565353
    agesq100 |   .0337349   .0021674    15.56   0.000     .0294869    .0379828
      female |   .0446246   .0097811     4.56   0.000     .0254539    .0637953
   homeowner |   .5353464   .0144859    36.96   0.000     .5069546    .5637382
    lnequinc |   .3284021   .0142074    23.11   0.000      .300556    .3562482
      lninc2 |   .0316093   .0024536    12.88   0.000     .0268004    .0364183
        empl |   .1930021   .0227458     8.49   0.000     .1484212    .2375831
       unemp |  -.4110227    .033462   -12.28   0.000     -.476607   -.3454384
     retired |    .367516   .0302527    12.15   0.000     .3082219    .4268102
        sick |  -.4145216    .034477   -12.02   0.000    -.4820953   -.3469479
-------------+----------------------------------------------------------------
finfut       |
       age10 |  -.5780115   .0216185   -26.74   0.000     -.620383   -.5356399
    agesq100 |   .0387898   .0020306    19.10   0.000     .0348098    .0427697
      female |  -.0496022   .0109948    -4.51   0.000    -.0711515   -.0280528
   homeowner |  -.1140781   .0154134    -7.40   0.000    -.1442879   -.0838684
    lnequinc |  -.0176062   .0111469    -1.58   0.114    -.0394537    .0042413
      lninc2 |   .0029657   .0013554     2.19   0.029     .0003093    .0056222
        empl |   .1149161   .0247606     4.64   0.000     .0663862    .1634459
       unemp |   .3558604   .0389355     9.14   0.000     .2795483    .4321726
     retired |   -.047609   .0301214    -1.58   0.114    -.1066459    .0114279
        sick |  -.2262982    .036674    -6.17   0.000    -.2981779   -.1544185
-------------+----------------------------------------------------------------
athrho       |
       _cons |   .0814084   .0068055    11.96   0.000     .0680698     .094747
-------------+----------------------------------------------------------------
      /cut11 |  -2.323442   .0574357                     -2.436014    -2.21087
      /cut12 |  -1.637693   .0564934                     -1.748418   -1.526968
      /cut13 |  -.6068782   .0560619                     -.7167575   -.4969989
      /cut14 |   .3814269   .0558543                      .2719545    .4908993
      /cut21 |  -2.697588   .0615936                     -2.818309   -2.576867
      /cut22 |  -.9829583   .0596949                     -1.099958   -.8659585
-------------+----------------------------------------------------------------
         rho |    .081229   .0067606                      .0679648    .0944645
------------------------------------------------------------------------------
LR test of indep. eqns. :           chi2(1) =   179.72   Prob > chi2 = 0.0000
```

We then fit the same model, using each of the five copula options with Gaussian marginals, using the `mix(none)` option. This is followed by a generalization using the `mix(equal)` option to allow the assumption of marginal normality to be relaxed, while enforcing the same mixture distribution for each of the residuals. The resulting maximized likelihood values and Akaike Information Criteria are shown in the first two blocks of columns of Table 1. They strongly suggest that non-normality is present, with the combination of Clayton copula and 2-component normal mixture giving much higher likelihood values than bivariate ordered probit.

Table 1 Likelihoods for alternative specifications

Copula	Mixture type					
	`None`		`Equal`		`Mix2`	
	lnL	*AIC*	*lnL*	*AIC*	*lnL*	*AIC*
Gaussian	-89,564.8	179184	-89,285.5	178631	-88,943.7	177948
Clayton	-89,480.6	179015	-89,194.3	178449	-88,850.2	177763
Frank	-89,557.7	179169	-89,279.3	178619	-89,939.9	178198
Gumbel	-89,613.0	179280	-89,334.4	178729	-88,993.0	179328
Joe	-89,647.7	179361	-89,368.8	178798	-89,032.3	179353

We can also attempt to generalize the residual distribution further by relaxing the constraint that both residuals have the same mixture structure. For each choice of copula, we encounter the same convergence problem, illustrated by the following example using the Clayton copula:

```
. //      Try Clayton model with unrestricted mixtures
. bicop finnow finfut age10 agesq100 female homeowner lnequinc lninc2 ///
>         empl unemp retired sick, cop(clayton) mix(both) mlo(iterate(15) difficult)
LogL for independent ordered probit model -89654.704

initial:       log likelihood = -107176.64
...
Iteration 15:  log likelihood = -89901.677  (not concave)
convergence not achieved

Bivariate copula model for ordered variables (copula: clayton, mixture: both)
```

```
                                                  Number of obs   =      40294
                                                  Wald chi2(20)   =   12725.93
Log likelihood = -89901.677                       Prob > chi2     =     0.0000
```

	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
finnow						
age10	-.1716123	.0062308	-27.54	0.000	-.1838244	-.1594002
agesq100	.0191943	.000655	29.30	0.000	.0179105	.020478
female	.0149546	.0053418	2.80	0.005	.0044848	.0254243
homeowner	.2376988	.0073846	32.19	0.000	.2232252	.2521723
lnequinc	.1523295	.0049841	30.56	0.000	.1425609	.1620982
lninc2	.0145851	.0006055	24.09	0.000	.0133983	.0157718
empl	.0631583	.0107133	5.90	0.000	.0421606	.084156
unemp	-.2066636	.0144717	-14.28	0.000	-.2350276	-.1782996
retired	.1428658	.0144134	9.91	0.000	.1146161	.1711154
sick	-.2033765	.0157936	-12.88	0.000	-.2343314	-.1724215
finfut						
age10	-.4447745	.0207268	-21.46	0.000	-.4853982	-.4041507
agesq100	.0305336	.0020795	14.68	0.000	.0264579	.0346094
female	-.0393579	.0118494	-3.32	0.001	-.0625824	-.0161335
homeowner	-.0852661	.0133927	-6.37	0.000	-.1115153	-.0590168
lnequinc	-.009991	.0109312	-0.91	0.361	-.0314158	.0114338
lninc2	.0029836	.0014142	2.11	0.035	.0002118	.0057555
empl	.0848232	.0235034	3.61	0.000	.0387575	.1308889
unemp	.2703249	.0344152	7.85	0.000	.2028724	.3377775
retired	-.0358353	.0300046	-1.19	0.232	-.0946433	.0229727
sick	-.1791444	.036143	-4.96	0.000	-.2499834	-.1083053
/cut1_1	-1.157563	.	.	.	.	.
/cut1_2	-.8553972	.0070799	-120.82	0.000	-.8692736	-.8415208
/cut1_3	-.4053125	.0146067	-27.75	0.000	-.4339411	-.3766839
/cut1_4	.0669366	.0241591	2.77	0.006	.0195856	.1142876
/cut2_1	-2.095953	.0560381	-37.40	0.000	-2.205786	-1.986121
/cut2_2	-.5338557	.0552698	-9.66	0.000	-.6421826	-.4255289
/depend	.1477293	.0072101	20.49	0.000	.1335977	.1618609
/pu1	-12.80001	14.19792	-0.90	0.367	-40.62743	15.02742
/mu2	-.001457	.	.	.	.	.
/su2	.4725846	.0099411	47.54	0.000	.4531003	.4920689
/pv1	51.2	.	.	.	.	.
/mv2	-.175	.	.	.	.	.
/sv2	.7	.	.	.	.	.
theta	.1591991	.008358			.1429329	.1756967
pi_u_1	2.76e-06	.0000392			2.27e-18	.9999997
pi_u_2	.9999972	.0000392			2.98e-07	1
mean_u_1	527.753	.			.	.
mean_u_2	-.001457	.			.	.
var_u_1	2799.702	.			.	.
var_u_2	.2233362	.0093961			.2052999	.2421318
pi_v_1	1	.			.	.
pi_v_2	5.81e-23	.			.	.
mean_v_1	1.02e-23	.			.	.
mean_v_2	-.175	.			.	.
var_v_1	1	.			.	.
var_v_2	.49	.			.	.

Warning: convergence not achieved

Extensive experimentation with alternative starting values leads to the same difficulty: very slow movement of the optimizer towards a region of the parameter space where residual U_{i1} has a marginal normal distribution but the marginal distribution for residual U_{i2} is a normal mixture.[1] We resolve this problem by using the `mixture(mix2)` options, giving the likelihood values and AIC appearing in column 3 of Table 1. The Clayton specification with normal mixture for the distribution of residual U_{i2} seems clearly the best choice of model.

Note that, as part of our computation strategy, we carry out ten (curtailed) optimization runs from random perturbations of the equality-constrained estimates, and then use the best of these points for the final run. We recommend always using a preliminary search of this kind to reduce the risk of reaching an inferior local optimum. The full results for the best-fitting Clayton copula are as follows:

```
. //      prepare start values for restricted version
. matrix b=e(b)

. matrix b=beqmix_clayton[1,1..27],b[1,31..33]

. //      try 10 runs of 15 iterations each from randomly-perturbed start vectors
. matrix ttt=b

. local maxll=minfloat()

. set seed 67553

. forvalues r=1/10 {
  2.         quietly bicop finnow finfut age10 agesq100 female homeowner lnequinc lninc2 ///
>                empl unemp retired sick, cop(clayton) mix(mix2) from(ttt, skip) ///
>                mlo(iterate(15) difficult) vce(cluster hidp)
  3.         if e(ll)>`maxll´ {
  4.                 matrix maxpar=e(b)
  5.                 local maxll=e(ll)
  6.         }
  7.         forvalues s=1/30 {
  8.                 matrix ttt[1,`s´]=b[1,`s´]+(runiform()-0.5)/3
  9.         }
 10. }

. di in red "Best point reached: max loglikelihood = " `maxll´
Best point reached: max loglikelihood = -88850.148

. di in red "Parameter values..."
Parameter values...
```

1. The same difficulty arises also for other choices of copula function, but the highest likelihood values are again achieved for the Clayton copula.

```
. //      attempt to refine best point
. bicop finnow finfut age10 agesq100 female homeowner lnequinc lninc2 empl ///
>   unemp retired sick, cop(clayton) mix(mix2) from(maxpar, skip) vce(cluster hidp) ///
>   mlo(nonrtolerance tolerance(1e-5) ltolerance(1e-5) iterate(150) difficult)
LogL for independent ordered probit model -89654.704
Bivariate copula model for ordered variables (copula: clayton, mixture: mix2)
                                                Number of obs   =      40294
                                                Wald chi2(20)   =    7475.97
Log pseudolikelihood = -88850.146               Prob > chi2     =     0.0000
                              (Std. Err. adjusted for 26594 clusters in hidp)

                             Robust
                 Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]

finnow
      age10  -.2820093   .0210865   -13.37   0.000    -.3233381   -.2406806
   agesq100   .0321319    .002148    14.96   0.000     .0279219    .0363419
     female   .0375323   .0097288     3.86   0.000     .0184641    .0566004
  homeowner   .5328363   .0143427    37.15   0.000      .504725    .5609475
   lnequinc   .3275872   .0140158    23.37   0.000     .3001167    .3550576
     lninc2   .0312958   .0024163    12.95   0.000     .0265599    .0360317
       empl   .1863432   .0226306     8.23   0.000     .1419881    .2306983
      unemp  -.4105893   .0331605   -12.38   0.000    -.4755826   -.3455959
    retired   .3786815    .029997    12.62   0.000     .3198885    .4374744
       sick  -.4068714   .0341174   -11.93   0.000    -.4737402   -.3400026

finfut
      age10  -.1425606   .0138589   -10.29   0.000    -.1697235   -.1153976
   agesq100    .003294   .0015013     2.19   0.028     .0003515    .0062365
     female  -.0524056   .0056586    -9.26   0.000    -.0634962    -.041315
  homeowner  -.0648677   .0072929    -8.89   0.000    -.0791615    -.050574
   lnequinc   .0023395   .0061869     0.38   0.705    -.0097865    .0144656
     lninc2   .0023049    .000921     2.50   0.012     .0004998      .00411
       empl    .037065   .0112487     3.30   0.001     .0150179     .059112
      unemp   .1674668   .0165921    10.09   0.000     .1349469    .1999867
    retired  -.0258336   .0168612    -1.53   0.125    -.0588809    .0072138
       sick  -.1149761   .0181668    -6.33   0.000    -.1505824   -.0793698

