



The World's Largest Open Access Agricultural & Applied Economics Digital Library

This document is discoverable and free to researchers across the globe due to the work of AgEcon Search.

Help ensure our sustainability.

Give to AgEcon Search

AgEcon Search

<http://ageconsearch.umn.edu>

aesearch@umn.edu

*Papers downloaded from **AgEcon Search** may be used for non-commercial purposes and personal study only. No other use, including posting to another Internet site, is permitted without permission from the copyright owner (not AgEcon Search), or as allowed under the provisions of Fair Use, U.S. Copyright Act, Title 17 U.S.C.*

No endorsement of AgEcon Search or its fundraising activities by the author(s) of the following work or their employer(s) is intended or implied.

THE STATA JOURNAL

Editor

H. Joseph Newton
Department of Statistics
Texas A&M University
College Station, Texas 77843
979-845-8817; fax 979-845-6077
jnewton@stata-journal.com

Editor

Nicholas J. Cox
Department of Geography
Durham University
South Road
Durham DH1 3LE UK
n.j.cox@stata-journal.com

Associate Editors

Christopher F. Baum
Boston College

Nathaniel Beck
New York University

Rino Bellocco
Karolinska Institutet, Sweden, and
University of Milano-Bicocca, Italy

Maarten L. Buis
Tübingen University, Germany

A. Colin Cameron
University of California–Davis

Mario A. Cleves
Univ. of Arkansas for Medical Sciences

William D. Dupont
Vanderbilt University

David Epstein
Columbia University

Allan Gregory
Queen's University

James Hardin
University of South Carolina

Ben Jann
University of Bern, Switzerland

Stephen Jenkins
London School of Economics and
Political Science

Ulrich Kohler
WZB, Berlin

Frauke Kreuter
University of Maryland–College Park

Peter A. Lachenbruch
Oregon State University

Jens Lauritsen
Odense University Hospital

Stanley Lemeshow
Ohio State University

J. Scott Long
Indiana University

Roger Newson
Imperial College, London

Austin Nichols
Urban Institute, Washington DC

Marcello Pagano
Harvard School of Public Health

Sophia Rabe-Hesketh
University of California–Berkeley

J. Patrick Royston
MRC Clinical Trials Unit, London

Philip Ryan
University of Adelaide

Mark E. Schaffer
Heriot-Watt University, Edinburgh

Jeroen Weesie
Utrecht University

Nicholas J. G. Winter
University of Virginia

Jeffrey Wooldridge
Michigan State University

Stata Press Editorial Manager
Stata Press Copy Editor

Lisa Gilmore
Deirdre McClellan

The *Stata Journal* publishes reviewed papers together with shorter notes or comments, regular columns, book reviews, and other material of interest to Stata users. Examples of the types of papers include 1) expository papers that link the use of Stata commands or programs to associated principles, such as those that will serve as tutorials for users first encountering a new field of statistics or a major new technique; 2) papers that go “beyond the Stata manual” in explaining key features or uses of Stata that are of interest to intermediate or advanced users of Stata; 3) papers that discuss new commands or Stata programs of interest either to a wide spectrum of users (e.g., in data management or graphics) or to some large segment of Stata users (e.g., in survey statistics, survival analysis, panel analysis, or limited dependent variable modeling); 4) papers analyzing the statistical properties of new or existing estimators and tests in Stata; 5) papers that could be of interest or usefulness to researchers, especially in fields that are of practical importance but are not often included in texts or other journals, such as the use of Stata in managing datasets, especially large datasets, with advice from hard-won experience; and 6) papers of interest to those who teach, including Stata with topics such as extended examples of techniques and interpretation of results, simulations of statistical concepts, and overviews of subject areas.

For more information on the *Stata Journal*, including information for authors, see the webpage

<http://www.stata-journal.com>

The *Stata Journal* is indexed and abstracted in the following:

- CompuMath Citation Index[®]
- Current Contents/Social and Behavioral Sciences[®]
- RePEc: Research Papers in Economics
- Science Citation Index Expanded (also known as SciSearch[®])
- Scopus[™]
- Social Sciences Citation Index[®]

Copyright Statement: The *Stata Journal* and the contents of the supporting files (programs, datasets, and help files) are copyright © by StataCorp LP. The contents of the supporting files (programs, datasets, and help files) may be copied or reproduced by any means whatsoever, in whole or in part, as long as any copy or reproduction includes attribution to both (1) the author and (2) the *Stata Journal*.

The articles appearing in the *Stata Journal* may be copied or reproduced as printed copies, in whole or in part, as long as any copy or reproduction includes attribution to both (1) the author and (2) the *Stata Journal*.

Written permission must be obtained from StataCorp if you wish to make electronic copies of the insertions. This precludes placing electronic copies of the *Stata Journal*, in whole or in part, on publicly accessible websites, file servers, or other locations where the copy may be accessed by anyone other than the subscriber.

Users of any of the software, ideas, data, or other materials published in the *Stata Journal* or the supporting files understand that such use is made without warranty of any kind, by either the *Stata Journal*, the author, or StataCorp. In particular, there is no warranty of fitness of purpose or merchantability, nor for special, incidental, or consequential damages such as loss of profits. The purpose of the *Stata Journal* is to promote free communication among Stata users.

The *Stata Journal*, electronic version (ISSN 1536-8734) is a publication of Stata Press. Stata, Mata, NetCourse, and Stata Press are registered trademarks of StataCorp LP.

Impact of interventions on discrete outcomes: Maximum likelihood estimation of the binary choice models with binary endogenous regressors

Michael Lokshin
The World Bank
Washington, DC
mlokshin@worldbank.org

Zurab Sajaia
The World Bank
Washington, DC
zsajaia@worldbank.org

Abstract. In this article, we describe the `switch_probit` command, which implements the maximum likelihood method to fit the model of the binary choice with binary endogenous regressors.

