The Stata Journal (2016)
16, Number 1, pp. 96–111

bireprob: An estimator for bivariate random-effects probit models

Alexander Plum
Otto von Guericke University Magdeburg
Magdeburg, Germany
alexander.plum@ovgu.de

Abstract. I present the `bireprob` command, which fits a bivariate random-effects probit model. `bireprob` enables a researcher to estimate two (seemingly unrelated) nonlinear processes and to control for interrelations between their unobservables. The estimator uses quasirandom numbers (Halton draws) and maximum simulated likelihood to estimate the correlation between the error terms of both processes. The application of `bireprob` is illustrated in two examples: the first one uses artificial data, and the second one uses real data. Finally, in a simulation, the performance of the estimator is tested and compared with the official Stata command `xtprobit`.

Keywords: st0426, bireprob, bivariate random-effects probit, maximum simulated likelihood, Halton draws

1 Introduction

When modeling a process (for example, the risk of becoming unemployed at a certain time point), one must distinguish between two types of error terms: an individual-specific time-invariant error term and a time-specific shock. When applying a random-effects estimator, one assumes that the persistent unobservable difference between the individuals is normally distributed. Two (seemingly unrelated) processes might be linked with each other by the correlation of their unobservables. For instance, an individual who is more likely to become unemployed (because of constraints in his or her ability, for example) might also be more likely to live in a poor household, and the differences between the individuals might be persistent over time. On the other hand, a time-specific shock that increases the risk of becoming unemployed might also increase the risk of becoming poor.[1]

In the past years, the number of journal articles accounting for correlation in the unobservables between two (seemingly unrelated) nonlinear processes has increased noticeably. Alessie, Hochguertel, and van Soest (2004) investigate the ownership dynamics of stocks and mutual funds; Devicienti and Poggi (2011) investigate the interrelation between poverty and social exclusion; Biewen (2009) and Ayllón (2015) investigate the dynamic relationship between unemployment and poverty; Stewart (2007) and Knabe and Plum (2013) investigate the interrelation between unemployment and low-pay; Miranda (2011) investigate the relationship between education and migration in

1. In example 2, I show that the `bireprob` command can be applied to any two-level equation system.

Mexico; Clark and Etilé (2006) investigate the spousal correlation in smoking behavior and Haan and Myck (2009) investigate the interrelation between poor health and unemployment.

To analyze the relationship between two (seemingly unrelated) nonlinear processes in Stata, one can use Hole's (2007) `mixlogit` command. However, `mixlogit` does not account for correlation in the time-specific shocks between two processes. In general, this is not possible for multilevel logistic regressions. Another possibility is shown by Ayllón (2014), who uses the statistical tool aML with Stata. aML is a multilevel multiprocessor estimator that estimates the correlation of random-effects error terms for various levels and different models. It is not restricted to a two-equation system. However, data need to be prepared carefully. With `bireprob`, a user-friendly estimator is presented that estimates two (seemingly unrelated) nonlinear processes and accounts for the correlation in the time-specific and individual-specific error terms.

The remainder of this article is structured as follows: Section 2 presents the bivariate random-effects probit model. Section 3 explains why quasirandom numbers (Halton draws) are used for simulation. Section 4 introduces the command `bireprob`. Sections 5 and 6 present examples of the application of the `bireprob` command. Section 7 shows the performance of `bireprob` in a simulation. The last section concludes.

2 The bivariate random-effects probit model

Assume that the observed binary outcome variables y_{1it} and y_{2it} are defined by the following latent-response models:

$$y_{1it} = \mathbf{1}\left(x'_{1it}\boldsymbol{\beta_1} + \nu_{1it} > 0\right)$$
$$y_{2it} = \mathbf{1}\left(x'_{2it}\boldsymbol{\beta_2} + \nu_{2it} > 0\right)$$

The subscript i refers to the panel variable (for example, individual or firm) with $i = 1, \ldots, N$ and t identifies the time point (for example, month or year) with $t = 1, \ldots, T$. The dependent variable y_{1it} is explained by the explanatory variables x_{1it}, and the dependent variable y_{2it} is explained by the explanatory variables x_{2it}. Furthermore, ν_{jit} refer to the process-specific error terms with $j \in (1, 2)$. It is assumed that ν_{jit} consists of an individual-specific time-invariant error term α_{ji} and of a time-specific idiosyncratic shock u_{jit}; thus $\nu_{jit} = \alpha_{ji} + u_{jit}$:

$$y_{1it} = \mathbf{1}\left(x'_{1it}\boldsymbol{\beta_1} + \alpha_{1i} + u_{1it} > 0\right)$$
$$y_{2it} = \mathbf{1}\left(x'_{2it}\boldsymbol{\beta_2} + \alpha_{2i} + u_{2it} > 0\right)$$