     /cut1_1  -2.291702   .0571651   -40.09   0.000    -2.403744   -2.179661
     /cut1_2  -1.608285   .0562034   -28.62   0.000    -1.718442   -1.498128
     /cut1_3  -.5783802   .0557599   -10.37   0.000    -.6876676   -.4690928
     /cut1_4   .4105508   .0555529     7.39   0.000     .3016691    .5194325
     /cut2_1  -1.946939   .0352732   -55.20   0.000    -2.016073   -1.877805
     /cut2_2   .1300269   .0348113     3.74   0.000     .0617981    .1982558
     /depend   .1245541   .0071889    17.33   0.000     .1104642    .1386441
        /pv1  -.9111758   .0271512   -33.56   0.000    -.9643912   -.8579603
        /mv2   .6028447   .0087637    68.79   0.000     .5856682    .6200212
        /sv2   .3167733   .0074526    42.51   0.000     .3021665    .3313801

      theta   .1326433   .0081424                       .1167964    .1487152
     pi_v_1   .2867593   .0055532                       .2759999    .2977657
     pi_v_2   .7132407   .0055532                       .7022343    .7240001
   mean_v_1  -1.499423   .0189674                      -1.536598   -1.462247
   mean_v_2   .6028447   .0087637                       .5856682    .6200212
    var_v_1    .085475   .0360389                         .01484    .1561099
    var_v_2   .1003453   .0047216                       .0913046    .1098128

Wald test of equality of coefficients chi2(df = 10 )=  6249.520 [p-value= 0.000 ]
. estimates store clayton
```

To show the differences in results that can follow from using `bicop` rather than `bioprobit`, we use the predict command to construct predictions for expectations of the change in FWB conditional on current reported FWB. These are sample means of estimates of $Pr(Y_2 = s|Y_1 = r, X_i)$. The following code computes these for $s = 1$ (expected worsening of FWB) and $s = 3$ (expected improvement) and for all $r = 1...5$, summarizing the relationship by plotting them against r.

```
. gen tee=_n if _n<=5
(40289 missing values generated)
. foreach c in clayton gaussian {
  2.         gen up`c´=.
  3.         gen down`c´=.
  4.         forvalues t=1/5 {
  5.                 estimates restore `c´
  6.                 capture drop tmp*
  7.                 predict tmp if e(sample), pcond2 outcome(`t´,3)
  8.                 predict tmp1 if e(sample), pcond2 outcome(`t´,1)
  9.                 quietly summ tmp, meanonly
 10.                 quietly replace up`c´=r(mean) if tee==`t´
 11.                 quietly summ tmp1, meanonly
 12.                 quietly replace down`c´=r(mean) if tee==`t´
 13.         }
 14. }