Keywords: `st0233`, `switch_probit`, endogenous variables, maximum likelihood, limited-dependent variables, binary choice models, impact evaluation, marginal treatment effect

1 Introduction

In this article, we describe the implementation of a maximum likelihood (ML) estimator of the parameters of binary choice models with endogenous regressors. In these models, a switching equation sorts individuals over two different states (with only one regime or outcome observed for each individual). The econometric problem of fitting a model with endogenous switching with binary endogenous regressors arises in a variety of settings in the modeling of the effects of fertility and migration on female labor force participation (LFP), modeling of housing demand, and in the modeling of markets in disequilibrium. For example,

- The paper by [Aakvik, Heckman, and Vytlačil \(2000\)](#) formulates an econometric framework for studying the impact of interventions on discrete outcomes when responses to treatment vary among observationally identical persons. Using this framework, the paper evaluates the effect of Norwegian Vocational Rehabilitation training programs on employment outcomes for women. The results demonstrate the positive effect of the training programs on employment after controlling for observable characteristics of the applicants. However, after controlling for unobserved characteristics of the applicants, the average treatment effect became negative, indicating that participants in the program have lower employment rates than nonparticipants.

- Carrasco (2001) estimates the causal effect of fertility on LFP of females in the United States using the 1986–1989 rounds of the University of Michigan Panel Study of Income Dynamics (PSID). The paper finds that the probability of LFP in women falls more in the model that accounts for endogenous fertility than in the model with exogenous fertility. The paper points to a downward bias induced by the exogeneity assumption of children variables that introduces a spurious positive correlation between fertility and LFP decisions.
- Lokshin and Glinskaya (2009) assess the impact of male migration on the labor market behavior of females in Nepal. The results indicate that male migration has a negative impact on the level of LFP by the women left behind. The paper finds evidence of substantial heterogeneity (based both on observable and unobservable characteristics) in the impact of male migration.

Models with endogenous switching can be fit one branch (selection equation and outcome equation) at a time (two `heckprob` estimations) or by simultaneous ML estimations (see [R] `biprobit` and [R] `heckprob`). However, both of these methods are inefficient, and the `biprobit` command is restrictive in that it assumes an equality of coefficients in the outcome equations for both treatment regimes. In addition, these approaches require potentially cumbersome adjustments to derive consistent standard errors. The `switch_probit` command, on the other hand, implements the full information ML method to simultaneously estimate the binary selection and the binary outcome parts of the model to yield consistent standard errors of the estimates. This approach relies on an assumption of joint normality of the error terms in the selection and outcome equations. The `switch_probit` command also derives the average treatment effects—the average effects of treatment on the treated and on the untreated—and the marginal treatment effects.

2 Model

Consider a model that describes the behavior of an agent with two binary outcome equations and a criterion function T_i that determines which regime the agent faces. T_i can be interpreted as a treatment. A motivating example is the effect of husband migration on wife's LFP. Here the treatment (migration of a husband) and the outcome (whether the wife works outside the home) can take one of the two potential values:

$$T_i = 1 \quad \text{if } \gamma Z_i + \mu_i > 0 \quad (1)$$

$$T_i = 0 \quad \text{if } \gamma Z_i + \mu_i \leq 0$$

$$y_{1i}^* = \beta_1 X_{1i} + \varepsilon_{1i} \quad y_{1i} = I(y_{1i}^* > 0) \quad (2)$$

$$y_{0i}^* = \beta_0 X_{0i} + \varepsilon_{0i} \quad y_{0i} = I(y_{0i}^* > 0) \quad (3)$$

Observed y_i is defined as

$$\begin{aligned} y_i &= y_{1i} & \text{if } T_i &= 1 \\ y_i &= y_{0i} & \text{if } T_i &= 0 \end{aligned}$$

where y_{1i}^* and y_{0i}^* are the latent variables (wife's propensity for LFP) that determine the observed binary outcomes y_1 and y_0 (whether the wife works or not); X_1 and X_0 are vectors of weakly exogenous variables; Z is a vector of variables that determines a switch between the regimes; β_1 , β_0 , and γ are vectors of parameters; and μ_i , ε_{1i} , and ε_{0i} are the error terms. Assume that μ_i , ε_{1i} , and ε_{0i} are jointly normally distributed, with a mean-zero vector and correlation matrix

$$\Omega = \begin{pmatrix} 1 & \rho_0 & \rho_1 \\ & 1 & \rho_{10} \\ & & 1 \end{pmatrix} \quad (4)$$

where ρ_0 and ρ_1 are the correlations between ε_0 , μ and ε_1 , μ , and ρ_{10} is the correlation between ε_0 and ε_1 .

Because y_{1i} and y_{0i} are never observed simultaneously, the joint distribution of $(\varepsilon_0, \varepsilon_1)$ is not identified, and consequently, ρ_{10} cannot be estimated. We assume that $\rho_{10} = 1$ (γ is estimable only up to a scalar factor). This model is identified by nonlinearities of its functional form. The log-likelihood function for the simultaneous system of equations [(1)–(3)] is

$$\begin{aligned} \ln(\mathfrak{F}) &= \sum_{T_i \neq 0, y_i \neq 0} w_i \ln \{ \Phi_2(X_{1i}\beta_1, Z_i\gamma, \rho_1) \} \\ &+ \sum_{T_i \neq 0, y_i = 0} w_i \ln \{ \Phi_2(-X_{1i}\beta_1, Z_i\gamma, -\rho_1) \} \\ &+ \sum_{T_i = 0, y_i \neq 0} w_i \ln \{ \Phi_2(X_{0i}\beta_0, -Z_i\gamma, -\rho_0) \} \\ &+ \sum_{T_i = 0, y_i = 0} w_i \ln \{ \Phi_2(-X_{0i}\beta_0, -Z_i\gamma, \rho_0) \} \end{aligned}$$

where Φ_2 is the cumulative function of a bivariate normal distribution and w_i is an optional weight for observation i . To ensure that estimated ρ_1 , ρ_0 are bounded between -1 and 1 , the ML directly estimates $\text{atanh}(\rho)$:

$$\text{atanh } \rho_j = \frac{1}{2} \ln \left(\frac{1 + \rho_j}{1 - \rho_j} \right) \quad j = 0, 1 \quad (5)$$

After estimating the model's parameters, the following statistics can be calculated (Aakvik, Heckman, and Vytlacil 2000):

- The effect of the treatment on the treated, or the expected effect of the treatment on individuals with observed characteristics x who participated in the program (TT):¹

$$\begin{aligned} \text{TT}(x) &= \Pr(y_1 = 1|T = 1, X = x) - \Pr(y_0 = 1|T = 1, X = x) \\ &= \frac{\Phi_2(X_1\beta_1, Z\gamma, \rho_1) - \Phi_2(X_0\beta_0, Z\gamma, \rho_0)}{F(Z\gamma)} \end{aligned} \quad (6)$$

where F is a cumulative function of the univariate normal distribution.