Because of normalization of the error terms, it is assumed that the individual-specific time-constant error terms are normally distributed, $\alpha_j \sim (0, \sigma^2_{\alpha_j})$, and that the idiosyncratic shocks are standard normally distributed, $u_j \sim (0, 1)$. The ratio of the time-constant individual-specific error term and composite error term is

$$\lambda_j = \text{corr}(\nu_{jit}, \nu_{jis}) = \frac{\sigma^2_{\alpha_j}}{\sigma^2_{\nu_j}}$$

for $t \neq s$.[2] Furthermore, it is assumed that both processes are interrelated by the correlation of their error terms:

$$\text{corr}(\nu_{1it}, \nu_{2is}) = \begin{cases} \rho_\alpha \sigma_{\alpha_1} \sigma_{\alpha_2} + \rho_u & \text{if } s = t \\ \rho_\alpha \sigma_{\alpha_1} \sigma_{\alpha_2} & \text{if } s \neq t \end{cases}$$

The individual likelihood function is the product of the joint probability of the observed binary outcome variable $\{P_i(\alpha_1, \alpha_2)\}$ and the joint density of the random-effects error terms $\{f_2(\alpha_1, \alpha_2; \mu_\alpha)\}$,

$$L_i = \int_{\alpha_1} \int_{\alpha_2} P_i(\alpha_1, \alpha_2) f_2(\alpha_1, \alpha_2; \mu_\alpha) \mathrm{d}\alpha_1 \mathrm{d}\alpha_2$$

with μ_α referring to the covariance of the random-effects error terms ($\mu_\alpha = \rho_\alpha \sigma_{\alpha_1} \sigma_{\alpha_2}$). Because it is assumed that the joint density of the random-effects error terms follows a bivariate normal distribution, the joint probability of the observed binary outcome variables is

$$P_{it}(\alpha_1, \alpha_2) = \Phi_2 \left\{ k_1 \left(x'_{1it} \boldsymbol{\beta_1} + \alpha_{1i} \right), k_2 \left(x'_{2it} \boldsymbol{\beta_2} + \alpha_{2i} \right), k_1 k_2 \rho_u \right\}$$

with

$$k_j = \begin{cases} 1 & \text{if } y_j = 1 \\ -1 & \text{else} \end{cases}$$

$\Phi_2[\cdot]$ is the bivariate normal cumulative distribution function. In general, the bivariate normal cumulative distribution function takes the following form (Greene 2012),

$$\Phi_2(x_1, x_2, \rho_u) = \int_{-\infty}^{x_2} \int_{-\infty}^{x_1} \phi_2(z_1, z_2, \rho_u) \mathrm{d}z_1 \mathrm{d}z_2$$

with the density

$$\phi_2(x_1, x_2, \rho_u) = \frac{e^{(-1/2)(x_1^2 + x_2^2 - 2\rho_u x_1 x_2)/(1-\rho_u^2)}}{2\pi(1-\rho_u^2)^{1/2}}$$

The sample likelihood now takes the following form:

$$L = \prod_{i=1}^{N} \int_{\alpha_1} \int_{\alpha_2} \left\{ \prod_{t=1}^{T} P_{it}(\alpha_1, \alpha_2) \right\} f_2(\alpha_1, \alpha_2; \mu_\alpha) \mathrm{d}\alpha_1 \mathrm{d}\alpha_2 \tag{1}$$

However, (1) cannot be solved analytically; therefore, the random-effects error terms must be integrated out. Strategies such as applying (adaptive) Gaussian quadrature or simulation belong to the most common approaches. For simulation, draws from random numbers are needed to simulate the bivariate normal distribution of the random-effects error terms. R uniformly distributed random draws r_j on the interval [0,1) are

2. Note that in `xtprobit`, this ratio is labeled by ρ, whereas ρ refers in this model to the correlation of the error terms.

taken and then transformed by the inverse cumulative standard normal distribution $\widetilde{\alpha}_j^r = \Phi^{-1}(r_j)$ (see figure 1). Thereafter, the Cholesky decomposition of the variance–covariance matrix of the bivariate normal distribution $\Sigma_\alpha = CC'$, with C being a lower triangular matrix, is integrated into the routine and updated during each iteration. The maximum simulated likelihood (MSL) is

$$\text{MSL} = \prod_{i=1}^{N} \frac{1}{R} \sum_{r=1}^{R} \left\{ \prod_{t=1}^{T} P_{it}(\alpha_1^r, \alpha_2^r) \right\}$$

The link between the transformed initial draws and the bivariate normally distributed numbers is

$$\alpha_1^r = \sigma_{\alpha_1} \widetilde{\alpha}_1^r$$
$$\alpha_2^r = \sigma_{\alpha_2} \rho_\alpha \widetilde{\alpha}_1^r + \sigma_{\alpha_2} \sqrt{1-\rho_\alpha^2}\, \widetilde{\alpha}_2^r$$

Because random numbers for the simulation are needed, quasirandom numbers are applied. Quasirandom numbers are based on prime numbers and are also called Halton draws. In section 3, I briefly introduce Halton draws and explain why they are applied instead of pseudorandom numbers.