line upgaussian upclayton tee, graphregion(fcolor(white) ilcolor(white) ///
icolor(white) lcolor(white) ifcolor(white)) ///
msymbol(none) xtitle("Current financial wellbeing") ytitle("Pr(better)") ///
yscale(titlegap(5)) xscale(titlegap(2)) xtick(1(1)5) xlabel(1(1)5) ///
legend(col(2) label(1 "Bivariate ordered probit") ///
label(2 "Generalized model")) lpattern(solid longdash) lcolor(black red)
line downgaussian downclayton tee, graphregion(fcolor(white) ilcolor(white) ///
icolor(white) lcolor(white) ifcolor(white)) ///
msymbol(none) xtitle("Current financial wellbeing") ytitle("Pr(worse)") ///
yscale(titlegap(5)) xscale(titlegap(2)) xtick(1(1)5) xlabel(1(1)5) ///
legend(col(2) label(1 "Bivariate ordered probit") ///
label(2 "Generalized model")) lpattern(solid longdash) lcolor(black red)
```

Figures 1 and 2 show these plots, compared with the corresponding probabilities from the bivariate ordered probit model. The most striking feature is that the generalized `bicop` model suggests considerably more pessimistic expectations conditional on a low current level of FWP, particularly in terms of the expectation of further worsening. Note that the data come from a period of government austerity targeted particularly on welfare recipients following a deep recession, so these very pessimistic predictions are not implausible.

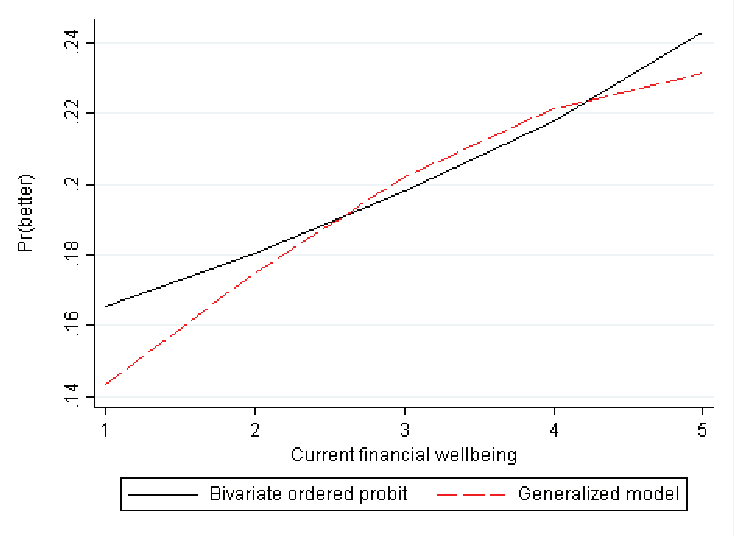

Figure 1: Predicted probability of expectation of better FWB conditional on current FWB

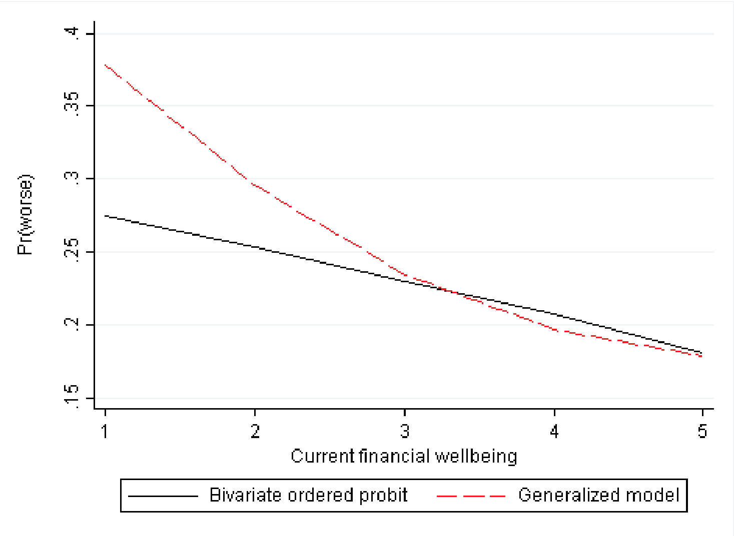

Figure 2: Predicted probability of expectation of worse FWB conditional on current FWB

6 Acknowledgements

This work was supported by the Medical Research Council under grant MR/L022575/1 and by the Economic and Social Research Council through the UK Centre for Longitudinal Studies (grant RES-586-47-0002). The views expressed in this article, and any errors or omissions, are those of the authors only.

7 References

Knies, G., ed. 2014. *Understanding Society The UK Household Longitudinal Study Waves 1-4, User Manual*. Colchester: ISER, University of Essex.

McLachlan, G. J., and D. Peel. 2000. *Finite mixture models*. New York: Wiley.

Pudney, S. E. 2011. Perception and retrospection: the dynamic consistency of responses to survey questions on wellbeing. *Journal of Public Economics* 95: 300–310.

Titterington, D. M., A. F. M. Smith, and U. E. Makov. 1985. *Statistical Analysis of Finite Mixture Distributions*. New York: Wiley.

Trivedi, P. K., and D. M. Zimmer. 2005. Copula Modeling: An Introduction for Practitioners. *Foundations and Trends in Econometrics* 1: 1–111.