- The effect of the treatment on the untreated (TU), which is the expected effect of the treatment on individuals with observed characteristics x who did not participate in the program:

$$\begin{aligned} \text{TU}(x) &= \Pr(y_1 = 1|T = 0, X = x) - \Pr(y_0 = 1|T = 0, X = x) \\ &= \frac{\Phi_2(X_1\beta_1, -Z\gamma, -\rho_1) - \Phi_2(X_0\beta_0, -Z\gamma, -\rho_0)}{F(-Z\gamma)} \end{aligned} \quad (7)$$

- The treatment effect (TE), which is the expected effect of the treatment for the person with observed characteristics x randomly drawn from the population:

$$\text{TE}(x) = \Pr(T = 1, X = x) - \Pr(T = 0, X = x) = F(X_1\beta_1) - F(X_0\beta_0) \quad (8)$$

- The marginal treatment effect (MTE), which is the effect of the treatment on individuals with observed characteristics x and unobserved characteristics $\bar{\mu}$:

$$\begin{aligned} \text{MTE}(x, \bar{\mu}) &= \Pr(T = 1|X = x, \mu = \bar{\mu}) - \Pr(T = 0|x = x, \mu = \bar{\mu}) \\ &= F\left(\frac{X_1\beta_1 + \rho_1\bar{\mu}}{\sqrt{1 - \rho_1^2}}\right) - F\left(\frac{X_0\beta_0 + \rho_0\bar{\mu}}{\sqrt{1 - \rho_0^2}}\right) \end{aligned} \quad (9)$$

- The average treatment effects (ATT, ATU, and ATE) for the corresponding subgroups of the population, which can be calculated by averaging (6) through (8) over the observations in the subgroups. For example, the average treatment effect on the treated (ATT), which is the mean effect of the treatment on those who actually participated in the program, is

1. The treatment effect statistics are defined only for the cases when the exogenous variables in (2) and (3) are the same, in other words, when $X_0 = X_1$. When $X_0 \neq X_1$, the treatment effect statistics are calculated based on the vector of explanatory variables, which is a union of variables in X_0 and X_1 . The coefficients corresponding to the variables that were not initially included in the sets of explanatory variables for either of the equations are set to zero.

$$\text{ATT} = \frac{1}{N_T} \sum_{i=1}^{N_T} \text{TT}(x_i)$$

where N_T is the number of observations with $T = 1$ (number of treated individuals).

2.1 Additional methods and formulas

The probability of being treated—the probability for a husband to migrate:

$$\Pr(T = 1|z) = F(z\gamma) \quad (10)$$

The probability of being treated and having a positive outcome—the probability for a husband to migrate and for his wife to work:

$$\Pr(T = 1, y = 1|X = x) = \Phi_2(Z\gamma, X_1\beta_1, \rho_1) \quad (11)$$

The probability of being treated and having a zero outcome—the probability for a husband to migrate and for his wife not to work:

$$\Pr(T = 1, y = 0|X = x) = \Phi_2(Z\gamma, -X_1\beta_1 - \rho_1) \quad (12)$$

The probability of not being treated and having a positive outcome—the probability for a husband not to migrate and for his wife to work:

$$\Pr(T = 0, y = 1|X = x) = \Phi_2(-Z\gamma, X_0\beta_0, -\rho_0) \quad (13)$$

The probability of not being treated and having a zero outcome—the probability for a husband not to migrate and for his a wife not to work:

$$\Pr(T = 0, y = 0|X = x) = \Phi_2(-Z\gamma, -X_0\beta_0 - \rho_0) \quad (14)$$

The probability of having a positive outcome conditional on being treated—the probability for a wife of a migrant to work, conditional on her husband being migrant:

$$\Pr(y = 1|T = 1, X = x) = \frac{\Phi_2(Z\gamma, X_1\beta_1, \rho_1)}{F(Z\gamma)} \quad (15)$$

The probability of having a positive outcome conditional on being not treated—the probability for a wife of a nonmigrant to work, conditional on her husband being nonmigrant:

$$\Pr(y = 1|T = 0, X = x) = \frac{\Phi_2(-Z\gamma, X_0\beta_0, -\rho_0)}{F(-Z\gamma)} \quad (16)$$

3 The `switch_probit` command

3.1 Syntax

`switch_probit` is implemented as a d2 ML evaluator that calculates the overall log likelihood along with its first and second derivatives. The command allows for weights, clustering, robust standard errors, and the full set of options associated with Stata's ML procedures. The generic syntax² for the command is as follows:

```
switch_probit (depvar1 varlist1) [(depvar0 varlist0)] [if] [in] [weight],
    select(depvar_s varlist_s) [options]
```

`pweights`, `fweights`, and `iweights` are allowed; see [U] **11.1.6 weight**.

depvar1 is a binary outcome variable in regime 1. *varlist1* is a vector of explanatory variables in the equation explaining outcome in regime 1 [equation (2)]. *depvar0* and *varlist0* are, correspondingly, the binary outcome variable and the set of explanatory variables in regime 0 [equation (3)]. *depvar_s* is a binary dependent variable in selection equation (1), and *varlist_s* is a set of explanatory variables in selection equation (1).

In cases when the explanatory variables in the binary outcome equations are the same and there is only one dependent variable, only one equation needs to be specified. Alternatively, when the exogenous variables are different in outcome equations [(2) and (3)] and the dependent variables are different between the two outcome equations, both equations must be specified.

3.2 Options

`select(depvar_s varlist_s)` gives the specification of switching (1) for T_i . *varlist_s* might include the set of instruments that help identify the model. It is an integral part of the `switch_probit` estimation and is required. A full specification of explanatory variables is required for the selection equation (1); in other words, both instruments and exogenous variables must be specified in *varlist_s*. If there are no instrumental variables in the model, the model will be identified by nonlinearities.

`noconstant` suppresses the constant terms.

`offset1(varname)`, `offset0(varname)`, and `offset_s(varname)` include variables in each equation with coefficients constrained to 1.

For more information, see [R] **estimation options**.

`constraints(numlist | matname)` applies linear constraints to the fitted model.