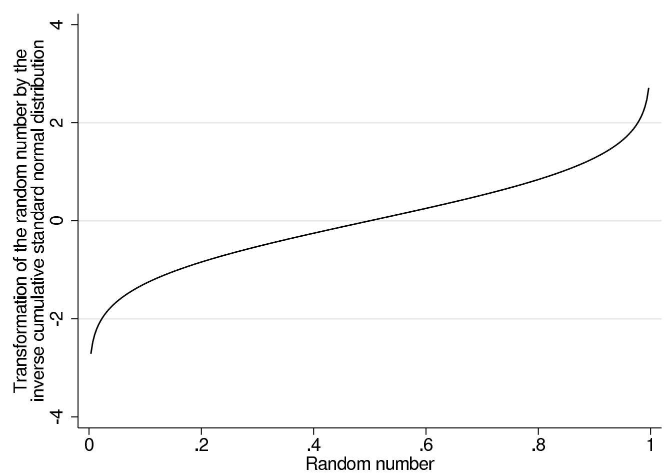

Figure 1. Transformation of the random number

3 Halton draws

Stata offers two possibilities for generating uniformly distributed random numbers. One possibility is to generate pseudorandom numbers by using the `runiform()` function. Another possibility is to generate quasirandom numbers such as Halton draws, which can be generated by using the `mdraws` command (Cappellari and Jenkins 2006). Halton draws are based on prime numbers and are often applied in the context of simulated maximum likelihood.[3] The advantage of Halton draws is that, compared with pseudorandom numbers, they have certain characteristics that make them more appropriate in the context of MSL:

1. They exhibit better coverage of the normal distribution, especially in the case of low numbers of observations (see figure 2).
2. Negatively correlated draws help to minimize the variance of the MSL maximand (Train 2009).

Therefore, for simulating bivariate normal distributions of the random-effects error terms, the `bireprob` estimator uses Halton draws.

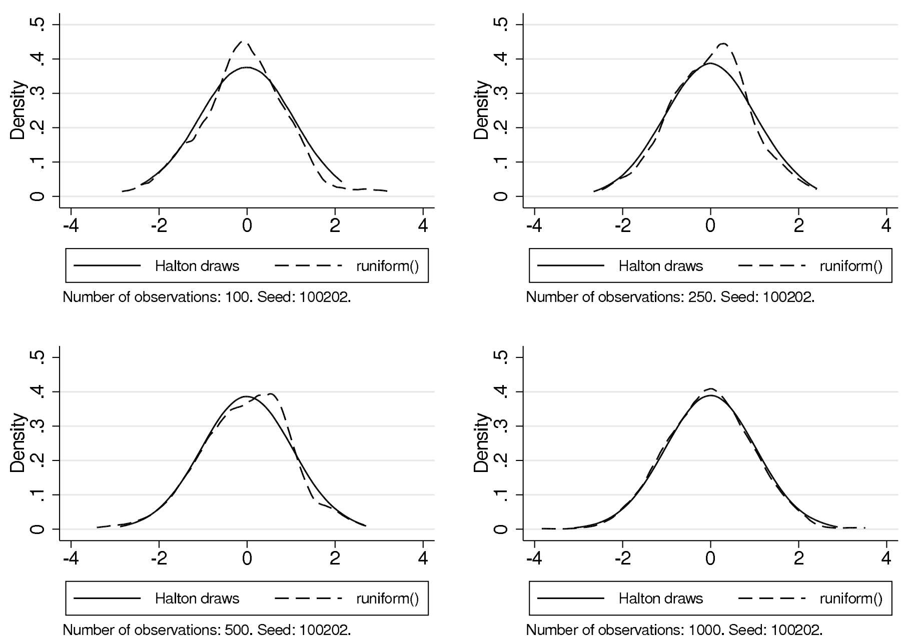

Figure 2. Coverage of different random-number generators

4 The bireprob command

The `bireprob` command fits a bivariate random-effects probit model that considers correlation in the random-effects error terms and in the idiosyncratic shocks. Note

3. For example, `mixlogit` (Hole 2007), `redpace` (Stewart 2006), `petpoisson` (Miranda 2012), or the Heckman estimator based on multivariate normal probabilities (Plum 2014).

that the `mdraws` command must be installed before using `bireprob`. `bireprob` checks whether `mdraws` is installed and, if it is not, will exit with a note to install the missing package.

4.1 Syntax

`bireprob` *depvar1 indepvars1* `(`*depvar2 indepvars2*`)` [*if*] [*in*] [`,` `draws(`#`)` `burn(`#`)` `primes(`*matname*`)` `from(`*matname*`)` `nosigma` `noalpha` `mutual`]

4.2 Options

`draws(`#`)` specifies the number of Halton draws needed for the simulation of the random effects. The default is `draws(10)`.

`burn(`#`)` specifies the number of initial elements of the Halton sequence to be dropped for burn in. The default is `burn(15)`. For details, see Cappellari and Jenkins (2006).

`primes(`*matname*`)` specifies a 1×2 matrix *matname* containing the primes to be used for the Halton sequences. The numbers specified must be integers. If `primes()` is not specified, the following primes are used: 2 and 3.

`from(`*matname*`)` lowers computational time by specifying a matrix *matname* that contains reasonable starting values for each equation. `bireprob` does not quietly fit a random-effects probit model by using `xtprobit`.