`collinear` keeps collinear variables in the equations. By default, only noncollinear explanatory variables are used.

2. The syntax of `switch_probit` is similar to the syntax of the `movestay` command (Lokshin and Sajaia 2004).

`robust` specifies that the Huber/White/sandwich estimator of the variance is to be used in place of the conventional ML variance estimator. `robust` combined with `cluster()` further allows observations that are not independent within cluster (although they must be independent between clusters). If you specify `pweights`, then `robust` is implied. See [U] **20.16 Obtaining robust variance estimates**.

`cluster(varname)` specifies that the observations are independent across groups (clusters) but not necessarily within groups. *varname* specifies to which group each observation belongs; for example, `cluster(personid)` refers to data with repeated observations on individuals. Specifying `cluster()` affects the estimated standard errors and variance-covariance matrix of the estimators but not the estimated coefficients. `cluster()` can be used with `pweights` to produce estimates for unstratified cluster-sampled data. Specifying `cluster()` implies `robust`.

`level(#)` specifies the confidence level, as a percentage, for confidence intervals. The default is `level(95)` or as set by `set level`; see [U] **20.7 Specifying the width of confidence intervals**.

`noskip` specifies that a full ML model with only a constant for the regression equation be fit. This model is not displayed but is used as the base model to compute a likelihood-ratio test for the model test statistic displayed in the estimation header. By default, the overall model test statistic is an asymptotically equivalent Wald test that all the parameters in the regression equation are zero (except the constant). For many models, this option can substantially increase estimation time.

maximize_options control the maximization process; see [R] **maximize**. With the possible exception of `iterate(0)` and `trace`, you should specify these options only if the model is unstable. The maximization uses the `difficult` option by default. This option need not be specified.

3.3 Saved results

`switch_probit` saves the following in `e()` (* indicates saved parameters specific for `switch_probit`):

Scalars

<code>e(N)</code>	number of observations
<code>e(k)</code>	number of parameters
<code>e(k.eq)</code>	number of equations in <code>e(b)</code>
<code>e(k.eq.model)</code>	number of equations in overall model test
<code>e(k.aux)</code>	number of auxiliary parameters
<code>e(k.dv)</code>	number of dependent variables
<code>e(df.m)</code>	model degrees of freedom
<code>e(ll)</code>	log likelihood
<code>e(ll_0)</code>	log likelihood, constant-only model (<code>noskip</code> only)
<code>e(N.clust)</code>	number of clusters
<code>e(chi2.c)</code>	value of the χ^2 test on the equality of both correlation coefficients (ρ_1, ρ_0) to 0
<code>e(p.c)*</code>	probability of rejecting the χ^2 test
<code>e(p)</code>	significance of comparison test
<code>e(rho1)*</code>	estimated coefficient of correlation between the error terms of the selection equation and the outcome equations in regime 1
<code>e(rho0)*</code>	estimated coefficient of correlation between the error terms of the selection equation and the outcome equations in regime 0
<code>e(rank)</code>	rank of <code>e(V)</code>
<code>e(rank0)</code>	rank of <code>e(V)</code> for constant-only model
<code>e(ic)</code>	number of iterations
<code>e(rc)</code>	return code
<code>e(converged)</code>	1 if converged, 0 otherwise

Macros

<code>e(cmd)</code>	<code>switch_probit</code>
<code>e(depvar)</code>	name of dependent variable
<code>e(wtype)</code>	weight type
<code>e(wexp)</code>	weight expression
<code>e(title)</code>	title in estimation output
<code>e(clustvar)</code>	name of cluster variable
<code>e(offset1)</code>	offset for selection equation
<code>e(offset2)</code>	offset for equation 1
<code>e(offset3)</code>	offset for equation 0
<code>e(chi2type)</code>	Wald or LR; type of model χ^2 test
<code>e(chi2.ct)</code>	Wald or LR; type of model χ^2 test corresponding to <code>e(chi2.c)</code>
<code>e(vce)</code>	<code>vcetype</code> specified in <code>vce()</code>
<code>e(vcetype)</code>	title used to label Std. Err.
<code>e(opt)</code>	type of optimization
<code>e(ml.method)</code>	type of <code>ml</code> method
<code>e(user)</code>	name of likelihood-evaluator program
<code>e(technique)</code>	maximization technique
<code>e(crittype)</code>	optimization criterion
<code>e(properties)</code>	<code>b V</code>
<code>e(predict)</code>	program used to implement <code>predict</code>

Matrices

<code>e(b)</code>	coefficient vector
<code>e(ilog)</code>	iteration log (up to 20 iterations)
<code>e(gradient)</code>	gradient vector
<code>e(V)</code>	variance-covariance matrix of the estimators

Functions

<code>e(sample)</code>	marks estimation sample
------------------------	-------------------------

4 Postestimation

The `predict` command can follow `switch_probit` to calculate the predictive statistics. The statistics could be both in and out of the sample; type “`predict ... if e(sample) ...`” to generate statistics for observations in the estimated sample only.

```
predict [type] newvar [if] [in] [, statistic]
```

One of the following *statistics* may be specified with the `predict` command after `switch_probit`:

`p11`, the default, calculates the probability of being treated and having a positive outcome [equation (11)].

`p10` calculates the probability of being treated and having a zero outcome [equation (12)].

`p01` calculates the probability of not being treated and having a positive outcome [equation (13)].

`p00` calculates the probability of not being treated and having a zero outcome [equation (14)].

`psel` calculates the probability of being treated [equation (10)].

`pcond1` calculates the probability of a positive outcome conditional on being treated [equation (15)].

`pcond0` calculates the probability of a positive outcome conditional on not being treated [equation (16)].

`zb` calculates the probit linear prediction for the selection equation.

`xb1` calculates the linear prediction based on the coefficients of the outcome equation in regime 1.

`xb0` calculates the linear prediction based on the coefficients of the outcome equation in regime 0.

`stdpsel` calculates the standard error of the linear prediction of the selection equation.

`stdp1` calculates the standard error of the linear prediction of regime 1.

`stdp0` calculates the standard error of the linear prediction of regime 0.

`tt` calculates the treatment effect on the treated [equation (6)].

`tu` calculates the treatment effect on the untreated [equation (7)].

`te` calculates the treatment effect [equation (8)].

`mte` calculates the marginal treatment effect [equation (9)].