`nosigma` specifies that the estimator not control for correlation in the idiosyncratic shock.

`noalpha` specifies that the estimator not control for correlation in the random-effects error terms.

`mutual` specifies that the two dependent variables, y_1 and y_2, be mutually exclusive (for instance, the three labor market positions high paid, low paid, and unemployed [Stewart 2007]). When one applies this option, y_2 is considered only if $y_1 = 0$. `bireprob` checks whether both dependent variables are mutually exclusive and, if they are not, exits. When one applies this restriction, a notification will be displayed.

5 Example 1

In the first example, I use artificial data to introduce the `bireprob` command.

5.1 Constructing the dataset

At first, an artificial dataset that contains 500 individuals is constructed. An individual identifier (`id`) is generated that is based on the consecutive number of the respective observation; thus `id` $= 1, \ldots, 500$.

```
. version 13
. local obs=500
. local per=5
. set obs `obs´
number of observations (_N) was 0, now 500
. set seed 987654321
. generate id=_n
```

The two dependent variables, y_{1it} and y_{2it}, are defined in the following way:

$$y_{1it} = \mathbf{1}\left(1.5x_1 + \alpha_{1i} + u_{1it} > 0\right)$$
$$y_{2it} = \mathbf{1}\left(-2x_1 + 3x_2 + \alpha_{2i} + u_{2it} > 0\right)$$

The two random-effects error terms are standard normally distributed; hence, $\alpha_{ji} \in (0,1)$. They are negatively correlated with $\rho_\alpha = -0.3$. The idiosyncratic shocks, which are also standard normally distributed with $u_{ji} \in (0,1)$, are positively correlated with $\rho_u = 0.5$. In the next step, the random-effects error terms are generated with the help of the `drawnorm` command. Note that the variance–covariance matrix must be specified before applying `drawnorm`. Then, the dataset is expanded to a panel dataset with five time periods per individual (note that the number of time periods is defined in the local '`per`').

```
. matrix C = (1, -.3 \ -.3, 1)
. drawnorm re1 re2, n(`obs´) corr(C)
. expand `per´
(2,000 observations created)
```

For each individual, the time-point identifier `tper` is generated. Moreover, the two explanatory variables x_1 and x_2 are generated. For defining the idiosyncratic shocks u_{1it} and u_{2it}, one again applies the `drawnorm` command. Thereafter, the two outcome variables are generated; they become 1 if the value exceeds 0, and 0 otherwise.

```
. by id, sort: generate tper=_n
. generate x1=invnormal(runiform())
. generate x2=invnormal(runiform())
. matrix C = (1 , .5 \ .5 , 1)
. local obs=`obs´*`per´
. drawnorm u1 u2, n(`obs´) corr(C)
. sort id (tper)
. by id: generate y1=(1.5*x1 + re1 + u1>0)
. by id: generate y2=(-2*x1 + 3*x2 + re2 + u2>0)
```

5.2 Estimation

Before applying the `bireprob` command, one must use `xtset` to declare the *panelvariable* and the *timevariable*. In this example, the *panelvariable* is `id`, and the *timevariable* is `tper`. The panel is strongly balanced; hence, each individual is observed for the same number of time points. However, `bireprob` is not restricted to balanced panels and can also be applied to unbalanced panels (see section 6).

Then, the `bireprob` command is applied. In this application, the first dependent variable is y_1, and the explanatory variable is x_1. In parentheses, the first variable indicates the second dependent variable, which is y_2, and x_1 and x_2 are used as explanatory variables. Furthermore, 50 Halton draws are chosen for the estimation (in section 5.4, I show how the results are affected by the number of Halton draws).

```
. xtset id tper
       panel variable:  id (strongly balanced)
        time variable:  tper, 1 to 5
                delta:  1 unit
. bireprob y1 x1 (y2 x1 x2), draws(50)
Dependent variable (1st equation): y1
Dependent variable (2nd equation): y2
Explanatory variables (1st equation): x1
Explanatory variables (2nd equation): x1 x2
Estimating 1st equation with xtprobit.
Estimating 2nd equation with xtprobit.
Generating 50 Halton draws with prime numbers 2 and 3. 15 Halton draws are
> burned in.
Estimating a bivariate random-effects probit model
Iteration 0:   log likelihood = -1731.9335
Iteration 1:   log likelihood = -1718.5778
Iteration 2:   log likelihood = -1718.5062
Iteration 3:   log likelihood = -1718.5062
Bivariate Random-effects Probit Model, 50 Halton draws
                                               Number of obs     =      2,500
                                               Wald chi2(1)      =     484.82
Log likelihood = -1718.5062                    Prob > chi2       =     0.0000

------------------------------------------------------------------------------
             |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
y1           |
          x1 |   1.502347   .0682307    22.02   0.000     1.368617    1.636077
       _cons |  -.0869098    .059284    -1.47   0.143    -.2031042    .0292847
-------------+----------------------------------------------------------------
y2           |
          x1 |  -2.006314   .1310377   -15.31   0.000    -2.263143   -1.749485
          x2 |   2.918305   .1854261    15.74   0.000     2.554876    3.281733
       _cons |   .0454579   .0645171     0.70   0.481    -.0809933     .171909
-------------+----------------------------------------------------------------
 /logitlam_1 |    .053315   .0696122     0.77   0.444    -.0831224    .1897524
 /logitlam_2 |  -.0264849   .1163182    -0.23   0.820    -.2544644    .2014946
     /atsiga |  -.2961899   .0932455    -3.18   0.001    -.4789478   -.113432
     /atsigu |    .443295   .1044243     4.25   0.000     .2386271    .6479629
-------------+----------------------------------------------------------------
     alpha_1 |   1.112523   .1548903     7.18   0.000     .8468388    1.461561
     alpha_2 |   .9484087   .2206344     4.30   0.000     .6011392    1.496291
   rho_alpha |   -.287822   .0855209    -3.37   0.001    -.4454005   -.112948
   rho_sigma |   .4163719   .0863207     4.82   0.000     .2341986     .570297
```

At the start of the estimation procedure, the `bireprob` command displays the dependent and the independent variables of the first and of the second equation. In the next two steps, the `bireprob` command quietly fits a random-effects probit model for each equation by using `xtprobit`. The estimated coefficients of the explanatory variables and the variances of the random-effects error terms are used as starting values for the `bireprob` command. As a starting value for ρ_α and ρ_u, 0 is chosen. Then, the Halton draws are generated, in this case 50 draws per individual. Because no prime numbers are defined, the prime numbers 2 and 3 are used. Moreover, the number of Halton draws that should be burned in is not defined. Therefore, the default number of initial draws dropped per dimension, 15, is used. Finally, the bivariate random-effects probit model is fit.

Looking at the output table and comparing the coefficients with the true values, we can see that the estimated coefficients are close to the true values. In the four last lines, the variances of the random effects and the correlation parameters are displayed. Referring to the variances of the random-effects error terms, $\sigma^2_{\alpha_1} = 1.11$ and $\sigma^2_{\alpha_2} = 0.95$, we see that both are close to 1. Furthermore, a negative correlation parameter of the random-effects error terms is found, $\rho_\alpha = -0.29$, and a positive correlation of the idiosyncratic shocks is found, $\rho_u = 0.42$. Moreover, all estimated coefficients are significantly different from 0 at the 1% level.

5.3 Predicted probabilities

I now show how to predict probabilities. In this example, we are interested in calculating the probability that $y_1 = 1$ and $y_2 = 1$ simultaneously when $x_1 = x_2 = 1$. Before calculating the predicted probabilities, we should note that the variances of the composite error terms are not standard normally distributed ($\sigma^2_{\nu_j} \neq 1$); therefore, the coefficients must be rescaled by $\sqrt{1/\sigma^2_{\nu_j}}$ (Arulampalam 1999). In general, the predicted probabilities that $y_1 = y_2 = 1$ are calculated as follows:

$$\widehat{p} = \Phi_2 \left\{ \left(x'_{1it}\widehat{\beta}_1 \right) \sqrt{\frac{1}{\widehat{\sigma}^2_{\nu_1}}}, \left(x'_{2it}\widehat{\beta}_2 \right) \sqrt{\frac{1}{\widehat{\sigma}^2_{\nu_2}}}, \widehat{\rho}_u \right\}$$

Note that the coefficients referring to the variance, $\sigma^2_{\alpha_1}$ and $\sigma^2_{\alpha_2}$, are included in the estimator as the square root of their logarithm; thus $\ln\left(\sqrt{\sigma^2_{\alpha_j}}\right)$. The correlation parameters, ρ_u and ρ_α, are included as the inverse hyperbolic tangent in the estimator. Furthermore, the variances of the idiosyncratic shocks are equal to 1; thus $\sigma^2_{\nu_j} = \sigma^2_{\alpha_j} + 1$. The predicted probabilities are calculated with the `nlcom` command, and the probability that both dependent variables are equal to 1 given that $x_1 = x_2 = 1$ is 0.67.

```
. nlcom (pred: binormal(
> (([y1]_b[x1]*1 + [y1]_b[_cons])*sqrt(1/((exp(_b[/logitlam_1]))^2+1))),
> (([y2]_b[x1]*1 + [y2]_b[x2]*1 + [y2]_b[_cons])*
> sqrt(1/((exp(_b[/logitlam_2]))^2+1))), tanh(_b[/atsigu])))
        pred:  binormal( (([y1]_b[x1]*1 + [y1]_b[_cons])*
> sqrt(1/((exp(_b[/logitlam_1]))^2+1))), (([y2]_b[x1]*1 +
> [y2]_b[x2]*1 + [y2]_b[_cons])*sqrt(1/((exp(_b[/logitlam_2]))^2+1))),
> tanh(_b[/atsigu]))
```

	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
pred	.6667429	.0215528	30.94	0.000	.6245003	.7089855

5.4 Sensitivity analysis

To illustrate how the number of Halton draws affects the estimation results, I repeat the estimation with different numbers of Halton draws: I start at 25 draws and increase the random numbers successively by an additional 5 draws until reaching 250 draws. The effect of the number of Halton draws on the simulated log likelihood, the estimated variances of the random effects ($\sigma^2_{\alpha_j}$) and the correlation parameters (ρ_α, ρ_u) is shown in figure 3. We can see that the simulated log likelihood changes only slightly depending on the number of Halton draws and that the variances $\sigma^2_{\alpha_j}$ and the correlation parameters ρ_α and ρ_u are also on the same level.