5 Example

We illustrate the use of the `switch_probit` command by looking at the problem of estimating the impact of a husband's migration on wife's LFP. A typical empirical specification for such a model might be the following:

$$M_i^* = Z_i\gamma + \mu_i \quad M_i = I(M_i^* > 0) = I(Z_i\gamma + \mu_i > 0) \quad (17)$$

$$\text{LFP}_{i1}^* = X_i\beta_1 + \varepsilon_{1i} \quad \text{LFP}_{i0}^* = X_i\beta_0 + \varepsilon_{0i} \quad (18)$$

$$\begin{aligned} W_i &= I(\text{LFP}_{i1}^* > 0) = I(X_i\beta_1 + \varepsilon_{1i} > 0) \quad \text{if } M_i = 1 \\ W_i &= I(\text{LFP}_{i0}^* > 0) = I(X_i\beta_0 + \varepsilon_{0i} > 0) \quad \text{otherwise} \end{aligned} \quad (19)$$

Here M_i^* is a latent continuous variable that determines the propensity of a husband to migrate; LFP_{i0}^* and LFP_{i1}^* is the latent continuous propensity of a wife to work outside the home if her husband migrates (subscript 1) or stays home (subscript 0); Z_i is a vector of characteristics that influences the migration decision; X_i is a vector of characteristics that is thought to influence the wife's LFP decision. β_1 , β_0 , and γ are vectors of parameters, and μ_i , ε_{1i} , and ε_{0i} are the disturbance terms. The observed wife's LFP, W_i , is a dichotomous realization of latent variable LFP_{i1}^* if a husband migrates and of latent variable LFP_{i0}^* if he does not migrate.

The assumption that is often made in this type of model is that the wife's decision to participate in the labor market is endogenous to her husband's migration decision. Some unobserved characteristics that influence the probability of a husband to migrate could also influence the decision of his wife to work or not. Neglecting these selectivity effects is likely to produce biased estimates of the impact of the husband's migration on the wife's LFP. The simultaneous ML estimation of (17), (18), and (19) with the proper instrumentation of the migration decision might correct such a bias.

The data from this example are a nonrandom subsample of the data from the 2004 round of the Nepal Living Standards Survey (for example, Lokshin and Glinskaya [2009]). The migration indicator `migrates` takes on value 1 if the husband migrates and 0 if he stays in the native country. The dependent variables in the wife's LFP equations, (19), are binary indicators of whether a wife works if her husband migrates (`works_1`) or whether she works if her husband stays (`works_0`). The set of exogenous variables in the LFP regressions (19) includes such wife's characteristics as her age, age-squared, educational dummies (`wedu_2`–`wedu_5`), and regional dummies (`reg2`–`reg6`). The omitted category for educational dummies is "illiterate", and higher-index dummies correspond to higher levels of wife's education. In addition to these variables, the migration equation (17) includes an instrument—`pmigrants`—to improve identification. A proportion of migrants in a ward is believed to influence the husband's migration decision but not to affect the wife's LFP decision.

The ML estimation of this specification using the `switch_probit` command on `switch_probit_example.dta` is shown below:

```
. use switch_probit_example
. switch_probit works age age2 wedu_2-wedu_5 hhsz hhsz2 reg_*,
> select(migrant age age2 wedu_2-wedu_5 hhsz hhsz2 reg_* pmigrants)
```

Fitting probit model for migrant=1:

(output omitted)

Fitting full model:

Iteration 0: log likelihood = -5631.5963

(output omitted)

Iteration 6: log likelihood = -5588.5707

Switching probit model

Number of obs = 5426

Wald chi2(14) = 319.25

Prob > chi2 = 0.0000

Log likelihood = -5588.5707

	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
migrant						
age	-.053828	.0101656	-5.30	0.000	-.0737521	-.0339039
age2	.0738756	.0133015	5.55	0.000	.0478051	.0999461
wedu_2	.0451867	.0639763	0.71	0.480	-.0802046	.170578
wedu_3	.1416557	.0674213	2.10	0.036	.0095124	.2737991
wedu_4	.197108	.0611319	3.22	0.001	.0772916	.3169243
wedu_5	.0142019	.0994006	0.14	0.886	-.1806197	.2090234
hhsz	-.0556154	.0140021	-3.97	0.000	-.0830591	-.0281718
hhsz2	.0013728	.000646	2.13	0.034	.0001068	.0026389
reg_2	.7102527	.082358	8.62	0.000	.548834	.8716713
reg_3	.873601	.0915923	9.54	0.000	.6940835	1.053119
reg_4	.6790904	.0865635	7.84	0.000	.509429	.8487518
reg_5	.7782603	.0931983	8.35	0.000	.5955949	.9609257
reg_6	.9390181	.0849033	11.06	0.000	.7726107	1.105425
pmigrants	.9877826	.1358246	7.27	0.000	.7215712	1.253994
_cons	-.2400678	.2142056	-1.12	0.262	-.659903	.1797675
works_1						
age	.1251623	.0247011	5.07	0.000	.0767491	.1735755
age2	-.1696758	.0327141	-5.19	0.000	-.2337943	-.1055572
wedu_2	-.2971512	.154475	-1.92	0.054	-.5999166	.0056142
wedu_3	-.0435156	.1531251	-0.28	0.776	-.3436352	.2566041
wedu_4	-.0806875	.141102	-0.57	0.567	-.3572423	.1958673
wedu_5	.4507575	.2118544	2.13	0.033	.0355305	.8659846
hhsz	-.1056211	.0539037	-1.96	0.050	-.2112704	.0000283
hhsz2	.0022055	.0033594	0.66	0.511	-.0043789	.0087898
reg_2	-.2725182	.2955016	-0.92	0.356	-.8516906	.3066542
reg_3	-.7467022	.3581774	-2.08	0.037	-1.448717	-.0446874
reg_4	-.4249299	.2932972	-1.45	0.147	-.9997818	.1499221
reg_5	-.3665319	.3253127	-1.13	0.260	-1.004133	.2710693
reg_6	-.1816081	.3456906	-0.53	0.599	-.8591491	.495933
_cons	-2.041429	.7257811	-2.81	0.005	-3.463934	-.618924
works_0						
age	.0888403	.0130782	6.79	0.000	.0632076	.1144731
age2	-.1230907	.0174413	-7.06	0.000	-.157275	-.0889063
wedu_2	.1174843	.077252	1.52	0.128	-.0339267	.2688954
wedu_3	.0595546	.0853855	0.70	0.486	-.1077979	.2269072
wedu_4	.0320592	.0765806	0.42	0.675	-.1180359	.1821544
wedu_5	.2318881	.0982554	2.36	0.018	.039311	.4244652
hhsz	-.0530284	.0204413	-2.59	0.009	-.0930925	-.0129642
hhsz2	.0011186	.0009241	1.21	0.226	-.0006926	.0029297