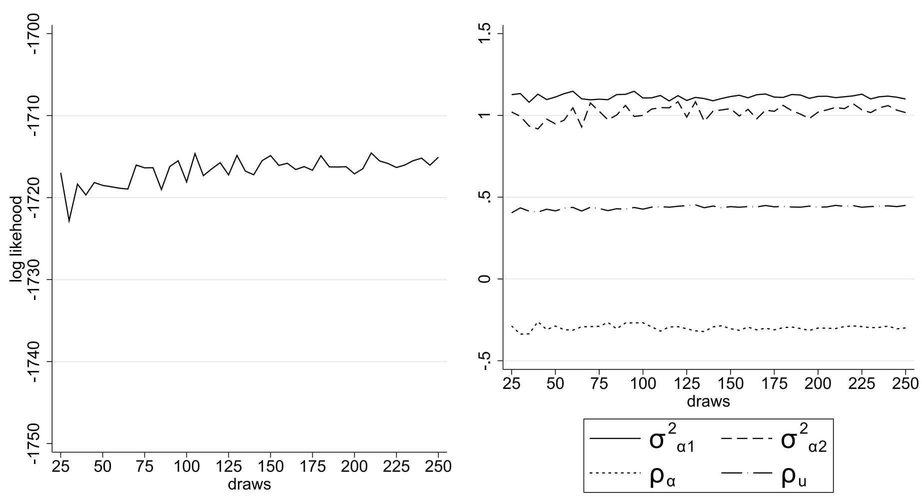

Figure 3. Sensitivity analysis

6 Example 2

In this example, I show how `bireprob` can be used with real data in the context of a two-level equation system with mutually exclusive dependent variables. For the illustration, data about teachers' evaluations of pupils' behavior are used. These data were also used by Haan and Uhlendorff (2006) and Hole (2007).[4] There are three different types of schools (`tby`), and the analysis focuses on whether there is some unobserved heterogeneity between those schools. The sample comprises 48 schools (`scy3`) and 1,313 pupils. The school is the panel variable, and the pupils are treated as the time variable.

```
. use jspmix, clear
. tabulate tby, gen(y)

        tby |      Freq.     Percent        Cum.
------------+-----------------------------------
          1 |        329       25.06       25.06
          2 |        678       51.64       76.69
          3 |        306       23.31      100.00
------------+-----------------------------------
      Total |      1,313      100.00
. by scy3, sort: generate tper=_n
. xtset scy3 tper
       panel variable:  scy3 (unbalanced)
        time variable:  tper, 1 to 85
                delta:  1 unit
```

Because each student can be at only one type of school, this variable is mutually exclusive. If $y_1 = 1$, the pupil is attending a school of the first category. If $y_1 = 0$ and $y_2 = 1$, the pupil is attending a school of the second category. If $y_1 = y_2 = 0$, the pupil is attending a school of the third category. Therefore, when applying `bireprob`, we choose the `mutual` option to indicate that y_2 is considered only if $y_1 = 0$. For the estimation, we take 50 Halton draws. Furthermore, we control for correlation only in the random effects. Thus we use the `nosigma` option. Following Haan and Uhlendorff (2006) and Hole (2007), we take a single explanatory variable: the gender of the pupil, `sex`.

4. In both articles, multinomial logistic regressions are applied.

```
. bireprob y1 sex (y2 sex), mutual nosigma draws(50)
  (output omitted)
Bivariate Random-effects Probit Model, 50 Halton draws
                                                Number of obs     =      1,313
                                                Wald chi2(1)      =      26.53
Log likelihood =  -1300.332                     Prob > chi2       =     0.0000

             |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
y1           |
         sex |  -.4173062    .081013    -5.15   0.000    -.5760886   -.2585237
       _cons |  -.5215657   .0840545    -6.21   0.000    -.6863096   -.3568219
-------------+----------------------------------------------------------------
y2           |
         sex |  -.3295871   .0867177    -3.80   0.000    -.4995506   -.1596237
       _cons |   .7069166   .0740025     9.55   0.000     .5618744    .8519588
-------------+----------------------------------------------------------------
  /logitlam_1|  -.8405676   .1689955    -4.97   0.000    -1.171793   -.5093425
  /logitlam_2|  -1.573497   .3180058    -4.95   0.000    -2.196777    -.950217
     /atsiga |   .5123391   .3707035     1.38   0.167    -.2142265    1.238905
-------------+----------------------------------------------------------------
     alpha_1 |   .1861625   .0629213     2.96   0.003     .0959829    .3610694
     alpha_2 |   .0429811   .0273365     1.57   0.116     .0123567    .1495037
   rho_alpha |   .4717657   .2881987     1.64   0.102    -.2110084    .8451429
```

The results indicate that there is some evidence for correlation in the random effects; however, ρ_α is not significantly different from 0 at the 10% level.

7 Simulation

Finally, I test the performance of `bireprob` by a simulation and compare the results with those of `xtprobit`. Again I use artificial data. To emphasize the necessity to control for correlation in the unobservables, I choose a dynamic model in which the current outcome depends on the outcome in the previous period. In the economic literature, an often-examined example is state dependence in unemployment (among others, see Arulampalam, Booth, and Taylor [2000]). Furthermore, the current outcome depends on the past outcome of the second dependent variable and vice versa. For example, while the first dependent variable is unemployment, the second dependent variable could be bad health. Past unemployment could significantly increase the risk of suffering from bad health. The same is true in the opposite direction: bad health not only increases one's risk of being affected by bad health in the future but also makes it more likely for one to become unemployed. In the simulation, the underlying model takes the following structure:[5]

$$
\begin{aligned}
y_{1it} &= \mathbf{1}\left(1y_{1it-1} - 1y_{2it-1} + \alpha_{1i} + u_{1it} > 0\right) \\
y_{2it} &= \mathbf{1}\left(1y_{2it-1} - 1y_{1it-1} + \alpha_{2i} + u_{2it} > 0\right)
\end{aligned}
$$

Both random-effects error terms are standard normally distributed and positively correlated with $\rho_\alpha = 0.7$. Not controlling for correlation in the random effects would lead

5. The respective do-file can be found in the supplement.

to an overestimation of the variances of the random effects and to an underestimation of the lagged dependent variables' coefficients.

The artificial dataset consists of 500 individuals observed for 5 subsequent time points. Therefore, the panel is strongly balanced. The outcome in the initial period is randomly assigned; thus we do not have to control for the "initial conditions problem" (Heckman 1981). The above equation system is estimated in total 100 times by `xtprobit` and by `bireprob`[6] (with 100 Halton draws). In each round, a new random draw of the distribution of the random effects and the idiosyncratic shocks is generated. The mean over all 100 estimations of the coefficients and the standard errors can be found in table 1. The first column of table 1 shows that when one does not control for correlated random effects, the coefficients are on a much lower level in absolute terms. However, when one does control for correlated random effects, the coefficients are much closer to the true value.

Table 1. Simulation results

Coefficients	`xtprobit`		`bireprob`†	
	Coefficient	Standard error	Coefficient	Standard error
y_1				
y_{1it-1}	0.935***	0.063	1.011	0.062
y_{2it-1}	−0.804***	0.067	−1.000	0.070
y_2				
y_{1it-1}	−0.802***	0.067	−0.997	0.070
y_{2it-1}	0.937***	0.063	1.011*	0.062
$\sigma^2_{\alpha_1}$	1.123***		1.001	
$\sigma^2_{\alpha_2}$	1.133***		1.010	
ρ_α	−		0.697	
Observations	100		100	
Log likelihood	−2 451.452		−2 410.959	

† 100 Halton draws with prime numbers 2 and 3.
* Coefficient statistically significantly different from the true value at the 0.10 level; ** at the 0.05 level; *** at the 0.01 level.

We can also conclude this when comparing the distribution of the coefficients in figure 4 (the solid vertical line refers to the true value). Furthermore, note that the variances of the random effects are greater in the first model than when controlling for the correlation between the random effects. Moreover, whether the means of the coefficients and the variances are significantly different from the true values is tested. Referring to the `xtprobit` model, we see that every coefficient and variance is significantly different from the true value at the 1% level. However, in the `bireprob` model,

6. It is not controlled for correlation in the idiosyncratic shocks.

only the lagged dependent y_{2it-1} of the second equation is significantly different from the true value at the 10% level. For the remaining estimated coefficients and variances, no significant difference from the true value is detected.