reg_2	-.5203729	.0806754	-6.45	0.000	-.6784939	-.362252
reg_3	-1.036087	.0916373	-11.31	0.000	-1.215693	-.8564816
reg_4	-.7069518	.082838	-8.53	0.000	-.8693112	-.5445923
reg_5	-1.03612	.1017284	-10.19	0.000	-1.235504	-.8367363
reg_6	-.5616403	.0946533	-5.93	0.000	-.7471574	-.3761232
_cons	-1.64381	.2658748	-6.18	0.000	-2.164915	-1.122705
<hr/>						
/athrho1	-.231402	.3879283			-.9917276	.5289235
/athrho0	-.8448182	.5563851			-1.935313	.2456764
<hr/>						
rho1	-.2273583	.3678756			-.758098	.4845578
rho0	-.6883527	.2927534			-.9591607	.2408502
<hr/>						
LR test of indep. eqns. (rho1=rho0=0):chi2(2) = 4.96 Prob > chi2 = 0.0838						
<hr/>						

The results of the husband's migration equation are reported in the section of the output headed "migrant". The results of the wife's LFP equation in the regime where her husband migrates are reported in the "works_1" section, and the wife's LFP equation in the regime where her husband stays is outputted in the "works_0" section.

The variables `/athrho1` and `/athrho0` are ancillary parameters used in the ML procedure. `/athrho1` and `/athrho0` are the transformations of the correlation coefficients as in (5).

The correlation coefficients `rho1` and `rho0` are both negative but are significant only for the correlation between the error terms in the equation determining the husband's migration and the wife's LFP equation if her husband stays home.

The likelihood-ratio test for joint independence of the equations is reported in the last line of the output. The test rejects the H_0 that $\rho_0 = \rho_1$: Prob > $\chi^2 = 0.08$.

We can now derive the effect of a husband's migration on his wife's LFP by interpreting migration as a treatment, (6), using the `predict` command:

```
. predict tt, tt
. summarize tt if (migrant == 1)
```

Variable	Obs	Mean	Std. Dev.	Min	Max
tt	1694	.1160357	.0705859	.0046564	.4448425

Women living in migrant-sending households had 11.6 percentage points or about 90% lower probability of LFP compared with the counterfactual scenario of women living in nonsending households.

6 Simulation Monte Carlo study

In this section, we provide a brief discussion of sensitivity of our estimator for model identification and the assumptions about the distribution of the error terms. To illustrate the properties of our estimator, we conduct Monte Carlo simulations across a range of model specifications.

The baseline data-generating process of (1) through (4) has the following form:

$$\begin{aligned}T_i &= 1 && \text{if } \gamma_1 x_{1i} + \gamma_2 x_{2i} + \gamma_3 x_{3i} + \gamma_4 z_i + \mu_i > 0 \\T_i &= 0 && \text{otherwise} \\y_{1i}^* &= \beta_{11} x_{1i} + \beta_{12} x_{2i} + \beta_{13} x_{3i} + \varepsilon_{1i} && y_{1i} = I(y_{1i}^* > 0) \\y_{0i}^* &= \beta_{01} x_{1i} + \beta_{02} x_{2i} + \beta_{03} x_{3i} + \varepsilon_{0i} && y_{0i} = I(y_{0i}^* > 0)\end{aligned}$$

(20)

We generate x_1 , x_2 , x_3 , and z as independent standard normal random variables. The true values of coefficients β and γ are shown in the top panel of table 1.

Table 1. Monte Carlo simulations for different model specifications

Selection equation				Binary outcome $T = 1$			Binary outcome $T = 0$		
	Coeff.	Std.	% test	Coeff.	Std.	% test	Coeff.	Std.	% test
		Dev.	rejection		Dev.	rejection		Dev.	rejection
<i>True coefficients</i>									
x1/x1	1.000			1.000			1.000		
x2/x1	1.000			2.000			2.000		
x3/x1	1.000			3.000			3.000		
z/x1	1.000								
<i>Normally distributed errors; model is identified through the instrument</i>									
x1/x1	1.000			1.000			1.000		
x2/x1	0.989	0.037	5.12	1.978	0.062	5.24	1.975	0.137	5.81
x3/x1	1.031	0.041	5.03	2.960	0.100	3.92	2.991	0.220	5.73
z/x1	0.987	0.034	4.81						
<i>Normally distributed errors; model is identified through nonlinearities only</i>									
x1/x1	1.000			1.000			1.000		
x2/x1	0.988	0.037	7.15	1.981	0.072	98.63	2.031	0.171	100.0
x3/x1	1.030	0.041	19.51	2.966	0.116	100.0	3.103	0.318	100.0
z/x1	0.989	0.034	99.7	−0.011	0.046	100.0	−0.058	0.141	100.0
<i>Nonnormal errors; model is identified through the instrument</i>									
x1/x1	1.000			1.000			1.000		
x2/x1	0.989	0.031	3.77	2.087	0.098	7.45	1.949	0.075	4.49
x3/x1	1.031	0.033	5.32	3.239	0.146	11.72	3.015	0.117	5.36
z/x1	0.996	0.028	7.71						
<i>Nonnormal errors; model is identified through nonlinearities only</i>									
x1/x1	1.000			1.000			1.000		
x2/x1	0.991	0.031	4.95	2.069	0.193	100.00	2.320	0.260	4.34
x3/x1	1.032	0.033	5.36	3.199	0.401	100.00	3.781	0.494	6.73
z/x1	0.997	0.029	8.43	0.021	0.192	98.87	−0.406	0.239	11.28

We conduct Monte Carlo simulations for four scenarios on samples of 10,000 observations with 1,000 repetitions. In all simulations, we show the ratio of γ_i to β_i ($i = 1, \dots, 3$) to ensure the comparability of the estimation results across different model specifications.

In the first scenario, shocks μ_i , ε_{1i} , and ε_{0i} are generated as standard trivariate normal; the instrument z is excluded from the outcome equations. The estimates of the ratios correspond well to the true coefficients.

In the second scenario, with the same error distribution, we add instrument z into the outcome equations. In this specification, the model is identified through nonlinearities of the functional form. The estimated ratios of coefficients are still close to the true ratios. The coefficient on the instrumental variable is insignificant in both outcome equations. The standard errors of the estimates in the outcome equations are larger compared with the instrument-identified specification. Note that Wald tests at the 5% level always reject the true null hypothesis for 5% of the parameter estimates. The weak identification offered by function-form identification makes the large-sample properties of the estimator worthless in this case.

In the third scenario, the errors are χ^2 distributed and instrument z is excluded from the outcome equations. The estimated coefficients are now further away from the true coefficients compared with the first scenario.

Comparable results are observed in the fourth scenario with nonnormal errors, although the precision of the estimates deteriorates significantly in this case. The simulation based on different functional forms for the nonnormal distribution of the shocks in (20) produces similar estimates.

The results of our simulations indicate that the estimator described in this paper is relatively robust in terms of identification of the model. These findings are consistent with conclusions of [Wilde \(2000\)](#) that “in recursive multiple-equation probit models with endogenous dummy regressors no exclusion restrictions for the exogenous variables are needed if there is sufficient variation in the data”.

We also evaluate the performance of our estimator in terms of predicting the ATE and ATT effects. The data-generating process in these simulations is similar to the data-generating process described in (20), but in addition to presenting the results based on 10,000 observations, we generate ATT and ATE for the sample sizes ranging from 200 to 30,000 observations.

Figure 1 shows the results of Monte Carlo simulations of ATE and ATT for the specification with normally distributed error terms and 1,000 repetitions. The simulations demonstrate a good performance of the ML algorithm described in this paper when the errors in (1), (2), and (3) are jointly normally distributed. Even for smaller sample sizes, the method produces efficient and unbiased estimates of ATE and ATT effects.

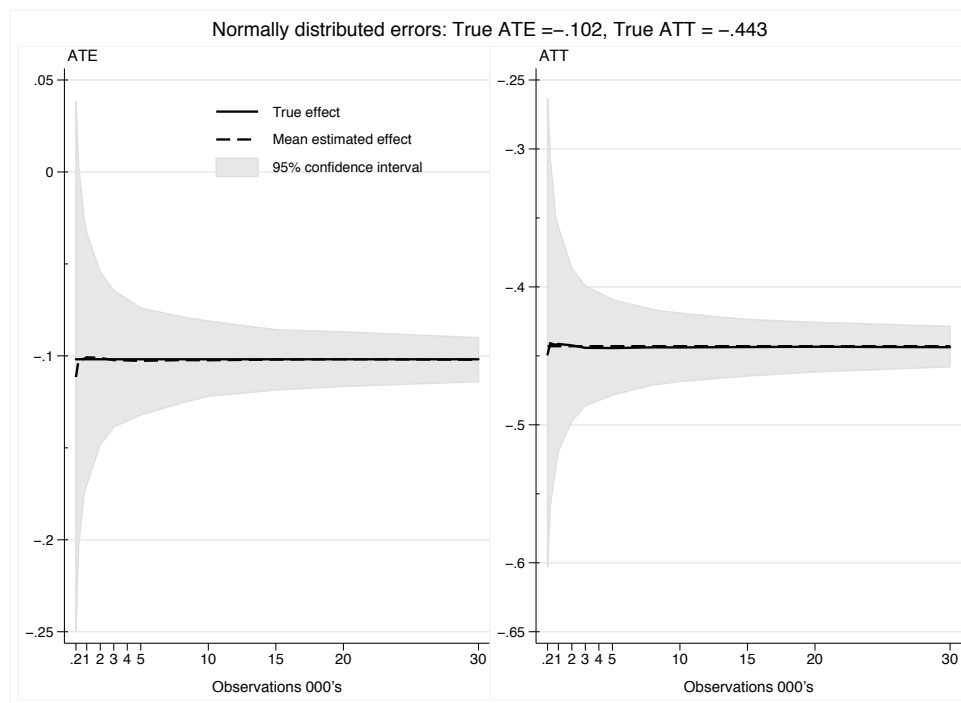


Figure 1. The results of Monte Carlo simulations (1,000 repetitions) of ATE and ATT effects; specification with normally distributed errors

Figure 2 presents the results of Monte Carlo simulations of ATE and ATT for the specification where the error terms are nonnormally distributed. The violation of the normality assumption results in biased estimates for both ATE and ATT effects. The bias is larger for estimations based on smaller sample sizes.

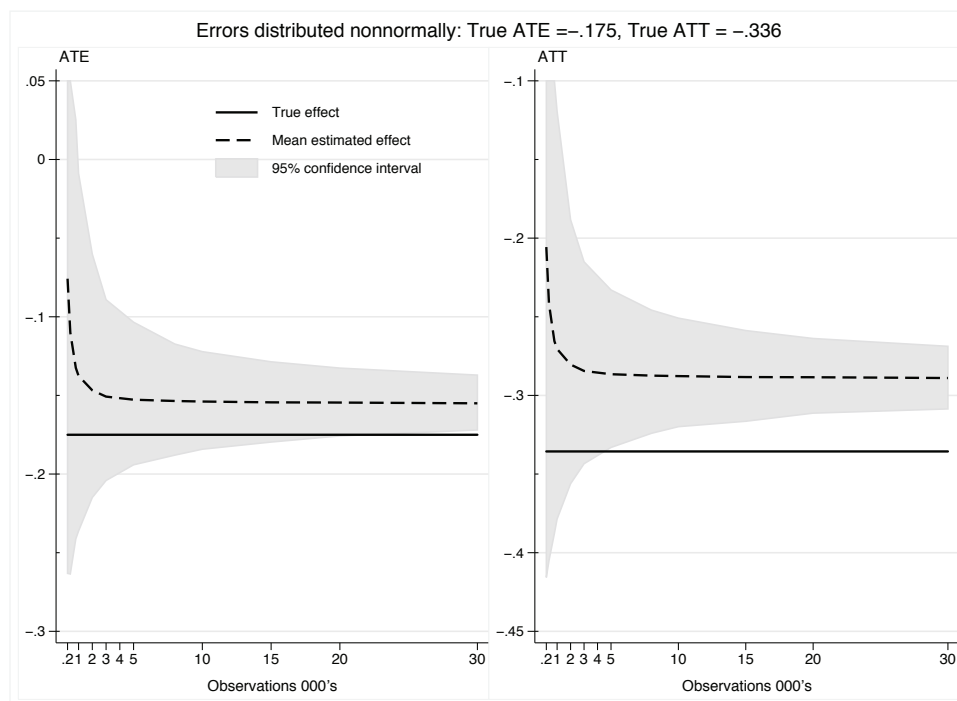


Figure 2. The results of Monte Carlo simulations (1,000 repetitions) of ATE and ATT effects; specification with nonnormally distributed errors