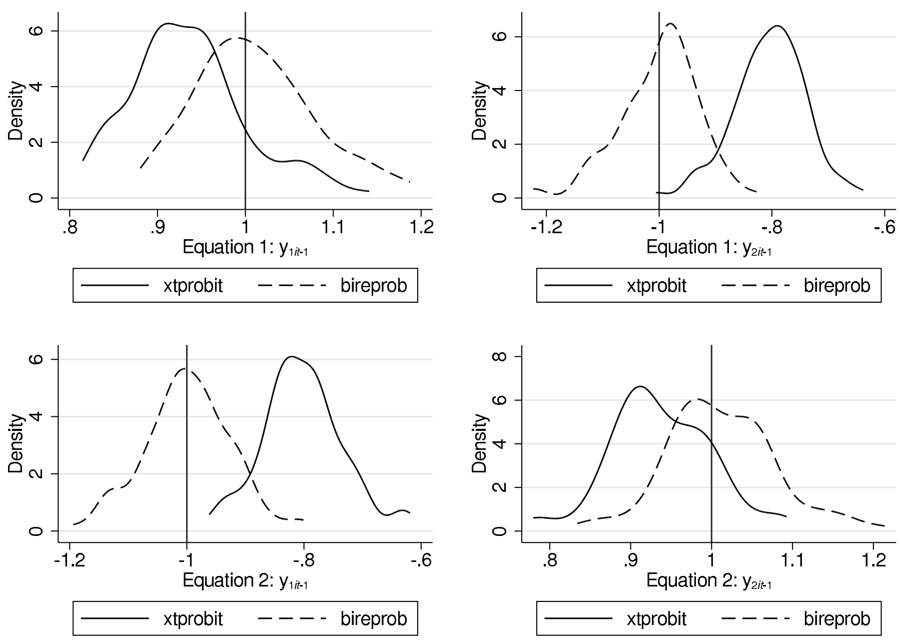

Figure 4. Distribution of the estimated coefficients

8 Conclusion

In this article, I presented the `bireprob` command. `bireprob` fits a bivariate random-effects probit model and allows one to control for correlation in the random-effects error terms and in the idiosyncratic shocks. The advantage of this estimator is that compared with existing estimators, such as `mixlogit` or aML, it requires no specific data preparation. After presenting the command and its options, I gave two examples and a simulation: the first example is based on artificial data, and the second example on real data. The simulation showed that not controlling for correlation in the random effects might cause biased estimation results. An open research task remains in lowering computational time.

9 Acknowledgments

I thank Sara Ayllón and an anonymous referee for helpful comments. Moreover, I acknowledge the financial support of the German Research Foundation (DFG, project KN 984/1-1).

10 References

Alessie, R., S. Hochguertel, and A. van Soest. 2004. Ownership of stocks and mutual funds: A panel data analysis. *Review of Economics and Statistics* 86: 783–796.

Arulampalam, W. 1999. A note on estimated coefficients in random effects probit models. *Oxford Bulletin of Economics and Statistics* 61: 597–602.

Arulampalam, W., A. L. Booth, and M. P. Taylor. 2000. Unemployment persistence. *Oxford Economic Papers* 52: 24–50.

Ayllón, S. 2014. From Stata to aML. *Stata Journal* 14: 342–362.

———. 2015. Youth poverty, employment, and leaving the parental home in Europe. *Review of Income and Wealth* 61: 651–676.

Biewen, M. 2009. Measuring state dependence in individual poverty histories when there is feedback to employment status and household composition. *Journal of Applied Econometrics* 24: 1095–1116.

Cappellari, L., and S. P. Jenkins. 2006. Calculation of multivariate normal probabilities by simulation, with applications to maximum simulated likelihood estimation. *Stata Journal* 6: 156–189.

Clark, A. E., and F. Etilé. 2006. Don't give up on me baby: Spousal correlation in smoking behaviour. *Journal of Health Economics* 25: 958–978.

Devicienti, F., and A. Poggi. 2011. Poverty and social exclusion: Two sides of the same coin or dynamically interrelated processes? *Applied Economics* 43: 3549–3571.

Greene, W. H. 2012. *Econometric Analysis*. 7th ed. Upper Saddle River, NJ: Prentice Hall.

Haan, P., and M. Myck. 2009. Dynamics of health and labor market risks. *Journal of Health Economics* 28: 1116–1125.

Haan, P., and A. Uhlendorff. 2006. Estimation of multinomial logit models with unobserved heterogeneity using maximum simulated likelihood. *Stata Journal* 6: 229–245.

Heckman, J. J. 1981. The incidental parameters problem and the problem of initial condition in estimating a discrete time-discrete data stochastic process. In *Structural Analysis of Discrete Data with Econometric Applications*, ed. C. F. Manski and D. McFadden, 179–195. Cambridge: MIT Press.

Hole, A. R. 2007. Fitting mixed logit models by using maximum simulated likelihood. *Stata Journal* 7: 388–401.

Knabe, A., and A. Plum. 2013. Low-wage jobs—Springboard to high-paid ones? *Labour* 27: 310–330.

Miranda, A. 2011. Migrant networks, migrant selection, and high school graduation in México. *Research in Labor Economics* 33: 263–306.

———. 2012. petpoisson: Stata module to estimate an endogenous participation endogenous treatment poisson model by MSL. Statistical Software Components S457393, Department of Economics, Boston College. https://ideas.repec.org/c/boc/bocode/s457393.html.

Plum, A. 2014. Simulated multivariate random-effects probit models for unbalanced panels. *Stata Journal* 14: 259–279.

Stewart, M. B. 2006. Maximum simulated likelihood estimation of random-effects dynamic probit models with autocorrelated errors. *Stata Journal* 6: 256–272.

———. 2007. The interrelated dynamics of unemployment and low-wage employment. *Journal of Applied Econometrics* 22: 511–531.

Train, K. E. 2009. *Discrete Choice Methods with Simulation*. 2nd ed. Cambridge: Cambridge University Press.

About the author

Alexander Plum is a research assistant at the Chair of Public Economics at the Otto von Guericke University Magdeburg. His main research interests are labor economics and applied econometrics.