Our final simulation results examine the validity of confidence intervals depending on the assumptions about the joint distribution of the error terms in (20). Figure 3 shows the coverage rates for ATE and ATT effects. The coverage rates are constructed for the specifications with normally (solid line) and nonnormally distributed error terms (dotted line) in (20). The simulations are based on the bootstrap estimations of the confidence intervals for ATE and ATT effects for 1,000 replications for 1,000 Monte Carlo repetitions (that is, 1,000,000 ML model estimation for each sample size). The size α confidence interval is reported as the interval between the $\alpha/2$ and $1 - \alpha/2$ quantiles of the simulated draws of ATE and ATT.

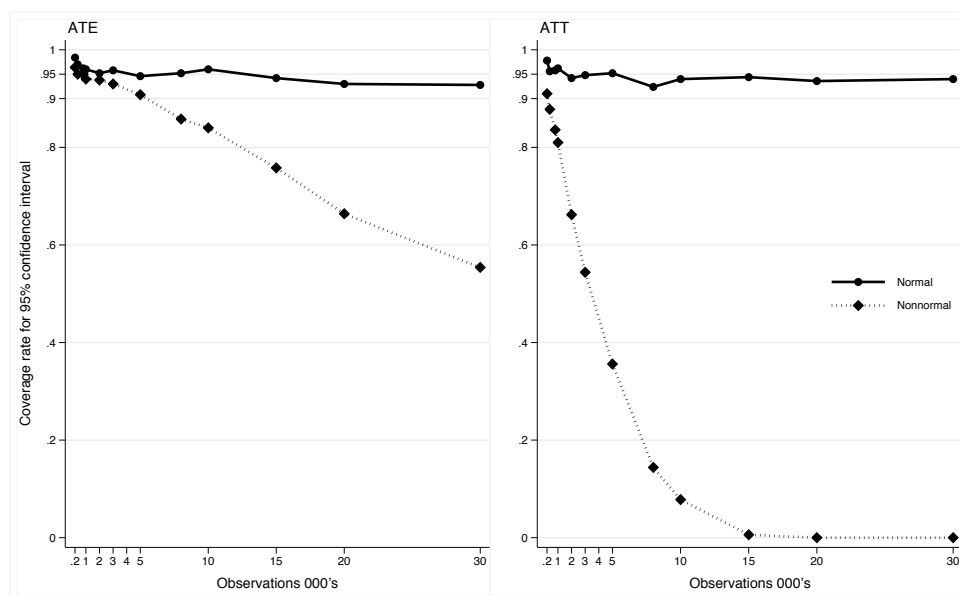


Figure 3. Coverage rates for ATE and ATT effects

The left panel of figure 3 shows the coverage rates for 95% confidence intervals of ATE for samples of a different size. The coverage rates for the confidence intervals estimated based on normally distributed errors are close to the nominal 95%. The coverage rates for nonnormal specification demonstrate undercoverage that increases with the sample size. For example, while the coverage rates for samples of up to 2,000 observations are close to the nominal 95%, the coverage rates drop to about 54% for the simulations based on the sample with 30,000 observations. This decline in the coverage rates for nonnormal specification are consistent with the estimation bias and the narrower confidence intervals shown in figure 2.

The right panel of figure 3 presents the coverage rates for 95% confidence intervals of ATT estimates. Similarly to ATE, the coverage rates for ATT estimated under normality assumptions are close to nominal. For the simulations based on nonnormal errors, the undercoverage is more severe compared with ATE: the 95% coverage rates decline rapidly from about 90% for the small samples to 0% for the samples of 15,000 observations and larger. Again these results are consistent with the bias and pattern of confidence intervals for ATT shown in the right panel of figure 2.

7 Conclusion

This article describes a Stata implementation of an ML estimator for the parameters of a binary response model with endogenous switching. The `switch_probit` command

extends the set of Stata ML algorithms for estimation of the models with endogenous switching (for example, `movestay` by Lokshin and Sajaia [2004]). We think that the ability of the new command to produce estimates of the treatment impact for different population subgroups could be useful in applied studies of impact evaluation.

The results of our Monte Carlo simulations indicate that while the estimator performs well under the assumption of normally distributed error terms, it produces biased estimates if the normality assumptions are violated. Researchers who suspect that the normality assumptions are not likely to hold might want to use other, semiparametric or nonparametric, methods of estimation for such models.

8 References

- Aakvik, A., J. J. Heckman, and E. J. Vytlačil. 2000. Treatment effects for discrete outcomes when responses to treatment vary among observationally identical persons: An application to Norwegian vocational rehabilitation programs. NBER Technical Working Paper No. 262. <http://www.nber.org/papers/t0262>.
- Carrasco, R. 2001. Binary choice with binary endogenous regressors in panel data: Estimating the effect of fertility on female labor participation. *Journal of Business and Economic Statistics* 19: 385–394.
- Lokshin, M., and E. Glinskaya. 2009. The effect of male migration on employment patterns of women in Nepal. *World Bank Economic Review* 23: 481–507.
- Lokshin, M., and Z. Sajaia. 2004. Maximum likelihood estimation of endogenous switching regression models. *Stata Journal* 4: 282–289.
- Wilde, J. 2000. Identification of multiple equation probit models with endogenous dummy regressors. *Economics Letters* 69: 309–312.

About the authors

Michael Lokshin is an adviser in the Research Department of the World Bank.

Zurab Sajaia is an economist in the Research Department of the World Bank.